Mauricio Christian Villarroel Montoya
Jesus College

Doctor of Philosophy
Trinity Term, 2017

Non-contact vital sign monitoring in the clinic

Current monitoring systems available to track changes in the vital signs of patients (such as heart rate, respiratory rate or peripheral oxygen saturation (SpO_2)) require adhesive electrodes or sensors to be in direct contact with the subjects. Most patients find the probes difficult to attach or wear over prolonged periods of time. The ideal technology would involve sensors with no direct contact with the patient. Non-contact sensing provides several advantages, such as: no subject participation to set the equipment up, no skin preparation or irritation, decrease in the risk of infection, and the potential to be seamlessly integrated into the patient's lifestyle.

It has been previously established that the colour and volume changes in superficial blood vessels during the cardiac cycle can be measured using a digital video camera over short time periods and under tightly-controlled conditions with relatively still and healthy subjects. Using an off-the-shelf camera and standard ambient light, this thesis proposes data fusion algorithms to compute estimates of heart rate and respiratory rate, and also to detect changes in SpO_2 in a real hospital scenario, without interfering with regular patient care.

A clinical study was designed to record a wide range of physiological values from patients undergoing haemodialysis treatment in the Renal Unit of the Churchill Hospital in Oxford. Video was recorded for a total of 104 dialysis sessions from 40 patients during the course of one year, comprising an overall video recording time of approximately 370 hours. Reference values were provided by two devices in regular clinical use.

Algorithms are proposed to analyse the input video to identify time periods for which the location of the patient is known in the video frame, discarding periods for which there is high activity or which are affected by motion artefacts. The accuracy of heart rate estimation, for periods during which the subject is stable, is comparable to that of the two standard pulse oximeters used in the clinical study (positioned at the finger and earlobe respectively). The mean absolute error between the heart rate estimates from the camera and the average from two reference pulse oximeters was 2.8 $\mathrm{beats/min}$ for over 65% of the time. The mean absolute error between the respiratory rate estimates from the camera and the reference values (computed from the Electrocardiogram and a thoracic expansion sensor - chest belt) was 2.1 $\mathrm{breaths/min}$ for over 69% of the time for which the reference signals were valid. By calibrating the camera data with the reference pulse oximeters, changes in SpO_2 could also be tracked during time periods with minimal patient motion.

To increase the robustness of the algorithms, novel methods were devised for cancelling out aliased frequency components caused by the artificial light sources in the hospital, using auto-regressive modelling and pole cancellation. Maps of the spatial distribution of heart rate and respiratory rate information were developed from the coefficients of the auto-regressive models. Most of the periods for which the camera could not produce a reliable heart rate estimate were under 10 minutes, this opens the possibility to continuously monitor heart rate with a clinically relevant frequency.

Non-contact vital sign monitoring in the clinic



Mauricio Christian Villarroel Montoya

Jesus College

University of Oxford

Supervised by

Prof. Lionel Tarassenko

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term, 2017

Declaration

I declare that this thesis is entirely my own work and, except where stated, describes my own research.

Mauricio Christian Villarroel Montoya

Jesus College

To my beautiful wife Henssy and my wonderful children Jose and Mia

Acknowledgements

The seed was planted several years ago by my good old friend Patrick McSharry. By showing me college brochures and those wonderful sunny pictures of Oxford during summertime, the desire and dream to one day be walking by the streets of this incredible city increased.

My journey into biomedical research started and spanned across two continents by the enormous gamble of who I now considered an intimate friend, Gari Clifford. Without his initial spark and continuous push, I wouldn't be where I am today ... I owe you one my friend!. Roger Mark and George Moody, from the Laboratory for Computational Physiology at MIT, played a pivotal role in my career by providing me an invaluable experience in a world-class research group. Li-Wei Lehman, Wei Zong, Qiao Li, Sherman Jia, James Sun, Tin Kyaw and other friends at LCP also greatly contributed to my research knowledge.

This thesis and all the work that followed it was the result of the inspiration from Lionel Tarassenko, who I not only consider my supervisor but a role model to follow. It was a real privilege to work beside him. Since my arrival to England, Lionel showed me how high-quality research is done and has become an example on how you should conduct yourself in life. Through his leadership, what started as a dream between the two of us several years ago, it spawned to create the camera research team to include vital members who I enjoyed working with: Alessandro Guazzi, Joao Jorge, Jonathan Daly and lately Sitthicock Chaichulee.

Chris Pugh and David Meredith were instrumental in the design and execution of the clinical study for this thesis and provided great medical insight into human physiology to help solve several problems during my research. David Clifton was always a reference and a source of knowledge any time I ran into bottlenecks. Andrew Zisserman provided indispensable expertise and support. The data used for this thesis would not have existed if it wasn't for the dedication and hard work of Sheera Sutherland, by far the best nurse I have worked with. I would like thank all the staff in the Renal Unit in the Churchill hospital who were very supportive of our study from the start. I would also like to thank to all the patients who agreed to participate in the study, to the ones who are still with us and the ones who sadly passed away.

A big thanks to all the friends I made throughout the years in the BSP group, IBME, engineering and across Oxford in general. Specially Jullien Oster, Christina Orphanidou, Dave Wong, Vernon Bailey, Vanessa Wilkins, Marco Pimentel, Mauro Santos, Clare Macewen, Yasmina Borhani and all of the people who contributed to my life.

Throughout the endless nights and long weekends of hard work, my family was always supporting me, making bigger sacrifices than me. To my companion of decades, my wife Henssy; my son Jose and daughter Mia ... A big thanks to the three of you ..."Los amo mucho mis amores".

This thesis was created using the LaTeX typesetting system.

Abstract

Current monitoring systems available to track changes in the vital signs of patients (such as heart rate, respiratory rate or peripheral oxygen saturation (SpO_2)) require adhesive electrodes or sensors to be in direct contact with the subjects. Most patients find the probes difficult to attach or wear over prolonged periods of time. The ideal technology would involve sensors with no direct contact with the patient. Non-contact sensing provides several advantages, such as: no subject participation to set the equipment up, no skin preparation or irritation, decrease in the risk of infection, and the potential to be seamlessly integrated into the patient's lifestyle.

It has been previously established that the colour and volume changes in superficial blood vessels during the cardiac cycle can be measured using a digital video camera over short time periods and under tightly-controlled conditions with relatively still and healthy subjects. Using an off-the-shelf camera and standard ambient light, this thesis proposes data fusion algorithms to compute estimates of heart rate and respiratory rate, and also to detect changes in SpO_2 in a real hospital scenario, without interfering with regular patient care.

A clinical study was designed to record a wide range of physiological values from patients undergoing haemodialysis treatment in the Renal Unit of the Churchill Hospital in Oxford. Video was recorded for a total of 104 dialysis sessions from 40 patients during the course of one year, comprising an overall video recording time of approximately 370 hours. Reference values were provided by two devices in regular clinical use.

Algorithms are proposed to analyse the input video to identify time periods for which the location of the patient is known in the video frame, discarding periods for which there is high activity or which are affected by motion artefacts. The accuracy of heart rate estimation, for periods during which the subject is stable, is comparable to that of the two standard pulse oximeters used in the clinical study (positioned at the finger and earlobe respectively). The mean absolute error between the heart rate estimates from the camera and the average from two reference pulse oximeters was 2.8 $\mathrm{beats/min}$ for over 65% of the time. The mean absolute error between the respiratory rate estimates from the camera and the reference values (computed from the Electrocardiogram and a thoracic expansion sensor - chest belt) was 2.1 $\mathrm{breaths/min}$ for over 69% of the time for which the reference signals were valid. By calibrating the camera data with the reference pulse oximeters, changes in SpO_2 could also be tracked during time periods with minimal patient motion.

To increase the robustness of the algorithms, novel methods were devised for cancelling out aliased frequency components caused by the artificial light sources in the hospital, using auto-regressive modelling and pole cancellation. Maps of the spatial distribution of heart rate and respiratory rate information were developed from the coefficients of the auto-regressive models. Most of the periods for which the camera could not produce a reliable heart rate estimate were under 10 minutes, this opens the possibility to continuously monitor heart rate with a clinically relevant frequency.

List of publications

The following is a list of articles, book chapters and patents that were published as a result of the research carried out for this thesis.

Articles:

- Mauricio Villarroel, João Jorge, Chris Pugh and Lionel Tarassenko. Non-contact vital sign monitoring in the clinic. 12th IEEE International Conference on Automatic Face & Gesture Recognition, 278-285, 2017.
- João Jorge, Mauricio Villarroel, Sitthichok Chaichulee, Alessandro Guazzi, Gabrielle Green, Kenny McCormick and Lionel Tarassenko. Non-contact monitoring of respiration in the neonatal intensive care unit. 12th IEEE International Conference on Automatic Face & Gesture Recognition, 286-293, 2017.
- Sitthichok Chaichulee, Mauricio Villarroel, João Jorge, Carlos Arteta, Gabrielle Green, Kenny McCormick, Andrew Zisserman and Lionel Tarassenko. Multi-task Convolutional Neural Network for Patient Detection and Skin Segmentation in Continuous Non-contact Vital Sign Monitoring. 12th IEEE International Conference on Automatic Face & Gesture Recognition, 266-272, 2017.
- Alessandro Guazzi, Mauricio Villarroel, João Jorge, Jonathan Daly, Matthew Frise, Peter Robbins and Lionel Tarassenko. Non-contact measurement of oxygen saturation with an RGB camera. Biomedical optics express, 3320-3338, 2015.
- Mauricio Villarroel, Alessandro Guazzi, João Jorge, Sara Davis, Peter Watkinson, Gabrielle Green, Asha Shenvi, Kenny McCormick, and Lionel Tarassenko. Continuous non-contact vital sign monitoring in neonatal intensive care unit. Healthcare technology letters, 87-91, 2014.
- Lionel Tarassenko, Mauricio Villarroel, Alessandro Guazzi, João Jorge, David Clifton, Chris Pugh. Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. Physiological measurement, 807-831, 2014.

Book chapters:

- Peter Charlton, Mauricio Villarroel and Francisco Salguiero. Waveform Analysis to Estimate Respiratory Rate. In Secondary Analysis of Electronic Health Records, pp. 377-390. Springer International Publishing, 2016.

Patents:

- Alessandro Guazzi, Mauricio Villarroel and Lionel Tarassenko. Method and apparatus for vital signs measurement. Patent Application WO2016193735 A1. Filed in June 2015.
- David Clifton, Mauricio Villarroel and Lionel Tarassenko. Analysing video images of a subject to identify spatial image areas which contain periodic intensity variations. WO 2014125250 A1. Filed in February 2013.
- Mauricio Villarroel, Lionel Tarassenko and David Clifton. Method and System for Signal Analysis. Patent Application No. 1302451.8. Filed in February 2013.
- Lionel Tarassenko, Mauricio Villarroel and David Clifton. Remote monitoring of vital signs. WO2013027027 A3. Filed in August 2011.

Table of contents

Table of contents xi

List of figures xvi

List of tables xxi

List of abbreviations xxiii

1 Introduction 1
- 1.1 Monitoring of vital signs 2
  - 1.1.1 Heart rate 2
  - 1.1.2 Respiratory Rate 3
  - 1.1.3 Peripheral oxygen saturation (SpO_2) 4
- 1.2 Clinical need 5
- 1.3 Objectives 5
- 1.4 Contributions of thesis 6
- 1.5 Outline of thesis 7

2 Non-contact vital sign monitoring 9
- 2.1 Introduction 9
- 2.2 Overview of methods 9
- 2.3 Non-contact ECG 10
- 2.4 Radar 11
  - 2.4.1 Vital-sign monitoring with CW radar 11
  - 2.4.2 Advantages of UWB radar 12
  - 2.4.3 Vital-sign monitoring with UWB radar 13
- 2.5 Laser-based techniques 14
- 2.6 Conclusion 14

3 Photoplethysmography and oxygen monitoring 15
- 3.1 Introduction 15
- 3.2 The importance of oxygen monitoring 15
- 3.3 Oxygen transport physiology 16
  - 3.3.1 Respiratory process 17
  - 3.3.2 Circulatory process 20
  - 3.3.3 Haemoglobin as oxygen carrier 22
- 3.4 Blood oxygen measurement 23
  - 3.4.1 Oxygen saturation in a blood sample 23
  - 3.4.2 Oxyhaemoglobin dissociation curve 23
  - 3.4.3 Methods for measuring blood oxygen content 24
- 3.5 Oximetry 25
  - 3.5.1 Principles of oximetry 25
  - 3.5.2 Haemoglobin absorption spectra 25
  - 3.5.3 Light absorbance 26
  - 3.5.4 Absorbance in haemoglobin solutions 28
  - 3.5.5 In vivo oximetry 28
- 3.6 Pulse oximetry 29

3.6.1 Principles of pulse oximetry 29
3.6.2 Wavelength selection criteria 29
3.6.3 The photoplethysmogram (PPG) 30
3.6.4 Methods of operation 32
3.6.5 Locations for oxygen saturation measurement 32
3.6.6 SpO_2 estimation 33
3.6.7 Heart rate estimation from the PPG 35
3.6.8 Respiratory rate estimation from the PPG 36
3.6.9 Limitations and sources of error 38
3.7 Conclusion 43

4 Photoplethysmographic imaging (PPGi) 45
4.1 Introduction 45
4.2 Imaging sensors 45
4.2.1 CCD imaging sensors 46
4.2.2 CMOS imaging sensors 47
4.2.3 Choice of imaging sensor 48
4.2.4 The imaging pipeline 48
4.3 Video-based vital-sign monitoring 48
4.3.1 Introduction 48
4.3.2 Review of methods 49
4.3.3 Use of thermal cameras 53
4.4 Challenges 53
4.5 Conclusion 54

5 Dialysis study 55
5.1 Introduction 55
5.2 Haemodialysis 55
5.2.1 Current haemodialysis monitoring 56
5.3 Clinical study design 57
5.3.1 Research ethics application 57
5.3.2 Patient consent 57
5.3.3 Exclusion criteria 58
5.3.4 Patient confidentiality 58
5.3.5 Patient privacy during recording 58
5.4 Instrumentation 59
5.4.1 Hidalgo's Equivital EQ02 LifeMonitor 60
5.4.2 Video and pulse oximeter recording software 60
5.4.3 Data storage 65
5.4.4 Time synchronisation 66
5.5 Patient population 66
5.5.1 Demographics 66
5.5.2 Vital signals 67
5.6 Conclusion 69

6 Heart rate estimation using a video camera 71
6.1 Introduction 71
6.2 Overview of the estimation process 71
6.3 Reference heart rate 73
6.3.1 Dual PPG monitoring during haemodialysis 73
6.3.2 Heart rate measurement error 74

6.3.3 Computing the reference heart rate ... 77
6.3.4 Results ... 79
6.3.5 Conclusion ... 80
6.4 Video analysis ... 80
6.4.1 Face detection ... 82
6.4.2 Face tracking ... 86
6.4.3 Activity ... 86
6.4.4 Extraction of stable periods ... 87
6.5 Deriving the PPGi signal from video ... 88
6.5.1 PPGi extraction ... 88
6.5.2 Subject region of interest (ROI_S) ... 90
6.5.3 Background region of interest ... 93
6.6 Beat-by-beat quality assessment ... 93
6.6.1 Bayesian change point detection ... 94
6.6.2 Digital filtering ... 96
6.6.3 Onset detection ... 97
6.6.4 Data windowing ... 99
6.6.5 Template generation ... 99
6.6.6 Multi-scale dynamic time warping ... 100
6.6.7 Beat quality assessment ... 108
6.6.8 Summary ... 110
6.7 Heart rate estimation methods ... 111
6.7.1 Beat counting ... 111
6.7.2 Fast Fourier Transform (FFT) ... 112
6.7.3 AR dominant pole ... 114
6.7.4 AR pole cancellation (ARk) ... 116
6.7.5 Summary ... 118
6.8 Data fusion ... 119
6.8.1 Simple data fusion ... 119
6.8.2 Single Kalman filter ... 119
6.8.3 Multiple Kalman filters ... 123
6.8.4 Summary ... 124
6.9 Results ... 125
6.9.1 Reference signals ... 125
6.9.2 Video analysis ... 126
6.9.3 Heart rate estimation ... 128
6.10 Discussion ... 129
6.10.1 Heart rate frequency maps ... 129
6.10.2 Continuous heart rate monitoring ... 131
6.11 Conclusion ... 133

7 Reference respiratory rate **135**
7.1 Introduction ... 135
7.2 Respiratory rate provided by Equivital LifeMonitor ... 136
7.2.1 The reference source signals ... 136
7.2.2 Reference respiratory rate comparison ... 138
7.2.3 Conclusion ... 140
7.3 Overview of the estimation process ... 141
7.4 Respiration signal extraction from ECG ... 142
7.4.1 Overview ... 142

7.4.2 Beat detection 142
7.4.3 Beat quality assessment 146
7.4.4 Respiration signal extraction 150
7.4.5 Conclusion 153
7.5 Respiratory rate estimation 154
7.5.1 Overview 154
7.5.2 Preprocessing 154
7.5.3 Estimation methods 156
7.5.4 Combined respiratory rate 158
7.6 Results 159
7.7 Conclusion 160

8 Respiratory rate estimation using a video camera 161
8.1 Introduction 161
8.2 Overview of the estimation process 161
8.3 Regions of interest for respiratory rate estimation 163
8.3.1 Physiological video maps 163
8.3.2 Regions of interest selection 164
8.4 Respiration signal extraction from the face 165
8.4.1 Overview 165
8.4.2 Data windowing 165
8.4.3 Respiratory-Induced Amplitude Variation (RIAV) 165
8.4.4 Principal Component Analysis (PCA) 166
8.4.5 Respiration signal quality assessment 170
8.4.6 Conclusion 171
8.5 Respiration signal extraction from the upper torso 172
8.5.1 Overview 172
8.5.2 Data windowing 172
8.5.3 Respiration signal extraction 172
8.5.4 Respiration signal quality assessment 173
8.5.5 Conclusion 174
8.6 Respiratory rate estimation 174
8.7 Results 174
8.8 Discussion 175
8.9 Conclusion 176

9 Towards SpO_2 estimation using a video camera 177
9.1 Introduction 177
9.2 Overview of the estimation process 177
9.3 Reference SpO_2 179
9.3.1 Dual PPG monitoring during haemodialysis 179
9.3.2 Computing the reference SpO_2 184
9.3.3 Results 185
9.3.4 Conclusion 186
9.4 Skin colour analysis 187
9.4.1 The choice of wavelengths in a pulse oximeter 187
9.4.2 Colour reproduction in digital cameras 188
9.4.3 Colour balancing 190
9.4.4 Colour spaces for skin 193
9.5 Identifying changes in peripheral oxygen saturation 195

9.5.1 PPGi signal extraction 196
9.5.2 Computing the ratio of normalised signal amplitudes 197
9.5.3 SpO_2 calibration 198
9.6 Results 199
9.7 Discussion 201
9.8 Conclusion 203

10 Real-time implementation **205**
10.1 Introduction 205
10.2 Software architecture 205
10.2.1 Supported platforms 206
10.2.2 Programming languages and external libraries used 206
10.3 Extensions to the data acquisition system 207
10.3.1 Pulse oximeter capture 208
10.3.2 Video capture 208
10.4 User interface 208
10.4.1 Video rendering with OpenGL 209
10.5 Real-time vital-sign estimation 209
10.5.1 Image processing 210
10.5.2 Signal processing algorithms 210
10.6 Real-time vital-sign estimation on mobile devices 210
10.6.1 Image and signal processing 211
10.7 Conclusion 211

11 Conclusions and future work **213**
11.1 Introduction 213
11.2 Clinical study 214
11.2.1 Video recording 214
11.2.2 Data storage 214
11.2.3 Time synchronisation 215
11.3 Computing reference signals 215
11.3.1 Dual PPG monitoring during haemodialysis 215
11.3.2 Continuous monitoring of respiratory rate 216
11.4 Non-contact vital-sign estimation summary 217
11.4.1 Heart rate 217
11.4.2 Respiratory rate 219
11.4.3 Towards SpO_2 219
11.4.4 Real-time implementation 220
11.5 Future work 221
11.5.1 Patient identification 221
11.5.2 Body and skin segmentation 221
11.5.3 Physiological video maps 222
11.5.4 Continuous non-contact vital-sign estimation 222
11.6 Concluding remarks 222

A Autoregressive Modelling **223**

B Bayesian change point detection **225**

Bibliography **229**

List of figures

3.1 The respiratory cycle 17
3.2 Oxygen diffusion to blood 18
3.3 The circulatory system 21
3.4 The oxyhaemoglobin dissociation curve in a theoretically healthy subject 24
3.5 Extinction coefficients for various chemical compounds in the visible and infra-red spectrum 26
3.6 Transmission of light and the Beer-Lambert law 27
3.7 Oxyhaemoglobin (HbO_2) and deoxygenated haemoglobin (Hb) absorption spectra 29
3.8 The photoplethysmogram waveform 30
3.9 Simplified model of the pulsatile arterial blood in the PPG waveform 31
3.10 Methods of operation in pulse oximetry 32
3.11 Common places to measure oxygen saturation using a pulse oximeter 33
3.12 The photoplethysmogram 35
3.13 Modulations of the PPG waveform due to respiration 36

4.1 Colour filter layers in an imaging sensor 46
4.2 The CCD imaging sensor 46
4.3 The CMOS imaging sensor 47
4.4 The imaging pipeline 48
4.5 PPG signals captured simultaneously from a camera and a pulse oximeter 49
4.6 In-vitro arterial and venous phantom regions captured by an imaging sensor 50

5.1 A generic haemodialysis session 56
5.2 Typical dialysis data collection set-up 59
5.3 Hidalgo's Equivital EQ02 LifeMonitor 60
5.4 Overall software architecture for the data acquisition system 61
5.5 The Nonin 4100 pulse oximeter 62
5.6 The PointGrey Grasshopper2 video camera 63
5.7 Video storage cluster diagram 65
5.8 Distribution of physiological values in the dialysis study 67
5.9 Sample 4-hour time series collected for a typical dialysis session 68
5.10 Ear and finger PPG comparison 69

6.1 Overview of the heart rate estimation process. 72
6.2 Comparison of PPG waveform recordings from the ear and finger pulse oximeters . 73
6.3 Time delay comparison between the two reference pulse oximeters 74
6.4 Agreement between the original heart rate values provided by the two pulse oximeters 76
6.5 Computing the time alignment for the reference heart rate 78
6.6 Agreement between the heart rate values provided by the two pulse oximeters ... 80
6.7 Image examples from typical video recordings from the Renal Unit 81
6.8 Overview of the video analysis phase 82
6.9 Examples of some of the rectangular features used by the face detector 83
6.10 Computing the sum of the pixel intensities in a region from the integral image 84
6.11 Face detection for typical video recordings from the Renal Unit 84

6.12 Face registration examples 85
6.13 Face tracking example for a patient going from frontal face to right profile 86
6.14 Activity Index for a typical 4-hour dialysis session 87
6.15 Examples of rejected dialysis sessions 88
6.16 Normality test for the histograms from the subject regions of interest 89
6.17 Mean, median and variance from the ROI on the patient's face 89
6.18 Location and size of the subject region of Interest 90
6.19 Time series extraction from regions of interest of different size 91
6.20 PPGi extraction from a subject region of interest during a quiet period 91
6.21 Location and size of the grid for each subject region of interest 92
6.22 Background region of interest selection 93
6.23 Beat-by-beat signal quality assessment overview 94
6.24 Bayesian change point detection example 95
6.25 Output of the Bayesian change point detection algorithm 96
6.26 Band pass digital filter design 96
6.27 Steps to compute the Boxed Slope Sum Function (BSSF) 97
6.28 Peak and onset detection for a sample PPGi time series 98
6.29 Computing a valid template from a PPGi window 99
6.30 Computing an invalid template from a PPGi window 100
6.31 Local distance matrix computation for DTW 101
6.32 Time alignment for multi scale Dynamic Time Warping 102
6.33 Time alignment and accumulated cost matrix examples 103
6.34 Variations in amplitude for a typical PPGi signal 104
6.35 Pre-processing steps for Dynamic Time Warping 104
6.36 Comparison of distance measurements between a point and a line segment 105
6.37 Comparison of the piece-wise linear approximation algorithm between the PPGi signal and the window template 106
6.38 Local path constraints for DTW 107
6.39 Optimal path computed by the DTW algorithm using several error thresholds 108
6.40 Signal quality metrics for a sample good-quality beat 109
6.41 Signal quality metrics for a sample poor-quality beat 109
6.42 The output of the beat quality assessment for a sample PPGi window 110
6.43 Heart rate estimation for a 15-second good-quality PPGi window using the beat count method 111
6.44 Heart rate estimation for a good-quality PPGi signal using the beat count method . 112
6.45 Heart rate estimation for a 15-second good-quality PPGi window using the FFT method 112
6.46 Heart rate estimation for a 15-second poor-quality PPGi window using the FFT method 113
6.47 Heart rate estimation for a good-quality PPGi signal using the FFT method 113
6.48 Frequency resolution for the Discrete Fourier Transform 114
6.49 Computing the dominant pole for a typical 15-second good-quality window 115
6.50 Heart rate estimation for a good-quality PPGi signal using the AR dominant pole method 116
6.51 The autoregressive model pole cancellation method 117
6.52 Heart rate estimation for a good-quality PPGi signal using the ARk method 118
6.53 Summary of heart rate estimation using all the proposed methods on a single region of interest 118
6.54 Simple data fusion technique for heart rate estimation 120
6.55 1D Kalman filter data fusion technique for heart rate estimation 123

6.56 Multiple Kalman filter data fusion technique for heart rate estimation 124
6.57 Summary of the proposed data fusion technique for heart rate estimation 125
6.58 Comparison of the heart rate estimates provided by the pulse oximeters 126
6.59 Examples of rejected dialysis sessions 127
6.60 Heart rate estimation result for a sample 20-minute period 127
6.61 Agreement between the heart rate estimates from the camera and the reference values computed from the two pulse oximeters 128
6.62 Computing the dominant pole for a typical 15-second good quality window 130
6.63 Heart rate maps for a typical 15-second video window using 25x25 pixel regions of interest 130
6.64 Estimation gaps in heart rate estimation using the data fusion method 131
6.65 Example of heart rate physiological frequency map 132

7.1 Respiratory rate estimates provided by the Hidalgo's chest belt 136
7.2 ECG recording by Hidalgo's LifeMonitor during a single cardiac cycle 137
7.3 Respiratory rate estimates provided by the Hidalgo using the ECG 138
7.4 The agreement between the respiratory rate estimates provided by Hidalgo using the ECG and chest belt 139
7.5 Comparison of the two respiratory rate estimates provided by Hidalgo for a typical dialysis session 139
7.6 The agreement between the respiratory rate estimates provided by Hidalgo when the values are limited to the physiological range for adults 140
7.7 Overview of the respiratory rate estimation process 141
7.8 Digital filtering of the ECG waveform 142
7.9 Processing steps for the QRS detector based on digital filtering method 143
7.10 Digital filtering of the ECG waveform for the length transform method 144
7.11 Processing steps for the QRS detector based on the length transform method 145
7.12 Comparison of the bSQI and iSQI signal quality metrics 147
7.13 The kurtosis quality metric (kSQI) 148
7.14 Quality metric base on spectral distribution (sSQI) 149
7.15 Signal quality assessment for a sample two-channel ECG record 150
7.16 Respiration signal extraction comparison for a good-quality ECG signal 151
7.17 Extracting the respiratory signal from a good-quality ECG waveform 152
7.18 Extracting the respiratory signal from a ECG waveform using the EDR method 152
7.19 Extracting the respiratory signal from a an inverted ECG waveform using the EDR method 153
7.20 Digital filtering of the respiration waveform from the chest belt 154
7.21 Finding the peaks and troughs from a respiration signal extracted from the ECG using the EDR method 155
7.22 Spectral purity index for a respiration signal from the chest belt 156
7.23 Respiratory rate estimation using the breath counting method for a respiration signal extracted from the ECG using the EDR algorithm 157
7.24 Respiratory rate estimation using the AR method for a respiration signal extracted from the chest belt 157
7.25 Respiratory rate estimation from a respiration signal from the chest belt 158
7.26 The agreement between the respiratory rate estimates computed from the ECG and chest belt for the overall data set 159

8.1 Overview of the respiratory rate estimation process 162
8.2 Physiological video maps for heart rate and respiratory rate 163

8.3 Respiratory rate maps for a typical 15-second video window using 25x25 pixel regions of interest 164
8.4 Measuring the beat amplitude from a typical PPGi waveform 165
8.5 Extracting a respiration signal from a typical PPGi window 166
8.6 The first 5 principal components for a typical 30-second respiration signal 170
8.7 Quality assessment of a valid PCA component 171
8.8 Valid PCA components for a typical 30-second window 172
8.9 Extracting a respiration signal from the patient's upper torso 173
8.10 Assessing the quality of the first two PCA components for a typical 30-second window 173
8.11 Agreement between the respiratory rate estimates from the camera and the reference values 175

9.1 Overview of the process for identifying changes in peripheral oxygen saturation ... 178
9.2 A dialysis session for a 54-year old male for which there is a good agreement between the SpO_2 estimates 179
9.3 Recording from a 56-year old male with seven months on dialysis, diagnosed with Diabetic Nephropathy, Type 2 Diabetes and Hypertension 180
9.4 Recording from an 85-year old male with two and a half years on dialysis, diagnosed with Hypertension and given Valsartan and Amlodipine 181
9.5 Recording from a 75-year old female with two months on dialysis, diagnosed with diabetes and hypertension 182
9.6 Agreement between the original SpO_2 values provided by the two reference pulse oximeters for the entire dialysis dataset 183
9.7 Agreement between the original SpO_2 values provided by the two reference pulse oximeters for the entire dialysis dataset 184
9.8 Four-stage process for computing the reference SpO_2 values from the two pulse oximeters 185
9.9 Agreement between the SpO_2 values provided by the two pulse oximeters for the entire dialysis dataset 186
9.10 Typical wavelengths chosen in pulse oximetry 187
9.11 Different absorption characteristics of the HbO_2 and Hb in the green and red parts of the spectrum 188
9.12 PointGrey Grasshopper2 video camera system 188
9.13 The spectral irradiance for different common light sources 189
9.14 The effect on the skin colour recorded by a video camera fro changing illumination conditions 190
9.15 The CIE XYZ colour space 191
9.16 The result of the colour balancing algorithms for a video recording with small blue hues 192
9.17 The result of the colour balancing algorithms for a video recording with strong green and red components 193
9.18 Colour balancing the input RGB images 194
9.19 Comparison of the extraction of a pulsatile signal from a single region of interest .. 195
9.20 PPGi signal extraction from high-quality regions of interest 196
9.21 PPGi template computed for a PPGi window using the red and green colour channels 197
9.22 Computing the AC and DC components for a PPGi signal for the red and green colour channels 197
9.23 Calibration from the ratio of normalised signal amplitudes 198
9.24 Agreement between the SpO_2 estimates provided by the two reference pulse oximeters 199

9.25 Comparison between the SpO_2 estimates from the camera and the reference SpO_2 values for sample desaturation segments 200
9.26 Agreement between the SpO_2 estimates from the video camera and the reference values 200
9.27 Relationship between the average ratio R from the normalised signal amplitudes from the red and green colour channels and the linear regression parameters 201
9.28 Comparison between the SpO_2 estimates from the camera and the reference SpO_2 values 202
9.29 Relationship between the average ratio R from the normalised signal amplitudes from the red and green colour channels and the linear regression parameters 203

10.1 Overall software architecture for the real-time vital-sign estimation system..... 206
10.2 Real-time vital-sign estimation running on the Apple Macbook Pro..... 209
10.3 Real-time heart rate and respiratory rate estimation running on a mobile phone. .. 211

A.1 Autoregressive Model computation for a typical 15-second window 224

B.1 Example of a step change in the PPGi time series 225

List of tables

5.1 Status flags for data format 7 for the Nonin pulse oximeter 63
5.2 General features and specification for the PointGrey Grasshopper2 camera 64
5.3 A sub-set of camera configuration options for Point Grey cameras..... 64
5.4 Demographics for the population in the clinical study 66

6.1 Comparison of the heart rate estimates provided by the reference pulse oximeters 79
6.2 Heart rate estimation results 129

7.1 Error for the respiratory rate estimates provided by the Hidalgo LifeMonitor 140
7.2 Respiratory rate estimation comparison between the ECG and chest belt for the overall dataset 159

8.1 Comparison between the respiratory rate estimates from the video camera and the reference values 175

9.1 Comparison of the SpO_2 estimates provided by the two reference pulse oximeters 186
9.2 SpO_2 estimation results between the video camera and reference values provided by the pulse oximeters 201

10.1 Compilers and third party dependency libraries required to build the real-time vital-sign estimation system 207

List of abbreviations

AR	Autoregressive model.
CCD	Charged—coupled device, an imaging sensor technology.
CMOS	Complementary metal—oxide semiconductor, an imaging sensor technology.
CO_2	Carbon dioxide.
CW	Constant-frequency continuous-wave, a radar technology.
DTW	Dynamic Time Warping.
ECG	Electrocardiogram.
FFT	Fast Fourier Transform.
Hb	Deoxygenated haemoglobin.
$HbCO_2$	Carboxyhaemoglobin.
HbO_2	Oxyhaemoglobin.
HR	Heart rate.
ICA	Independent Component Analysis.
ICU	Intensive Care Unit.
IP	Impedance Pneumography.
LED	Light-emitting diode.
MAD	Mean absolute deviation.
MAE	Mean absolute error.
$metHb$	Methaemoglobin.
NCC	Normalised Colour Coordinate.
PCO_2	Partial pressure of carbon dioxide.
PLA	Piece-wise linear approximation.
PO_2	Partial pressure of oxygen.
PPG	The photoplethysmogram waveform.
PPGi	The photoplethysmographic imaging waveform.
RGB	Red, green and blue. The colour channels recorded by a camera.
ROI	Regions of interest.

ROI$_R$ Background region of interest.

ROI$_S$ Subject region of interest.

RR Respiratory rate.

SaO_2 Arterial oxygen saturation.

SpO_2 Peripheral oxygen saturation, as measured by a pulse oximeter.

SQI Signal quality index.

$sulfHb$ sulfhaemoglobin.

UWB Ultra-Wideband, a radar technology.

Chapter 1

Introduction

It is widely reported, notably by the World Health Organisation [314], that global changes in population demographics, the increased incidence of chronic diseases and the unmet needs for more personalised care are trends that need new integrated approaches to health and social care.

As a result of recent advances in medicine, the implementation of new technologies in healthcare, falling birth rates, access to better nutrition and improved living conditions, the average age of the population in most European countries is increasing and the proportion of older people is growing. It is estimated that the proportion of the population aged 65 or above will increase from 17% in 2007 to 21% by 2020 and 28% by 2040 in Europe [215]. With an increasingly ageing population, healthcare systems need to change from the treatment of acute life-threatening conditions to the management of a wide range of chronic diseases [12].

The future supply of informal care may also become insufficient because family structures in modern societies are changing, increasing the proportion of people living alone [75]. Policy-makers and healthcare providers are constantly being challenged to meet the demand for services by an ever-increasing numbers of patients, insufficient human and capital resources, leading to high workloads. To maintain high-quality care, frequent visits are required to establish whether patients are deteriorating or encountering problems at home.

The European Commission favours changing today's reactive model of care to an integrated framework that facilitates individuals and communities to continuously monitor and manage their health from the comfort of their homes or work places [138]. To support these efforts, the European Commission has funded several projects [227], including the implementation of electronic health record systems and the establishment of regional health information networks that connect all points of care such as hospitals, primary care, labs, pharmacies and homes, and support all stages of prevention, diagnosis, treatment and follow-up.

As the pressure to change the way in which healthcare is delivered has increased, the volume and complexity of the information have also risen [174]. New technologies are needed to ensure the maintenance of the patients or of the general population at home, without unnecessary and frequent visits from health providers, thereby avoiding costly unscheduled hospital admissions.

1.1 Monitoring of vital signs

The standard vital signs include temperature, heart rate (HR), respiratory rate (RR), blood pressure and, when appropriate, peripheral oxygen saturation (SpO_2). The routine measurement and interpretation of these vital signs is a core component of the physiological assessment of most patients [109, 265], as they can provide critical information about the underlying state of their health.

Of the standard vital signs, heart rate, respiratory rate and SpO_2 are of main interest for the research carried out for this thesis. A brief description of these three vital signs is presented in the next sections. A more in-depth study of the relevant physiology, the current technologies developed to monitor physiological parameters and common algorithms for vital-sign estimation are discussed in the next chapters.

1.1.1 Heart rate

Heart rate is a measure of the rate at which the heart beats. Appropriate monitoring of the heart's pumping mechanism is of vital importance as with each heart beat, blood is sent throughout the body carrying gases, nutrients, hormones and other substances used in metabolic processes by the cells.

Heart rate is measured in beats per minute ($\mathrm{beats/min}$). Its normal range, at rest, varies with age from 110-160 $\mathrm{beats/min}$ at birth to approximately 60-100 $\mathrm{beats/min}$ by late adolescence [93, 257]. During adulthood, heart rate varies widely between individuals depending on several factors such as physical activity, body composition, stress levels or disease. An unusual high heart rate (known as tachycardia) or low rate (known as bradycardia) may indicate an underlying clinical problem.

The electrocardiogram (ECG) is the gold standard method for measuring heart rate. It records the electrical activity of the heart with electrodes placed in contact with the patient's skin. Typical configurations for recording the ECG, depending on the clinical information desired, are either 3 electrodes on the chest, or 10 to 12 electrodes placed at various locations on the patient's upper body [64].

It is also possible to measure heart rate by using a pulse oximeter, a non-invasive optical sensor usually positioned on the fingers, toes or ear lobes [287, 328]. Pulse oximetry has made significant contributions to the non-invasive monitoring of physiological signals in a wide variety of clinical situations (see chapter 3). Pulse oximeters are regularly used in primary care [328], the emergency department [85] and at home [337].

Manual measurement of heart rate is common in clinical practice during initial assessment, either by using a stethoscope, or by palpating the wrist (radial artery) or the neck (carotid artery) [92]. These manual measurements have been shown to underestimate the true heart rate [93]. During a clinical competency assessment in a simulated environment, only 25% of nurses could palpate the heart rate [255].

1.1.2 Respiratory Rate

The respiratory system is a complex biological structure composed of multiple organs that enables us to breathe. It filters and transports air into the lungs, where gas exchange occurs through a process called diffusion, allowing the movement of oxygen into the lungs and the release of carbon dioxide. The complete process of gas exchange between the atmosphere and body cells is called respiration. The actions responsible for these air movements are called inspiration and expiration. The sequence beginning with inspiration, followed by the gas exchange in the lungs and ending with expiration, is known as the respiratory cycle. The frequency at which the respiratory cycle occurs is the respiratory rate.

Respiratory rate is recognised as an important vital sign since it has been found to be predictive of lower respiratory tract infections [294] and the evaluation of the severity of pneumonia [188]. It is also a risk factor for unplanned hospital readmissions [204] and mortality risk assessment for paediatric patients in an intensive care unit (ICU) [262]. Respiratory rate is measured in breaths per minute ($\mathrm{breaths/min}$). Similarly to heart rate, normal respiratory rate varies with age, with 30-40 $\mathrm{breaths/min}$ at birth, reducing to 15-20 $\mathrm{breaths/min}$ by late adolescence [257].

Inspiration and expiration movements cause changes in thoracic pressure leading to variations in the air volume within the lungs, which in turn cause expansion and contraction of the thorax that can easily be seen by eye. This effect is used in hospital settings such as general wards or emergency departments, where respiratory rate is typically measured by the clinical staff counting chest wall movements [293]. Although this method is simple and does not interfere with the breathing mechanisms, human observation of respiratory rate is time-consuming and the result is not always accurate, with errors reported up to 34% [172, 196, 289]. To obtain reasonably accurate results, Simoes et al [303] discussed the need to count chest wall movements for at least 60 seconds; in clinical practice however, counting periods of just 15 seconds or at most 30 seconds are more common [293]. The training of clinical staff to adequately measure respiratory rate is also an issue. Pearson et al [255] reported that only 67% of nurses were able to measure respiratory rate to within 10 $\mathrm{breaths/min}$ during an investigation in a simulated environment.

The variation in thorax volume when a patient is breathing can also be measured using a strain-gauge transducer typically held by elastic bands around the chest. The respiratory movements result in resistance changes of the strain gauge that correspond to the respiratory activity [96, 161]. However, the accuracy of measurements can be affected by body position, variability in the signal recorded between patients and even in measurements of the same patient at different times [21].

In addition to measuring the changes in circumference of the thoracic cavity, there are a number of other methods that estimate respiratory rate from variations in the airflow during expiration. These methods typically use a face mask, a nasal cannula or a mouth piece to collect the air. During expiration, airflow can be detected because the expired air is warmer, has higher humidity and contains more carbon dioxide than inspired air [96]. However, measuring airflow can affect respiratory rate. Baird et al [21] reported that breathing into a face mask resulted in an increased depth of breathing and a decrease in its frequency [21].

Impedance pneumography (IP) is a commonly-used technique to monitor a patient's respiratory rate in a hospital setting. It is implemented by using either two electrodes or four electrodes. The objective of this technique is to measure changes in the electrical impedance of the individual's thorax caused by the increased resistivity of lungs filled with air during inspiration. The electrical impedance is measured by injecting a high-frequency current into the thorax through surface electrodes [249]. A possible reason for its popularity in a hospital setting is that it can be measured simultaneously with the ECG, using the same electrodes, thereby eliminating the need to attach extra transducers [279].

1.1.3 Peripheral oxygen saturation (SpO_2)

Oxygen is a chemical substance essential to the functioning of the cells in the human body and, therefore, necessary to sustain life. It is important to monitor if all organs are receiving a sufficient supply of oxygen as it is being delivered to the body. The measurement of blood oxygenation, also known as oxygen saturation, is an important indicator of a patient's health. A prolonged lack of oxygen can rapidly cause permanent damage to cell tissue, leaving patients with devastating neurological handicaps, and has the potential to be life-threatening if cells having high metabolic rate in organs such as the brain, heart or the central nervous system are damaged (see chapter 3).

Several methods have been developed to analyse oxygen delivery. In principle, oxygen saturation can be measured in vitro (such as in haemolysed blood samples in a cuvette) or in vivo (measurements taken directly from the human body) [81, 253]. In-vitro chemical methods were developed during the second half of the twentieth century and are still in use for blood gas analysis as gold standards for oxygen content measurement. Chemically-based methods determine the oxygen content of blood from a sample by using chemical reactions to remove the oxygen from the blood. One of the disadvantages of these methods is that they can be slow and, therefore, measurements are performed infrequently, usually only once or twice a day in the intensive care unit (ICU).

The advent of oximetry as an optical method for measuring oxygen content in the blood developed rapidly during the 1930s and 1940s. Based on spectrophotometry, oximetry measures the amount of light absorbed by the various molecules that bind to oxygen in the blood [353]. Early oximeters were calibrated empirically and achieved an acceptable accuracy, but they were too cumbersome and expensive to use in a clinical environment [253].

Pulse oximetry revolutionised clinical practice, making it possible to measure, non-invasively and continuously, arterial oxygen saturation in the peripheral arteries [253]. Peripheral oxygen saturation gives a measure of the amount of oxygen carried by the arterial blood. It is given by the percentage of haemoglobin that is bound to oxygen with respect to the total haemoglobin. Unlike heart rate and respiratory rate, SpO_2 does not vary with age and a value between 95% and 100% is considered normal in both children and adults [257]. The common measurement sites for pulse oximetry are the finger, toe and ear lobe. Non-invasive and continuous monitoring of oxygen delivery is an invaluable aid to the physician as an indicator of the patient's health [81]. Pulse oximeters are small, cheap, portable, non-invasive and do not require any prior training for their use.

1.2 Clinical need

The establishment of protocols of care to track changes in vital signs (such as heart rate, respiratory rate or peripheral oxygen saturation) at home is essential for the early identification of patient deterioration and the assessment of response to treatment. Ultimately the aim is to reduce the number of unplanned hospital admissions for patients with moderate to severe long-term conditions such as Chronic Obstructive Pulmonary Disease (COPD) or heart failure.

Current monitoring systems available to track changes in the vital signs of patients at home o in the hospital require contact with the subject. Most patients requiring regular monitoring find the probes difficult to attach and use properly. The process of recording vital signs, even if it only takes a few minutes, becomes burdensome as it usually has to be performed on a daily basis. The low compliance of patients with wearing sensors ("wearables") is an obstacle to successful monitoring.

The ideal technology to estimate vital signs would involve sensors with no direct contact with the patient ("non-contact sensing"). Non-contact vital-sign monitoring provides several advantages over traditional methods because no subject participation is required to set the equipment up. It requires no skin preparation, causes no skin irritation, decreases the risk of infection, and has the potential to be seamlessly integrated into the patient's lifestyle.

The possibility of measuring multiple vital signs from a single non-contact sensor is also beneficial as it would improve user acceptability by reducing the burden on the patient to wear continuous vital-sign monitoring hardware, reducing cabling and decreasing the costs of implementing and maintaining new equipment.

Several technologies have been proposed for the non-contact monitoring of vital signs during the last decade. Chapter 2 discusses the most relevant techniques such as Doppler-based methods or ultra-wideband radar. With the cost of digital video cameras continuing to decrease as the technology becomes more ubiquitous, research in non-contact vital-sign monitoring has recently expanded through the use of off-the-shelf video cameras. Video cameras can be found in laptops, mobile phones, set-top boxes and television sets in patients' living rooms, opening up new possibilities for the monitoring of vital signs.

1.3 Objectives

The major objectives for the research carried out for this thesis, and as a consequence of the design of the clinical study that supported it, are:

- To investigate how accurately vital signs such as heart rate (HR), respiratory rate (RR) and changes in peripheral oxygen saturation (SpO_2) can be estimated using a standard colour camera and regular ambient light sources.
- To obtain calibration data for the development of vital-sign estimation algorithms from a hospital setting under real-world conditions, without interfering with regular patient care.

- The development of a reference video database from a hospital environment that includes a wide range for each physiological parameter to be estimated (HR, RR and SpO_2).
- To develop a non-contact *real-time* vital-sign estimation software package that is flexible, expandable and can run on both desktop PCs and mobile devices.

1.4 Contributions of thesis

Most of the current work in video-based non-contact vital-sign monitoring has so far been performed over short time periods (typically up to a couple of minutes per recording) and under tightly controlled conditions. The effects of light sources, subject motion and other external factors have been controlled and reduced. The recordings have been carried out from relatively young (usually between 20 and 30 years old) and healthy subjects volunteering for the studies.

The primary contribution of this thesis is the robust estimation of vital signs in a real hospital scenario, with sick patients being monitored without interfering with their regular care by the clinical staff. Videos were recorded for close to 370 hours from patients undergoing haemodialysis treatment in the Renal Unit of the Churchill Hospital in Oxford. The recordings include a variety of patients of different ages and degrees of sickness, with a wide range of values for each physiological variable to be estimated (HR, RR and SpO_2).

Subject motion presented a non-trivial challenge during the video recordings. Methods are proposed to find time periods during which the patients are relatively stable so that reliable vital-sign estimation is possible. Tracking the location of the patients during video recording is also important. Methods from the computer vision community are applied to register landmark features from the patients and track the location of regions of interest from which vital signs can be estimated.

Another factors to consider are the aliased components from artificial light sources such as fluorescent tubes, incandescent light bulbs or TV screens that affect the estimation process. We propose a method for cancelling frequency components using the poles of an autoregressive model from regions of interest in the human face and from the background [65].

Physiological maps of the spatial distribution of heart rate and respiratory rate data are constructed by processing and colour coding the parameters from the autoregressive models across the entire image frames [66]. The proposed methods are used to select the regions of interest that contain frequency content in the physiological range of interest and to identify areas in the video from which vital signs can be estimated most accurately. From the fact that the heart rate is a periodic signal derived from skin regions, the proposed physiological maps are extended for the detection of subjects in a video image.

Another major contribution of the thesis is the development of real-time software for time series acquisition and vital-sign estimation. The software was developed following appropriate medical and safety standards that enabled it to run on workstations and laptops with Linux, Microsoft Windows or Apple OS X operating systems. The software also supports the real-time estimation of vital signs on mobile devices such as the iPad or the iPhone.

1.5 Outline of thesis

Chapter 2 discusses previous research in the area of non-contact vital-sign monitoring, including techniques such as Doppler and ultra-wideband radar. Chapter 3 provides an overview of the relevant physiology background of oxygen monitoring and an in-depth description of the theory behind pulse oximetry. After introducing relevant background information on video and imaging technologies, a thorough description of non-contact methods for vital-sign monitoring using video cameras is presented in chapter 4. The clinical study undertaken as part of this thesis is presented in Chapter 5 with a detailed description of the patient population, the data collection process and an overview of the main vital signs used as the reference data.

The algorithms to estimate heart rate developed in this thesis are described in chapter 6. Chapter 7 analyses the validity of the respiratory rate values provided by the reference devices used in the clinical study and postulates that new reference respiratory rate values need to be computed as gold standard measures. The proposed algorithms to estimate respiratory rate from video cameras are described in detail in chapter 8. Chapter 9 discusses initial work towards the detection of changes in oxygen saturation using video cameras. Chapter 10 discusses the design and the implementation of real-time software for vital-sign estimation in both desktop workstations and mobile devices. Conclusions from the research carried out in this thesis and an outline for future work are presented in chapter 11.

Chapter 2

Non-contact vital sign monitoring

2.1 Introduction

Non-contact cardiac and respiratory activity monitoring is a challenging task. During the last decade, several techniques have been proposed for the non-contact monitoring of vital signs, providing several advantages over traditional approaches. This chapter presents a review of the recent work described in the literature.

2.2 Overview of methods

Methods for non-contact vital-sign monitoring can be divided into the following categories:

Electrical : A coordinated pattern of propagation of a wave of electrical current passing through the entire heart results in a measurable change in potential difference on the body surface of a patient. This resultant signal is known as the electrocardiogram (ECG) [64]. Non-contact methods for recording the ECG are described in the next section.

Body motion : Movements of the body associated with cardiac contraction and ejection of blood and with the deceleration of blood flow through the large vessels, provide information about the overall performance of the circulatory system. Methods include radar, laser and video-based motion analysis (see below).

Temperature : A thermal imaging camera measures the radiation emitted by objects in the long-infrared range of the electromagnetic spectrum, between 9 and 14 microns. Since the amount of radiation emitted by an object increases with temperature, thermography can estimate the distribution and changes in body temperature [131].

Perfusion : The variations in blood volume in a body segment during the cardiac cycle, modulate the reflection or transmission of light. When the skin is perfused, the pulsatile nature of the blood flow causes a rhythmic expansion and contraction of the arterial blood vessels in time with the heart beat. Optical methods to measure these changes include the traditional photoplethysmogram (PPG) and photoplethysmographic imaging (PPGi).

Photoplethysmography, photoplethysmographic imaging (PPGi) and the use of thermal video cameras to estimate physiological signs are discussed in more detail in the next chapters.

2.3 Non-contact ECG

A diagnostic-quality ECG monitoring system is typically based on 12-lead electrodes in good electrical contact with the skin, which is typically accomplished with a high-conductivity gel, at different positions on the chest. ECG electrodes can potentially increase the mechanical and biomedical stress of dermal tissue, reducing its acceptance by patients for long-term monitoring (over several hours).

Capacitive coupling electrodes, first introduced in the 1960s and 1970s [193, 273], do not require direct contact with the body [164, 189]. Without direct contact with the skin, some of the issues associated with adhesive electrodes such as skin infection and signal deterioration can be avoided, enabling the long-term monitoring of cardiac activity [59, 309].

ECG monitoring using capacitive-coupled electrodes can be embedded in objects of daily use at home or at the workplace, such as chairs, wheelchairs or beds [264]. Kim et al [163] proposed an ECG recording system on a toilet seat. An ECG recording device was developed for an office chair by Steffen et al [313] with electrodes in the seat and backrest. One of the major challenges in the design of capacitive coupling electrodes is the high contact impedance due to the indirect contact and the capacitive mismatch caused by motion artefacts [58].

The Ohio Transportation Consortium conducted pilot investigations on the development of an in-vehicle measurement system that not only monitors common physiological signals such as heart rate and respiratory rate, but can also record the ECG signals from an effective distance of up to 30 cm away from the human body [368]. The system detects the potential difference on the human body through capacitive coupling using a sensor made from conductive polymer materials.

In addition, capacitive sensing can also be used for other vital signs such as respiratory rate. Hoffman et al [128] developed a system based on textile force sensors weaved into clothing for the long-term monitoring of respiration. Monitoring of respiratory rate with an array of capacitive electrical field sensors under a sleeping mattress has also been developed [142, 349].

Research in non-contact ECG has increased over the last decade in diverse areas ranging from fitness monitoring to the management of chronic health conditions [59, 248, 350]. However, there still exist major challenges to overcome. A triboelectric charge is generated from the rubbing of the ECG electrodes when attached to clothing, an effect which is very difficult to predict. Electromagnetic noise from mains power, motion-related artefacts and other common-mode electromagnetic interference continue to be challenging problems to solve in practical applications [256].

Although a few attempts to increase the distance from the ECG electrodes to the subject being monitored have been proposed [368], most of the research in non-contact ECG is performed with electrodes located at a few millimetres from the individuals [57, 316], usually embedded into fabric or clothing. This constraint may well delay the widespread adoption of non-contact ECG monitoring.

2.4 Radar

Radar is an object-detection system that uses radio waves to determine the range, altitude, direction, or speed of objects. It was initially designed during the late 19^{th} century and greatly developed during World War II to detect aircraft, ships, guided missiles, motor vehicles, weather formations, and terrain [27, 69, 168, 250, 329].

In its basic form, a radar system has a transmit-receive switch, data control and processing elements. The transmitter emits a radio wave of known frequency and duration, called the radar signal, into free space in a predetermined direction. The radar signal propagates outwards, scattering from objects it encounters along the way. Part of the scattered signal is reflected back towards the receiver. Radar signals are reflected especially well by materials of high electrical conductivity, for example most metals, seawater and wet ground. The presence of the received signal can be detected and measurements of the object of interest can be made. Radar systems are able to operate during the day or at night, in different kinds of weather, can penetrate walls and have broad coverage.

For non-contact vital-sign monitoring, two main types of radar systems have been proposed: constant-frequency continuous-wave (CW) and ultra-wideband (UWB). In a CW radar system, a signal of known constant frequency is continuously transmitted towards the direction of the subject's chest from a transmitter usually located 1 or 2 metres away. As the chest moves away or towards the transmitter during inspiration and expiration respectively, the reflected radar signal changes frequency, as a result of the Doppler effect. The resulting signal modulated by the chest's periodic motion can be used, with appropriate baseband filters, to estimate respiratory rate. Heart rate can also be estimated by analysing smaller chest wall movements due to the changes in position and volume of the heart during a cardiac cycle. CW radar devices that are dedicated to take advantage of the Doppler effect are called Doppler radar systems.

Ultra-wideband radar systems transmit pulses with very short durations (less than 1 ns) and bandwidth greater than 0.5 GHz. UWB equipment is low cost and requires low power for its continuous operation, opening the possibility to design sensors that can be highly integrated, small and power-efficient. Section 2.4.2 discusses the advantages of ultra-wideband radar over conventional radar systems.

2.4.1 Vital-sign monitoring with CW radar

With methods based on constant-frequency continuous-wave-based methods, most of the motion in the radar's field of view comes from breathing and appears as the AC component of the reflected signal, while stationary objects in the background appear in the DC component and can be removed or cancelled out [322]. In practice, motion artefacts from the physical movement of the subject can result in interference in the radar's output signal and may even be at the same frequency as the heart rate or respiratory rate.

Mostafanezhad et al [233] presented a CW radar system which can determine the heart and respiratory rates of relatively still subjects without contact and through clothing or bedding. Ivashov et al [145], developed a CW radar system to derive respiratory rate from humans behind a wall.

Li [183] proposed a signal demodulation technique that cancels random body movements using multiple transceivers and antennas detecting signals from different body parts, improving the accuracy of CW radar systems.

Droitcour et al [84] designed a custom microwave radio for Doppler radar and further developed it into a single integrated CMOS chip system [83], making inexpensive mass production and miniaturisation of the system possible. The current system is reported to be able to estimate heart rate at ranges up to 1 m. The extraction of respiratory rate is reported as being reliable for up to 1.5 m. If the manufacturing costs could be decreased sufficiently, additional transceivers could be simultaneously used to overcome problems such as motion artefacts.

Fletcher and Han presented a low-cost dual-antenna radar system for vital-sign monitoring with potential applications to automotive driver safety systems, health monitoring, and security screening [95]. Although the use of multiple antennas helps reduce the background motion noise, the design is still only adequate for short-range recordings.

Several applications of Doppler-based systems have been investigated. Suzuki [319] developed a method to derive vital signs from subjects inside a moving ambulance using dual-frequency microwave radars. Uenoyama et al [336] designed a microwave antenna attached to the room ceiling to derive respiratory rate from patients in nursing care institutions. Heart rate estimation accuracy was reported to be improved by measuring it from the back of the body, under the bed during overnight studies [184]. Matsui et al [210] combined a Doppler blood-flow meter, microwave radar and a thermal camera to derive heart rate, respiratory rate and temperature to conduct the fast screening of airport passengers with infection. Yu and Li [370] designed an integrated low-cost hand-held apparatus for the detection of heart rate and respiratory rate using a radio frequency (RF) transceiver.

The published literature on Doppler radar systems for physiological monitoring states that subject motion remains a great challenge. The issue of penetrating non-metallic materials such as wood, plastic or thick mattresses commonly found in the home or in hospitals might be mitigated to some degree by using UWB radar technology (see below) with the addition of adaptive filters based on heart rate and respiratory frequencies such as those presented in [295].

2.4.2 Advantages of UWB radar

Ultra-wideband radar technology has several advantages for medical applications over conventional radar systems [6]:

Penetrating through obstacles : As opposed to ultrasound, which requires a line of sight and has a very short range of operation, radar systems use low-frequency radio frequency (RF) pulses with long wavelengths. UWB radar technology can penetrate effectively through different materials such as clothing or through obstacles such as walls and mattresses.

High precision : The UWB pulse has a very strong temporal and space resolving capability. For a 1 ns pulse, the multi-path resolving power is equal to 30 mm, which is suitable for localisation in medical applications.

Low electromagnetic radiation : Many studies have been published regarding the effects of animal and human exposure to radar, particularly UWB [63, 78, 147, 221, 296, 302, 345]. Research suggests that the health risks are potentially minimised due to the low level and non-ionizing nature of the radiation used. However, most research has been conducted for the 1 to 2 m range of distances between subject and radar system, without extensive analysis of the interference that can be caused by electronic equipment in a modern hospital setting or at home. Exposure to human-made sources of electromagnetic fields at short distances could also present a problem [360].

Power efficiency : Radar technologies such as UWB, contain very short pulses and use very low average power levels for radio transmission. A low processing energy consumption enables the usage of long-life battery-operated devices, thus making the technology suitable as a cost-effective solution for human vital-sign monitoring.

2.4.3 Vital-sign monitoring with UWB radar

The use of radar systems based on ultra-wideband to measure physiological signals dates back to the 1970s. Caro and Bloice [47] developed a radar transmitter and receiver operating in the 10 GHz range for the detection of apnoea in newborn infants, with a system located 2 to 5 cm above the incubator. Franks et al [99] compared respiratory rate estimation using a ultra-wideband (UWB) radar-based system with an air-filled mattress, under-mattress pressure sensor, a magnetometer and a capacitance-change detector. Respiration was monitored in infants for the first year of life during overnight recordings. Franks extended the work in [100] analysing overnight respiratory patterns and risk of sudden death in 25 high-risk and 42 low-risk infants. The recordings were analysed for respiratory frequency, total duration of apnoea, periodic breathing, duration of regular breathing, and amount of body movement.

McEwan, from the Lawrence Livermore National Laboratory, proposed an UWB radar system in which a large number of reflected pulses were averaged to produce a voltage that modulated an audio oscillator to produce a tone that corresponded to the heart motion [212] [213]. Staderini [312] further advanced McEwan's system by incorporating a model that considered the thickness, impedance, attenuation and wave speed of six tissue layers while the UWB pulse travelled through the human thorax from the skin to the heart.

Thiel et al [327] reported that UWB radar also has high spatial and temporal resolutions and does not have compatibility issues with established narrowband systems such as Magnetic Resonance Imaging (MRI). However some interference with global positioning systems (GPS) has been reported [127]. The frequency range from 100 MHz up to 10 GHz has been proposed for the combined use of UWB radar and MRI to monitor the motion of organs within the human body as well as obtaining images of internal structures to gain complementary information for improved functional diagnosis and imaging.

2.5 Laser-based techniques

Chest wall movements induced by the pumping action of the heart can be measured not only by radar systems, but also with light in the form of lasers [310]. Aubert et al [20] were among the first to design a system which measured cardiac-related displacement of the chest by recording the position of a laser dot on the body with a camera.

Research on laser-based systems for vital-sign monitoring expanded with the development of the Laser Doppler Vibrometer (LDV). By directing a laser beam onto a surface of interest, a LDV system can measure the vibration amplitude and frequency due to the motion of the surface. Morbiducci et al [231] used a LDV system to record the vibration of the chest wall and compute heart rate and heart rate variability. Cardiac-related skin velocity and displacement have also been recorded using lasers by Hong et al [130]. Pinto et al [259] used LDV to extract the carotid artery pulse waveform from adult volunteers simulating regular breathing and apnea.

Recent developments allow laser vibrometers to estimate heart beats, respiratory motion and gross physical activity, even in the presence of clothing [347]. Scalise et al [283] reported the measurement of respiratory rate in pre-term infants by using a laser Doppler vibrometer outside the incubator.

Another important method for physiological monitoring is the use of laser speckle contrast imaging (LSCI) as an optical imaging method with high spatial and temporal resolution that can be used to image dynamic changes in blood flow. The basic theory of laser speckle was developed in the 1960s and further developed during the 1970s and 1980s [39]. Typically, a LSCI system is composed of expensive scientific-grade instruments, Richards et al [271] proposed a low-cost laser speckle system using an off-the-shelf laser pointer and a consumer-grade webcam to image dynamic blood flow changes in a microfluid flow phantom and in vivo in a mouse model. Sirkis et al [304] further expanded a LSCI system to measure blood pulse wave velocity based on tracking the temporal changes of the pressure waveform produced from the reflected speckle patterns in four healthy human volunteers.

2.6 Conclusion

Most of the research in non-contact vital-sign monitoring reviewed in this chapter has been performed by recording data over short time periods, with relatively still and healthy subjects for whom physiological parameters do not vary greatly.

Most of the technologies reviewed require specialised and expensive hardware and software, minimising their potential application. In addition, radar-based or laser systems are complex and non-mobile, reducing their applicability to vital-sign monitoring in clinical settings.

Chapter 3

Photoplethysmography and oxygen monitoring

3.1 Introduction

Notwithstanding its toxic nature, oxygen is a gas fundamental for the function and survival of organisms. Even the simplest forms of life have evolved elaborate mechanisms and chemical processes to exploit the advantages of oxygen in their metabolism [198].

Over the course of the past decades, several technologies have been proposed to monitor the delivery of oxygen to tissue, organs and other body parts. Photoplethysmography, a low-cost optical technology, is often used non-invasively and in contact with the skin surface, and can measure blood volume changes in the vasculature of tissue [53]. There are extensive clinical applications for photoplethysmography, with several devices using its principles available commercially, such as pulse oximeters, instruments for vascular diagnostics and digital beat-to-beat blood pressure estimation systems [8]. Pulse oximetry and the monitoring of oxygen saturation are of major interest as they form the basis for the research presented in the following chapters.

This chapter begins by discussing the need for oxygen monitoring in section 3.2. Section 3.3 presents the relevant physiology of oxygen transport to the tissue through the combined work of the respiratory and circulatory system. Section 3.4 discusses basic concepts and prior work on oxygen measurement. Section 3.5 introduces oximetry as an optical method for oxygen content measurement in the human blood. Section 3.6 analyses pulse oximetry in depth, with its principles, modes of operations, limitations, sources of errors and a summary of methods that describe how peripheral oxygen saturation (SpO_2), heart rate (HR) and respiratory rate (RR) can be estimated from the signals recorded with a pulse oximeter. The chapter finishes with a discussion and conclusions in section 3.7.

3.2 The importance of oxygen monitoring

Oxygen is a chemical substance essential to the functioning of each cell in the human body and, therefore, necessary to sustain life. Cells require a constant supply of energy to maintain the

biological order that keeps them alive. During cellular respiration, glucose is broken down in the presence of oxygen, releasing energy that is captured by energy-carrying molecules such as adenosine triphosphate (ATP). Substances that can no longer be used by the organism as a consequence of cellular respiration, such as water and carbon dioxide (CO_2), are released as metabolic waste. Cells are able to capture almost half of the energy released [298], the rest escaping as heat, which helps maintain body temperature.

The physiological processes that exist for maintaining cell function, and therefore the survival of organisms, consist of complex control systems. The continuous monitoring of oxygen supply is of great importance. A prolonged lack of oxygen can rapidly cause permanent damage to cell tissue, leaving patients with devastating neurological handicaps. It has the potential to be life-threatening if cells having high metabolic rate in organs such as the brain, heart or the central nervous system are damaged [80]. The cerebral cortex has less than 1 min of survival time after total loss of oxygen delivery before permanent cellular damage occurs [198].

It is important to monitor if organs are receiving a sufficient supply of oxygen as it is being delivered to all the body parts. Several abnormal conditions exist, such as: lower than normal blood oxygenation (hypoxaemia), lower than normal tissue oxygenation (hypoxia) or the condition where blood in the system contains more than the normal amount of oxygen (hyperoxia) [198, 298].

Measuring abnormal levels of oxygen is important in many clinical situations due to the toxic nature of oxygen radicals [290]. As an example, tissue oxygenation and, consequently blood oxygen saturation, are of extreme importance during anaesthesia where the patient's cardiopulmonary system is placed in a state where it can no longer meet oxygen demands on its own through the use of anaesthetic agents, which decrease the respiratory drive and blood pressure [353].

Non-invasive and continuous monitoring of oxygen delivery is an invaluable aid to the physician as an indicator of the patient's health [81]. If oxygen delivery monitoring is supplemented with other physiological variables such as temperature, blood pressure, heart rate or respiratory rate; it can warn the clinical staff of critical physiological conditions before damage occurs [353].

3.3 Oxygen transport physiology

This section summarises the relevant physiology of the respiratory and circulatory systems relating to oxygen transport within the human body. This background is needed to review the development of several technologies presented later in the chapter such as oximetry, pulse oximetry and the fundamental theory behind photoplethysmography. The algorithms proposed throughout chapters 6 to 9 also benefit from an understanding of the relevant physiology.

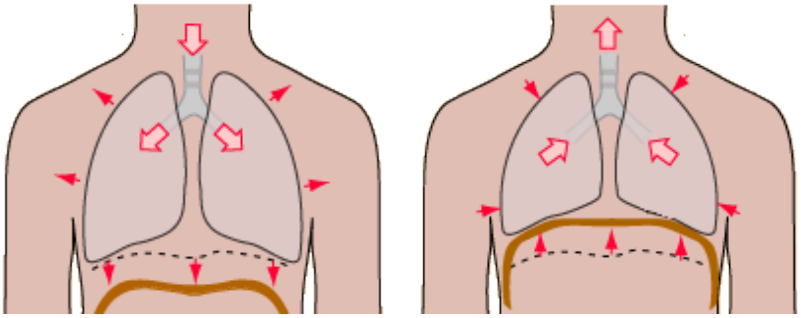
The process of oxygen delivery to cells requires the respiratory system to work in synchrony with the cardiovascular system. The organs of the respiratory system exchange gases between the blood and the air. The cardiovascular system uses blood to carry oxygen from the lungs and nutrients from the digestive organs to all body cells, where these substances are used in metabolic processes [298].

3.3.1 Respiratory process

The respiratory system groups organs and interconnections that filter air and transport it into the lungs. The primary role of the lungs is gas exchange through a process called diffusion [165], which allows the movement of oxygen into the body and the removal of carbon dioxide. The complete process of gas exchange between the atmosphere and body cells is called respiration.

Breathing, often called ventilation, is the initial step of moving air from the nasal cavity or mouth into and out of the lungs. The actions responsible for these air movements are called inspiration or inhalation and expiration or exhalation. The sequence of inspiration, gas exchange, followed by expiration is known as the respiratory cycle [298].

Inspiration and expiration movements cause expansion and contraction of the thoracic cavity that can be easily seen by eye. During inspiration, the diaphragm contracts downwards and the thoracic cavity expands, filling the lungs with air (see figure 3.1(a)). During expiration, the diaphragm relaxes upward and the thoracic cavity contracts, releasing carbon dioxide through the respiratory tract (see figure 3.1(b)). The algorithms proposed in chapter 8 take advantage of these physical changes to estimate respiratory rate.



(a) (b)

Figure 3.1: The respiratory cycle: (a) During inspiration, the diaphragm contracts downwards and the thoracic cavity expands, filling the lungs with air. (b) During expiration, after gas exchange takes place, the diaphragm relaxes upwards and the thoracic cavity contracts, releasing carbon dioxide through the respiratory tract (Source: HyperPhysics, Georgia State University).

Inspiration

Air in the earth's atmosphere is a mixture of gases such as nitrogen, oxygen, carbon dioxide and others retained by the earth's gravity. Atmospheric pressure due to the weight of the air is the force that moves air into the lungs. The amount of pressure each gas contributes is called the partial pressure of that gas and is proportional to its concentration in the total gas mixture.

For the respiratory process, two gases are of main importance: oxygen and carbon dioxide (CO_2). Dry air at sea level contains around 20.946% of oxygen and 0.0397% of carbon dioxide. Assuming that atmospheric pressure at sea level is 760 mmHg, the partial pressure of oxygen (PO_2) is therefore approximately 160 mmHg and the partial pressure of carbon dioxide (PCO_2) is close to 0.3 mmHg [198].

Although the percentage of oxygen in dry air at altitudes higher than sea level is constant, the drop in atmospheric pressure decreases the partial pressure of oxygen and therefore gas exchange within the lungs can be compromised. The pressure exerted by the weight of the atmosphere, and hence the partial pressure of inspired oxygen, decreases in a linear fashion as altitude increases. At 5500 m, the pressure falls to approximately 50% of that of the sea level, and further drops to 30% at 8900 m (the height of the summit of Mount Everest) [254].

Inspiration begins by nerve impulses contracting muscle fibres in the diaphragm. As the diaphragm moves downward, the external intercostal muscles also contract, raising the rib cage and expanding the thoracic cavity (figure 3.1(a)). As the thoracic cavity expands, the pressure inside the lungs decreases. Atmospheric pressure due to the weight of the air is greater on the outside, forcing air into the respiratory tract from the nasal cavity or mouth through the larynx, trachea, bronchial tree and finally into the alveoli (microscopic air sacs within the lungs where oxygen diffusion to blood occurs - see figure 3.2), filling the lungs with air.

Expiration

As the neurons in the rhythmicity area of the brain stem stop sending nerve impulses, the diaphragm, the pectoral muscles and intercostal muscles relax. The elastic tissues of the lungs and thoracic cage, which were stretched during inspiration, suddenly recoil. Surface tension collapses alveolar walls, decreasing the amount of air space and causing air to flow out of the lungs, releasing carbon dioxide (figure 3.1(b)).

Oxygen diffusion to blood

Diffusion of a gas is the tendency of atoms, molecules, and ions in a liquid or air solution to move from areas of higher concentration of partial pressure to areas of lower partial pressure, thus becoming more evenly distributed, or more diffused.

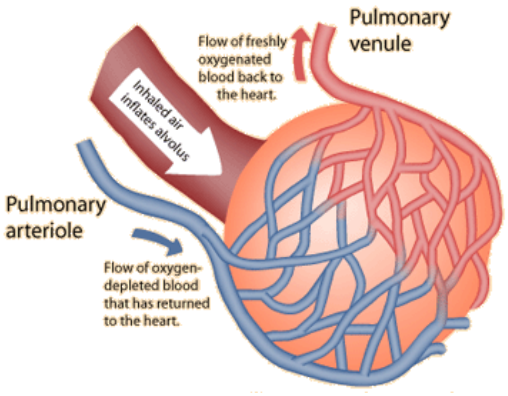


Figure 3.2: Oxygen diffusion to blood: a capillary network covers the surface of the alveolus to facilitate oxygen and carbon dioxide exchange (Source: HyperPhysics, Georgia State University).

Oxygen diffuses from the air within the lungs between the alveoli and the capillaries (see figure 3.2). Air in the alveoli has a higher partial pressure of oxygen and a lower partial pressure of carbon dioxide compared to the equivalent pressures in blood. This pressure difference causes

diffusion to occur. The net movement of carbon dioxide is towards the alveoli and the net movement of oxygen towards the blood [298]. The oxygenated blood then returns to the heart via a pulmonary venule to be pumped out to the rest of the body. Carbon dioxide is then released during expiration. The exchange of gases between the air in the lungs and the blood is sometimes called external respiration.

Diffusion is most efficient over short distances and is enhanced over large surface areas. The distance between cells in the alveoli and capillaries is about 1 $\mu\mathrm{m}$. The pulmonary capillary bed is the largest vascular bed in the body. The alveoli provide a large surface area of thin epithelial cells; if the roughly 300 million alveoli present in the human lung were spread out, they would cover an area of approximately 70 to 80 m^2 [165].

The concentration of a gas dissolved in a solution is directly proportional to the partial pressure of that gas above the solution [298]. Therefore, to approximate the concentrations of oxygen and carbon dioxide in the blood, the literature often refers to computing partial pressure of oxygen (PO_2) and partial pressure of carbon dioxide (PCO_2).

Ventilatory control

Normal breathing is a rhythmic involuntary action that starts early during pregnancy by a group of neurons located in the medulla oblongata in the fetal brain stem called the respiratory centre. [198]. The ventilatory process is closely controlled by a combination of chemical and physical reflexes of four major components: the respiratory centre, central chemoreceptors, peripheral chemoreceptors, and pulmonary mechanoreceptors/sensory nerves.

The respiratory centre periodically triggers nerve impulses that signal the diaphragm and other inspiratory muscles to contract during inspiration (rhythmicity area) [165]. These impulses begin weakly, increase for about two seconds, and stop abruptly causing expiration to occur. The rhythmicity neurons remain inactive while expiration occurs, subsequently triggering new inspiratory impulses so that the respiratory cycle repeats. The respiratory centre also controls the rate and amplitude of the ventilatory pattern by analysing information from other brain centres and chemoreceptors.

The central chemoreceptors, located in the ventral area of the medulla oblongata near the vagus nerves, are very sensitive to changes in partial pressure of carbon dioxide and blood pH. If PCO_2 concentration rises, the central chemoreceptors signal the respiratory centre, which in turn increases the respiratory rate. As alveolar ventilation intensifies, more carbon dioxide is exhaled and the blood PCO_2 concentration returns towards normal levels.

The peripheral chemoreceptors are located in the aortic arch close to the heart and in the walls of the carotid sinuses within the neck. When PO_2 levels decrease, impulses are transmitted to the respiratory centre, causing the increase of respiratory rate and tidal volume. Peripheral chemoreceptors do not initiate warning signals until the PO_2 reaches a very low level in the blood, suggesting that carbon dioxide, not oxygen, plays a major role in ventilatory control [298].

The pulmonary mechanoreceptors help to regulate the depth of breathing. They signal the respiratory centre to shorten the duration of inspiration as the lung tissues are stretched. This action prevents over-inflation of the lungs during forceful breathing.

Factors affecting breathing

Ventilation is a rhythmic involuntary activity that lasts over the entire life span of an individual. Thus, one of its goals, from a mechanical perspective, is to minimise work. A complex process adjusts the respiratory muscles to adapt to stimuli that alter the respiratory requirements or influence the performance of the respiratory muscles such as changes of posture or varying degrees of exercise.

Although breathing is involuntary, it may also be voluntarily over-ridden or be temporarily interrupted by sporadic events such as swallowing, sneezing or coughing. Respiration during speech also varies fundamentally to resting respiration [71], changing the duration and velocity of inspiration to that of expiration.

From a physiological perspective, ventilation maintains blood gases levels, regulates arterial PCO_2 and helps to maintain acid-base balance in the brain. Emotional state or external stimulation may also alter the normal respiratory patterns. Drastic changes in respiratory rate are common responses to fear, anger, shock, excitement or pain.

Although normal breathing is a rhythmic involuntary act, the factors described above make it challenging to have an accurate estimation of respiratory rate. This is of particular interest when assessing the health of sick patients in hospital who present abnormal respiratory patterns, as for the population undergoing haemodialysis treatment described in chapter 5. The techniques presented in chapter 8 address some of these challenges.

3.3.2 Circulatory process

The cardiovascular system includes the heart and blood vessels that help force blood throughout the body. Blood transports gases, nutrients, hormones and other substances used in metabolic process to all body cells.

Overview of blood circulation

Blood transports oxygen and carbon dioxide between the lungs and the body cells (see figure 3.3). Oxygenated blood leaves the lungs and enters the heart's left atrium through the pulmonary veins. The left atrium contracts and the blood moves to the left ventricle through the left atrio-ventricular orifice. Freshly oxygenated blood is pumped from the heart's left ventricle to the aorta and its major arterial branches.

Arteries are elastic vessels optimised to rapidly transport blood away from the heart under high pressure and distribute it to the limbs and other organs. Major arteries include the carotid, which takes blood to the brain; the brachial arteries, which take blood to the arms; the thoracic artery, which takes blood to the thorax and bifurcates to reach the liver, kidneys and stomach; and the iliac artery, which takes blood to the lower limbs. The arteries subdivide successively into thinner tubes and smaller vessels called arterioles to help reach more deeply into the muscles and other organs in the body. The smaller arterioles branch into capillaries where oxygen and carbon dioxide exchange occurs.

As blood moves through the aorta and its arterial branches, blood vessels change their structure. The aorta is a predominantly elastic artery, the peripheral arteries become more muscular and

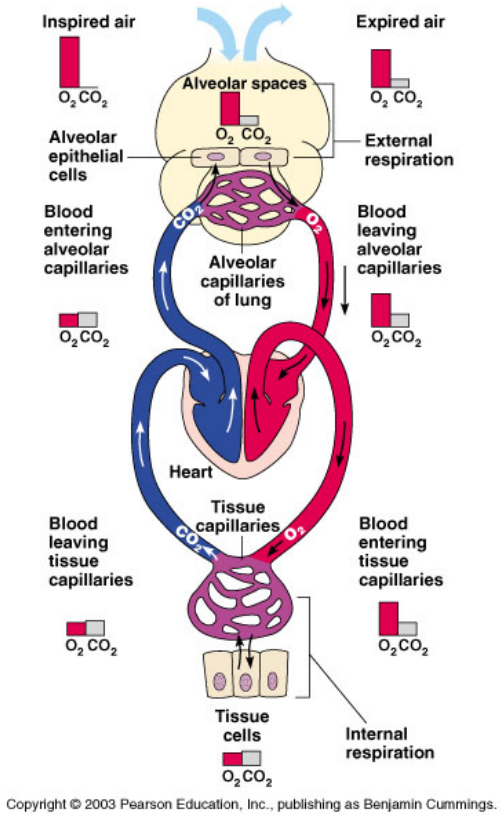


Figure 3.3: The circulatory system: oxygenated blood leaves the lungs and is distributed by the arteries (in red colour) around the body in a process called diffusion. De-oxygenated blood makes its way back in the veins (in blue colour) towards the lungs (Source: Pearson Education).

the arterioles have a predominantly muscular layer. These changes in structure in the arterial branches also affect frictional resistance to blood flow and pressure. Resistance is relatively small in large arteries as the pressure is only slightly less than in the aorta. Resistance reaches a maximal level in the arterioles causing pressure to drop.

Throughout the arterial branches there is also a change of blood flow from pulsatile to a steady flow. Pulsatile arterial blood flow is caused by the periodic ejection of blood from the heart. Blood flow reaching the capillaries is affected mainly by two causes: distensibility of the arteries and frictional resistance in the small arteries and arterioles [165]. The algorithms described in chapter 6 take advantage of the peripheral blood flow pulsatility.

Blood low in oxygen content leaves the capillaries and connects to the venules, making its way back towards the heart through veins of increasing size, finally entering the right atrium through the venae cave. The right atrium contracts and bloods enters the right ventricle chamber, from where it is pumped through the pulmonary arterial system at a mean pressure of about one seventh of that in the systemic arteries [165]. The blood then enters the capillaries associated with the alveoli in the lungs, where carbon dioxide in the blood is released and oxygen is diffused into the blood stream. Oxygen-rich blood returns to the heart via the pulmonary vein.

Oxygen diffusion to tissue

Gas exchange between the blood and cell in the tissue, often called internal respiration (see figure 3.3), occurs by diffusion in the capillaries, which are extensions of the arterioles that connect with the venules.

Diffusion is most efficient over short distances. Capillaries are the smallest diameter blood vessels with walls composed of a single thin layer of endothelium. These one-cell size walls help oxygen to diffuse from the blood into the surrounding body cells of tissue in exchange for carbon dioxide. Other products of the active cellular metabolism are also exchanged.

3.3.3 Haemoglobin as oxygen carrier

The cardiovascular system provides oxygen at adequate rates to all the body tissues to support their oxygen consumption for survival. The rate of diffusion of gases such as oxygen in liquids is slow. This would be an issue as blood is composed of about 50% plasma, which is 95% water. To solve the problem, the addition of haemoglobin molecules to the red blood cells effectively improves oxygen diffusion and transport. Haemoglobin can typically carry 19 ml of oxygen per 100 ml of blood, which is about 65 times the amount that could be carried dissolved in the plasma alone [253]. Most of the oxygen (around 98%) is carried in the blood stream bound to haemoglobin; the remainder is dissolved in plasma.

Each red cell has approximately 265 million haemoglobin molecules, each consisting of two types of components called heme and globin. Globin is a protein that contains 574 amino acids in four polypeptide chains associated with a heme group. Each heme group contains an atom of iron that can bind with one oxygen molecule [298]. Therefore, each haemoglobin molecule is capable of bonding with up to four oxygen molecules.

Oxygen bound to haemoglobin forms a new compound called oxyhaemoglobin (HbO_2). As discussed in previous sections, the partial pressure of oxygen (PO_2) determines the amount of oxygen molecules that can be bound with haemoglobin. The higher the PO_2, the more oxygen molecules will be combined with haemoglobin, until the haemoglobin molecules become saturated with oxygen. A haemoglobin molecule without an association to oxygen is often referred to as deoxygenated haemoglobin (Hb).

Haemoglobin molecules that are able to bind reversibly with oxygen and help release it in the systemic capillaries, such as HbO_2 and Hb, are called *functional haemoglobins*. Haemoglobin molecules that are either unable to bind reversibly to oxygen or do not support the transport of oxygen to the tissues are called *dysfunctional haemoglobins*.

The four most common dysfunctional haemoglobins are [81, 353]: carboxyhaemoglobin ($HbCO_2$), formed when carbon monoxide is bound to haemoglobin; methaemoglobin ($metHb$), when blood is exposed to various drugs or oxidising agents; sulfhaemoglobin ($sulfHb$), produced by the reaction of oxyhaemoglobin and hydrogen sulfide; and carboxysulfhaemoglobin, the result from the reaction between sulfhaemoglobin and carbon monoxide.

3.4 Blood oxygen measurement

3.4.1 Oxygen saturation in a blood sample

In healthy individuals, functional haemoglobins bind with the oxygen diffused into the blood stream and release it into the capillaries. Therefore, the oxygen content in a blood sample can be expressed as *functional oxygen saturation* (functional SO_2) [80], measured as the proportion of oxyhaemoglobin in the total functional haemoglobins:

$$Functional\ SO_2 = \frac{HbO_2}{HbO_2 + Hb} * 100\% \tag{3.1}$$

The functional oxygen saturation of explicitly arterial blood is called *functional arterial oxygen saturation* (functional S_aO_2). In healthy adults, arterial oxygen saturation is usually between 96% and 100%. Saturation levels below 90% could indicate levels of risk [253].

Not all haemoglobin molecules are functional haemoglobins. The fraction of oxygenated haemoglobin from the total haemoglobin in a blood sample is called the *fractional oxygen saturation* (Fractional SO_2) and is defined by:

$$Fractional\ SO_2 = \frac{HbO_2}{total\ haemoglobin} * 100\% \tag{3.2}$$

where *total haemoglobin* represents a combination of all the different species of haemoglobin present in a blood sample.

3.4.2 Oxyhaemoglobin dissociation curve

As discussed in previous sections, the concentration of a gas dissolved in a solution is directly proportional to the partial pressure of that gas in the solution; thus the total amount of oxygen in arterial blood is directly proportional to the partial pressure of oxygen (PO_2).

The relationship between arterial oxygen saturation (SaO_2) and partial pressure of oxygen (PO_2) is described visually by the oxyhaemoglobin dissociation curve, shown in figure 3.4 for a theoretically healthy subject. It describes the mechanism by which oxygen is bound and released in blood. In clinical practice, it is common to measure either PO_2 or SaO_2 and estimate the other. The relationship between oxygen saturation and PO_2 is not constant, even for an individual [81]. Three main important physiological factors affect possible changes in the dissociation curve: blood pH, temperature, and the concentration of 2,3-diphosphoglycerate (DPG).

Changes in blood pH are directly proportional to the affinity between PO_2 and SaO_2 and inversely proportional to the concentration of carbon dioxide (CO_2), a phenomenon known as the Bohr effect [34]. A decrease in blood pH results in haemoglobin molecules releasing more oxygen, reducing the affinity and thereby causing the dissociation curve to shift to the right. In contrast, an increase in blood pH results in haemoglobin molecules binding with more oxygen, increasing the affinity and causing a shift to the left in the dissociation curve. The concentration of CO_2 can also affect the acidity in the blood, an increase in CO_2 can result in a decrease in blood pH. Conversely, a decrease in CO_2 can result in an increase in blood pH.

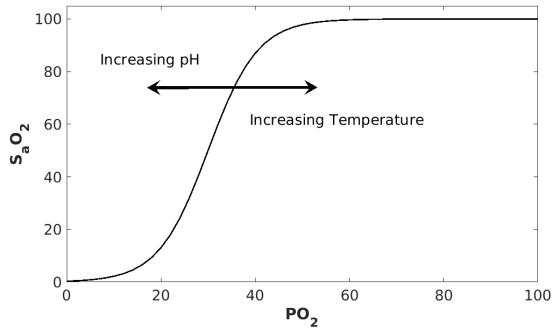


Figure 3.4: The oxyhaemoglobin dissociation curve in a theoretically healthy subject (Source: De Kock [81]).

Changes in temperature are inversely proportional to changes in the dissociation curve [284]. Hyperthermia, elevated body temperature, causes a rightward shift in the curve, decreasing the affinity. Hypothermia, low body temperature, causes a leftward shift, increasing the affinity between partial pressure of oxygen and arterial oxygen saturation.

The effects of 2,3-diphosphoglycerate (DPG), a molecule involved in the control of affinity between haemoglobin and oxygen, is more complex [339]. DPG is an organophosphate (a molecule containing one or more phosphate ester groups) created in the red blood cells during glycolysis, the first step in the breakdown of glucose to extract energy for cellular metabolism. DPG is thought to be an important adaptive mechanism, because its production increases when tissue oxygen levels decrease during several chronic conditions, such as: hypoxaemia, chronic lung disease, anaemia, septic shock and congestive heart failure [68]. High levels of DPG shift the curve to the right, while low levels of 2,3-DPG cause a leftward shift, decreasing the affinity between partial pressure of oxygen and arterial oxygen saturation.

3.4.3 Methods for measuring blood oxygen content

The measurement of the content of oxygen carried by blood is an important indicator of a patient's health (see section 3.2). Several methods have been developed to analyse oxygen delivery. During the early 1900's clinicians depended on cyanosis, an abnormal diffused blue discolouration of the skin caused by the excessive concentration of deoxygenated haemoglobin in the blood [199]. Cyanosis was shown to be unreliable, being only useful when 15% or between 2 g to 3 g of the haemoglobin is desaturated and blood haemoglobin concentration is not below normal limits [70].

In-vitro chemical methods were developed during the second half of the twentieth century, which proved to be more reliable and deterministic. Chemically-based methods determine the oxygen content of blood from a sample by using chemical reactions to remove the oxygen from the blood. Some methods like the Clarke electrode [353] and transcutaneous PO_2 sensor [332] are still in use for blood gas analysis and are considered as gold standards for the measurement of partial pressure of oxygen. Some blood gas analysers incorporate a co-oximeter that directly measures the various forms of haemoglobin including oxyhaemoglobin, carboxyhaemoglobin and methaemoglobin [353]. One of the disadvantages of these methods is that they tend to be slow.

The advent of oximetry as an optical method for measuring oxygen content in the blood developed rapidly during the 1930s and 1940s. Pulse oximetry revolutionised clinical practice [253], bringing the possibility of measuring, non-invasively and continuously, arterial oxygen saturation in the peripheral arteries.

3.5 Oximetry

Oximetry groups several well established techniques for the optical measurement of oxygen saturation in the blood [81, 253]. In principle, oxygen saturation can be measured in vitro (such as blood samples haemolysed in a cuvette) or in vivo (measurements taken directly from the human body). Optical analysis of oxygen saturation in tissue was introduced by Nicolai in 1932 to record the progressive colour change of the hand below a tourniquet as a measure of the rate of tissue oxygen consumption [240]. Millikan introduced the term “oximeter” in 1942 [224] when designing a device attached to the ear to measure, continuously, oxygen saturation in arterial blood.

3.5.1 Principles of oximetry

Oximetry is based on spectrophotometry, a method to measure the amount of light a chemical substance absorbs. Analytical techniques and the principles of spectrophotometry were initially developed by Bunsen and Kirchhoff [253] in the second half of the nineteenth century.

Every chemical substance absorbs, transmits or reflects light over a range of wavelengths. The absorptivity of the chemical substance, also known as its extinction coefficient, determines the fraction of light absorbed by the substance. The extinction coefficient of a chemical compound can be measured by a spectrometer and graphed at various wavelengths to produce a spectrum [353]. This spectrum is unique for any substance.

It has been well known since the 1930s that variations in blood volume in a body segment during each cardiac cycle, modulate the reflection or transmission of visible and infra-red light through that body segment [353]. An oximeter is a spectrometer with multiple light sources to illuminate the tissue and a photo-detector to measure the variations in light intensity associated with changes in oxygen content of the blood.

Oximetry operates on the principle that tissue is relatively transparent to both red and near infra-red light, therefore blood absorbs visible and infra-red light more than the surrounding tissue in a body segment. The light absorbance for the different haemoglobin species also have different optical spectra [362], as shown in figure 3.5(a).

3.5.2 Haemoglobin absorption spectra

The light-absorbing characteristics of the haemoglobin molecule change with its chemical binding to other molecules and with the wavelength of the incident light. Oxyhaemoglobin and deoxygenated haemoglobin absorb most of the light passing through blood [353]. However, haemoglobin may combine with other substances such as carbon monoxide or hydrogen sulfide, which change its colour and hence the transmitted light.

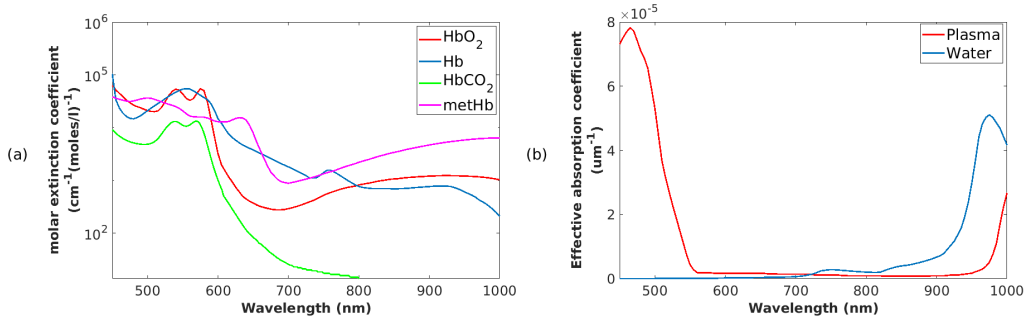


Figure 3.5: Light attenuation for various chemical compounds in the visible and infra-red spectrum. (a) Extinction coefficients for the four most common haemoglobin species: oxyhaemoglobin (HbO_2), deoxygenated haemoglobin (Hb), carboxyhaemoglobin ($HbCO_2$) and methaemoglobin ($metHb$). (b) Effective absorption coefficients for water and plasma (Data source: [216, 251, 269, 301, 373]).

The chemical binding of the different haemoglobin species also changes the physical properties of haemoglobin. Figure 3.5(a) shows the extinction coefficients of the common types of functional and dysfunctional haemoglobins in the visible and infra-red parts of the spectrum. The absorbance of light in the red region of the spectrum is much higher for deoxygenated haemoglobin than for oxyhaemoglobin. The extinction coefficients for both species of haemoglobin are similar at the isosbestic point (around 805 nm). Deoxygenated haemoglobin is more transparent to light from the infra-red region than oxyhaemoglobin [269, 301, 373].

Water and plasma are two major factors that affect light absorption in human blood [216, 251]. Both substances have a relatively flat absorption in the red to infra-red range of the spectrum from approximately 550 nm until about 850 nm. Water and, to a lesser extend plasma, strongly absorb light in the longer infra-red wavelengths, as shown in figure 3.5(b).

3.5.3 Light absorbance

The Beer-Lambert law helps to describe the attenuation of a collimated beam of light in a non scattering medium containing an absorbing substance. If monochromatic incident light of intensity I_0 enters an absorbing medium, part of this light is transmitted through the medium while another part is absorbed [353]. The intensity I of light travelling through the medium decreases exponentially with its distance, as defined by:

$$I = I_0 e^{-\epsilon(\lambda)cd} \tag{3.3}$$

where $\epsilon(\lambda)$ is the extinction coefficient of the absorbing substance in the medium, which is constant at a given wavelength λ; the concentration of the absorbing substance c is constant in the medium, and d is the optical path length through the medium (see figure 3.6). Equation (3.3) assumes that the sum of transmitted and absorbed light is equal to the incident light. Therefore the transmittance T of light travelling through a medium containing an absorbing substance is defined as the ratio of transmitted light I to the incident light I_0:

$$T = I/I_0 = e^{-\epsilon(\lambda)cd} \tag{3.4}$$

The unscattered absorbance A, often referred to as the optical density of a medium, is defined as the negative natural logarithm of the transmittance of light:

$$A = -lnT = \epsilon(\lambda)cd \tag{3.5}$$

The Beer-Lambert law can be extended to include more than one absorbing substance in the medium. The resulting total absorbance A_t is the sum of the individual absorbing characteristics for each substance:

$$A_t = \epsilon_1(\lambda)c_1d_1 + \epsilon_2(\lambda)c_2d_2 + ... + \epsilon_n(\lambda)c_nd_n = \sum_{i=1}^{n} \epsilon_i(\lambda)c_id_i \tag{3.6}$$

where $\epsilon_i(\lambda)$ and c_i represent the extinction coefficient and concentration of substance i respectively. The optical path length d_i through the absorbing substance may change for different substances in the medium.

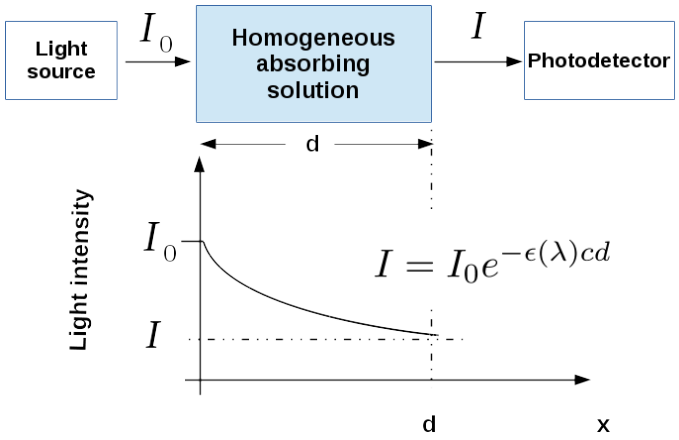


Figure 3.6: The Beer-Lambert law states that the light intensity travelling through a homogeneous absorbing solution falls exponentially with the increasing distance d, the concentration of the absorbing substance c and the extinction coefficient ϵ of the substance in the solution.

The Beer-Lambert law can be used to determine the unknown concentrations of N different absorbing substances (such as different types of haemoglobin molecules) in a homogeneous medium (such as blood) if the absorbance of light is measured at N different wavelengths and the extinction coefficients of the substances are known. In practice, however, Kramer [169] showed that the Beer-Lambert law cannot be applied to whole blood, because the optical path changes and the concentration of haemoglobin varies. Nevertheless Kramer also showed that, in principle, the Beer-Lambert law could be applied to approximate haemoglobin concentration for arterial blood in glass cuvettes with known characteristics and fixed dimensions [353]. Measurements improve if the red blood cells have been broken up (lysed) and no light scattering takes place.

3.5.4 Absorbance in haemoglobin solutions

Haemoglobin is the main light absorber in human blood, therefore the functional oxygen saturation can also be expressed as the ratio of concentrations of oxyhaemoglobin (c_{HbO2}) and deoxygenated haemoglobin (c_{Hb}):

$$Functional\ SO_2 = \frac{c_{HbO_2}}{c_{HbO_2} + c_{Hb}} * 100\% \tag{3.7}$$

The concentration for each of these two main haemoglobins can therefore be related to the functional oxygen saturation SO_2 by:

$$c_{HbO_2} = SO_2(c_{HbO_2} + c_{Hb}) \tag{3.8}$$

$$c_{Hb} = (1 - SO_2)(c_{HbO_2} + c_{Hb}) \tag{3.9}$$

The total absorbance A_t can be derived by substituting the concentrations of oxyhaemoglobin and deoxygenated haemoglobin in equations (3.8) and (3.9) respectively into equation (3.6), assuming that the optical path length d is the same for HbO_2 and Hb.

$$A_t = \Big[\epsilon_{HbO_2}(\lambda)SO_2 + \epsilon_{Hb}(\lambda)(1 - SO_2)\Big](c_{HbO_2} + c_{Hb})d \tag{3.10}$$

Therefore, the total absorbance can be expressed as a function of known concentrations of haemoglobin for a given functional oxygen saturation, the extinction coefficients and the length of the optical path. Zijlstra et al [372] computed the extinction coefficients of Hb and HbO_2 for the most commonly wavelengths used by oximeters.

3.5.5 In vivo oximetry

The Beer-Lambert law is not applicable to the non-invasive estimation of oxygen content in whole blood from a light signal which passes through living tissue such as the finger or ear lobe. The light is not only absorbed but is also scattered by tissue, dark skin, and other blood haemoglobin (capillary and venous blood haemoglobin). There is both reflection of the light at the surface of the medium and scattering of the light in the medium.

Early designs for in vivo oximeters include those of Squire [311], Goldi [110] and Wood [358]. They proposed oximeters that compress the ear lobe so all the blood is squeezed out of tissue to give a bloodless reference. Absolute saturation values were measured when blood was re-admitted into the tissue. These instruments were unstable and the technology of the photo-detectors and light sources was not reliable. Millikan [224] obtained measurements from arterial blood by heating the ear lobe to increase arterial blood flow in tissue due to thermal dilation of the vessels. Hewlett Packard extended this design into an 8-wavelength oximeter using a multi-coloured filter wheel capable of measuring different haemoglobin species [253]. This instrument was calibrated empirically and achieved an acceptable accuracy, but it was far too cumbersome and expensive to be used in a clinical environment.

3.6 Pulse oximetry

Pulse oximetry was introduced during the late 1970s as an optical method for measuring blood oxygen saturation non-invasively and continuously. It is recognised worldwide as the standard of care in anaesthesiology, widely used in intensive care, operating rooms, emergency departments, general wards, neonatal intensive care units, sleep laboratories and home care [353]. It provides early information on problems in the delivery of oxygen to tissue.

3.6.1 Principles of pulse oximetry

Pulse oximetry is based on the theory of oximetry. In its basic form, a pulse oximeter requires only a pair of light sources to illuminate the tissue and a photo-detector to measure the small variation in light intensity associated with changes in blood volume.

Pulse oximeters rely on two basic principles: The light absorbance of oxyhaemoglobin and deoxygenated haemoglobin is markedly different in the red and infra-red wavelength range, as previously shown in figure 3.5(a) and more closely in figure 3.7); arterial blood is pulsatile in nature in contrast to venous blood (see section 3.3.2).

The pulsatile nature of arterial blood results in a waveform, the photoplethysmogram (PPG), that allows the absorbance effects of arterial blood to be identified from those of non-pulsatile venous blood and other body tissue, as shown in figure 3.8.

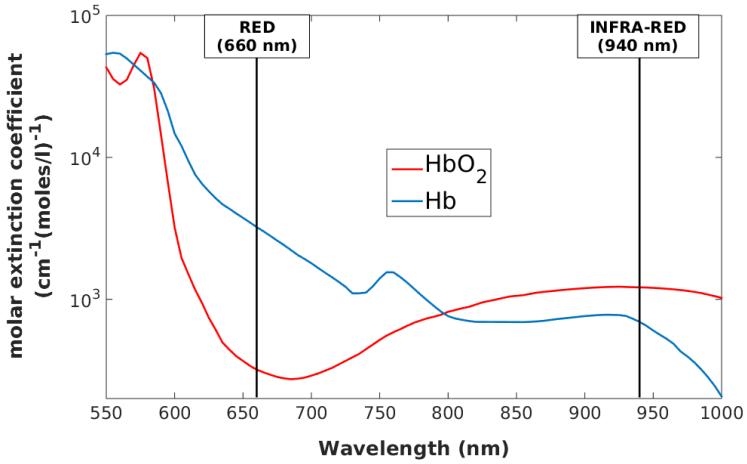


Figure 3.7: Oxyhaemoglobin (HbO_2) and deoxygenated haemoglobin (Hb) exhibit markedly different absorption characteristics in the red and infra-red parts of the spectrum.

3.6.2 Wavelength selection criteria

The depth to which light penetrates tissue for a given intensity of optical radiation depends on the operating wavelength. To measure arterial oxygen saturation, the light source needs to penetrate the tissue sufficiently. Skin penetration increases as the wavelength increases in the visible range, is strong in the near infra-red region, and decreases in the infra-red range [8].

The typical wavelengths chosen for the two light sources in pulse oximetry are around 660 nm (red) and close to 940 nm (infra-red) [362], as shown in figure 3.7. There are many practical and physiological reasons for this choice of wavelengths. Wavelengths shorter than 600 nm are suboptimal due to light being mostly absorbed by melanin, the pigmentation of human skin [151]. As the main component of tissue is water, light is absorbed strongly in the ultraviolet range (wavelengths less than 400 nm) and infra-red (above 1000 nm).

Significant differences exist in the light absorption characteristics between oxyhaemoglobin and deoxygenated haemoglobin at 660 nm and 940 nm. The absorption spectra are also relatively flat at these wavelengths, helping to minimise errors due to peak characteristics of light sources [353]. The absorbance of light at a wavelength of 940 nm increases with a higher oxygen saturation. At 660 nm the absorbance of light decreases rapidly with an increasing oxygen saturation. At the isosbestic wavelength (around 805 nm), the extracted signal is mostly unaffected by changes in oxygen saturation [113].

3.6.3 The photoplethysmogram (PPG)

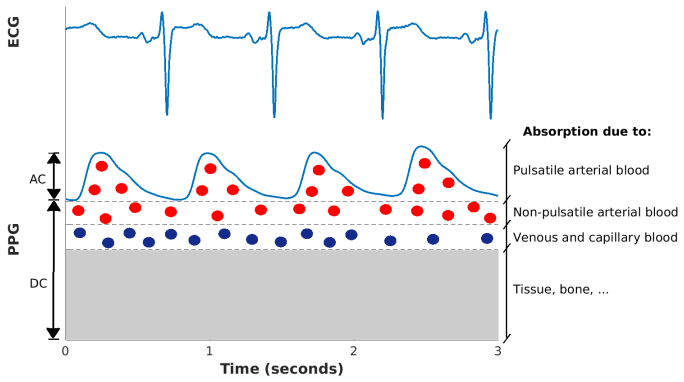


Figure 3.8: The photoplethysmogram (PPG) waveform (bottom plot) contains an alternating AC signal mostly due to pulsatile arterial blood in synchrony with each heart beat, compared with the ECG in the top plot. The AC signal is superimposed on a steady-state DC signal that relates to non-pulsatile arterial blood, venous blood, tissue and other components affecting light absorption.

The photoplethysmogram (PPG) measures the changes in absorption of light travelling though biological tissue such as the finger or ear lobe. It originates primarily from variations in arterial blood volume during the cardiac cycle as the blood is pumped into the arteries and throughout the body. Arteries are elastic vessels that expand their diameter as the blood volume increases during systole and the pressure onto their walls becomes greater. As the heart relaxes during diastole, the pressure on the arteries decreases and their diameter reduces. This pulsatile effect occurs mainly in the arteries and arterioles but not in the veins, as the pressure in veins is very low (see section 3.3.2).

The photo-detector in a pulse oximeter records the light transmitted through the tissues and allows for the differentiation between an alternating AC signal and a steady-state DC signal. The

AC signal is caused by the absorption by pulsatile arterial blood in synchrony with each heart beat. The steady-state DC signal is caused by the absorbance of venous blood, a constant amount of arterial blood and other non-pulsatile components such as skin pigmentation, bone and tissue (see figure 3.8). The AC signal in the PPG waveform usually does not exceed 1% or 2% of the constant absorbance of the total DC components [353].

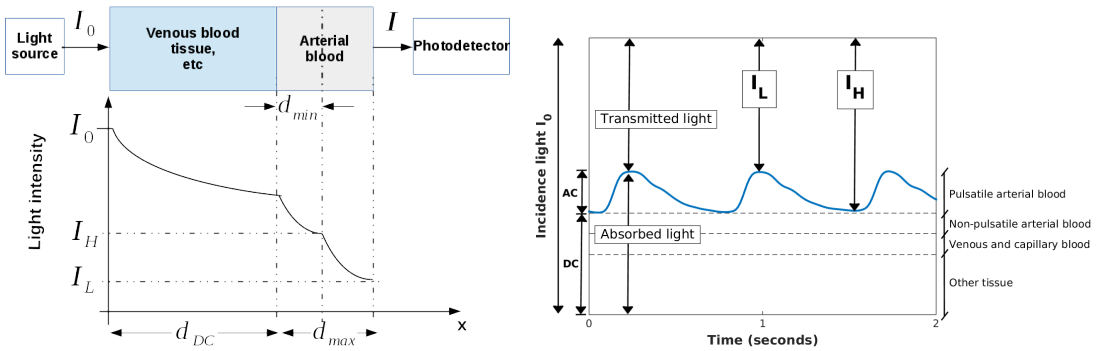


Figure 3.9: The DC components (such as venous blood, tissue and non-pulsatile arterial blood) absorb a constant amount of the incident light I_0. During diastole the optical path length through the arteries is minimum d_{min} and the light intensity is maximal I_H. During systole, the optical path reaches a maximum d_{max} and the haemoglobin in the arteries absorbs a maximum amount, causing the intensity to decrease to a minimum level of I_L. (a) Simplified model, (b) the photoplethysmogram (PPG) waveform.

Given an incident light I_0, the alternating changes in blood volume effectively changes the length of the optical path the light has to travel. When the heart muscle relaxes during diastole allowing its chambers to fill with blood, the intensity of the transmitted light passing through the tissue is high "I_H" and the diameter of the arterial vessels is minimal "d_{min}" as arterial pressure is low. The main light absorbents during diastole are the DC components (see figure 3.9). When blood volume increases during systole, the length of the optical path increases to d_{max} and the intensity of the transmitted light is reduced to I_L as more absorbents are present in the blood stream. The resulting light intensity can be expressed by the following equations:

$$I_H = I_0 e^{-\epsilon_{DC}(\lambda)c_{DC}d_{DC}} e^{-[\epsilon_{Hb}(\lambda)c_{Hb}+\epsilon_{HbO_2}(\lambda)c_{HbO_2}]d_{min}} \tag{3.11}$$

$$I_L = I_0 e^{-\epsilon_{DC}(\lambda)c_{DC}d_{DC}} e^{-[\epsilon_{Hb}(\lambda)c_{Hb}+\epsilon_{HbO_2}(\lambda)c_{HbO_2}]d_{max}} \tag{3.12}$$

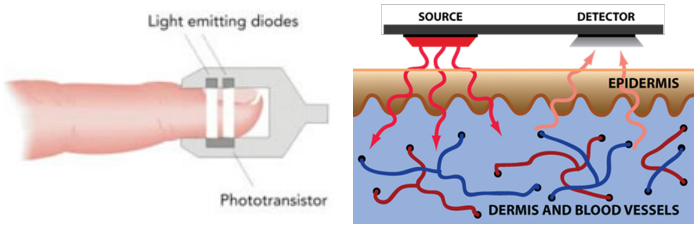
where $\epsilon_{DC}(\lambda)$ is the extinction coefficient of the DC components at a given wavelength λ. c_{DC} and d_{DC} are the concentration and optical path length of the DC components respectively.

The light intensity I of the AC component of the PPG signal can be defined as a function of the diameter of the arteries and arterioles. During one cardiac cycle this diameter changes from d_{min} to d_{max}. Therefore, the Beer-Lambert law can be defined as a function of I_H and Δd, the part of the diameter that changes from 0 to $d_{max} - d_{min}$:

$$I = I_H e^{-[\epsilon_{Hb}(\lambda)c_{Hb}+\epsilon_{HbO_2}(\lambda)c_{HbO_2}]\Delta d} \tag{3.13}$$

3.6.4 Methods of operation

Over the course of the development of the technology for pulse oximetry, two modes of operation have often been used in clinical practice: transmission and reflectance.



(a) (b)

Figure 3.10: Methods of operation in pulse oximetry. (a) Transmission mode, in which the optical elements are located on opposite sides of a body part (Source: Parin Dedhia, Cornell). (b) Reflectance mode, in which the optical elements are located in the same plane, on the same body surface (Source: Nonin).

Transmission mode pulse oximetry

In transmission mode pulse oximetry [54], the photo-detector is located on the opposite side from the two light sources. The sensors are usually applied to fingers, toes or ear lobes, as shown in figure 3.10(a).

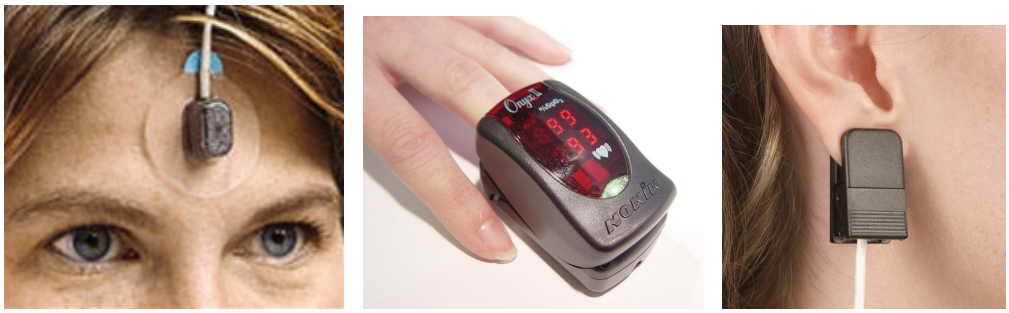
Reflectance mode pulse oximetry

Reflectance-mode pulse oximeters, which have the light sources and detector side by side on the same plane of the body surface (see figure 3.10(b)), are also occasionally used. These work because some of the light is back-scattered by the tissues. The main use of this type of pulse oximeters is in the forehead [4], for clinical applications in which peripheral shut-down occurs, for example during vascular surgery [351] or in severe shock [238]. During peripheral shut-down, conventional transmittance-mode pulse oximetry loses accuracy, and sometimes fails outright because of the reduction in blood flow in the extremities.

3.6.5 Locations for oxygen saturation measurement

The main peripheral body sites where the PPG signal can be obtained and the SpO_2 measured are the tissue pads of the ears, fingers and toes where superficial vasculature is easily accessible and abundant (figure 3.11) [8]. Several other body sites have also been proposed, such as the supraorbital artery above each eye [178], as well as measurements on the skin above key arterial landmarks [194]. Each body site has different PPG characteristics; the pulses vary in amplitude, shape and variations over time [335].

Many of the studies reported in the literature use only one pulse oximeter attached to a single body site. By studying the pulses obtained simultaneously from different sites, important information about the peripheral circulation may be extracted [8]. Simultaneous measurements



(a) (b) (c)

Figure 3.11: Common places to measure oxygen saturation using a pulse oximeter. Probes can be located on: (a) forehead, (b) finger, (c) ear lobe. (Source: Nonin)

are also likely to provide a better assessment of key cardiovascular parameters such as heart rate, body temperature or blood pressure over sequential measurements from a single body site.

Studies investigating dual PPG systems, mostly between two fingers from the right and left arms or between a finger and the ear lobe, have been widely reported [72, 115, 241, 245, 263]. Three-channel PPG systems have been described by Evans et al [89]. Four-channel PPG systems have also been proposed in [42] and [87]. Allen et al [7, 9, 10, 11] have extensively studied a six-channel PPG research system placed on the right and left ear lobes, index fingers or thumbs, and great toes, emphasising the importance of multi-site photoplethysmography to analyse the physiological differences between right and left body sites.

3.6.6 SpO_2 estimation

Current pulse oximeters estimate the arterial oxygen saturation of the blood by measuring light absorbance at two wavelengths. Because of this, the pulse oximeter is only able to account for oxygenated and reduced haemoglobins. The term peripheral oxygen saturation (SpO_2) refers to arterial oxygen saturation measurements determined by pulse oximetry in the peripheries such as the finger or toe.

Normalisation

The light sources used in pulse oximeters may emit light with varying intensities and the sensitivity of the photo-detector can differ. Moreover, the tissue absorption and optical path length can vary between patients [81]. Therefore, the light intensities measured by the pulse oximeter have to be normalised, dividing the transmitted light intensities by their individual maximum peaks ($I_{H,R}$ for the red wavelength and $I_{H,IR}$ for the infra-red wavelength). From equation (3.13) we derive:

$$I_n = \frac{I}{I_H} = e^{-[\epsilon_{Hb}(\lambda)c_{Hb}+\epsilon_{HbO_2}(\lambda)c_{HbO_2}]\Delta d} \tag{3.14}$$

The normalised signals have similar intensities during diastole. The AC components of the normalised signals from the red and infra-red wavelengths represent changes by arterial pulsatile blood and can be compared with each other.

Ratio of normalised signals

A ratio of the AC components for the red (R) and infra-red (IR) wavelengths can be computed using the total absorbance (see equation (3.5)) of the normalised signal during diastole. This will result in absorbance that depends only on pulsatile arterial blood.

$$R = \frac{A_{t,R}}{A_{t,IR}} = \frac{ln(I_{L,R})/I_{H,R}}{ln(I_{L,IR})/I_{H,IR}} \tag{3.15}$$

where $A_{t,R}$ and $A_{t,IR}$, as before, are the total absorbance at the red and infra-red wavelengths respectively.

Using equation (3.14), and the optical path length d_R and d_{IR} for the red and infra-red wavelengths respectively, the ratio can be derived as

$$R = \frac{[\epsilon_{Hb}(\lambda_R)c_{Hb} + \epsilon_{HbO_2}(\lambda_R)c_{HbO_2}]\Delta d_R}{[\epsilon_{Hb}(\lambda_{IR})c_{Hb} + \epsilon_{HbO_2}(\lambda_{IR})c_{HbO_2}]\Delta d_{IR}} \tag{3.16}$$

From equation (3.10), and assuming that the optical path lengths for red and infra-red are equal:

$$R = \frac{\epsilon_{Hb}(\lambda_R)[\epsilon_{HbO_2}(\lambda_R) - \epsilon_{Hb}(\lambda_R)]SaO_2}{\epsilon_{Hb}(\lambda_{IR})[\epsilon_{HbO_2}(\lambda_{IR}) - \epsilon_{Hb}(\lambda_{IR})]SaO_2} \tag{3.17}$$

Equation (3.17) can be rewritten in a form where the oxygen saturation is a function of the calculated ratio R:

$$SaO_2 = \frac{\epsilon_{Hb}(\lambda_R) - \epsilon_{Hb}(\lambda_{IR})R}{\epsilon_{Hb}(\lambda_R) - \epsilon_{HbO_2}(\lambda_R) + [\epsilon_{HbO_2}(\lambda_{IR}) - \epsilon_{Hb}(\lambda_{IR})]R} * 100\% \tag{3.18}$$

Therefore, oxygen saturation in arterial blood can be derived theoretically by computing the ratio R of normalised light absorbances at the two chosen red and infra-red wavelengths.

Calibration curves used in pulse oximetry

Pulse oximeters usually assume that there are no dysfunctional haemoglobins in the arterial blood and measure the blood oxygen saturation as functional SaO_2, see equation (3.1). When the amount of dysfunctional haemoglobin is very small, both readings are comparable [253]. When the patient has large amounts of dysfunctional haemoglobin, such as increased levels in carboxyhaemoglobin caused by carbon monoxide poisoning or smoking, these measurements vary widely [353].

Because arterial oxygen saturation cannot be calculated using the Beer-Lambert law due to the multiple scattering by the red blood cells, pulse oximeter manufacturers estimate it using calibration tables empirically determined from data acquired in studies with human volunteers [353], measuring the ratio R from a pulse oximeter and the actual arterial oxygen saturation using an accurate method such as an oximeter (from blood samples withdrawn from the volunteers). These calibration tables are used to find the relationship between R and SaO_2, Mendelson and Kent [217] describe an equation for the calibration curve, based on the Beer-Lambert law:

$$S_pO_2 = \frac{k_1 - K_2R}{k_3 - K_4R} \tag{3.19}$$

where the extinction coefficients from equation (3.18) are replaced by the constants k_i. These constants are determined by clinical studies to give the curve with a best fit to the in vitro measured data. Fine and Weinreb [91] proposed the use of another polynomial for calibration of the ratio R:

$$S_pO_2 = k_1 + K_2R + k_3R^2 \tag{3.20}$$

3.6.7 Heart rate estimation from the PPG

The increase in arterial blood volume is synchronous with the cardiac activity, as shown in figure 3.12, whereby each peak in the PPG signal corresponds to a cardiac pulse in the ECG waveform (Consecutive heart beats give the peaks at P_1 and P_2 in figure 3.12). Thus, the heart rate can be computed from the AC component of the PPG signal as follows:

$$HR = 60 * \frac{1}{T_{P1} - T_{P2}} \, [beats/min] \tag{3.21}$$

where T_{P1} and T_{P2} are the timestamps expressed in seconds for peaks P_1 and P_2 respectively in figure 3.12.

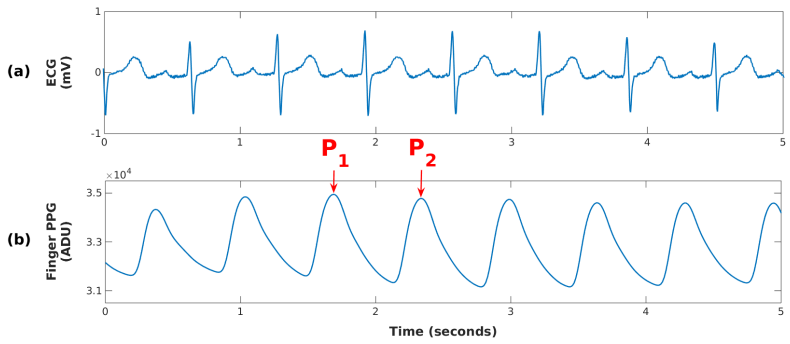


Figure 3.12: Simultaneously recorded (a) ECG and (b) photoplethysmogram waveform from the ear lobe for a patient undergoing haemodialysis treatment. The time interval P_1 and P_2 is 2.4s - 1.8s = 0.6s, hence, the heart rate (computed from equation (3.21)) is 92 beats/min.

Pulse oximeters therefore compute and display the heart rate alongside the peripheral oxygen saturation (SpO_2). Movement artefacts and a number of cardiac conditions can challenge the reliability of heart rate computation. Nakajima et al [236] developed digital filtering algorithms to separate cardiac components from the ear PPG signal. Johansson et al [149] acquired ECG and PPG data for over 8 hours from newborn infants in a neonatal care unit and estimated heart rate from it. Yu et al [369] used a support vector machine classifier for the automatic assessment of the reliability of reference heart rates from vital-sign monitors incorporating both ECG and PPG signals. In sleep studies, Foo and Wilson [97] estimated the heart rate with the addition of an accelerometer to detect movement.

Extensive online PPG databases exist. Of those, PhysioNet (`http://physionet.org`) is widely used in the signal processing community as it contains multiple recordings from real patients, including many instances of motion artefacts.

3.6.8 Respiratory rate estimation from the PPG

Oxygen delivery is of important concern for the well-being of the patient in many clinical settings. An abnormal respiratory rate is often an indicator of critical illness that may precede or occur simultaneously with changes in heart rate, SpO_2, blood pressure or other vital signs [220, 292]. Respiratory rate has also been shown to be an important predictor of serious events such as cardiac arrest or admission to an intensive care unit [76].

While heart rate and peripheral oxygen saturation are measured continuously by a pulse oximeter as a result of the analysis of the photoplethysmogram (PPG) waveform, continuous estimation of respiratory rate remains unreliable and is often not included in commercial pulse oximeters [220]. Respiratory rate is typically measured either by the clinical staff counting chest wall movements for periods of 15 or 30 seconds [293] or by impedance pneumography, a non-invasive technique that works by injecting a small high-frequency current into the patient's thorax through a set of electrodes (often the ECG electrodes) [249].

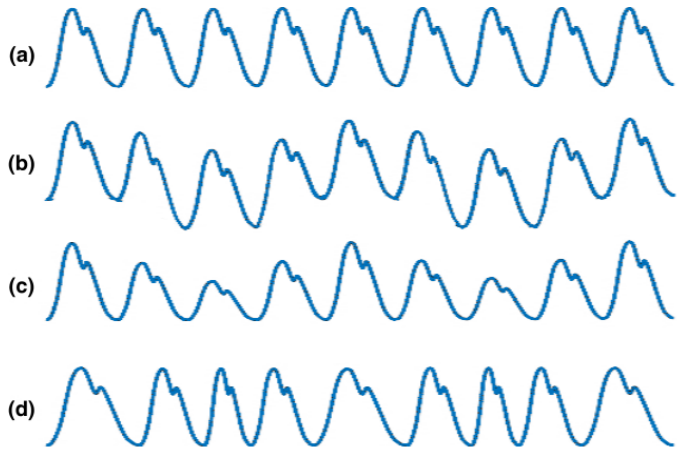


Figure 3.13: Modulations of the PPG waveform due to respiration: (a) unmodulated PPG waveform, (b) respiratory-induced intensity variation (RIIV), (c) respiratory-induced amplitude variation (RIAV), (d) respiratory-induced frequency variation (RIFV) (Source: [3])

The addition of a reliable estimation of respiratory rate from the PPG signal would improve user acceptability by reducing the burden of wearing additional continuous vital-sign monitoring hardware, reducing cabling and decreasing the costs of implementing and maintaining new technology.

Respiration causes variation in the peripheral circulation, making it possible to design signal processing algorithms to estimate respiratory rate using a pulse oximeter. The respiratory cycle modulates the PPG in three fundamental ways [3, 67] (shown in figure 3.13):

Respiratory-induced intensity variations (RIIV) : A modulation of the DC component of the PPG waveform, believed to be caused by changes in venous return throughout the respiratory cycle [3]. During inspiration, a decrease in intra-thoracic pressure results in a decrease in central venous pressure increasing venous return, the opposite occurring during expiration [149, 242].

Respiratory-induced amplitude variation (RIAV) : The amplitude of the PPG signal is defined as the difference between corresponding peaks and troughs. RIAV is thought to be caused by a decrease in stroke volume from the left ventricle during changes in intra-thoracic pressure in the inspiratory cycle.

Respiratory-induced frequency variation (RIFV) : A change of instantaneous heart rate during the respiratory cycle, also known as the respiratory sinus arrhythmia (RSA). Heart rate increases during inspiration and decreases during expiration [3].

Methods for estimating respiratory rate from the PPG signal take advantage of one or more of the effects discussed above and are shown in figure 3.13. Johansson et al [149], designed and evaluated a new device to estimate heart rate and respiratory rate continuously from eight-hour recordings of six infants in a neonatal intensive care unit. By filtering the PPG signal using a 16^{th} order Bessel filter with an adaptive pass-band, the respiratory rate was estimated using a zero-crossing method. The estimated respiratory rate was reported to have 2.7% false negative and 1.5% false positive breaths comparing to the reference respiratory signal (electrical impedance plethysmography).

Digital filtering methods have also been used in the adult population. Nilsson et al [244] recorded the PPG waveform from 12 patients with a mean age of 66 years. The reported sensitivity and specificity was 86% and 96% respectively in comparison to the reference respiratory signal from an end-tidal CO_2 device.

Nakajima et al [236] study 11 healthy subjects at rest and during exercise. Using adaptive filtering techniques, heart rate and respiratory rate were separated from the PPG signal and the results were compared with other conventional methods. The maximum error of heart rate and respiratory rates was 10 $\mathrm{beats/min}$ and 7 $\mathrm{breaths/min}$ respectively.

The Fast Fourier Transform (FFT) was used in [155] to obtain the frequency content of each of the respiratory-induced variations shown in figure 3.13 and combining the results to present a reliable respiratory rate estimation on 29 children and 13 adults. For spontaneous breathing and positive pressure ventilation under anaesthesia, Nilsson et al [243] showed that the RIIV waveform recorded with a reflection-mode pulse oximeter located on the forearm was correlated in time with the changes in central and peripheral venous pressure, arterial pressure, tidal volume and respiratory rate.

Fleming et al [92, 93, 94] used autoregressive models to estimate the respiratory rate from the PPG waveform and compared their results to those obtained with wavelet decomposition, showing that their proposed method outperformed existing techniques and had the potential to be used for real-time estimation.

Shah et al [293] also used autoregressive models and developed an automated method for respiratory rate estimation from the PPG waveform using both the amplitude modulation in figure 3.13(b) and the heart rate variability associated with respiration in figure 3.13(d). By identifying high-quality PPG segments through the use of beat-by-beat analysis and template matching, new respiratory rate estimates were combined with other vital signs (heart rate, SpO_2 and temperature) in a data fusion system to be used in both primary care and the Emergency Department for the assessment of children suspected of having a serious infection.

Clifton et al [67] showed that, when applied to large quantities of data acquired in hospitals from patients with a diverse set of clinical conditions, the current techniques for the estimation of respiratory rate from the PPG waveform are not sufficiently reliable. By pre-processing the PPG waveform to identify high-quality and low-quality segments and perform a fusion of several techniques, a more robust estimation of respiratory rate in adult patients was shown to be possible.

3.6.9 Limitations and sources of error

Pulse oximetry has proven to be an invaluable tool for the clinician as a non-invasive, accurate and safe method for the measurement of arterial oxygen saturation [205]. Nevertheless, there are several factors that can affect the accuracy of the readings provided by a pulse oximeter that need to be considered before its use. A review of the limitations of the technology can be found in [25, 205, 287, 352]. Some of the potential sources of errors are summarised below.

Accuracy at different levels of oxygen saturation

Section 3.6.6 highlighted that the relationship between the true oxygen saturation of an individual and the ratio of the red and infra-red signals recorded by a pulse oximeter cannot be established using a linear or logarithmic model such as the Beer-Lambert law. Hence, manufacturers use calibration curves derived empirically to compute the SpO_2 estimates.

Calibration data is often obtained by exposing healthy volunteers to multiple steady-state episodes of induced hypoxia for a range of oxygen saturation values typically higher than 75% or 80%, a safe lower limit of tolerable steady-state hypoxia [152, 291]. The accuracy of pulse oximeters therefore depends on the accuracy of the calibration curves used. Thus, it is commonly accepted in the clinical community that the accuracy of pulse oximeters is good for a normal range of saturation values and for periods during which a good-quality signal can be recorded. However, for lower saturation values, the readings of a pulse oximeter are less accurate.

Normal saturation values, as measured by a gold-standard device such as a co-oximeter, are typically defined as having a fractional oxygen saturation (see equation (3.2)) between 90% and 97% [352]. The performance of pulse oximeters on healthy individuals for this range is generally good. Morris et al [232] evaluated 15 pulse oximeters; the mean errors were no greater than 2.3% in all cases, The standard deviation was between 2% and 3% for 14 devices, and only one pulse oximeter had a standard deviation greater than 3%. Nickerson et al [239] compared four pulse oximeter manufacturers with two conventional arterial oximeters in five healthy subjects, the absolute mean error ranging from 0.4% to 2.6% with a standard deviation of 2%.

The accuracy of pulse oximeters has also been evaluated in critically ill patients. Louw et al [195] analysed 120 patients admitted to the ICU. The mean difference was less than 1% with a standard deviation of 2.1% when oxygen saturation was 90% or above; however, only 323 data points were recorded for periods during which a reliable pulse oximeter signal could be obtained and patients were not suspected to have significant levels of methaemoglobin or carboxyhaemoglobin. The authors concluded that large differences might occur between SpO_2 and the reference SaO_2 for critically ill patients, with a SpO_2 greater than 94% recommended to ensure a SaO_2 of 90%.

Similar results were obtained by Webb et al [352] for intensive care patients. When restricting the analysis for patients with good perfusion and "normal saturation" range, the mean error was 1.0% and standard deviation less than 2%.

For "high saturation" values, given by a Fractional SO_2 greater than 97%, pulse oximeters are often limited by their software with a tendency to overestimate SpO_2. Webb et al [352] reported that for oxygen saturation values between 97% and 98%, as measured by a co-oximeter, pulse oximeters reported a value of 100% for 87% of the cases. This might be explained by the oxygen dissociation curve (see figure 3.4); for oxygen saturation values greater than 90%, a large increase in partial pressure of oxygen (PO_2) results in a small increase in SaO_2.

The accuracy of commercially-available oximeters for "low saturation" values, defined as a fractional SO_2 less than 90%, is poor. In critically ill patients, poor agreement between the oximeter and a co-oximeter has been observed [152], with errors ranging from -12% to 18% and a tendency by pulse oximeters to systematically underestimate SaO_2. Even if a one-point measurement of SaO_2 can be made, pulse oximeters cannot reliably predict continuous changes in SaO_2 [195].

Taylor et al [324] evaluated five pulse oximeters on patients with arterial lines in the operating theatre who had saturation values as low as 70%. For 63% of the readings, pulse oximeters underestimated SaO_2 by between -1% and -12%, and for 37% of the cases, they overestimated SaO_2 with errors of at least 1% and as much as 18%.

Similar results were obtained by Mihm et al [223] from monitoring 18 patients with respiratory distress, the pulse oximeters underestimated 45% of the values for which the SaO_2 was under 80% saturation and overestimated them for 55% of the cases.

Low arterial perfusion

Pulse oximetry relies on the pulsatile nature of arterial blood and assumes that the absorbance changes of the surrounding tissue, venous blood and other absorbents are minimal. There are several factors that can reduce vascular pulsation, and therefore, reduce the accuracy of pulse oximeter readings [287]. Hypothermia (low body temperature), hypotension (a sustained low blood pressure), hypovolaemia (a decrease in blood volume), peripheral vascular disease or the use of vasopressors (medicines that contract blood vessels to increase blood pressure) can negatively affect blood perfusion and make it difficult for the sensor to distinguish the true signal from the background noise.

Clayton et al [61] evaluated the performance of 20 pulse oximeters attached to the fingers of 40 patients with poor peripheral perfusion after being transferred from the operating theatre to

the cardiac surgery recovery ward. When compared to arterial blood samples analysed with a co-oximeter, most pulse oximeters failed to give a reading for several patients due to the poor quality of the signal. Only 5 pulse oximeters could measure the oxygen saturation on all 40 patients. The mean difference for the values reported ranged from 0.1% to 4.5%. 16 pulse oximeters overestimated SaO_2 with a mean error of 2.7% whereas four pulse oximeters underestimated SaO_2 with a mean error of 4.5%.

Rapid changes in oxygen saturation

Section 3.6.3 showed that the AC component in the PPG waveform, caused by the absorption of pulsatile arterial blood in synchrony with each heart beat, usually does not exceed 1% or 2% of the total absorbance of the steady-state DC component. Thus, pulse oximeter manufacturers process very small-amplitude optical signals and attempt to reject a variety of artefacts [159]. This process often requires the averaging of data over several seconds.

During data acquisition, pulse oximeters analyse the PPG signal not only to assess the quality of each pulse but also to estimate the amount of light intensity sufficient for transmission through the tissue. Moreover, pulse oximeters are usually calibrated over varying steady-state levels of saturation; therefore their response for periods during which rapid changes in oxygen saturation occur may be compromised.

Severinghaus et al [291] evaluated the performance of pulse oximeters from six manufacturers for healthy volunteers with an age range from 18-64 years old. The volunteers were subjected to 30-45 second periods of profound hypoxia at 40-70% saturation followed by rapid resaturations. The authors reported substantial errors, ranging from 4.4% to 21.2%, the most common problem being the pulse oximeters underestimating SaO_2.

The location of the measurement probes

Measurement discrepancies can occur from devices attached to different body sites. Clayton et al [62] evaluated the performance of ten pulse oximeters attached to the finger, ear, nose and forehead for 120 patients two hours after cardiac surgery. The earlobe was rubbed with an alcohol swab for 30 seconds to increase local circulation. Only 3 out of 10 finger probes and two ear probes could provide measurements for all patients. For periods during which there was good arterial perfusion, only two ear probes and two finger probes were within 4% saturation of the reference values for 95% of the readings, as advertised by the manufacturers and required by the "ISO 80601: Particular requirements for basic safety and essential performance of pulse oximeter equipment" standard [143]. The probes attached to the nose and forehead performed poorly, with an error ranging between 3.6% to 10.5% saturation.

The response time to periods of significant desaturation varies depending on the location of the probe. Severinghaus et al [291] reported a delay of about 24 to 35.1 seconds for the finger probes as opposed to only 9.6 to 19.8 seconds for the ear probes. Kagle [153] reported a similar average delay of about 24 seconds between the finger and ear probes.

Probes attached to the nose and forehead were also reported to respond faster than the finger probes to changes in saturation [352]. The slower response time for the finger probes might be caused by the greater transit time for blood to reach the finger when compared to the ear, nose or forehead [62]. Similarly, Kelleher [159] reported that probes attached to the earlobe have been found, in some cases, to produce a better signal than finger probes.

Motion artefacts

Although most pulse oximeters can tolerate small amounts of movement, motion artefacts can cause intermittent changes in the readings, a reduction in the accuracy of the estimates, loss of signal or false desaturation alarms [23].

In a randomised evaluation study of 20,802 patients across five hospitals in Denmark [228], as many as 56% of the cases for which pulse oximetry was stopped was due to patient motion. The study also reported that pulse oximeters failed to record data for 7.8% of patients with severe systemic disease, challenging the use of pulse oximetry to monitor sicker patients.

Hanning et al [118] reported that most of the motion artefacts in a clinical setting occurred as a result of the movement of the probe in relation to the skin. They suggested that probes and their connecting wiring should be adequately secured to the skin or that self-adhesive probes should be used to minimise the effects of movement for prolonged monitoring.

Dyshaemoglobinaemias

An abnormal amount of dysfunctional haemoglobins in the blood stream or the presence of disorders in which haemoglobin molecules are altered and prevented from carrying oxygen (dyshaemoglobinaemias) may lead to erroneous measurements.

Pulse oximeters can consistently overestimate the oxygen saturation when the individual is exposed to carbon monoxide, increasing the levels of carboxyhaemoglobin ($HbCO_2$). Buckley [43] reported that pulse oximeters are unreliable for subjects with a history of heavy smoking, failing to report oxygen saturation values lower than 96% even with levels of carboxyhaemoglobin as high as 44%. Barker et al [24] reported that as the levels of carboxyhaemoglobin were varied from 0% to 70% (lethal for most humans), pulse oximeters continue to read SpO_2 values greater than 90%, while the actual oxyhaemoglobin (HbO_2) fell below 30%.

The treatment of methaemoglobinaemia, a blood disorder in which an abnormal amount of methaemoglobin ($metHb$) is produced, can cause a drop in SpO_2 to values around 65% for several minutes [25, 205].

Sulfhaemoglobinaemia, a condition in which there is an excess of sulfhaemoglobin ($sulfHb$) in the blood, have not been studied extensively. Nadarajah et al [16] reported that low saturation readings were obtained from a pulse oximeter when a high concentration of $sulfHb$, as much as 28%, was present in the blood as a consequence of long-term prescription of metoclopramide, causing cyanosis.

Venous shunting

Pulse oximeters attempt to measure oxyhaemoglobin saturation assuming that the pulsatile volume change in the PPG waveform is caused by arterial blood inflow. Occasionally, small shunts can open up in the blood vessels of the skin that bypass the capillaries and send arterial blood directly into the venous system (called arterio-venous anastomoses - AVA). It is thought that this is part of the cooling/warming mechanism in humans, and varying amounts of these shunts are present in the blood vessels of the earlobe, the finger and toes [55].

The shunting of arterial blood into the veins may cause some pulsatility in the venous system and mislead the pulse oximeter into assuming that some of this poorly-saturated venous blood is in fact arterial. In these cases, the SpO_2 reading is derived from light interacting partly with arterial blood and partly with venous blood. The result leads to SpO_2 values which are lower than the "true" arterial oxygen saturation [55, 219]. Evidence of the effects of arterio-venous anastomoses on the finger tip for 20 healthy subjects was also reported by Kim et al [162].

Skin damage

Pulse oximeters only provide data for a few square centimetres of the tissue in contact with the probe. While this technology supplies important information, there are several conditions for which dermal contact is undesired [51], such as damaged skin caused by burns or lacerations, or simply by poor patient hygiene.

Light sources

Although pulse oximeters correct for ambient light, inaccuracies in SpO_2 readings have been reported with fluorescent and xenon arc lamps [152]. Swedlow [320] advised to cover the sensor with an opaque material to obtain accurate measurements of SpO_2. Interference by infrared heat lamps was also reported by [40].

The light emitting diodes (LEDs) operating at the two wavelengths used in pulse oximetry are required to emit light with a narrow band [81]. LED technology is mostly reliable, nevertheless, errors may arise from manufacturing problems. A small error in the wavelength of the LED can be translated into a large error in the absorbance, particularly for the steep slopes in the oxyhaemoglobin dissociation curve [333].

Colour dyes

There are several colour and pigments that can interfere with the estimates provided by a pulse oximeter. Intravenous dyes such as methylene blue (a medication used as a treatment of methaemoglobinaemia), indocyanine green (used for computing cardiac output) and indigo carmine (an organic salt commonly used as food colorant) can falsely cause low SpO_2 readings, these effects can persist for up to 20 min [152]. Artificial nail materials might also cause inaccurate SpO_2 readings as reported by [205]. Cote et al [74] applied several opaque nail polish colours to 14 healthy subjects. No significant errors were found for red and purple nail colours, but the errors for black, green and blue were 3%, 5% and 6% respectively [152].

Skin colour

Although it is thought that skin colour does not affect the readings of pulse oximeters, errors are reported in the literature, primarily on subjects with a dark skin. Cecil et al [49] evaluated the performance of two pulse oximeters on 152 patients, 15 of whom were black. A larger error was reported for black patients that might be due to an increased LED power output. Similarly, during an exercise testing study of 187 pulmonary patients, Ries et al [274] reported higher instances of poor quality signal and greater inaccuracies in the SpO_2 estimates for black subjects. In critically ill patients, a bias of more than 4% has been observed to occur more frequently in black (27%) than in white patients (11%) [152].

Limited knowledge of technique

Many users have only a limited understanding of pulse oximetry. One survey revealed that 30% of physicians and 93% of nurses thought that the oximeter measured partial pressure of oxygen. A more recent audit demonstrated that less than 50% of nurses and physicians were able to identify that motion artefacts, arrhythmias, and nail polish can affect the accuracy of pulse oximeters [152].

3.7 Conclusion

Pulse oximetry is a low-cost optical technique that has revolutionised clinical practice by enabling the non-invasive and continuous measurement of peripheral oxygen saturation (SpO_2). Pulse oximeters often also measure the heart rate from the PPG waveform, but the estimation of respiratory rate remains a research technique.

A pulse oximeter only measures peripheral oxygen saturation and heart rate from a single location. There have been a limited number of studies (for example those by Allen at all discussed in section 3.6.5) that quantify the effects of recording the PPG waveform at different body sites. These showed regional variations, for example between the left and right side of the body (ears, fingers or toes). A global analysis of the complex physiological variability of an individual would be desirable, but impractical with a contact sensor such as the pulse oximeter.

Modern pulse oximeters are non-invasive devices using small probes that have been used to record data from patients for several hours in a clinical setting. Long-term vital-sign monitoring, for example at home, is still an issue due to the contact nature of the devices. Most patients find the probes difficult to attach and use properly. The low compliance of patients with wearable sensors is an obstacle to successful continuous monitoring.

Chapter 4

Photoplethysmographic imaging (PPGi)

4.1 Introduction

As discussed in chapter 2, research in non-contact vital-sign monitoring technology has greatly expanded in recent years. Nevertheless, most of the techniques presented in chapter 2 still require specialised hardware and software.

In contrast to other devices, video cameras are ubiquitous. Cameras can be found from surveillance systems, traffic control or being embedded into mobile phones. The International Telecommunications Union (ITU), in its annual report, states that there is now one cell phone for every one and a half person in the planet [144].

Decades of extensive research from the computer vision community have helped to develop imaging systems that are capable of complex computations (such as face detection, identity access control or object tracking), are interactive (such as motion/gesture and body tracking in games) and can perform complex 3D reconstruction operations.

This chapter introduces the imaging sensor technology found in most consumer-grade video cameras in section 4.2. Section 4.3 summarises recent methods used for the non-contact vital-sign monitoring of heart rate, respiratory rate and peripheral oxygen saturation. Major challenges affecting video-based vital-sign monitoring are discussed in section 4.4. The chapter finishes with a discussion and conclusions in section 4.5.

4.2 Imaging sensors

An important component in any digital video camera system is its imaging sensor. Standard off-the-shelf consumer-level digital camera imaging sensors are colour blind and only convert the light intensity reaching their surface into electrical current. As the human retina has cells that are more sensitive to green light, most digital imaging sensors make use of a Bayer colour filter array (see figure 4.1) with a pattern that favours the green luminance-sensitive elements in their photo sensors [28]. The information for the red and blue channels for each pixel is typically computed as a weighted average of the surrounding pixels and do not represent the real values at these two given wavelengths.

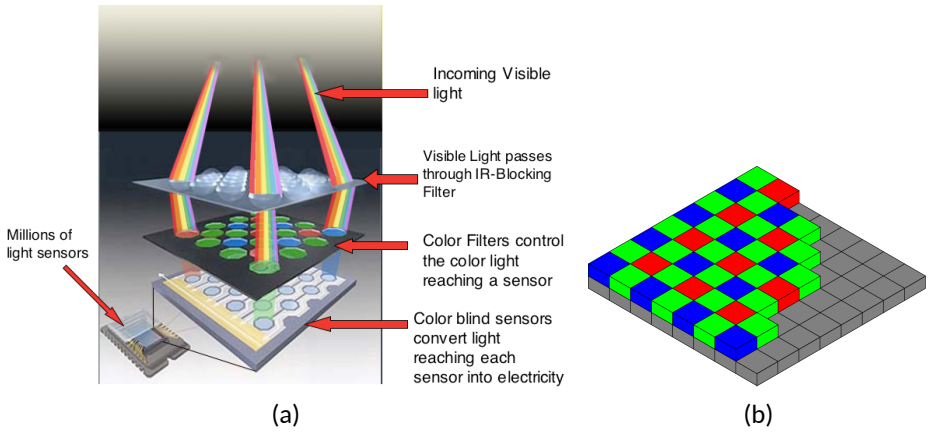


Figure 4.1: Colour filter layers in an imaging sensor (source: photoaxe.com) (a) Path of light in a typical colour imaging sensor, (b) Typical Bayer filter pattern: 50% green, 25% red and 25% blue.

Most imaging sensors for consumer devices or scientific applications in today's market are either charge—coupled device (CCD) or complementary metal—oxide semiconductor (CMOS). Each imaging sensor technology has its advantages and disadvantages. Both still play a significant role in the imaging industry today.

4.2.1 CCD imaging sensors

Charge—coupled device (CCD) imaging sensors were developed in the 1970s by Boyle and Smith [37] at AT&T Bell Labs while they were searching for a silicon-based electrical equivalent of the magnetic bubble memory. CCD imaging sensors are very responsive to visible light and, with appropriate fabrication and optimisation, can also respond well to extreme ultraviolet light and soft X-rays [346].

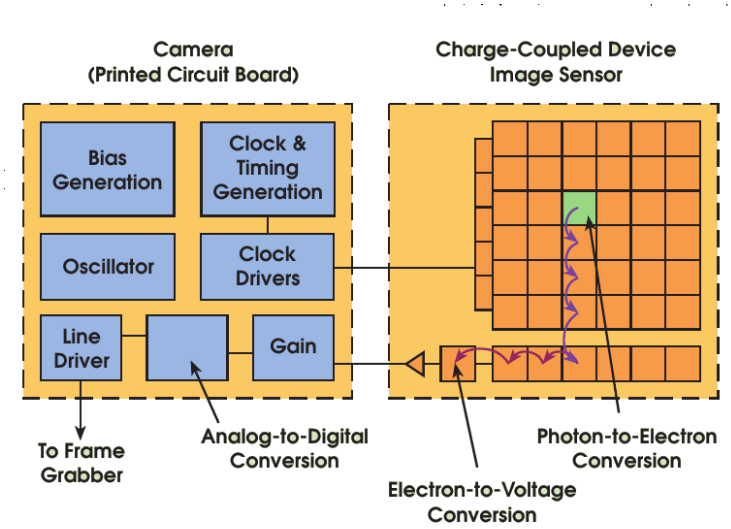


Figure 4.2: On a CCD sensor, most functions take place on the camera's printed circuit board. The circuitry can be changed without redesigning the imaging sensor. (Source: Litwiller [190]).

Figure 4.2 shows the basic structure of a CCD imaging sensor, formed from an array of electrodes running orthogonally to a series of isolated charge transfer channels. Following an exposure, the imaging area electrodes are pulsed or "clocked" to transfer the integrated image charge pattern down the array one line at a time. The lowest line is transferred into a serial readout register that runs orthogonally to the imaging area transfer channels or columns. This register is clocked separately and at a higher rate allowing each pixel to be read out sequentially through a charge detection amplifier. Once the complete line has been read out the imaging area electrodes can be clocked again so that the next image line is transferred into the readout register [346]. This sequence is repeated until all image lines have been read out.

On a CCD, most functions take place on the camera's printed circuit board. If the application demands a change, the designer can modify the processing electronics without redesigning the imaging sensor.

4.2.2 CMOS imaging sensors

Complementary metal—oxide semiconductor (CMOS) sensors, most of which today can be referred to as CMOS Active Pixel Sensors (APS), exploit the same technology used to manufacture integrated circuits such as microprocessors [37], whereby several millions of transistors can be integrated on a single silicon circuit. Instead of converting the light wave into an electrical charge on a different chip, a CMOS sensor converts the photons into electrons by processing the data at each pixel level. As a result, these sensors are usually faster than CCD technology. CMOS sensors also combine, inside each photo-detector, charge-to-voltage conversion with transistors providing buffering and addressing capability. This unique feature allows for a memory-like organisation, as seen in figure 4.3.

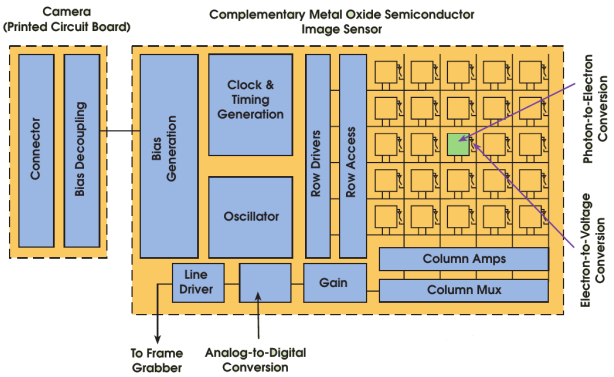


Figure 4.3: A CMOS imaging sensor converts charge to voltage at the pixel level, and most functions are integrated onto the chip. This makes image functions less flexible but, for applications in rugged environments, a CMOS camera can be more reliable. (Source: Litwiller [190]).

The relatively recent emergence of CMOS sensors has challenged CCD technology in the consumer markets of mobile phones and digital cameras, in which miniaturisation and lower power consumption are key requirements.

4.2.3 Choice of imaging sensor

Traditionally, CCD sensors have been the first-choice technology for high-end applications [202] because they can provide high sensitivity, low noise performance and good dynamic range as a result of their inherently high fill factor and full well capacity. On the other hand, CMOS detectors have intrinsic advantages, such as low power consumption, higher readout rate, radiation hardness and camera-on-a-chip integration. This makes them well suited for machine vision applications.

Manufacturers are constantly improving the performance of both imaging technologies. Bradley et al [46] conclude that for low-resolution imaging (Video Graphics Array (VGA) and below) CCD and CMOS sensor technologies are converging to be practically indistinguishable solutions in terms of performance, size and cost.

CMOS sensors generally offer higher frame rates than comparable CCD sensors. For vital sign estimation, we are interested in monitoring signals with frequencies in the physiological range for heart rate (typically from 42 to 240 beats/minute, i.e. 0.7 Hz to 4 Hz) and for respiratory rate (typically from 6 to 42 breaths/minute, i.e. 0.1 Hz to 0.7 Hz), therefore there is no need for high frame rates.

4.2.4 The imaging pipeline

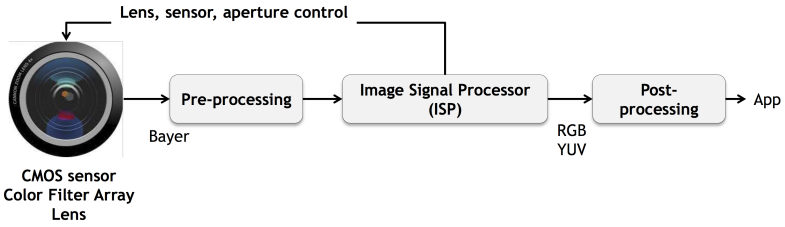


Figure 4.4: The imaging pipeline. Most cameras include circuitry for automatically adjusting exposure, focus, white balance and other settings usually not available to third-party applications (source: The Khronos Group).

Imaging sensors are not often used in isolation, but rather as part of a more complex package in a video camera. To compensate for variable changes in the scene, most consumer-level video cameras modify the sensor raw data output to adjust for exposure, focus, white balance and other settings [179, 348]. This imaging pipeline, as shown in figure 4.4, includes specialised circuitry for image signal processing and is not usually available to third-party applications.

4.3 Video-based vital-sign monitoring

4.3.1 Introduction

With the cost of digital video cameras continuing to decrease as the technology becomes more ubiquitous, research in non-contact vital-sign monitoring has greatly expanded in recent years. Most of the current algorithms proposed are based on photoplethysmography methods, that is

they attempt to record the blood volume changes associated with the cardiac cycle using a photo sensor. Hence, the term photoplethysmographic imaging (PPGi) is used throughout this thesis.

A summary of several non-contact vital-sign monitoring methods can be found in [5, 171]. The following sections present a review of the different methods for estimating vital signs using video cameras relevant to the work of this thesis.

4.3.2 Review of methods

The possibility of measuring back-scattered light remotely using a camera (rather than a reflectance probe in contact with the body) is first discussed in the scientific literature around 2005 by Wieringa et al [356]. The authors describe how reflectance-mode plethysmographic signals were acquired for seven healthy volunteers using a CMOS camera positioned 0.7 m away from the left inner arm (near the wrist). The arm was exposed to diffuse non-coherent monochromatic light at three different wavelengths (660 nm, 810 nm and 940 nm). The camera-derived signals all contained a respiration-correlated component, as well as smaller amplitude variations at the cardiac frequency. Despite the paper's title "A First Step Toward SpO_2 Camera Technology", there were no experimental results to measure oxygen saturation.

The basis for the estimation algorithms in most of the current techniques involve extracting a statistical metric (usually the mean of the light intensity of pixels) for each video frame and colour channel (red, green and blue) over one or more regions of interest defined in the human face, as in [321]. Different parts of the body such as the head, neck, arm, and palm regions have also been used in the evaluation of photoplethysmographic imaging [180].

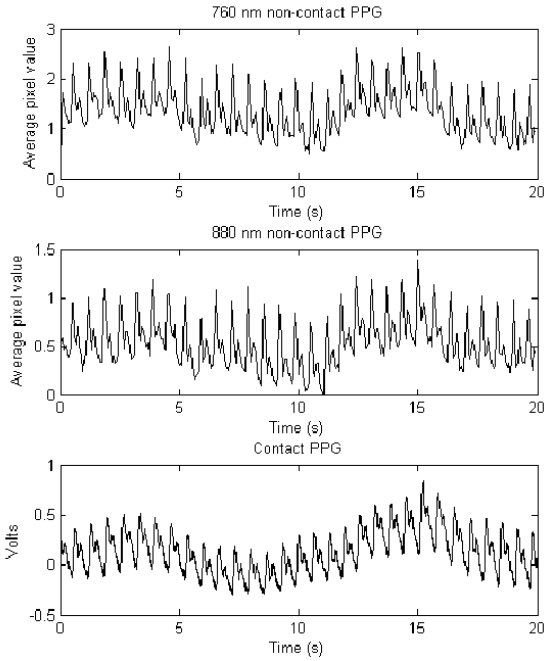


Figure 4.5: PPG signals captured simultaneously by the camera-based, non-contact device at 760 nm and 880 nm, and from the conventional finger pulse oximeter, from a 21-year-old male test subject (source: Humphreys [134]).

Humphreys et al [134, 135] investigated the feasibility of remote non-contact plethysmography and pulse oximetry, designing a device comprised of a high-quality CMOS camera and a dual wavelength array of light emitting diodes (LEDs) (at 760 nm and 880 nm). By alternately illuminating a region of tissue (usually the hand) with each wavelength of light and detecting the backscattered light with the camera at a rate of 16 frames per second per wavelength, two PPG signals were simultaneously captured. A sample output is shown in figure 4.5. The determination of oxygen saturation required calibration of the device for each test subject, which makes its implementation difficult. The choice of background illumination also had an effect on the estimation results, as the quality of the reflected light captured by the video camera is different when the sensor is not on the skin as with the traditional pulse oximeter probe.

Wieringa et al [357] designed an in-vitro phantom, perfused with human blood by a heart-lung machine to examine the feasibility of pulse oximetry by a monochrome CMOS camera with an apochromatic lens and a 3-wavelength LED ring light. 3 video sets were sequentially recorded from the phantom at 60, 80, 100 and 120 beats/min, followed by standard reflectance mode pulse oximetry. Each acquisition at 120 beats/min was followed by laboratory blood gas analysis which provided the gold standard measure. As shown in figure 4.6, the resulting images clearly revealed distinct arterial and venous phantom regions.

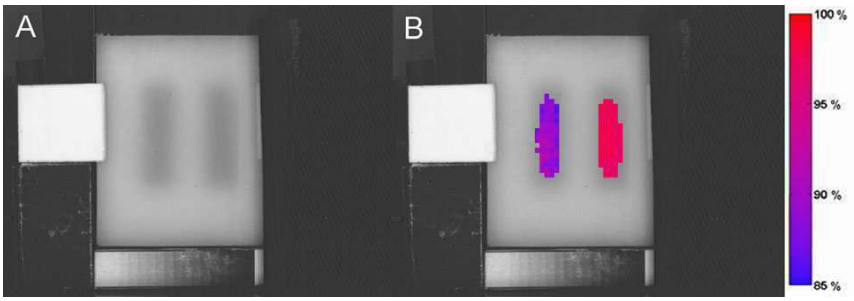


Figure 4.6: In-vitro arterial and venous phantom regions captured by a CMOS imaging sensor with an apochromatic lens. (A) near infra-red 810 nm reference, (B) dual chamber phantom filled with human blood (venous blood on the left chamber and arterial blood on the right); colour scale: red represents high saturation, blue represents low saturation values (source: Wieringa et al [357]).

Although the early work concentrated on the feasibility of non-contact oxygen saturation measurement using illuminated tissue, the focus then switched to heart rate and respiratory rate estimation with no illumination other than ambient light.

Verkruysse et al [341] showed, for the first time, that PPG signals could be remotely acquired from the human face with a simple, digital, consumer-level camera as the detector more than 1 m away; daylight was used as the illumination source in combination with normal artificial fluorescent light. A region of interest (ROI), usually from the forehead, was selected from images of the faces of human volunteers. The raw signal, calculated as the average of all pixel values in the ROI, was band-pass filtered using a 4^{th}-order Butterworth filter with cut-off frequencies of 0.8 and 6 Hz, corresponding to a heart rate range of 48 to 360 beats/min respectively.

Verkruysse's publication presented evidence that the reflectance signals were PPG signals by showing that signals corresponding to movement of facial areas with no exposed skin (edge of the face and hair above the ear) were not predominantly at the heart rate frequency. The green channel was found to provide the strongest plethysmographic signal, corresponding to an absorption peak by oxyhaemoglobin. The red and blue channels were also shown to contain plethysmographic information. The paper showed how heart rate could be extracted from the frequency content of these images using the Fast Fourier Transform (FFT) for 10-second windows, and suggested that respiratory rate might be computed using a region of interest which encompasses the entire face.

Poh et al [260] described the recording of videos from facial regions from 12 human volunteers with varying skin colours (Asians, Africans and Caucasians). They used as their detector the standard webcam embedded in a MacBook Pro laptop (Apple Inc.), positioned at approximately 0.5 m away from the subjects. The studies were all conducted indoors during the day-time, with a varying amount of sunlight as the only source of illumination. The authors' hypothesis was that volumetric changes in the facial blood vessels during the cardiac cycle modify the path length of the incident ambient light. The red, green and blue (RGB) sensors in the webcam amplified a mixture of the reflected plethysmographic signal along with other sources of fluctuations in light due to artefacts caused by motion and changes in ambient light conditions. A rectangular bounding box was set up covering most of the skin area of the face, the ROI then being defined using the full height of the box but only 60% of the width.

As in the ultra-wideband radar-based techniques described in chapter 2, correction for patient motion presents a non-trivial challenge in real-life video recordings. The time series for the three colours were decomposed into three independent source signals using the Joint Approximation Diagonalization of Eigen-matrices (JADE) algorithm for independent component analysis (ICA). An FFT was then applied to the strongest plethysmographic component (over 30-second windows) and the largest spectral peak in the region from 0.75 to 4 Hz was deemed to be the cardiac frequency.

Poh expanded his earlier work [261] to estimate the respiratory rate from healthy volunteers with a well-known indirect method [282] based on heart rate variability (HRV). The peaks of the PPG waveform (corresponding to the dominant ICA component) were identified to derive a time-series of inter-beat intervals. The spectrum of this unevenly-sampled time series was computed using the Lomb periodogram [229] with the low-frequency (LF) component and the high-frequency (HF) (respiratory) component calculated as the areas under the curve within the 0.04-0.15 Hz and 0.15-0.4 Hz bands, respectively. Scatter plots of heart rate and respiratory rate for the data acquired from the volunteers demonstrated good correlations between the webcam-derived estimates and those from a finger pulse oximeter probe (for heart rate) and a chest belt for respiratory rate.

Tsouri et al [334] investigated a constrained independent component analysis approach to improve the performance of Poh's algorithm [261]. Tests were carried out on forty-five volunteers relatively still, with one-minute video recordings in each case.

Schmitz in his Master's thesis [286], explored the concept of a video camera system to derive heart rate from one or multiple subjects using the ambient visible light spectrum (red, green and blue channels). The green channel was found to be the most suitable for heart rate estimation.

Respiratory rate estimation from HRV works well in healthy young volunteers, but it can be less accurate in elderly subjects, especially those with chronic diseases, most of which depress autonomic function [82]. However, breathing is associated with movement of the upper thorax and regions of the face. The changes in the amplitude of the PPGi waveform caused by breathing synchronous motion can be extracted by applying digital filtering algorithms and performing spectral analysis (see chapter 8). In addition, respiratory rate can also be estimated directly using motion-tracking algorithms, well known in the computer vision community. Wu et al [361] proposed a novel Eulerian Video Magnification (EVM) technique to track and amplify the motion-related changes caused by breathing in videos of human subjects, including neonates, recorded under normal lighting conditions.

Motion-induced *cardiac*-synchronous changes in the reflected ambient light measured up to a distance of 1 m away were reported by Balakrishnan [22]. These were thought to be caused by the cyclical movement of blood from the heart to the head via the carotid arteries giving rise to periodic head motion at the cardiac frequency (ballisto-cardiographic changes). The cardiac-synchronous changes in the ambient light reflected from the face appear to be the sum of three components, each of which is caused by the inflow of blood into the head during the cardiac cycle: volume changes in the facial blood vessels [260], changes in the colour of these vessels [361], and ballisto-cardiographic changes [22].

In parallel with this work in the US, Philips in Europe has developed similar methods based on FFT analysis of the remote PPG signal to estimate heart rate and motion analysis to estimate respiratory rate. This led to the release of a software application for the iOS operating system for the consumer market in 2011. There is also a family of patent applications which describe aspects of non-contact vital-sign monitoring such as the identification of people in the image [50] or the automatic selection of the measurement zone [148].

Further work from Philips researchers includes the use of multi-spectral illumination and adaptive filtering techniques to minimise the effect of motion artefacts [51]. Related theses from members of the Video & Imaging Processing Group at Philips Research Laboratories in Eindhoven focused on the issues of motion estimation and compensation [280, 285]. De Hann et al [79] proposed techniques to deal with the periodic motion arising from exercise. Colour difference ratios are used in an attempt to eliminate specular reflection components and more generally minimise the effect of motion.

Christinaki et al [60] recorded two-minute videos for seven healthy participants to evaluate the performance of three blind source separation algorithms for the estimation of the heart rate: JADE, the FastICA and RobustICA. The authors reported better performance when using the FastICA algorithm.

There are several papers in the literature which describe variations on the theme of remote PPG signal analysis to derive heart rate and occasionally respiratory rate, using for example Principal Component Analysis [182], ICA and time-frequency analysis [318] or the continuous wavelet transform [36]. However, since the early work of [356] and [134], with illumination at specific wavelengths, there appears to have been little progress in estimating oxygen saturation with non-contact methods and ambient light.

Kong et al [166] used two CCD cameras, each with a narrow-band filter (at 660 nm and 520 nm) mounted in front of the camera, but this is a more complicated arrangement than the single camera set-up which is the focus of this thesis.

4.3.3 Use of thermal cameras

A thermal camera uses an imaging sensor capable of capturing infra-red radiation at wavelengths as long as 14 $\mu\mathrm{m}$, as opposed to conventional cameras that capture ambient light (wavelengths from 450 nm to 750 nm). Thermal imaging sensors are usually monochromatic and the temperature measurement is often displayed as a colour scale with the warmest areas in white, intermediate temperatures in red or yellow, cooler areas in blue and the coolest areas in black.

Thermal cameras, depending on the construction of their imaging sensor, are mainly divided in two types: Cooled or uncooled. Cooled thermal camera sensors are often integrated in a vacuum-sealed case cooled to very low temperatures (below -150^{o}C). Uncooled imaging sensors are often based on a microbolometer, a small vanadium oxide resistor with a large temperature coefficient on a silicon element with large surface area, low heat capacity and good thermal isolation [359].

The use of thermal cameras to derive vital signs was explored in [365], in which infra-red light modulation emitted near superficial blood vessels or a nasal area was measured by the camera. Sergey et al [56] used thermal imaging to derive the arterial pulse from the subject's neck, automatically detecting regions of interest by analysing the areas of the skin where heat pulsations were most pronounced.

Tracking the human breathing signal using infra-red imaging is discussed in [371] based on a mean shift localisation based particle filtering algorithm. In [103, 104, 317], techniques to estimate blood flow velocity and vessel location using thermal imaging are discussed in detail.

4.4 Challenges

Most of the previous work in non-contact vital-sign monitoring has been performed over short time periods, under tightly controlled conditions with relatively still subjects. Subject motion presents a non-trivial challenge in real-life video recordings and is part of the focus of this thesis (see chapter 6). Tracking the location of the subject being monitored is also a non-trivial task for the non-contact methods presented in this chapter.

Another factor to consider is the interference from ambient and artificial light sources such as fluorescent, incandescent lights or TV that might affect the estimation process. In his thesis [207], Birgitta states that the colours of objects perceived by a colour camera are dependent on the illumination conditions. When the prevailing illumination condition does not correspond to the one used in the white balancing of the camera, the object colours can change their appearance due to the lack of colour constancy capabilities. Chapter 6 presents methods developed to deal with some of these problems.

4.5 Conclusion

An advantage of vital-sign monitoring using video cameras is that it can benefit from decades of extensive research from the computer vision community to help overcome the challenges presented above. Examples of the applications of such techniques are presented in later sections of this thesis.

Although the methods for non-contact vital-sign monitoring introduced in this chapter show promise, most of the current studies have been performed in well-controlled environments. The use of these methods in real-life scenarios such a hospital ward or at home requires further research. External factors such as different illumination conditions, external sources of electromagnetic noise, subject motion or change in camera perspective have an effect on the estimation of physiological signals such as heart rate, respiratory rate or oxygen saturation.

Similarly, most studies reported in the literature are usually carried out with healthy volunteers for whom physiological parameters do not vary greatly, more so if the subjects are at rest during the recording. While studies of healthy individuals can potentially produce usable data, the robustness of algorithms for estimating vital sigs is more difficult to achieve when processing the data from patients with a variety of conditions. Carefully collected hospital databases that include a diverse set of patients with a wide range of values for each physiological parameter are needed to advance such research.

Chapter 5

Dialysis study

5.1 Introduction

A major goal of this thesis is to obtain calibration data for vital-sign estimation algorithms from a real-world hospital setting. For this purpose, video was recorded from patients undergoing haemodialysis treatment in the Renal Unit of the Churchill Hospital in Oxford. Dialysis is a process by which blood is removed from the patient, filtered, and then replaced back into the body. Haemodialysis, together with kidney transplant and peritoneal dialysis, are collectively known as renal replacement therapy (RRT) methods. These methods are commonly used in the United Kingdom (UK) for the treatment of end-stage renal disease (ESRD) [219], the last stage of chronic kidney disease occurring when the kidneys can no longer meet the daily demands to remove waste products and water from the body.

According to the last report from the UK Renal Registry [19], there were 888 patients per million population receiving renal replacement therapy in 2013, a 69% increase from 2000. Although kidney transplant is the most common treatment (52%), haemodialysis accounts for 41.6% of the cases and peritoneal dialysis for the rest. ESRD is a predominantly adult disease [219], the median age of patients receiving RRT being 58.4 years, haemodialysis 66.9 years, peritoneal dialysis 63.7 years and requiring transplant 52.8 years.

One of the advantages of recording videos from patients undergoing haemodialysis treatment is that, in a relatively short amount of time (a typical dialysis session lasts 4 hours), these patients experience a wide range of physiological values (see section 5.5). This therefore helps to validate vital-sign estimation algorithms.

5.2 Haemodialysis

When the kidneys are no longer working adequately, waste products and fluid build up in the blood stream. If untreated, this can cause severe kidney failure, also known as renal failure or end-stage kidney disease, a life-threatening condition. Once severe kidney failure has been diagnosed, the patient will normally go through a form of dialysis treatment. Dialysis partially takes over the role of the failing kidneys to remove the fluid and waste that build up over time.

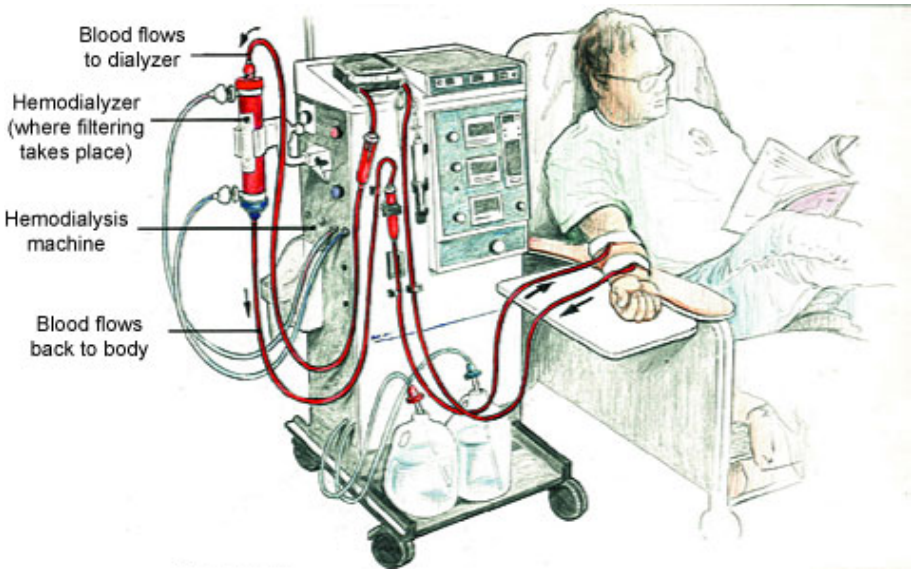


Figure 5.1: A generic haemodialysis session. Haemodialysis uses a pair of catheters in the forearm to circulate and clean the blood from the body through special artificial kidneys, known as dialysers or a kidney machine. (Source: Bhutan Kidney Foundation)

Haemodialysis, uses a pair of surgically placed catheters (often located in the forearm) to circulate blood, at a low flow rate, from the body through special artificial kidneys, known as dialysers or a kidney machine (see figure 5.1). The waste products and toxins are removed and the clean blood is returned to the body.

Haemodialysis treatments are most often performed in a clinical centre, but can also be done at home. The process takes several hours, with the patient usually lying on a bed or sitting on a chair while connected to the kidney machine.

5.2.1 Current haemodialysis monitoring

Patients on haemodialysis attend the dialysis unit two or three times a week for a treatment session lasting for approximately 4 hours each visit. Patients are weighed at the beginning and at the end of each dialysis session. Their blood pressure and heart rate are measured intermittently and their temperature may be taken once or twice during the dialysis process. Changes in their haematocrit concentration can be tracked in real-time using a Critline^{TM} (Hema Metrics, Massachusetts, USA) monitor. For the most vulnerable patients these observations are supplemented with ECG and respiratory rate monitoring.

The disparate data streams described above are recorded intermittently and manually by the nursing staff along with comments from patients to assess the physiological status of each patient. At any one time, one or more of these variables may be out of the ordinary for either real or artefactual reasons, and the staff have to use their knowledge, experience and past data on a given patient to assess the significance of these individual deviations. Often, this data is used in hindsight to understand why a patient became unwell, rather than prospectively to anticipate deterioration and prevent it, which would clearly be preferable.

The charting of the patient data is cumbersome and somewhat haphazard. Large amounts of physiological data are collected or displayed but later discarded. The process of charting and interpreting monitored data and deciding whether a corrective action is required is intermittent, manual and dependent on the workload, skills and experience of the person taking the observations.

The monitoring apparatus already in use does contain alarms for each variable being monitored. Monitors alarm quite frequently, often in response to transient artefacts or stable deviations from the norm. Unfortunately these false alarms often lead to the alarm system being silenced by the user and more significant departures from normal behaviour are overlooked.

5.3 Clinical study design

The study for which all the data analysed in this thesis was acquired, was designed to run in parallel with regular dialysis treatment sessions, which minimises the inconvenience to patients of having to give additional time for research purposes. No patient care decisions were made on the basis of the data collected during the study.

The interventions in the study were minimally invasive and similar in nature to those occasionally undertaken when more intensive monitoring is used with patients who are acutely unwell. As this requires a degree of exposure to place electrodes and monitoring probes, patient dignity and privacy were respected at all times. A precise description of which interventions were planned, and what these would entail, was discussed with the patient before the start of any involvement in the study.

5.3.1 Research ethics application

The Oxford University Hospitals (OUH) Trust is one of the largest teaching trusts of the National Health Service (NHS) in the UK. The OUH Trust requires that any clinical research studies involving patients, healthy volunteers or human tissues must be compliant with relevant governance and regulations. Therefore, a research application was submitted to the Oxford University Clinical Trials and Research Governance (CTRG) and permission was granted to perform the study (reference number 11/SC/0207).

5.3.2 Patient consent

Patients may be sensitive about being photographed, even though the video camera images were only used by the research team for the specified analysis and in an anonymised format. Individual patient consent was therefore required for this study. Patients were given full details by a member of the research team during one of their dialysis treatment episodes and they were then approached on subsequent treatment episodes to answer any questions that may have arisen and discuss whether they would be prepared to enter the study.

Once they were fully informed, patients agreeing to enter the study signed a consent form. Patients were able to drop out of the study at any time they wished, without giving any reason for their decision and without any modification to their clinical care plan.

5.3.3 Exclusion criteria

Although one of the goals of the study was to obtain physiological data from a wide range of patients, the following individuals were excluded from the study:

- Patients under 18 years old.
- Patients unable to tolerate continuous monitoring.
- Patients scheduled for kidney transplant.
- Patients who were unable to understand the study.
- Patients who chose not to participate in the study.
- Patients who are prisoners.
- Patients participating in other studies that might lead to confounding effects on either study.

5.3.4 Patient confidentiality

Clinical research studies may be audited by representatives of the University for compliance with the Department of Health Research Governance Framework for Health and Social Care. Therefore, the study was compliant with the Data Protection Act and adheres to the National Health Service (NHS) code of confidentiality, which requires patient data to be anonymised as soon as it is practical to do so.

Potential participants were identified by members of clinical staff with appropriate access to their clinical information. The clinical staff ensured that the participant's anonymity was maintained when he or she entered the study. Issues regarding data collection, storage and transfer were carefully considered in the study design. Methods by which appropriately anonymised data is collected and stored securely were introduced. Access to the data was only granted to predesignated members of the research team, and only anonymised data was released to approved collaborators.

Participants were only identifiable by their research participant number. An electronic database of anonymised clinical information was provided to the University members of the Research Team for further analysis, but identifiable personal information such as name, address or health care records were kept confidential.

5.3.5 Patient privacy during recording

The clinical study was designed to investigate the accuracy of vital-sign estimation algorithms analysing video recorded using a standard colour camera for the approximately 4 hours that a typical dialysis session lasts. It is conceivable that even though patients have consented to participate in the study and have agreed to be recorded for the entire dialysis session, there can exist periods for which the patient does not wish to be recorded. Common time periods of privacy concern included medical interventions, family visits or other personal reasons of the patient. Therefore, the clinical staff were allowed to move away the camera or stop video recording if the patient requested.

5.4 Instrumentation

The recording of patient data for a clinical study within a hospital presents many challenges. More specifically, the recording of the reference physiological signals (heart rate, respiratory rate and SpO_2 for the case of this thesis) to which the camera estimates will be compared to and evaluated against, is of great importance. Most of the studies reported in the literature use only one data source or reference device, such as one pulse oximeter recording heart rate or SpO_2 values from a single body site. There are limited studies of multi-site recordings [8] suggesting that, by studying recordings obtained simultaneously from different sites, not only important information about the peripheral circulation could be derived, but errors between measurements by the reference devices could be analysed. As such, the clinical study for this thesis was designed to have at least two reference sources for heart rate, respiratory rate and SpO_2 values.



Figure 5.2: A typical dialysis data collection set-up. The red circle shows the location of the video camera, the Nonin 4100 pulse oximeter attached to the patient's finger on the right hand and the Equivital EQ02 LifeMonitor located on the patient's chest.

Figure 5.2 shows the typical data collection set-up used for this study. Two reference devices in regular clinical use were chosen as the primary means to correlate the results of the analysis with the data extracted from the video recordings: Hidalgo's Equivital EQ02 LifeMonitor [124], shown as a chest best harness attached to the patient's chest in figure 5.2; and a separate Nonin 4100 pulse oximeter [140], typically attached to the patient's finger on the right hand as shown in figure 5.2. The Hidalgo equipment, together with all the software and support needed to extract the required physiological signals, was provided by the manufacturer. Custom software was developed from the ground up by the author to record data from the video camera and the Nonin pulse oximeter.

5.4.1 Hidalgo's Equivital EQ02 LifeMonitor

The Equivital EQ02 LifeMonitor [124] is a FDA 510(k) certified ambulatory multi-parameter vital signs telemetry device intended for the monitoring of adults (16-65 yr) in hospital care facilities, the home, workplace, and alternate care settings. The device is indicated for use as a general patient monitor, to provide physiological information as part of an occupational welfare monitoring system, and for general research and performance measurement purposes. Its usability and acceptability have been evaluated for the physiological monitoring of military personnel [326].



(a) (b) (c)

Figure 5.3: The Hidalgo's Equivital EQ02 LifeMonitor consists of (a) the battery-powered sensor electronics module (SEM) placed inside a (b) chest belt/harness. The SEM module contains not only the major circuitry to control the recording process but also serves as the storage module. (c) The chest harness contains an elastic thoracic expansion band to record the patient's respiration and also hosts the ECG electrodes (Source: Hidalgo).

As can be seen in figure 5.3, the device consists of a chest belt harness which contains skin electrodes, a thoracic expansion sensor and a battery-powered electronics module. The electronics module connects directly to the chest belt and records, digitises and transmits the physiological information wirelessly to a receiving display device.

The EQ02 LifeMonitor records two-channel electrocardiogram (ECG) waveforms at a sampling frequency of 256 Hz and a respiration signal (from the chest belt) at 25 Hz. The sensor also provides the following information: Heart rate computed from the ECG, reported every 5 seconds; Respiratory rate computed from the chest belt, reported every 5 seconds; Body orientation; Chest skin surface temperature; and indications and alerts if physiological variables exceed predefined boundaries.

The manufacturer customised the device to add a wired interface for an external Nonin pulse oximeter (Nonin OEM-III module) that was attached to the patient's ear lobe. The pulse oximeter module records a 16-bit photoplethysmogram (PPG) waveform at a sampling frequency of 75 Hz, and additional derived heart rate and oxygen saturation values at a rate of 1 estimate every 5 seconds.

5.4.2 Video and pulse oximeter recording software

This section describes the development of a data acquisition system used to capture video and pulse oximetry data to support the work on this thesis and future clinical studies in our research group. Custom user interfaces, operating system drivers and a common hardware abstraction layer were developed to follow the design guidelines for a clinical data collection system.

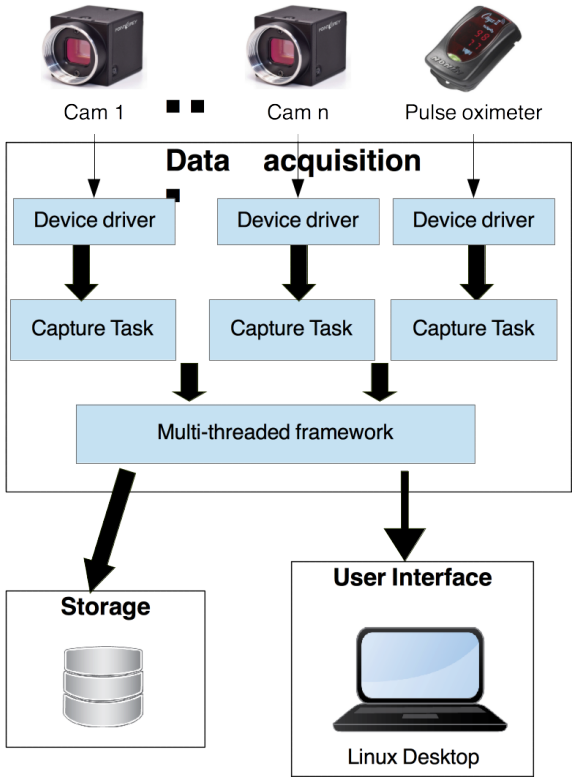


Figure 5.4: Overall software architecture for the data acquisition system

Framework for the development of medical device software

Software within a medical device or a device that interacts with medical devices should be developed and maintained within a quality and risk management system. For countries that are part of the European Union, the European Committee for Electrotechnical Standardization (CENELEC) approved in 2006 the "Medical device software - Software life-cycle processes (IEC 62304:2006)" standard [139], with Great Britain incorporating it as the British Standard "BS EN 62304:2006" [116]. As such, the data acquisition modules that interact with medical devices or the video camera were developed following the guidelines outlined in the "BS EN 62304:2006" British standard; so that the system, with minor modifications, may be submitted for medical software certification in the future.

The software was mainly written in the C/C++ programming languages, compliant with the latest "C11" and "C++14" ANSI standards. The advantage of these two languages is that there are publications and recommendations from international standard bodies that thoroughly address safety and best practices for critical systems across several industries. Therefore, the majority of the code developed also complies with the guidelines described by the three following documents: "2004 Motor Industry Software Reliability Association (MISRA) Guidelines for the use of the C Language in Critical Systems" [17], "2008 Motor Industry Software Reliability Association (MISRA) Guidelines for the use of the C++ Language in Critical Systems" [18], and "Joint strike fighter air vehicle c++ coding standards for the system development and demonstration program by Lockheed Martin Corporation, 2005" [73].

The above guidelines were an important factor that allowed the system to acquire vital signals reliably so that the image and signal processing algorithms that operate on the video camera data can produce estimates comparable to the reference clinical devices. The standards and guidelines can also help to develop portable code that runs on different platforms and form factors, from workstations to mobile devices.

Software architecture

The data acquisition system is a real-time software running in a Linux workstation, as shown in figure 5.4. It consists of mainly three modules: The data acquisition module interacts, at a low level, with devices such as cameras or pulse oximeters and also coordinates with the operating system to schedule tasks with the required priority; the storage module saves the incoming video frames and other data to disk; finally, the user interface is responsible for the display of information to the user on the workstation (the nurse in the Renal Unit).

Scheduling in the Linux kernel

A clinical data acquisition system does not necessarily depend on the fast central processing unit (CPU) computing speeds or the low latency between an action and a response, but rather its objective is to create a predictable and deterministic environment where users can run software in which they can calculate the actions of the system under a given load with deterministic results.

The standard Linux kernel provides basic UNIX operations for time handling. It only provides soft real-time implementations but has no guarantees for hard timing deadlines. Molnar and Gleixner developed a real-time patch and a generic clock event layer with a high-resolution timing support, so the Linux kernel could gain hard real-time capabilities [108, 214, 275]. Data capture from the video camera or pulse oximeter is carried out by high-performance threads that prioritise the correct sampling rate for the given device. These threads run with maximum priority and with real-time scheduling policies in a modified version of the Linux kernel.



Figure 5.5: The Nonin 4100 pulse oximeter. (Source: Nonin)

Pulse oximeter recording

The Nonin pulse oximeter model 4100 (see figure 5.5), is a clinically-approved device that can measure peripheral oxygen saturation, heart rate and also records a 16-bit PPG waveform. In collaboration with the manufacturer, we developed customised software targeted for real-time acquisition.

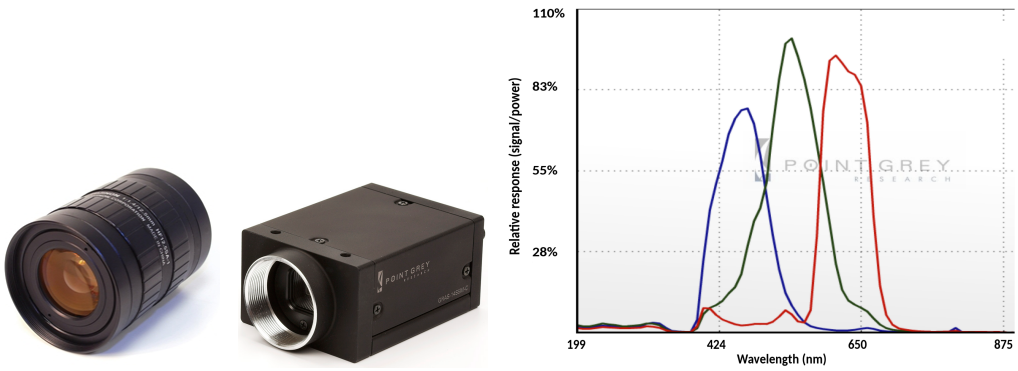
Table 5.1: Status flags for data format 7 for the Nonin pulse oximeter model 4100.

Flag	Description
SNSD	Sensor Disconnect: Sensor is not connected to oximeter
ARTF	Artefact: A detected pulse beat did not match the current pulse interval
OOT	Out Of Track: An absence of consecutive good pulse signals
SNSF	Sensor Fault: Sensor is providing unusable data for analysis
RPRF	Red Perfusion: Low signal quality
YPRF	Yellow Perfusion: Medium signal quality
GPRF	Green Perfusion: High signal quality

The pulse oximeter communicates with the data collection workstation using a Bluetooth connection. It uses the serial port profile (SPP), which defines the protocols and procedures used by devices using Bluetooth for RS232 (or similar) serial cable emulation.

The Nonin pulse oximeter can be configured for the transmission of physiological data in several formats. For our clinical study, data format 7 was used because it transmits a high-resolution 16-bit photoplethysmogram (PPG) signal at a sampling rate of 75 Hz, in conjunction with 4-beat averages of heart rate and SpO_2 at a sampling rate of 3 Hz. In addition to the physiological information, the Nonin device also provides access to several status flags that report on the condition of the pulse oximeter, such as: pulse quality, artefacts, sensor disconnect and data packets out-of-sync. Table 5.1 describes all the flags reported by the device.

Video recording



(a) (b) (c)

Figure 5.6: The video camera used in the study: (a) Fujinon HF12.5SA-1 lens, (b) PointGrey Grasshopper2 camera module. (c) Response curve for the red, green and blue wavelengths from the image sensor inside the Grasshopper2 camera: Sony ICX625 2/3" progressive scan CCD (Source: PointGrey).

Chapter 4 discussed the advantages and disadvantages of different sensor technologies. The imaging sensor is not the only factor to be considered, but rather is part of a more complex package. Collaboration with the camera system manufacturer is essential when designing clinical trials. Therefore, the Grasshopper2 camera from Point Grey Research [141] was chosen for this clinical study.

Figure 5.6 shows the camera module, lens system and the sensor response curve for the red, green and blue wavelengths. Note how the response curves for the three colour channels overlap, this might have a negative effect in the estimation of peripheral oxygen saturation (SpO_2) as discussed in chapter 9. Table 5.2 presents technical details on the imaging sensor used by the Point Grey video camera system.

Table 5.2: General features and specification for the PointGrey Grasshopper2 camera. (Source: PointGrey)

Item	Description
Imaging Sensor	Sony ICX625 2/3" progressive scan CCD
Image size (pixels)	2448 (H) x 2048 (V)
Pixel Size	3.45 μm x 3.45 μm
A/D Converter	AD9977 14-bit, dual-channel
Max frame rate	15 FPS
Video Data Output	8, 12, 16 and 24-bit digital data
Gain & Exposure	Automatic/Manual/One-Push
Lens Mount	C-mount
Interface	Gigabit Ethernet
Physical dimensions	44 (W) mm x 29 (H) mm x 58 (L) mm

Point Grey developed a feature-rich software development kit (SDK) that provides a common application programming interface (API) to configure and capture data from its cameras. A subset of the configuration options are shown in table 5.3. The SDK provides access to configure important parameters such as sensor gain, exposure time, custom frame rates and others.

Table 5.3: A sub-set of camera configuration options for Point Grey cameras.

Item	Description
Pixel format	RGB, Monochrome or Raw Bayer sensor data
Image size	Predefined or custom width and height
Time stamp	Frames can be time-stamped in-camera
Frame sequence	Each frame carries a unique sequence number
Brightness	Image luminance intensity
Exposure	The time and mode of image capture
White Balance	Colour balance adjustment
Hue	The gradation of the colours
Saturation	Colour intensity
Shutter	Frequency of the shutter being activated
Gain	Global or per colour channel sensor gain
Frame rate	Predefined or custom frame rate

The Point Grey camera used in the study has a 14-bit analogue-to-digital converter with the possibility of capturing raw Bayer sensor, providing a high dynamic range and enhancing the quality of the extracted photoplethysmographic imaging (PPGi) signal. Video data streaming is done through either Gigabit Ethernet (GigE) or USB3 interfaces, providing high data throughput and capturing video at faster frame rates than other typical cameras.

The Point Grey camera has several other advantages: each image frame is time-stamped by the camera and assigned a unique sequence number, increasing the robustness of the algorithms to detect skipped frames or periods during which the signal is corrupted.

5.4.3 Data storage

Processes regarding data collection, storage, data protection, privacy and transfer were carefully considered during the design of the clinical study. Methods and protocols by which appropriately anonymised data was collected and stored securely were incorporated and communicated to the nurses and other clinical staff involved in the recording process.

A storage cluster was acquired for the study and hosted in the Institute of Biomedical Engineering Data Centre. The storage cluster (figure 5.7) is comprised of a DELL MD3200® head node and several MD1200® storage nodes. Each MD1200® storage node contains 12 hard disks, each one with a capacity of 4 terabytes in a RAID 6 configuration with extra spare disks. This allows up to 3 simultaneous disk failures per node.

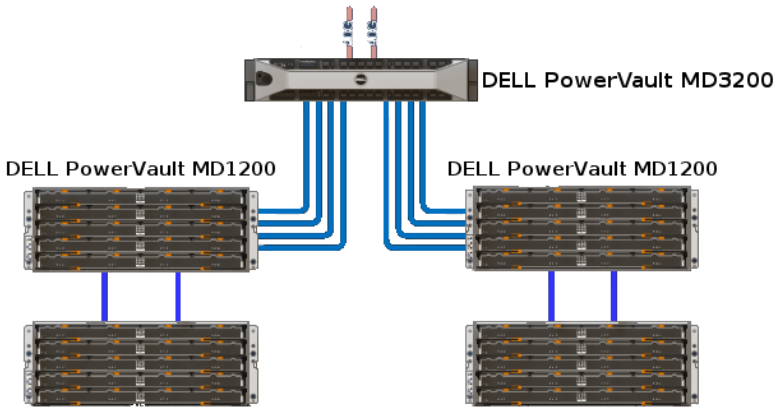


Figure 5.7: Video storage cluster diagram, comprised of a DELL MD3200® head node and several MD1200® storage nodes. Each MD1200® storage node contains 12 hard disks, each one with a capacity of 4 terabytes. (Source: DELL)

To write the captured data in real-time to persistent storage such as the internal hard disks in the recording workstation, high-performance storage tasks with robust buffering techniques were developed. As the monitoring equipment used in the clinical study produced large volumes of data, the storage software threads also run with maximum operating system priority. The captured data was saved as binary files in the respective source format of the given devices. The binary files were converted to a common format at a later stage to ease the development of algorithms.

The size of the data recorded from the Hidalgo equipment or the Nonin pulse oximeter was small when compared to the video camera. For a typical 4-hour dialysis session, uncompressed video recording at 15 frames per second (FPS) and full 24-bit resolution (8-bit for each RGB channel) at a resolution of 5 megapixels, required close to 3 terabytes of storage. This presented some challenges to the data integrity and long-term archival of the database.

For reliable video recording, a customised DELL Precision® workstation was built, housing several terabytes of storage in a RAID 5 configuration. Due to the high volume of data, the recordings from each dialysis session were individually copied to external encrypted hard disks and physically transferred on a regular basis from the hospital ward to a research storage cluster within the Institute of Biomedical Engineering in the University campus. Transferring the data electronically would have been logistically impractical.

5.4.4 Time synchronisation

For comparing the vital-sign estimation from the video camera with the estimates produced by the reference devices, accurate and reliable timestamps are necessary across all the three recording computers: the Hidalgo desktop PC, the Nonin computer and the camera workstation.

The network time protocol (NTP) [225] was used to synchronise the three computers. NTP is an internet standard protocol in regular operation over the last few years. It is used to synchronise computer clocks across thousands of hosts with high accuracy to within a few milliseconds. It is a robust protocol able to deal effectively with cases of disruptions on the network services.

The computer used for the Hidalgo system and configured by the research nurse at the beginning of every dialysis session was chosen as the time server. The camera and pulse oximeter workstations automatically synchronised their clocks with the Hidalgo computer at the beginning of every recording.

5.5 Patient population

5.5.1 Demographics

Table 5.4: Demographics for the population in the clinical study as of March 2017.

Item	Value
Total number of sessions	104
Total number of patients	40
Total video length	369.1 hrs
Average video length of a session[2]	3.5 ($\pm$0.8) hrs
Age (years)[2]	64.7 ($\pm$15.3)
Gender (males)[1]	36 (78.3%)
Height (cm)[2]	171.4 ($\pm$8.9)
Dry Weight (kg)[2]	77.1 ($\pm$15.3)
Pre-dialysis Weight (kg)[2]	78.8 ($\pm$17.1)
Post-dialysis Weight (kg)[2]	77 ($\pm$16.8)
Body Mass Index[2]	26.5 ($\pm$5.4)
Length on dialysis (months)[2,3]	30.2 ($\pm$23.3)

[1] N (percentage from total number of patients)

[2] mean ($\pm$ std)

[3] Before the start of the study

As shown in table 5.4, a total of 104 dialysis sessions from 40 patients were recorded from October 2011 to December 2012. The total length of video for all sessions is 369.1 hours, with the average session lasting 3.5 hours.

It is often the case that in dialysis studies, the patient population is comprised of elderly participants [173, 176]. The average age for a patient in the clinical study was 65 years, the majority being males (78%) with a mean body mass index (BMI) of 26.5 and have been under dialysis treatment for an average of 30 months before being recruited to the study. As of March 2017, out of the 40 patients, 18 have died (45%), 4 patients have received a kidney transplant (10%) and 18 continue to receive haemodialysis treatment in the hospital (45%).

The ethnicity of the patients enrolled in the study was predominantly white British, there were no patients of dark skin colour or high melanin content, therefore the effects of skin colour in the estimation of vital signs could not be analysed.

5.5.2 Vital signals

The distribution of values for the three physiological parameters of interest for this thesis (heart rate, respiratory rate and peripheral oxygen saturation (SpO_2)) are shown in figure 5.8 for all patients in the study. Figure 5.9 shows the recordings for the reference physiological signals for a typical 4-hour dialysis session.

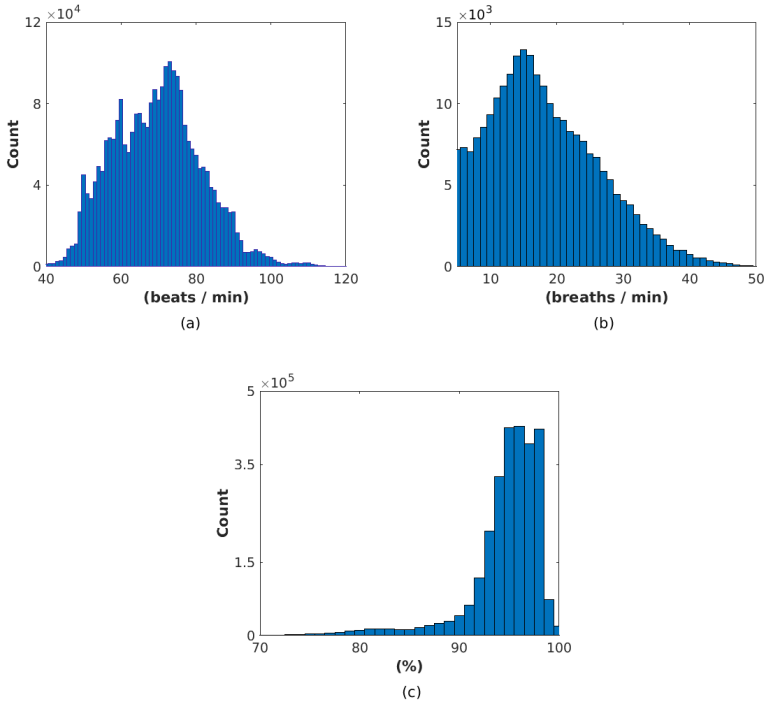


Figure 5.8: Distribution of physiological values recorded from the entire data set acquired in the dialysis study: (a) heart rate from the Nonin finger pulse oximeter, (b) respiratory rate from the Hidalgo chest belt, (c) SpO_2 from the Nonin finger pulse oximeter

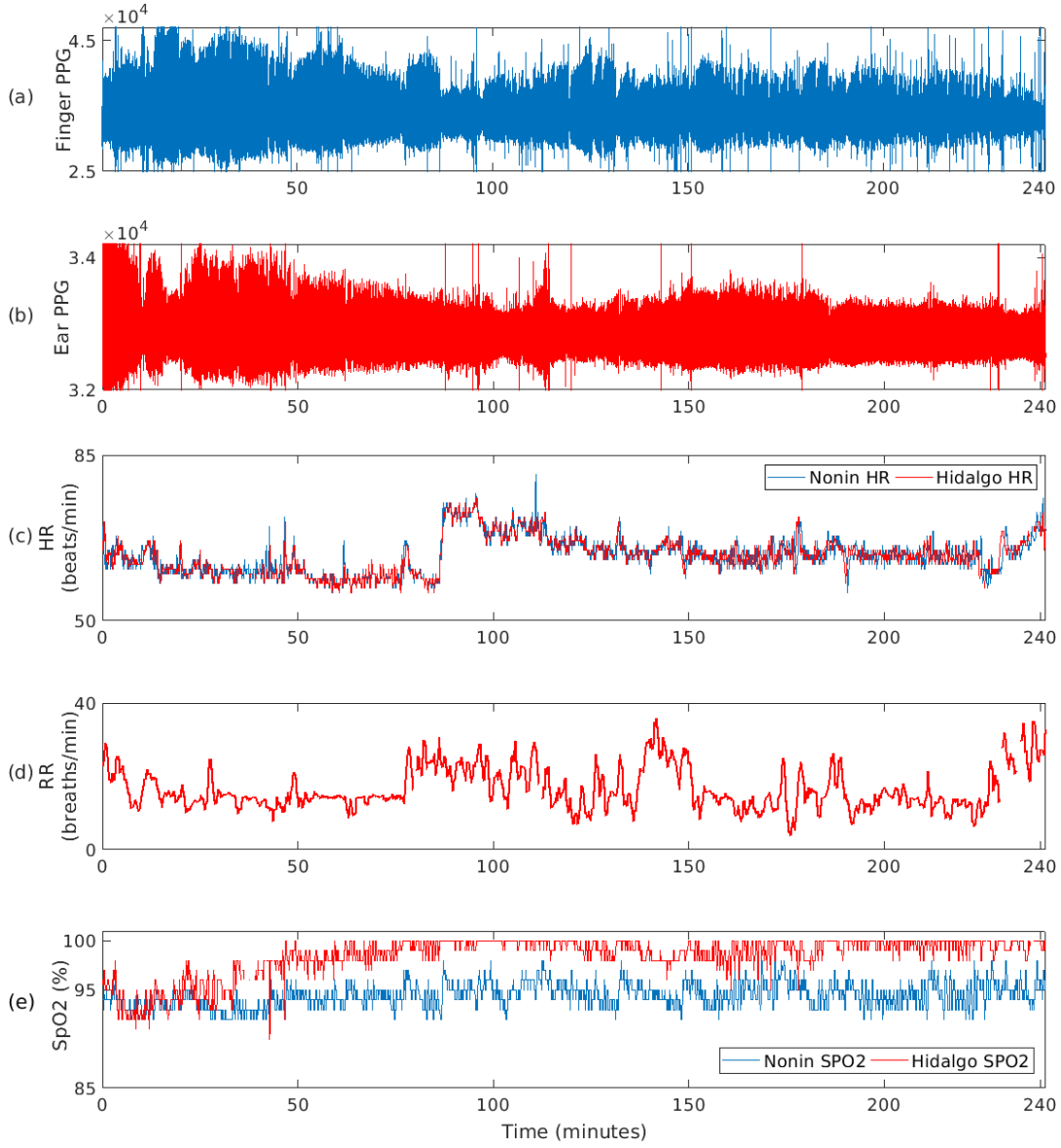
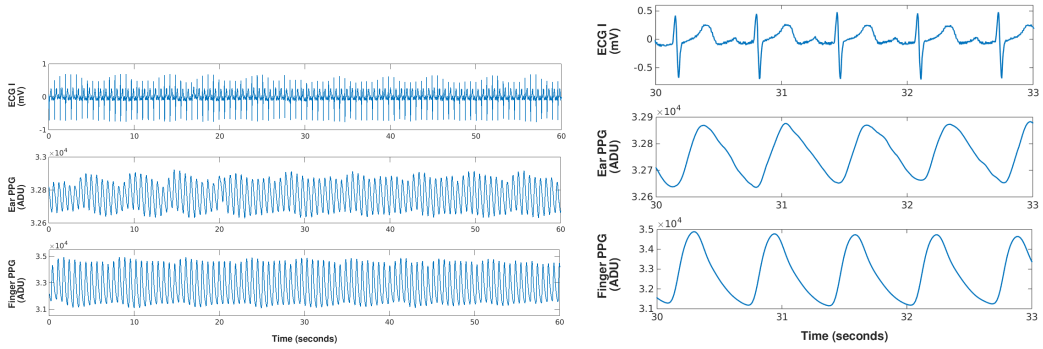


Figure 5.9: Sample 4-hour time series collected for a typical dialysis session. (a) PPG from the finger pulse oximeter, (b) PPG from ear pulse oximeter, (c) heart rate from ear (red) and finger pulse oximeters (blue), (d) respiratory rate from the Hidalgo chest belt, (e) SpO_2 from ear (red) and finger (blue) pulse oximeters.

Most studies reported in the literature use only one pulse oximeter, recording the PPG waveform from a single body site and, therefore, obtaining single measures of peripheral oxygen saturation and heart rate. There are limited studies of multi-site PPG recordings. Allen et al [8] suggested that studying pulses obtained simultaneously from different sites would allow important information about the peripheral circulation to be extracted. As discussed in the instrumentation section in this chapter, two transmission-mode pulse oximeters were used, one on the finger and other attached to the ear lobe. As seen in figure 5.10, there are some difference between the two waveforms.



(a) (b)

Figure 5.10: Photoplethysmogram (PPG) waveform recordings from two transmission-mode pulse oximeters attached to the ear lobe and finger for a typical patient undergoing haemodialysis treatment; the ECG is shown for reference. (a) 60-second recording. (b) 3-second zoomed-in section

5.6 Conclusion

This chapter has described the design and implementation of real-time software for data acquisition developed to run on a high-end Linux workstation. This software was used to collect all the video and pulse oximetry data for the dialysis clinical study.

As described in the previous chapters, most of the current work on non-contact vital-sign monitoring has been performed over short time periods, under tightly-controlled conditions with relatively still and healthy subjects. This chapter has presented the design and methodology for the development of a reference physiological database acquired in a real-life hospital environment without interfering with regular patient care. The database developed includes a diverse set of patients with varying values for each physiological parameter to be estimated: heart rate (HR), respiratory rate (RR) and peripheral oxygen saturation (SpO_2). The next chapter describes the algorithms for the estimation of heart rate from the video data stored in this database.

Chapter 6

Heart rate estimation using a video camera

6.1 Introduction

As discussed in chapters 2 and 4, several techniques have been proposed for the non-contact monitoring of heart rate. This chapter presents the methodology and algorithms that were developed for estimating heart rate using the video camera described in section 5.4.2 and visible ambient-light. The video data was recorded from patients undergoing haemodialysis treatment in the Renal Unit of the Churchill Hospital in Oxford (see chapter 5).

This chapter begins with an overview of the estimation methodology in section 6.2. Section 6.3 presents the analysis of the heart rate values provided by the two pulse oximeters used as reference devices. Section 6.4 describes the algorithms for tracking the location of patients and finding suitable regions in the images for heart rate estimation. Extracting a PPGi signal from the video camera data is described in section 6.5. The techniques to analyse the quality of the PPGi signal are discussed in section 6.6. Section 6.7 describes in detail the methods used for heart rate estimation from a single region of interest (ROI), whereas section 6.8 describes the techniques used to fuse the estimation results from multiple regions of interest. Section 6.9 presents the heart rate estimation results with respect to the two reference devices. Section 6.10.1 extends the estimation methods to derive heart rate physiology maps. Finally, conclusions are presented in section 6.11.

6.2 Overview of the estimation process

The overall process for heart rate estimation from the video camera is shown in figure 6.1. Before the camera data is processed, reference heart rate values are computed using the measurements reported by two pulse oximeters for time periods during which there is good agreement between the two devices. These reference values are used to assess the performance of the heart rate estimation algorithms from the video camera.

As our goal is to be unobtrusive and not affect regular patient care, heart rate estimation from the video camera begins by analysing the input video to identify time periods for which

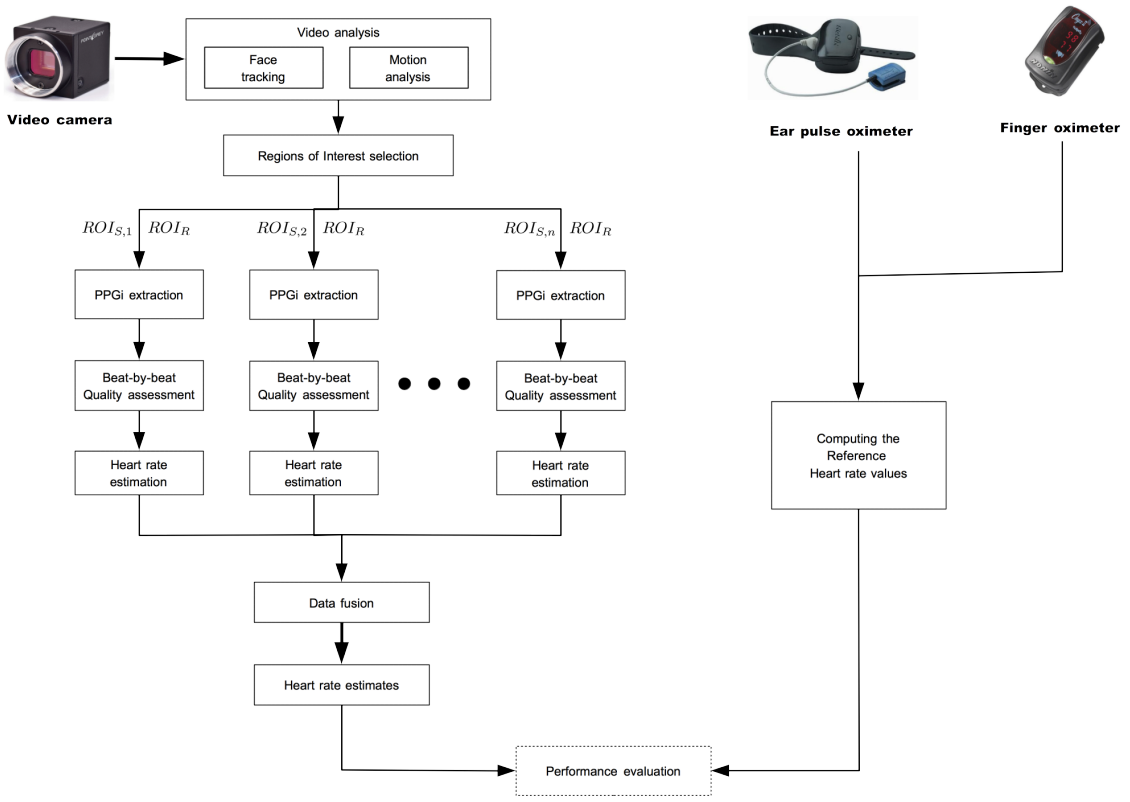


Figure 6.1: Overview of the heart rate estimation process.

the location of the patient is known, discarding periods for which there is high activity, motion artefacts arising from the regular interaction between the patient and the clinical staff, or the patient being awake and engaging in a conversation.

Once stable periods are identified from the video recordings, multiple regions of interest, labelled $\{ROI_{S,1}, ..., ROI_{S,n}\}$, are selected from the patient's face (typically from the cheek and forehead) for each video frame. One region of interest from the background (typically the wall), labelled ROI_R, is also selected to help reduce the effects of external illumination sources that are present in the Renal Unit.

Each PPGi signal is analysed independently using signal processing algorithms that include several stages: step changes in the camera signal are detected using a Bayesian change point detection algorithm; frequencies outside the heart rate physiological range are removed using a finite impulse response (FIR) band-pass digital filter; the quality of each pulse in the PPGi signal is evaluated using a template-based algorithm, as explained in section 6.6.

After the PPGi signals from the regions of interest have been processed and good-quality time periods identified, heart rate is subsequently estimated from each ROI using a time-domain method and three frequency-domain methods. To compute the overall heart rate estimate for each of the individual methods, the estimates from the regions of interest are combined using three data fusion techniques. The estimates from the camera are then compared with the reference heart rate values computed from the measurements provided by the two pulse oximeters.

6.3 Reference heart rate

6.3.1 Dual PPG monitoring during haemodialysis

The reference heart rate estimates for the data from the clinical study used for this chapter were provided by two transmission-mode pulse oximeters, one located on the finger and the other on the ear lobe (see chapter 5). Most of the studies reported in the literature use only one pulse oximeter as a reference device, recording the PPG waveform and heart rate estimates from a single site. There are limited studies of multi-site PPG recordings. Allen et al [8] suggested that, by studying pulses obtained simultaneously from different sites, important information about the peripheral circulation could be derived.

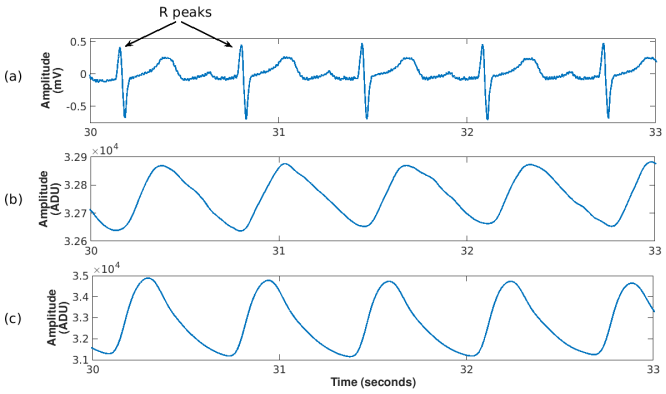


Figure 6.2: Comparison of PPG waveform recordings for a typical patient undergoing haemodialysis treatment using two transmission-mode pulse oximeters connected to (b) the ear lobe and (c) the finger. The characteristics of peripheral circulation can result in variations of timing, amplitude or both. The ECG is shown in (a) for reference, the positive-going spikes in the ECG waveform being known as the R peaks, corresponding to ventricular depolarisation.

Using heart rate estimates from more than one pulse oximeter introduces other factors to be considered. As freshly oxygenated blood is pumped from the heart's left ventricle to the aorta, it is distributed not only to the internal organs but also to the limbs throughout the body. Pulse oximeters take advantage of the pulsatile nature of arterial blood flow and are usually placed on peripheral sites such as the finger tip or ear lobe (see chapter 3). The characteristics of the peripheral circulation can result in variations in the PPG waveform recorded at different sites, including differences in timing, amplitude or both (as shown in figure 6.2). The differences in timing can negatively affect the comparison between the heart rate estimates derived from two pulse oximeters when probes are attached to two different sites.

There are other non-physiological factors that can affect the timing and values of the two recorded PPG waveforms and the heart rate estimates derived from them. Although both pulse oximeters used in our study were manufactured by Nonin, they correspond to different generations of hardware, using different versions of software and different connection technologies. The pulse oximeter attached to the ear lobe was connected to the Hidalgo ambulatory device through a physical wire using a serial connection, its PPG signal and heart rate estimates were further processed by Hidalgo's proprietary algorithms. In contrast, the raw PPG signal and the heart rate

estimates from the finger pulse oximeter were directly recorded and transmitted without further processing using a Bluetooth wireless connection to the computer which also controlled the video data acquisition (see chapter 5). As a consequence, heart rates estimates are reported at different time intervals: the Hidalgo device provides a heart rate measure every 5 seconds (corresponding to a data rate of 0.2 Hz), and the finger pulse oximeter provides a heart rate estimate three times a second (corresponding to a data rate of 3 Hz).

According to the finger pulse oximeter manufacturer (Nonin), each reported heart rate estimate is the result of averaging 8 beats, representing variable time periods depending on the patient's heart rate. There was no documentation available regarding the algorithm or window size used for reporting heart rate estimates from the ear pulse oximeter. It is important to also note that the heart rate estimates from the two pulse oximeters are reported as integers, but the rounding strategy is not documented by the manufacturer.

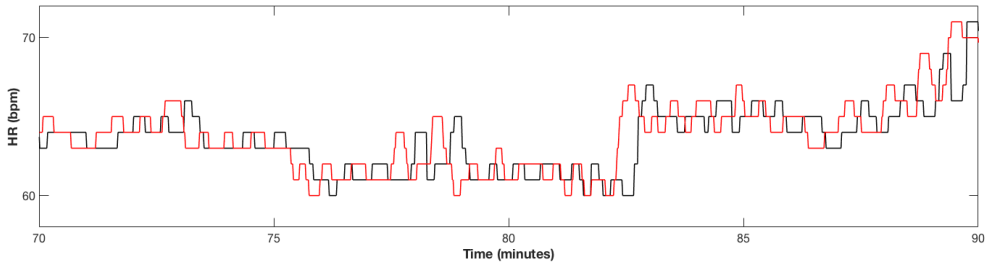


Figure 6.3: Time delay comparison between the two reference pulse oximeters for a 20-minute segment from a typical dialysis session. The ear pulse oximeter (in black) was delayed by approximately 1.2 minutes with respect to the pulse oximeter attached to the finger (in black).

Although time synchronisation was performed at the beginning of each dialysis session for all the recording equipment, the two pulse oximeters do not share the same global clock. Typical clocks are highly inaccurate, have different drift rates and are susceptible to changes in ambient temperature [120]. Because of the time needed to set up the ambulatory device worn by the patient, the Hidalgo equipment was usually synchronised with the reference time source (a separate computer used by the nurse to configure the medical equipment) around 30 minutes before the video camera recording started; moreover, the Hidalgo clock was only able to synchronise its clock with the reference time source to a precision of up to one minute (see chapter 5). The time alignment discrepancies can be seen in figure 6.3 for an example when the heart rate estimates from the ear pulse oximeter were found to be delayed by approximately 1.2 minute with respect to the pulse oximeter attached to the finger.

6.3.2 Heart rate measurement error

To measure the error between two time series $X = x_1, x_2, ..., x_n$ and $Y = y_1, y_2, ..., y_n$ of length N, two quantities are often used: the mean absolute error (MAE) and the mean absolute deviation (MAD). As its name suggests, the mean absolute error is an average of the absolute errors and is given by:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - x_i| \tag{6.1}$$

The mean absolute deviation is a measure of statistical dispersion, quantifying the average of the absolute deviations from a central point such as the mean. If $Z = z_1, z_2, ..., z_n$ with mean μ_Z represent the difference between the time series X and Y presented above, the mean absolute deviation is given by:

$$MAD = \frac{1}{N}\sum_{i=1}^{N}|z_i - \mu_Z| \tag{6.2}$$

The mean absolute error between the heart rate estimates as provided by the finger and ear pulse oximeters for the entire dialysis data set without further processing is 27.5 $\mathrm{beats/min}$ with a mean absolute deviation of 94.0 $\mathrm{beats/min}$. When the data set is restricted to contain only good-quality values (as measured by Nonin), the MAE and MAD reduces to 2.5 $\mathrm{beats/min}$ and 3.8 $\mathrm{beats/min}$ respectively. These errors are well within the maximum of $\pm 5\mathrm{beats/min}$ recommended by the Association for the Advancement of Medical Instrumentation (AAMI), and later turned into the "ANSI/AAMI EC13:2002 Cardiac monitors, heart rate meters, and alarms" standard [13].

The two metrics presented above quantify the overall error and dispersion between the values reported by the two pulse oximeters, but do not provide an insight into the underlying association between the measurements by the two devices. Correlation analysis is a statistical method frequently employed to evaluate the linear association between two variables. If correlation is found, it indicates that when there is a systematic change in one variable, there is also a systematic change in the other over a certain period of time. This linear dependence is often measured using the Pearson correlation coefficient [160]. Given two time series $X = x_1, x_2, ..., x_n$ and $Y = y_1, y_2, ..., y_n$ of length N, the correlation coefficient r is given by:

$$\begin{aligned} r &= \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{X_i - \mu_X}{\sigma_X}\right)\left(\frac{Y_i - \mu_Y}{\sigma_Y}\right) \\ &= \frac{cov(X,Y)}{\sigma_X \sigma_Y} \end{aligned} \tag{6.3}$$

where μ_X and σ_X are the mean and standard deviation of X respectively and μ_Y and σ_Y are the mean and standard deviation of Y. The correlation coefficient represents the ratio of covariance between the variables to the product of their standard deviations, ranging from -1.0 for a negative correlation to +1.0 for a positive correlation.

Typically, a scatter plot or correlation plot is used to show the relationship between the two variables by fitting a line to the values. The goodness of fit for the linear regression is given by the coefficient of determination r^2, representing the proportion of variance that the two variables have in common. Figure 6.4(d) shows the correlation plot for the heart rate estimates from the two pulse oximeters. The values from the two devices are very close to the regression line with a positive correlation coefficient of 0.91. Although a strong correlation exists between the measurements from the two devices, large non-physiological discrepancies are present.

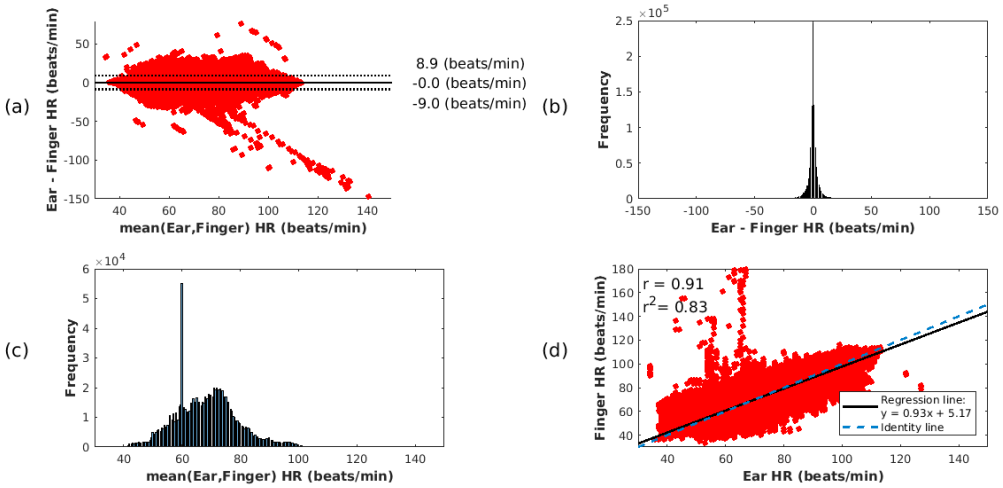


Figure 6.4: Agreement between the original heart rate values provided by the two pulse oximeters for the entire dialysis dataset. (a) The Bland-Altman plot presents no sensor bias but a large difference between the measurements. (b) The distribution of the differences between the two pulse oximeters. (c) The distribution of the mean values, spanning the heart rate physiological range for adults. (d) The correlation plot shows a high correlation between the two devices with a correlation coefficient of 0.91, even though large differences between the measurements exist.

Correlation quantifies the strength of the linear association between the heart rate estimates from the finger and ear pulse oximeters but not their differences. A high correlation does not necessarily imply that there is a good agreement between the two devices [107]. To describe the agreement between two quantitative measurements, the Bland-Altman plot was introduced in 1999 [32]. In a Bland-Altman plot, the differences between the measurements from two devices or methods are plotted against their mean. Plotting the difference against the mean allows the identification of possible relationships between the measurement errors and the true values. The mean of the differences between the devices represents the measurement bias. The Bland-Altman plot, shown in figure 6.4(a), presents no measurement bias between the devices but shows large differences. The distribution of the differences is presented in figure 6.4(b) and the distribution of the mean values from the two devices shown in figure 6.4(c).

It is also important to test if the differences between the two devices, shown in figure 6.4(b), are normally distributed. The Jarque-Bera test is a goodness-of-fit test to estimate if the histogram data have the skewness and kurtosis matching a normal distribution [146]. It is defined as follows:

$$JB = \frac{n}{6}\left(s^2 + \frac{(k-3)^2}{4}\right) \tag{6.4}$$

where n is the sample size, s is the sample skewness, and k is the sample kurtosis, defined by:

$$s = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{\frac{3}{2}}} \tag{6.5}$$

$$k = \frac{\hat{\mu}_4}{\hat{\sigma}^4} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{2}} \tag{6.6}$$

where $\widehat{\mu}_3$ and $\widehat{\mu}_4$ are the estimates of third and fourth central moments respectively, $\bar{x}$ is the sample mean and $\widehat{\sigma}^2$ is the estimate of the variance.

Given the error analysis presented above, a direct comparison (on a sample-by-sample basis) of the heart rate estimates between the video camera and each of the two pulse oximeters is potentially affected not only by physiological factors but also by the recording set-up and each of the manufacturer's processing context, leading to inaccurate performance results. Furthermore, the measurement of a physiological process implies some degree of error; when two sensing devices are used, neither provides an absolute correct measurement. Since the true value is not known, the mean of the two measurements is usually taken as the representative value [33]. Therefore, for the analysis described in this chapter, a new reference heart rate time series is computed from the average of the reported values from the two pulse oximeters. This new "ground truth" reference heart rate is used for comparison with the estimates derived from the video camera.

6.3.3 Computing the reference heart rate

To compute the new reference heart rate, a three-stage process is proposed. Initially, as the two pulse oximeters provide values at different time intervals, the heart rate estimates from each of the two pulse oximeters are re-sampled to 1 Hz to match the video camera estimates reported every second. Secondly, the two signals are time aligned by subtracting the sample delay as computed from the lag for which the normalised cross-correlation has the largest absolute value for the entire recording session. Finally, the new reference heart rate is computed, on a second-by-second basis, as the mean of the estimates from each pulse oximeter that do not vary by more than 5 $\mathrm{beats/min}$ as recommended by the "ANSI/AAMI EC13:2002 Cardiac monitors, heart rate meters, and alarms" standard [13].

The normalised cross-correlation is a standard method of estimating the time delay between two signals. Given two time series $X = x_1, x_2, ..., x_n$ and $Y = y_1, y_2, ..., y_n$ of length N, cross-correlation measures the similarity between X and shifted (in time) copies of Y as a function of the time displacement, known as the lag. The normalised cross-correlation, at time lag t, is defined by:

$$\begin{aligned} R_{xy}(t) &= \frac{E\{(x_i - \mu_X)(y_{i-t} - \mu_Y)\}}{\sigma_X \sigma_Y} \\ &= \frac{\sum_{i=1}^{N}[(x_i - \mu_X)(y_{i-t} - \mu_Y)]}{\sqrt{\sum_{i=1}^{N}(x_i - \mu_X)^2}\sqrt{\sum_{i=1}^{N}(y_{i-t} - \mu_Y)^2}} \end{aligned} \tag{6.7}$$

The time delay D_{xy} of the signal Y with respect to X is computed as occurring at the maximum of the normalised cross-correlation for all possible time lags t:

$$D_{xy} = \arg\max_t |R_{xy}(t)| \tag{6.8}$$

The heart rate of a patient in a hospital setting is not constant, but rather highly non-stationary. The sudden increase or decrease in the heart rate time series can negatively affect the computation of the time delay between the two pulse oximeters. To lessen these effects, a median filter is applied to the heart rate values from the two pulse oximeters before computing the cross-correlation. This median filter is only used to compute the time delay between the signals and not for computing the final reference heart rate against which the camera estimates are compared.

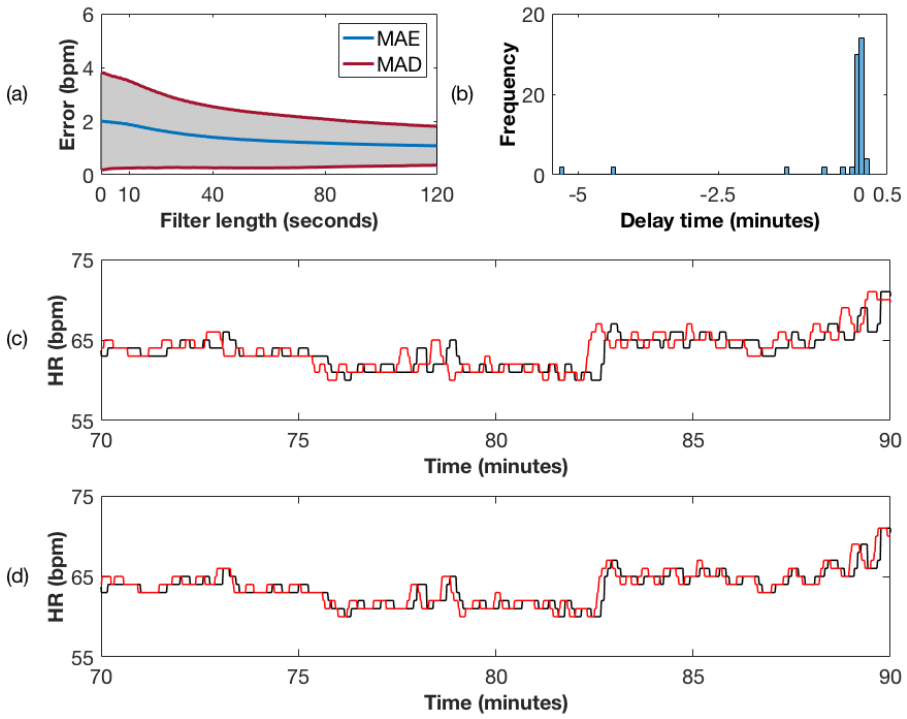


Figure 6.5: Time alignment between the heart rate estimates provided by the two reference pulse oximeters across the entire population in the study. (a) Error when the length of the median filtering applied is varied from 0 to 120 seconds. (b) A summary of the overall time delays across the data set. (c) Heart rate values as reported by the ear pulse oximeter (in black) and the finger pulse oximeter (in red) for a sample 20-minute window from a typical dialysis session, (d) time aligned signals after computing the delay from the normalised cross-correlation.

The length of the median filter applied to the heart rate estimates from the two pulse oximeters is chosen as a compromise between reducing the mean absolute error between the two reference devices and making the filter length comparable to the 15-second windows used to derive the camera estimates (as described in section 6.7 below).

Figure 6.5(a) shows the mean absolute error and the mean absolute deviation as the length of the median filter is varied from 0 to 120 seconds prior to time alignment. The overall time delays, when an 11-second median filter is used, across the population in the study is shown in figure 6.5(b) with most of the sessions having minimal time differences of under 0.3 minutes. A sample 20-minute segment is shown in figure 6.5(c) before time alignment, and the result of aligning the two time series is shown in figure 6.5(d).

6.3.4 Results

According to table 5.4 in chapter 5, 104 dialysis video sessions were recorded from 40 patient corresponding to a total recording time of 369.1 hours. For 12 sessions, the recording from at least one pulse oximeter was interrupted due to patient discomfort, medical intervention, equipment malfunction or other external factors. 8 sessions were discarded because the video camera recording equipment was interrupted by the clinical staff or the recording was not complete. Therefore, there were simultaneous recordings with both pulse oximeters and camera for only 84 sessions, consisting of a total recording time of 304.1 hours.

Table 6.1: Comparison of the heart rate estimates provided by the two reference pulse oximeters placed on the ear lobe and the finger for the overall patient population, comprising a total recording length of 304.1 hours.

Data	MAE ($\mathrm{beats/min}$)	MAD ($\mathrm{beats/min}$)	Time hours (%)
Original values, no processing	27.5	94.0	304.1 (100%)
Original values, good quality values	2.5	3.8	263.5 (86.6%)
Reference HR computation	1.2	1.3	242.7 (79.7%)

Table 6.1 shows the error between the heart rate estimates from the two pulse oximeters. The original values provided by the manufacturers have a large mean absolute error of 27.5 $\mathrm{beats/min}$ with a mean absolute deviation of 94.0 $\mathrm{beats/min}$. When comparing only the values for which the manufacturer reported as good-quality estimates, the error reduces to a MAE of 2.5 $\mathrm{beats/min}$ and a mean absolute deviation of 3.8 $\mathrm{beats/min}$ for 86.6% of the time, corresponding to approximately 263.5 hours. This is because for approximately 13.4% of the time, corresponding to 40.6 hours, the heart rate measurements from at least one of the pulse oximeters were flagged as errors by the manufacturer (see table 5.1 in section 5.4.2 for the status flags reported by the pulse oximeter).

When the heart rate estimates from the two pulse oximeters are aligned in time and the new reference heart rate is computed, only 79.7% (approximately 242.7 hours) of heart rate values differ by no more than 5 $\mathrm{beats/min}$ with a mean absolute error of 1.2 $\mathrm{beats/min}$ and a mean absolute deviation of 1.3 $\mathrm{beats/min}$.

Figure 6.6 shows in more detail the agreement of the heart rate estimates from the two pulse oximeters for the values used in computing the reference heart rate. The Bland-Altman plot, shown in figure 6.6(a), presents no sensor bias with narrow differences across the heart rate physiological range, implying a good agreement between the two measurements. The distribution of the differences, shown in Figure 6.6(b), is normally distributed. The distribution of the mean values in figure 6.6(c), shows that the values are within the expected physiological range for adults. The values from the two devices are now highly correlated with a positive correlation coefficient of 0.99 shown in figure 6.6(d).

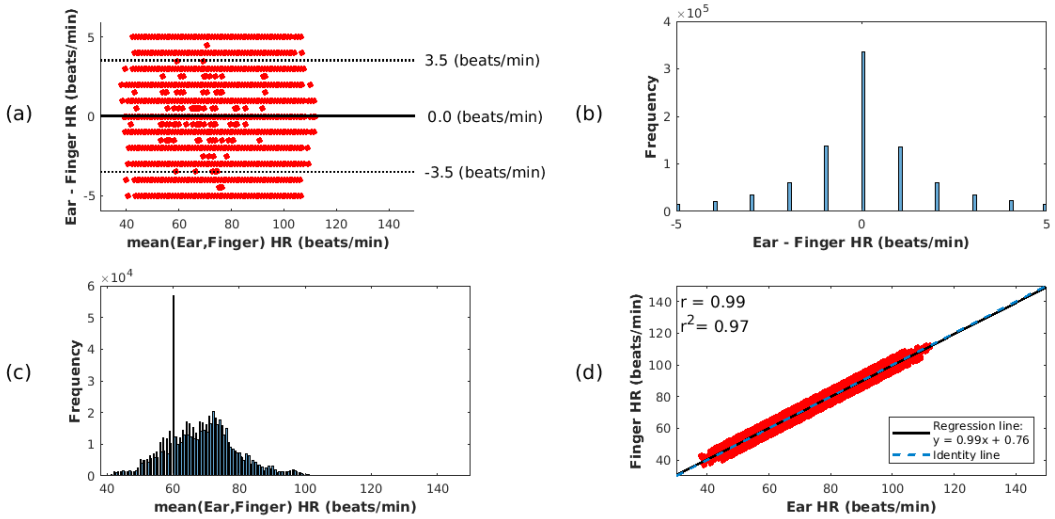


Figure 6.6: Agreement between the heart rate values provided by the two pulse oximeters used to compute the reference heart rate for the entire dialysis dataset, comprising a total recording time of approximately 304.1 hrs. (a) The Bland-Altman plot presents no sensor bias. (b) The distribution of the difference between the two pulse oximeters is normally distributed. (c) The distribution of the mean values show that most of the heart rate estimates are within the expected physiological range for adults. (d) The plot shows a strong correlation between the two devices with a correlation coefficient of 0.99.

6.3.5 Conclusion

Given the analysis described above, when the heart rate estimates from the two pulse oximeters are aligned in time, the measurements are within 5 $\mathrm{beats/min}$ on average, both time series have a high positive correlation and are in good agreement. Therefore, the proposed reference heart rate, computed from the mean of the estimates from the two pulse oximeters, can be used as the ground truth and be compared to the estimates derived from the camera.

6.4 Video analysis

Most of the previous work in non-contact vital sign monitoring has been performed over short time periods of typically a few minutes, recorded from healthy young volunteers in tightly-controlled conditions (see chapters 2 and 4 for more detail). While studies in such controlled environments can potentially be used to develop and test new technology, the robustness of algorithms for estimating the values of vital signs such as heart rate is challenged when processing video data recorded from subjects under real-life hospital conditions.

Figure 6.7 shows the typical set-up for recording video data during a typical 4-hour dialysis session, including some examples of common patient behaviour during the recordings. To estimate heart rate accurately, suitable time periods need to be identified from the videos during which the patient is stable and in the frame. Examples of such periods are when the patient's torso is frontal to the camera (figure 6.7(b)), the patient is sleeping quietly (figure 6.7(i)) or when minimal motion occurs (figure 6.7(d) if the image is rotated to portrait mode).

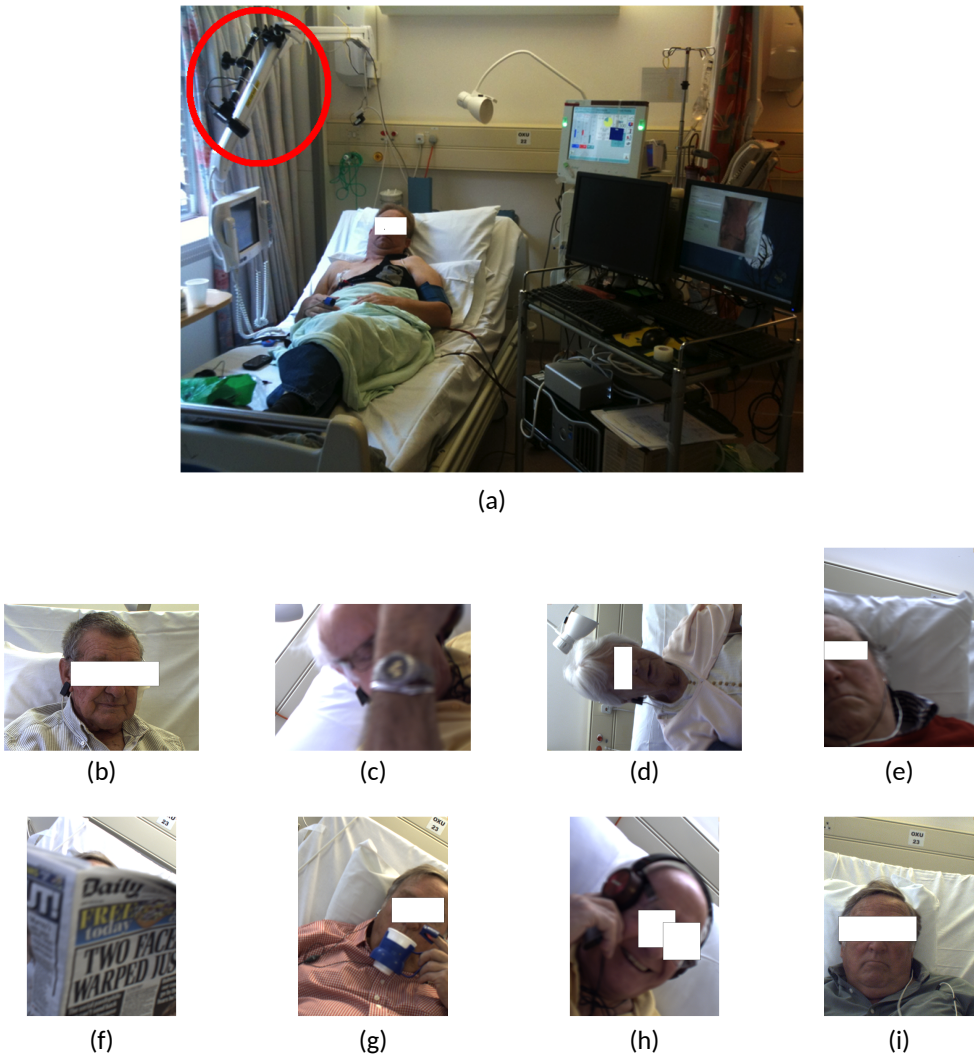


Figure 6.7: Video recordings from the Renal Unit: (a) typical data collection set-up, with the red circle showing the camera location. (b) through (i) are image examples from typical dialysis sessions where the patient is: (b) correctly in frame, (c) moving freely with limbs obstructing the frame, (d) being filmed with the camera at an angle, (e) sleeping with torso half frame, (f) reading the newspaper and obstructing the frame, (g) having tea, (h) listening to music and speaking on a mobile phone, (i) sleeping and correctly in frame.

One of the goals of our study was not to interfere with regular patient care and, as a result, the video recordings were affected by several external sources of distortion, such as:

- Regular interaction between the patient and the clinical staff.
- Camera angle changed to allow access for patient care (see figure 6.7(d)).
- Patients regularly change position during the 4-hour recording (see figure 6.7(c)).
- Patient torso or head moving out of camera frame (figure 6.7(e)).
- The use of electronic devices such as mobile phones, music players or tablets (figure 6.7(h)).
- Regular consumption of beverages such as water or tea (figure 6.7(g)).
- Objects obstructing the patient's face such as newspapers (figure 6.7(f)), magazines or limbs (figure 6.7(c)).

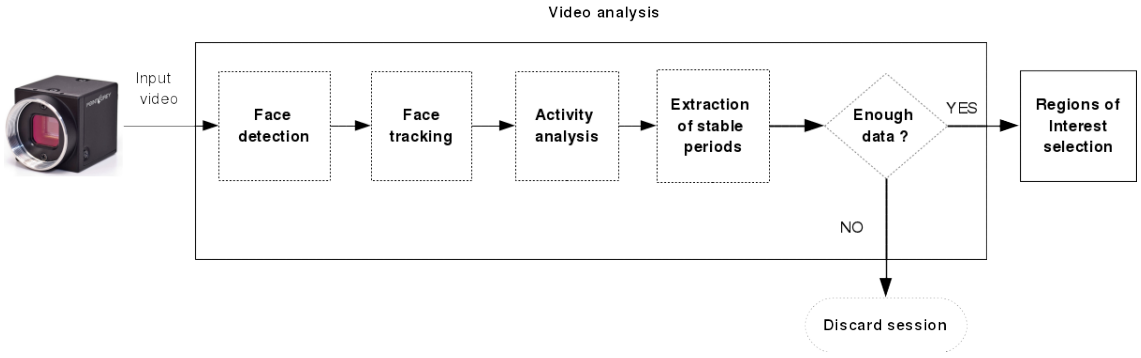


Figure 6.8: Overview of the video analysis phase. Suitable time periods for which the patient's face is located and minimal motion occurs are identified from the input video. The output is used by subsequent algorithms to select regions of interest.

Video analysis (see figure 6.8) starts with the task of detecting and tracking the patient's face using the methods described by Zisserman, Everingham and Sivic [90] and further improved in [305]. Subsequently, periods of high activity or motion are found by tracking the movements of the patient's face. The results of these tasks are combined to identify stable time periods within the video for which the location of the face is known. If the proportion of valid time periods is judged as sufficient, the output of this first video analysis phase is used by subsequent algorithms to select regions of interest from which heart rate can be estimated. On the contrary, if the proportion of valid time periods is insufficient, the session is discarded and no further analysis is performed (see section 6.4.4).

6.4.1 Face detection

Face detection is the first stage in the video processing pipeline. The goal of the face detection algorithm is to provide a rough estimate of the location and scale of face areas in the video frames. Face locations are then used as inputs for the tracking algorithms discussed in the next section.

Several algorithms for face detection are reported in the literature using cues such as skin colour, facial or head shape, facial appearance, or a combination of more than one technique [187]. Although these problems have received a lot of attention, face detection and facial feature extraction are still challenging, especially when illumination, subject expression and object occlusion vary considerably [305]. Furthermore, the process of determining the location of the patient's face from videos in a hospital environment presents other unique challenges, as shown in figure 6.7. Patients were usually active during the recordings, the clinical staff were constantly interacting with them and the camera was not necessarily positioned in a frontal plane, as access for patient care by the clinical staff was the over-riding priority.

Although the ideal face detector should be able to locate faces with high accuracy regardless of the position, orientation, age or expression of the patient, the speed performance of the face detection algorithm is also of great importance, given that the database collected for this thesis contains close to 370 hours of uncompressed video recorded at 5 megapixels per frame (see section 5.5).

The work described by Everingham et al [90] uses face detectors based on [342], capable of processing images rapidly and achieving a high detection rate. The input image is resized from the sensor's native resolution of 2448x2048 pixels by a factor of 6 to about 408x340 pixels using bicubic interpolation as a compromise between algorithm complexity and runtime performance, as discussed by Lehmann et al [181]. The resized image is then converted from RGB to greyscale where it is scanned at several possible locations and scales (from 300x300 down to 20x20 pixels) by sub-windows using several rectangular features (see figure 6.9) to classify each sub-window as either face or non-face. By transforming the input image to a representation called "integral image" (see the section below), these features can be computed rapidly.

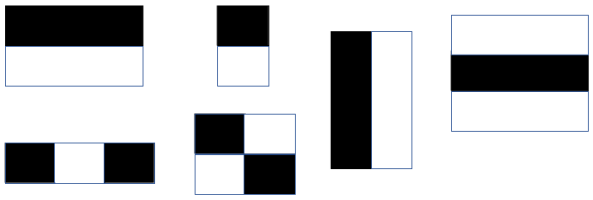


Figure 6.9: Examples of some of the rectangular features used by the face detector based on [342] to classify image sub windows as face or non-face regions. Commonly used features involve two, three and four-rectangular areas.

The face/non-face classifier is learned from training examples using a learning algorithm based on AdaBoost [101]. A cascade method of increasingly complex classifiers is used to quickly discard background regions of the image and spend more computation time on regions that are likely to contain faces.

Integral image

The sample rectangular features shown in figure 6.9 result in a single value computed by subtracting the sum of the pixel intensities in the white rectangle from the sum of the pixels in the black rectangle. To determine the presence or absence of these features at any location and scale in the whole image, the grey-scale image is transformed to an intermediate representation called the integral image. The integral image at location x, y contains the sum of the pixel intensities above and to the left of x, y inclusive, defined by:

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \tag{6.9}$$

where $ii(x, y)$ is the integral image and $i(x', y')$ are the pixel intensities from the original image. The integral image can be computed in a single pass with the following equation:

$$ii(x, y) = i(x, y) + ii(x - 1, y) + ii(x, y - 1) - ii(x - 1, y - 1)) \tag{6.10}$$

Once the integral is constructed, any rectangular features can be computed at any scale or location in constant time as shown in figure 6.10.

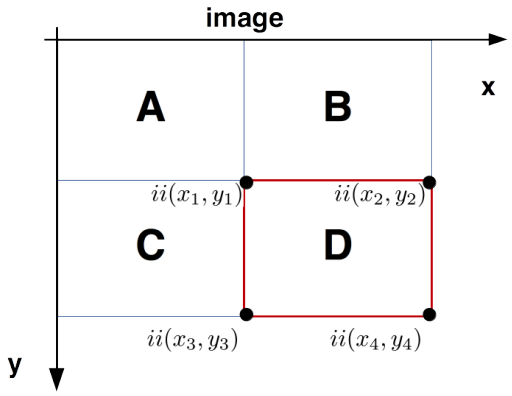


Figure 6.10: Computing the sum of the pixel intensities in rectangle D from the integral image can be done in constant time for any size and scale using the formula: $sum(D) = ii(x_4, y_4) - ii(x_2, y_2) - ii(x_3, y_3) + ii(x_1, y_1)$

Cascade classifier

The total number of possible sub-windows in the input image is large, and the total number of rectangular features in a sub-window is even larger. However, the number of potential faces in each video frame is usually one (the patient). Although the integral image helps to compute the features rapidly, a method to select a small number of critical features is needed to quickly reject sub-windows that do not contain a face and therefore not incur the penalty of having to compute every possible feature for every sub-window. These sets of minimal features are selected in the training phase using Adaptive Boosting (AdaBoost [101]) and adjusting the thresholds to minimise false negatives. AdaBoost combines many weak classifiers based on a single rectangular feature to create a final strong classifier.

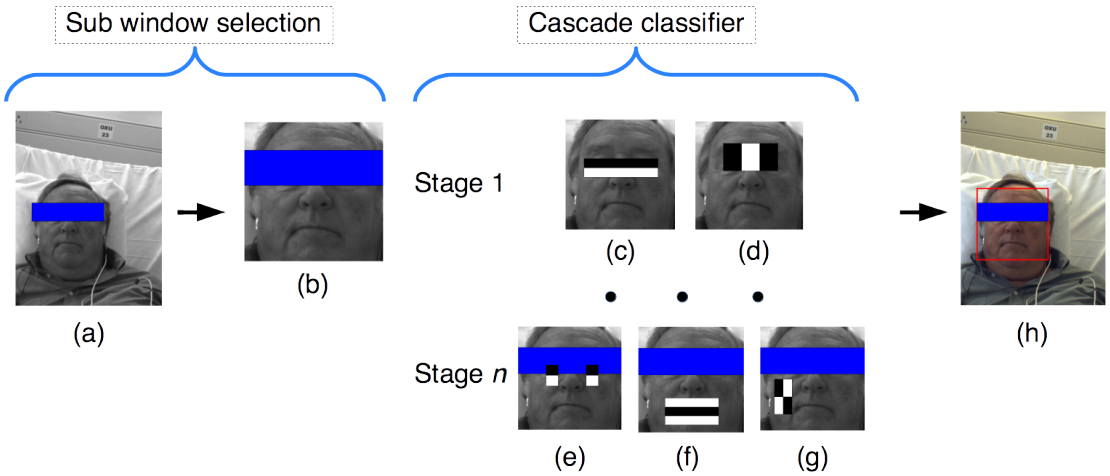


Figure 6.11: Face detection for typical video recordings from the Renal Unit (blue rectangle to protect patient privacy): (a) grey-scale input image, (b) sample sub window selected. Stage 1 of the classifier comprises looking for regions where (c) the eyes are darker than the region of the cheeks and (d) the eyes are darker than the nose bridge. Subsequent stages can look for other rectangular features such as (e) individual eyes, (f) mouth darker than the lips and chin or (g) patterns in the cheeks. The face detected area is shown in (h).

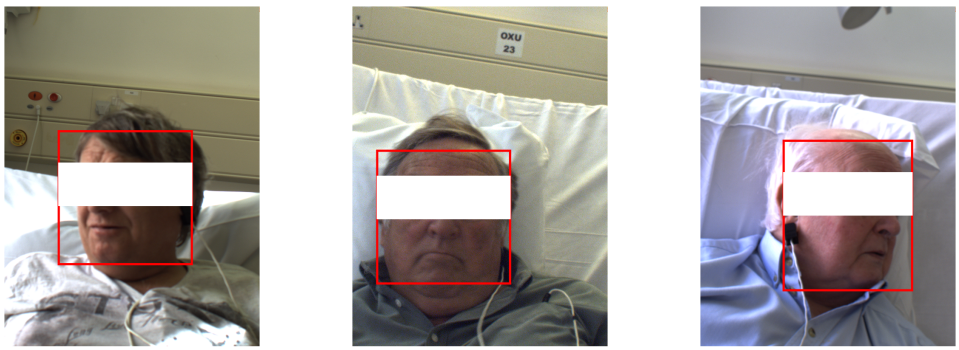
A series of classifiers is assembled in a cascade manner (see figure 6.11). As in [342], the first stage classifier is constructed from a two-feature classifier. The first feature relies on the property that the region of the eyes is often darker than the region of the upper cheeks (figure 6.11(c)). The second feature relies on the eyes being darker than the bridge of the nose (figure 6.11(d)). These features are comparatively large relative to the detection sub-window. If at any stage the feature does not pass the classifier, the sub-window is rejected and no further computation is performed.

Classifier training

Most of the video recordings from the dialysis clinical data set described in chapter 5 were recorded in similar conditions, with the camera pointing at the patient's upper torso from a similar distance. Although several frontal face detectors are described in the literature [187], the recording conditions in the dialysis unit makes the process of face detection challenging. Some of these factors that can potentially contribute to errors include the patients lying on a hospital bed at different angles towards the camera, the patients are awake and active, family members often visit and interact with the patients, the clinical staff are often performing several procedures throughout the day and the ambient light changes during video recording.

A new frontal cascade classifier was developed from a training set extracted from a subset of all the dialysis videos. 200 images from each of the 40 patients were manually extracted, totalling 4000 frontal face and 4000 non-face sample images, all of which were hand-labelled by the author. 1000 images were used for training and the remaining 3000 for testing. Training the classifier was performed using the open source tools provided by the OpenCV project[1].

A similar procedure was carried out to train the cascade classifier for detecting left-profile faces. 200 images were extracted for each patient and manually annotated by the author. Right-profile faces were detected by flipping the image horizontally and using the left-profile detector.



(a) (b) (c)

Figure 6.12: The result of the face detection algorithm (red square) using three types of detectors: (a) left profile, (b) frontal, (c) right profile. Right-profile faces were detected by flipping the image horizontally and using the left-profile detector.

[1] http://opencv.org

Figure 6.12 shows some examples of the output of the face detection algorithms. Some videos were manually rotated to help the algorithms to cope with changes in camera position by the clinical staff when they needed to carry out an obtrusive procedure, such as intermittent blood pressure measurement using an inflatable cuff.

6.4.2 Face tracking

The work described by Zisserman, Everingham and Sivic [90] and further improved in [305] is used in this thesis for the purpose of face tracking. Three face detectors are run in parallel for each video frame: one for approximately frontal faces, one for approximately $3/4$ view to full left-profile and one right-profile detector running on a mirrored image.

Faces detected in consecutive frames are connected using the Kanade-Lucas-Tomasi feature tracker [297]. For any pair of faces, the number of common feature point tracks normalised by the number of feature point tracks which are not shared by both faces, is used as an overlap score. If the overlap score is greater than 0.5, a match is declared. As in [305], the feature tracker is run from the beginning to the end of each dialysis video session, and symmetrically from the end to the beginning. Figure 6.13 shows a face tracking example for a patient going from frontal face to right profile within 30 seconds.

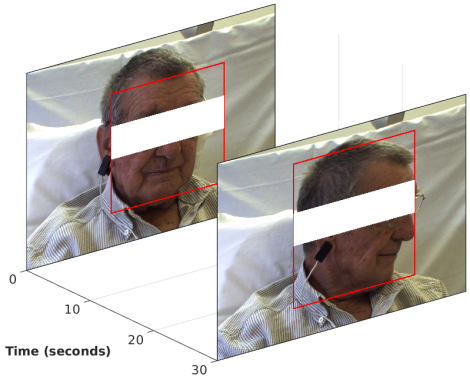


Figure 6.13: Face tracking example for a patient going from frontal face to right profile within 30 seconds.

6.4.3 Activity

To estimate heart rate, suitable time periods during which patient motion is minimal (known as stable periods) are needed, such as when the patient is sleeping, as seen in figures 6.14(b) and 6.14(d). To measure subject motion, the centroid of the detected face is tracked as the patient moves within each video frame. Subsequently, the Euclidean distance between the centroids of two successive frames is quantified. When the movement of the face centroid crosses a given threshold, the frame is considered as active and assigned an activity index of 0, otherwise it is assigned a value of 1. Figure 6.14(f) shows the output of the activity analysis for a typical 4-hour dialysis session.

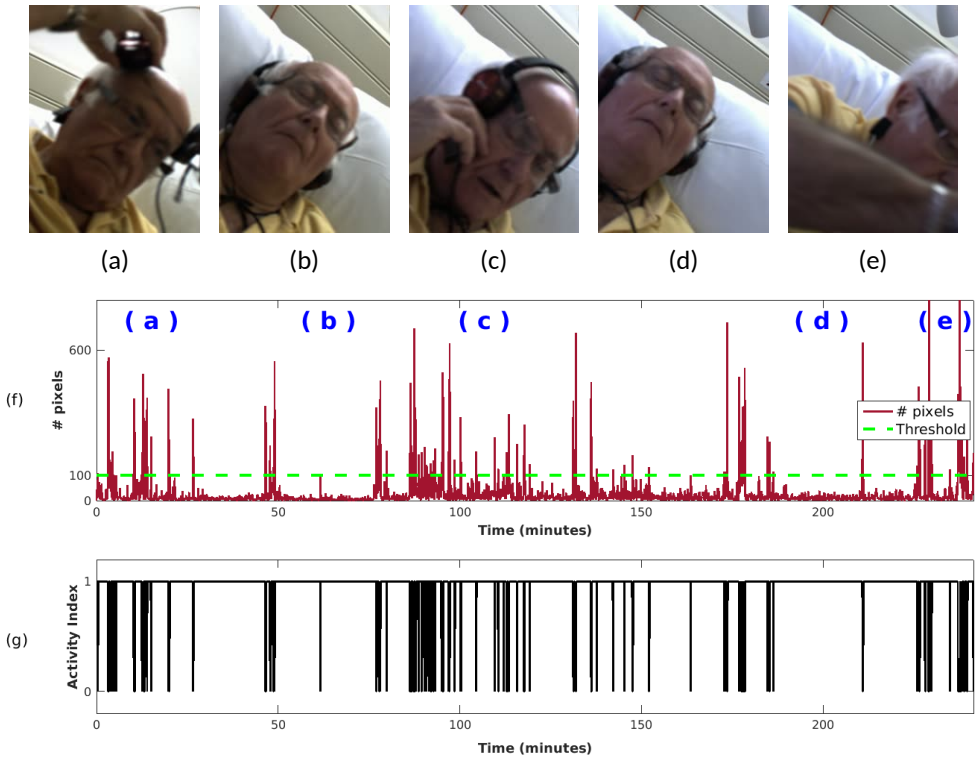


Figure 6.14: Activity Index for a typical 4-hour dialysis session: (a) minute 10: patient wearing headphones to start listening to music, (b) minute 60: sleeping (c) minute 106: awake and engaging in phone conversation, (d) minute 200: sleeping, (e) minute 240: movement and interaction with the clinical staff. (f) Pixels in motion from the activity analysis: shaded areas are labelled (in blue) to correspond to the figures on top. (g) Activity index; a value of 0 indicates a period of patient motion and a value of 1 indicates a stable period suitable for heart rate estimation.

As most of the video recordings from the dialysis clinical data set were recorded in similar conditions, the same threshold of 100 pixels is applied to compute the activity index. As shown in figure 6.14(g), the activity index is a binary value, where 0 indicates a period of patient motion and a value of 1 indicates a stable period suitable for heart rate estimation.

6.4.4 Extraction of stable periods

The outputs from the face-tracking algorithm and the activity analysis are combined to find, for all video sessions, suitable time periods for heart rate estimation during which the patient is stable and the location of the face is known. To prioritise the dialysis sessions for which heart rate is estimated, the sessions with suitable time periods that are less than one-third of the total length of the videos are discarded and no further analysis is done. The rest of the sessions are used in subsequent algorithms to estimate heart rate.

Figure 6.15 shows some sample images from the rejected dialysis sessions. The top row shows sessions for which the camera was moved away from the patient to ease access for interventions, family visits or other circumstances. The bottom row shows sessions for which the patient was only partially visible and therefore his face could not be detected.

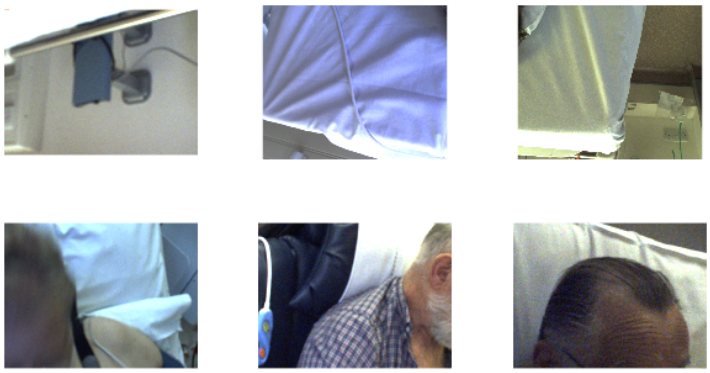


Figure 6.15: Examples of dialysis sessions that were rejected and therefore, heart rate estimation was not carried out. Top row shows sessions where the camera was moved away from the patient to ease access for interventions. Bottom row shows sessions where the patient was only partially visible and his face could not be detected.

6.5 Deriving the PPGi signal from video

The algorithms for heart rate estimation depend on the selection of two types of rectangular regions of interest (ROI) in the image:

- Subject region of interest (ROI_S): One or more areas on the patient's face, such as the cheek or forehead, for which heart rate can be measured by analysing the varying colour of the reflected light.
- Background region of interest (ROI_R): An area not containing any body part from the patient, such as the background wall, used to minimise the effects of external lighting sources (see section 6.7.4).

The extraction of the photoplethysmographic imaging (PPGi) signal from each subject region of interest involves computing a statistical measure of the light intensity of the pixels per colour channel, as described in the following section.

6.5.1 PPGi extraction

The mean is one of the statistical metrics that measure the central tendency of the values in a data set, by attempting to describe the most common value. Other measures of central tendency commonly in use are the mode and median.

As shown in figure 6.16(b), as the histogram of a region of interest approaches a normal distribution, the mean, median and mode become similar; they all represent the most typical value. However, as the histogram becomes multi-modal, the mean no longer represents the most common value for the region of interest (figure 6.16(a)) and its use may lead to inaccurate conclusions, as the modes representing the colour variations due to blood volume changes could not be easily determined.

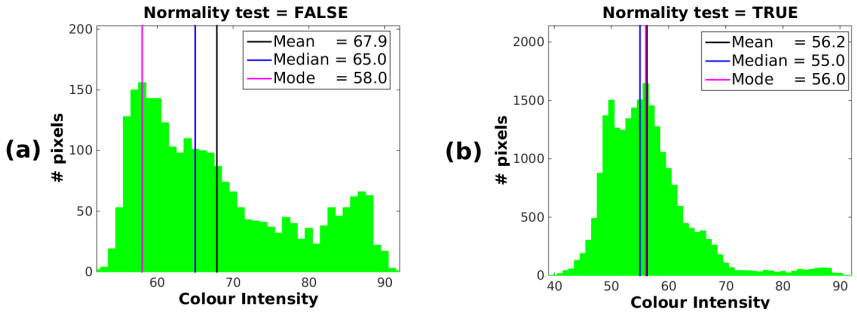


Figure 6.16: Jarque-Bera normality test for the histograms of two typical subject regions of interest extracted from the green channel data: (a) 50x50 pixel size, multi-modal histogram rejected by the test, hence this region of interest is not suitable for heart rate estimation. (b) 150x150 pixel size region of interest can potentially be used to estimate heart rate.

To determine if the histogram of a region of interest can be modelled by a normal distribution, the Jarque-Bera test described in section 6.3 is used. The Jarque-Bera test is a goodness-of-fit test to estimate if the histogram data has the skewness and kurtosis matching a normal distribution.

Figure 6.16 shows the results of applying the Jarque-Bera normality test to two regions of interest extracted from the green channel data of size 50x50 and 150x150 pixels. As expected, the normality test rejects the multi-modal histogram for the region of 50x50 pixels and that region is, as a consequence, not suitable for heart rate estimation. The variation in colour intensity in the 150x150 pixel size region can potentially be used to estimate heart rate.

Figure 6.17 shows the mean, median and variance pixel intensities computed for 15 seconds from a 150x150 region of interest extracted from the patient's face for each successive frame. Heart rate information can potentially be estimated using the time course of the mean, whereas the variations in the variance contains lower frequencies likely to be related to breathing. Therefore, the PPGi signal is taken as the time course of the mean for a given subject region of interest.

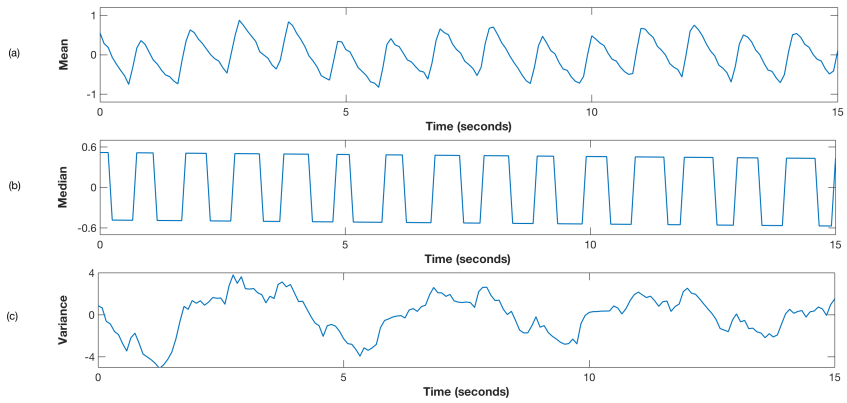


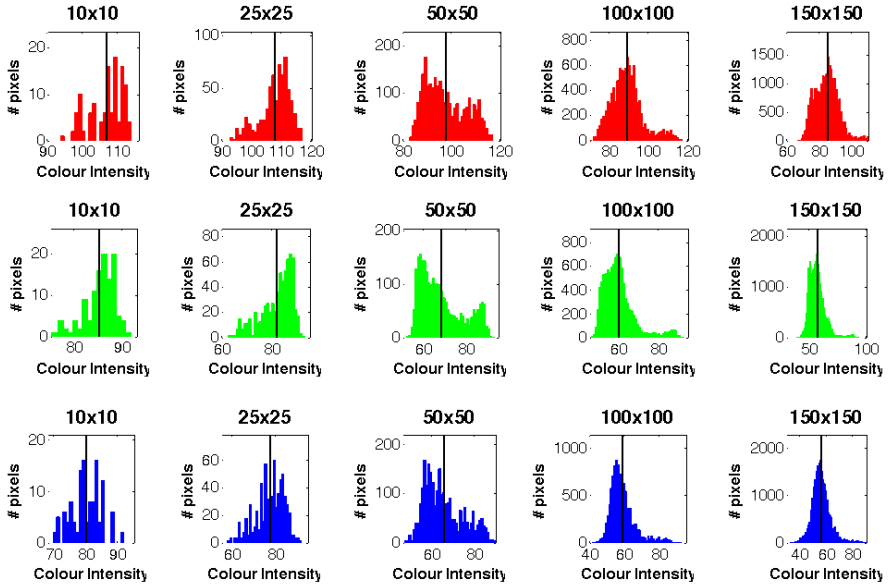
Figure 6.17: Computing (a) the mean, (b) the median and (c) the variance for each successive frame of a 15-second window from a 150x150 region of interest on the patient's face using the green channel. Heart rate information can potentially be estimated using the time course of the mean, whereas the variance presents lower frequencies likely to be related to breathing. Note that the signals have been de-trended to help the comparison.

6.5.2 Subject region of interest (ROI_S)

The process of estimating heart rate requires extracting the mean of the light intensity of pixels for each ROI$_S$. It is therefore important that the location and size of the region of interest are carefully selected. Figure 6.18 shows the histograms computed for the three colour channels from five regions of interest of different sizes, from 10x10 pixels up to 150x150 pixels, located on the patient's right cheek. From the normality test described in the previous section, the ROI$_S$ of size 150x150 pixel gives the best results and, therefore, is suitable for heart rate estimation.



(a)



(b)

Figure 6.18: Location and size of the subject region of Interest: (a) input image frame where the red squares represent the locations of the regions of interest under analysis located in the detected face area, (b) histograms for the regions of interest in (a), where each column represents ROI sizes from 10x10 pixels up to 150x150 pixels, and rows represent each RGB colour channel.

Figure 6.19 shows a 15-second time series for the mean colour intensity for the green channel from the different ROI sizes shown in figure 6.18, from 10x10 pixels up to 150x150 pixels. Heart rate can be extracted from the bigger ROI$_S$, whereas changes in intensity due to breathing are potentially more prominent in smaller ROI$_S$. Hence, a subject region of interest of size 150x150 pixels is chosen for heart rate estimation.

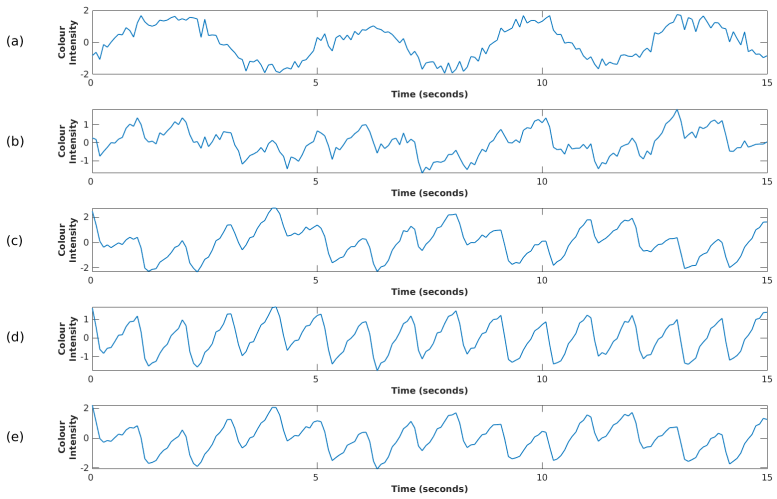


Figure 6.19: 15-second time series for the green channel extracted from the five ROI pixel sizes of figure 6.18 (to help visualisation, the signals have been detrended): (a) 10x10, (b) 25x25, (c) 50x50, (d) 100x100 and (e) 150x150 pixels. Heart rate can be extracted from the bigger ROIs, whereas changes in pixel intensity due to breathing are more prominent in the smallest ROI. Note that the signals have been de-trended to help the comparison.

Most off-the-shelf digital image sensors only record a monochromatic intensity of the incident light over a wide range of the visible spectra. The colour response of the individual pixels is controlled by the use of a Bayer filter array that filters the incident light at specific wavelengths (figure 5.6 shows the response curve for the red, green and blue channels for the image sensor used in the dialysis study). The Bayer filter array, often favours the green channel as the human retina has cells that are more sensitive to green light (see chapter 4). Although the transformation from monochrome light intensity recordings into a colour image is done by proprietary algorithms (often called debayering or demosaicing algorithms), the red and blue channels are usually weighted averages of the surrounding pixels including the green colour pixels [121].

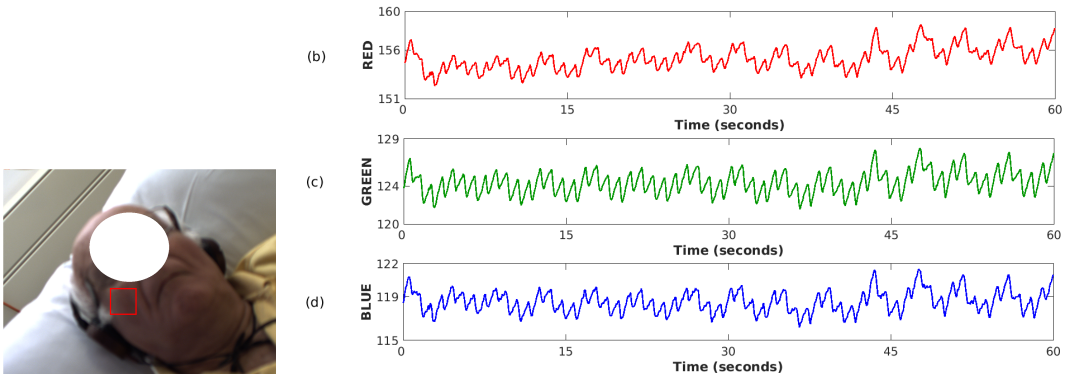
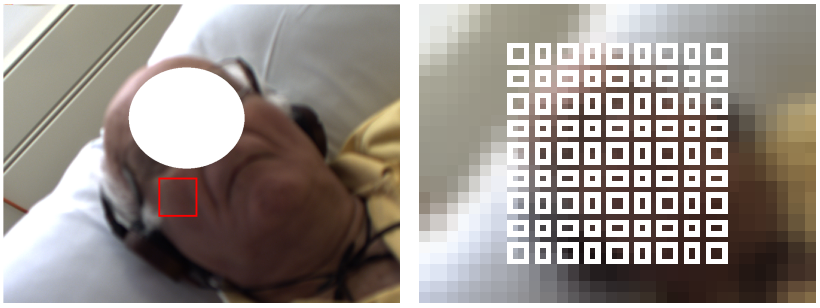


Figure 6.20: 60-second PPGi extraction for a region of interest from the patient's cheek. (a) image sample (the red square represents the location of the ROI) at the start of the extraction period, when the patient was quietly sleeping. The mean light intensity time series for the red, blue and green channels are shown in (b) through (d) respectively.

Demosaicing algorithms have the effect that the red and blue time series have a signal amplitude and pulsatile nature similar to those of the green channel, as can be seen in figure 6.20 for a 60-second sample time series for the red, green and blue channels from a region of interest on the patient's cheek (the red square in figure 6.20(a) represents the location of the ROI) during a quiet period when the patient was sleeping. Although similar pulsatile components for the three colour channels are shown in figure 6.20(b), 6.20(c) and 6.20(d), the vertical axis show that the AC variations are superimposed on different DC levels.



(a) (b)

Figure 6.21: Location and size of the grid for each subject region of interest: (a) sample image with only one ROI_S (the red square). (b) the 9x9 grid of all the regions of interest.

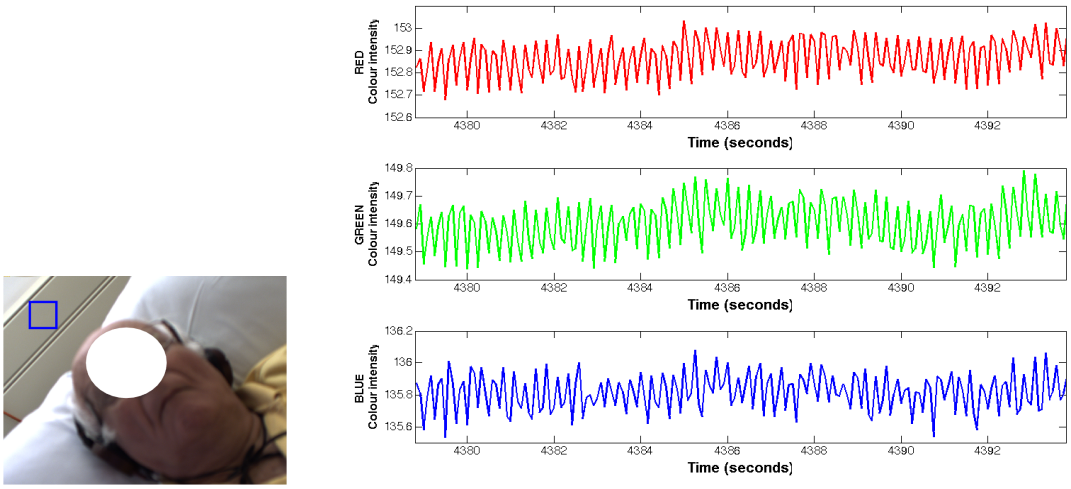
Crowe and Damianou [77] compared the wavelength dependence of the photoplethysmogram signal in the range of 450 nm to 1000 nm using both reflectance and transmission mode pulse oximeters attached to the fingertip. In their experiments, the authors found that wavelengths from 500 nm to 720 nm (green and red colours) follow the absorption spectrum of oxyhaemoglobin in "reflectance mode" pulse oximetry. Unsurprisingly in the signals extracted from the video, the red and green channels often have larger amplitudes and are slightly less noisy than the blue channel, with the green channel having the greatest amplitude (possibly due to the Bayer filter effects). Maeda et al [201] also compared the green signal with infra-red for skin and found a strong correlation between the green contact PPG signal and the R peaks from an ECG monitor (see figure 6.2 for a representative ECG waveform and its R peaks).

During quiet periods (as in figure 6.20) when the effects of motion are minimised and the ambient illumination does not vary greatly, all three colour channels can provide useful information for heart rate estimation. As a result, all the processing described in the rest of the chapter is carried out with all three colour channels.

Using the location of the face detected by the algorithms described in previous sections, and after choosing the size of the subject region of interest, a grid of multiple ROI_S is laid out evenly across the total area of the patient's face. The grid is composed of 81 ROI_S organised in a matrix of 9 rows and 9 columns, as shown in figure 6.21(b). A PPGi signal is computed from each of the regions of interest.

6.5.3 Background region of interest

The background region of interest (ROI_R) is used to minimise the effects of external lighting sources such as fluorescent lights falling both on the background and on the subject's face. Only one background region of interest is selected from the wall behind the hospital bed and above the patient's face (the blue square in figure 6.22), as it is an unobstructed clean area close to the patient. The size of the ROI_R is similar to that of the subject region of interest.



(a) (b)

Figure 6.22: The background region of interest (ROI_R) is selected from the wall behind the patient's bed. The blue square in (a) is of similar size as the subject region of interest. (b) 15-second window extracted from the selected ROI_R for the red, green and blue channels.

6.6 Beat-by-beat quality assessment

The PPGi signal extracted in the previous sections is regularly affected by the changes in ambient light conditions, poor blood perfusion or a variety of motion artefacts. Therefore, a beat-by-beat assessment of its quality is needed so that heart rate can be estimated using only good-quality pulses. Figure 6.23 presents an overview of the overall methodology. Sections 6.6.1 to 6.6.6 describe the various pre-processing steps and components of the algorithm. Section 6.6.7 discusses the final strategy to decide if a beat is labelled as a good or poor-quality beat.

A Bayesian change point detection algorithm is first run to find step changes due to illumination variations or sudden movements affecting the regions of interest and flag the corresponding PPGi pulses as invalid. To eliminate frequencies outside the heart rate physiological range, chosen to be 36 and 240 beats/min, the PPGi waveform is digitally filtered and subsequently the location of beat onsets and peaks are computed. Similar to Li et al [185], the input signal is then divided into 30-second running windows with an overlap of 5 seconds. For each window, a template is constructed from all the valid beats. If a template cannot be computed, the template from the previous window is used.

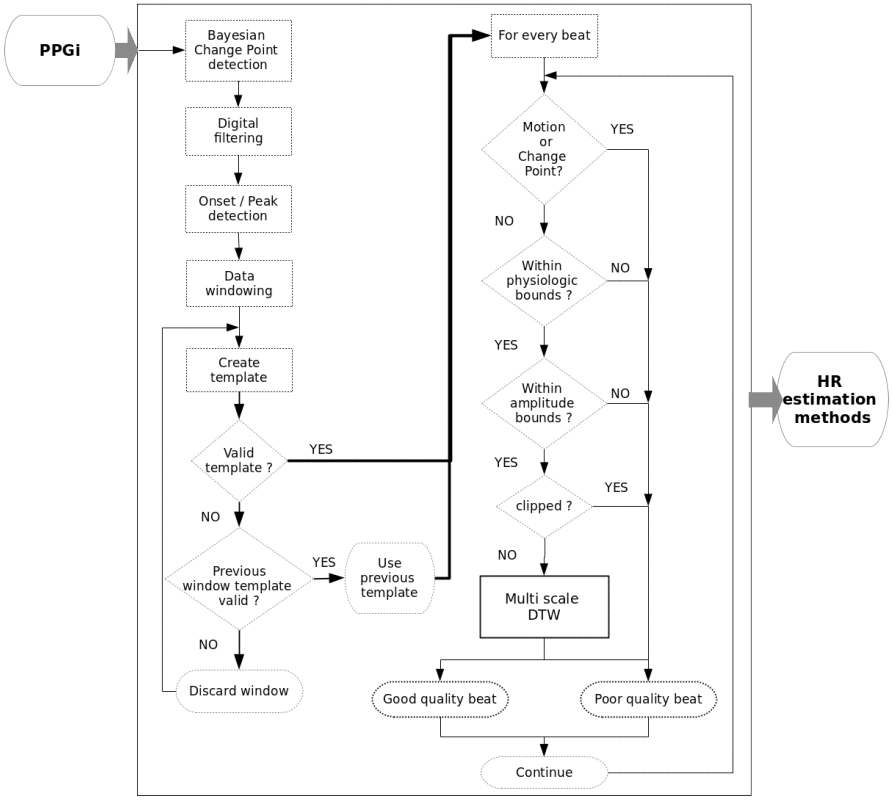


Figure 6.23: Beat-by-beat signal quality assessment overview. DTW is Dynamic Time Warping.

The template is used to analyse the morphology of each individual beat within the window. Firstly, beats corresponding to periods of motion (as identified by the activity index algorithm described in section 6.4.3) or occurring during a step change are flagged as invalid and are assigned a quality value of 0. Secondly, beats that are clipped or are outside a valid physiological or amplitude range are also flagged as invalid. Finally, Dynamic Time Warping is used to compute the minimum distance from the beat to the template. Beats for which this distance lies within a given threshold are ruled to be of good quality, otherwise they are flagged as invalid.

6.6.1 Bayesian change point detection

The PPGi signal, extracted from the regions of interest on the patient's face, is corrupted by sudden changes of posture, periods during which the patient is speaking, interacting with the clinical staff or other movement artefacts. Even during periods for which the patient is quiet (sleeping or listening to music), video cameras still automatically modify the gain for each colour channel to compensate for changes in the scene, such as sudden changes in the overall lighting (fluorescent lights turned on or off) or shadows. A typical illumination step change is shown in figure 6.24(a), corresponding to the time when a fluorescent light was turned on.

Change points or step changes, are abrupt variations or discontinuities in a data sequence. To estimate their location, a Bayesian change point detection algorithm based on Ruanaidh [276] and Adams [2] is applied to the PPGi signal. Given a data sequence x of N samples from two piece-wise constant inputs μ_1 and μ_2 with Gaussian noise added:

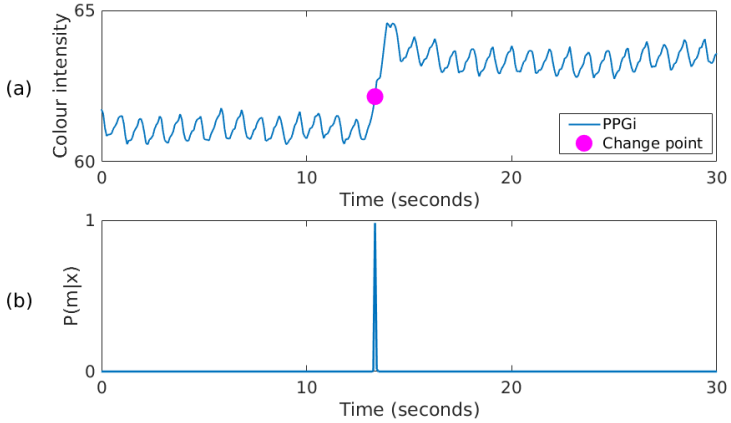


Figure 6.24: Bayesian change point detection algorithm applied to a 30-second window during which a fluorescent light was turned on producing a step change: (a) Input PPGi signal with the change point marked around second 13; (b) The probability of the change point.

$$x_i = \begin{cases} \mu_1 + \epsilon_i, & \text{if } i < m. \\ \mu_2 + \epsilon_i, & \text{otherwise.} \end{cases} \tag{6.11}$$

where the noise samples ϵ_i are assumed to be independent and m is the step change location. The likelihood of the data is given by the joint probability of the noise samples ϵ_i:

$$P(x|\{\mu_1\mu_2\sigma m\}) = \prod_{i=1}^{N} P(\epsilon_i) \tag{6.12}$$

where σ is the standard deviation of the Gaussian noise; μ_1, μ_2 and σm are the known time series parameters. The probability density function for the noise samples is defined by:

$$P(\epsilon) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\epsilon-\mu)^2}{2\sigma^2}} \tag{6.13}$$

Given Bayes' theorem, the posterior probability density of the signal parameters $\mu_1\mu_2\sigma m$ given the data sequence x can be computed by:

$$P(\{\mu_1\mu_2\sigma m\}|x) = \frac{P(x|\{\mu_1\mu_2\sigma m\})P(\{\mu_1\mu_2\sigma m\})}{P(x)} \tag{6.14}$$

Assuming all parameter values are equally likely a priori and by integrating over the parameters μ_1, μ_2 and σm , the posterior probability density can be modified to account only for the location of the step change m [31] by:

$$P(m|x) \propto \frac{1}{\sqrt{m(N-m)}} \left[\sum_{i=1}^{N} x_i^2 - \frac{1}{m}(\sum_{i=1}^{m} x_i)^2 - \frac{1}{N-m}(\sum_{i=m+1}^{N} x_i)^2 \right]^{-\frac{N-2}{2}} \tag{6.15}$$

Equation (6.15) computes the probability of a single step change in the given data window. More detailed background on Bayesian change point detection is presented in appendix B.

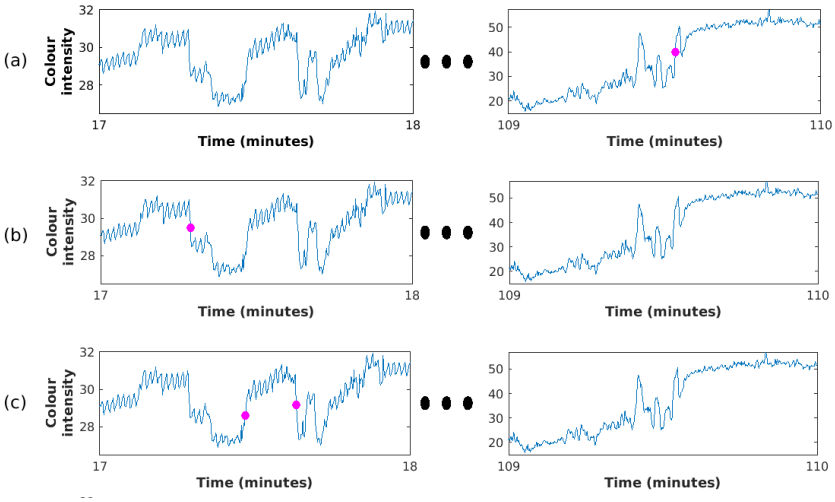
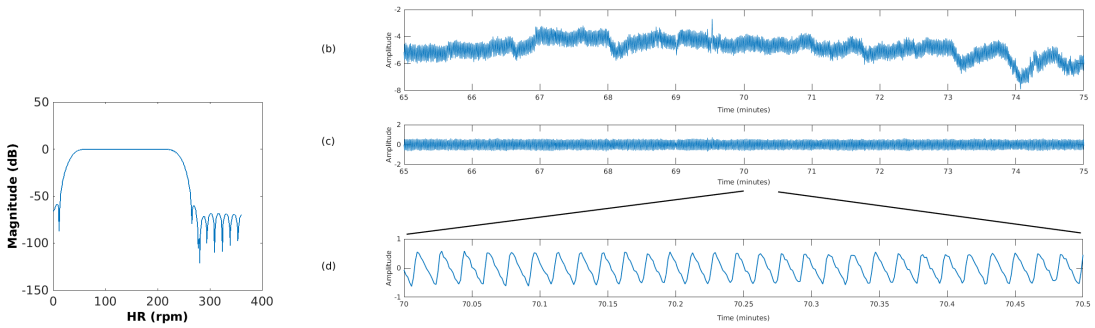


Figure 6.25: Bayesian change point detection algorithm applied to two one-minute PPGi segments. Three running windows of length (a) 30, (b) 15 and (c) 10 seconds. Slow changes can be detected in longer windows, whereas more abrupt changes can be detected in shorter windows.

As the beat analysis windows are of 30-second of length, several step changes can occur. Therefore, the change point algorithm is run for three time scales: 30, 15 and 10 seconds. Figure 6.25 shows two typical PPGi waveform segments of 1 minute in length each. Slow changes can be detected in longer windows, whereas more abrupt changes can be detected in shorter windows.

6.6.2 Digital filtering

The PPGi time series from each of the subject regions of interest contain both high and low frequency components that make heart rate estimation challenging. To eliminate frequencies outside the physiological heart rate range, chosen to be from 36 to 240 $\mathrm{beats/min}$, a finite-impulse response (FIR) band-pass filter with cut-off frequencies of 0.6 Hz and 4 Hz (shown in figure 6.26(a)) is applied. Figure 6.26(c) shows the result of the filter being applied to a 10-minute PPGi signal.



(a)

Figure 6.26: Digital filtering for a 10-minute PPGi time series from a subject region of interest using a finite impulse response (FIR) band-pass filter with cut-off frequencies of 0.6 Hz and 4 Hz corresponding to 36 and 240 beats/minute respectively. (a) Band pass digital FIR filter design. (b) Input PPGi signal, (c) filtered signal, (d) 30-second detail around minute 70.

6.6.3 Onset detection

The PPGi signal extracted from the video camera, as opposed to the photoplethysmogram waveform recorded from a contact pulse oximeter, can contain a high number of motion artefacts, can be affected by ambient light or other external artificial illumination sources, and the morphology of each pulse in the time series can be compromised by other sources of noise that make the detection of peaks and troughs challenging.

The detection of each pulse in the PPGi signal is essential for analysing its quality. Motivated by the work of Zong [374], a new algorithm was developed based on adaptively tracking the upslopes of a Boxed Sum Slope Function (uBSSF).

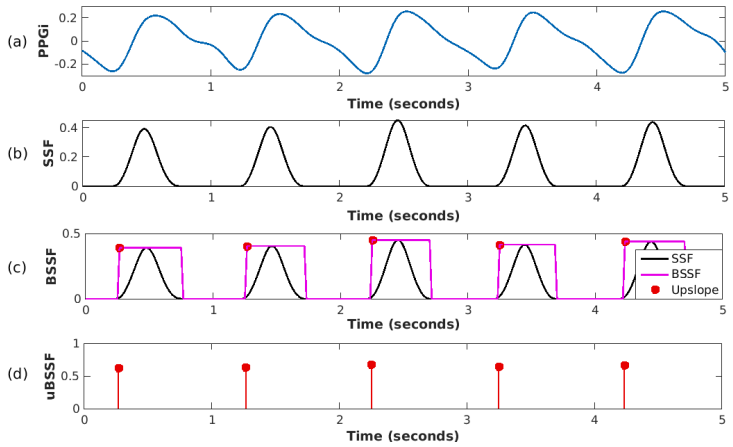


Figure 6.27: Steps to compute the Boxed Slope Sum Function (BSSF) for a 10-second PPGi time series: (a) Input PPGi, (b) Slope sum function, (c) The BSSF is scaled to match each pulse in the SSF function, (d) The BSSF upslopes

The first stage in the onset detection algorithm is to enhance the upslope of the PPGi signal. Given an input PPGi $x_n = x_1, x_2, ...x_n$, a time series containing only the positive slopes is constructed given by the equation:

$$u_k = \begin{cases} \Delta x_k & , \Delta x_k > 0 \\ 0 & , \Delta x_k \leq 0 \end{cases} \tag{6.16}$$

The time series u_k is summed over a running window of approximately 170 ms that corresponds to the typical duration of the upslope of the PPG signal for a typical adult with a normal resting heart rate of 60 $\mathrm{beats/min}$ [374]. This new signal is called the slope sum function (SSF) and is computed by the equation:

$$SSF: \quad z_i = \sum_{k=i-w}^{i} u_k \tag{6.17}$$

where w is the number of samples in the slope window (with a video frame rate of 12 FPS, w is approximately equal to 2, corresponding to a window length of approximately 166.7 ms). A relationship between the PPGi and the SSF is shown in figure 6.27(a) and 6.27(b).

A series of boxes are computed by finding the zero-crossing points at the beginning of the upslope and the end of the downslope of each pulse in the SSF signal. This new time series is called the Boxed Slope Sum Function (BSSF), as shown in figure 6.27(c). To find the beat onsets, instead of analysing the amplitudes of all the values in the BSSF signal, it is only required to analyse the upslope points (uBSSF) as shown in figure 6.27(d). This makes the computation more tractable across all the regions of interest extracted from the human face in the video images.

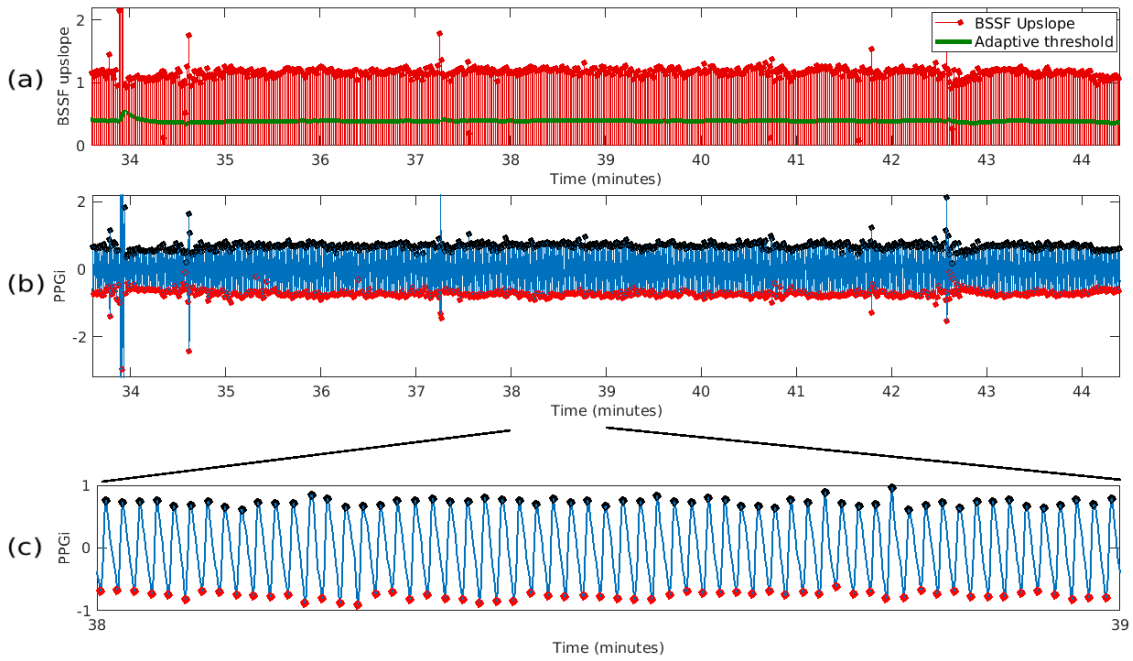


Figure 6.28: Peak and onset detection for a 10-minute sample PPGi time series: (a) The BSSF upslopes, (b) beat onset detection results, (c) Detail for a 1-minute window. The black dots correspond to the peaks of each beat, the red dots to the troughs (or onsets).

In general, the values of the uBSSF data points only correspond to the onsets of each pulse in the PPGi signal during periods of good quality (as shown in figure 6.27(d)), therefore a decision rule has to be established. Firstly, an initial amplitude threshold is computed as the median of the first 30-seconds from the uBSSF data points. This threshold value is adaptively updated as the uBSSF values increase or decrease. For each subsequent uBSSF data point, if its value is greater than the computed threshold, it is considered as a beat onset and the threshold is increased by 10% of the difference between the previous threshold and the current uBSSF data point. If the uBSSF data point is less than the threshold, the threshold is reduced by 10%. Once a beat onset is detected, the next uBSSF data point to be analysed is at least 167 ms later in the time series, corresponding to the interval for the minimum heart rate of 36 beats/minute.

The peaks of the PPGi waveform are computed as the maximum point between two beat onsets. Figure 6.28 shows the beat onset detection results for a 10-minute time series from a subject region of interest.

6.6.4 Data windowing

Once the beat onsets are located, the PPGi signal is then subdivided into 30-second running windows with an overlap of 5 seconds between them. The following sections describe the processing carried out for each window separately, with each PPGi pulse analysed independently.

6.6.5 Template generation

To determine the quality of each PPGi pulse within a given window, a template is first computed. The template serves as a representative beat. To compute the window template, a three-stage method is used (see figure 6.29). At every stage, all the pulses are aligned using their peaks as reference points. The average of all the valid beats, given the criterion of the current stage, is taken as the template for that stage. The method proceeds with the next stage only if at least 50% of the pulses are deemed to be valid using the criterion for that stage. If less than half of the beats are rejected, the window template is considered invalid and no further processing is done.

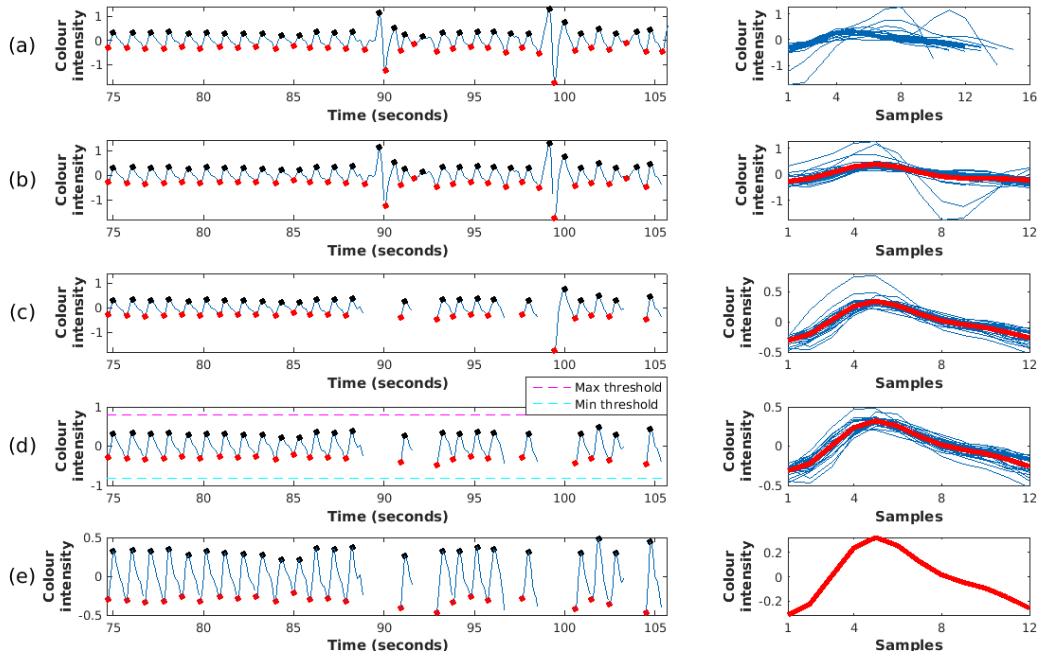


Figure 6.29: Computation for a valid template from a 30-second PPGi window. (a) Input PPGi time series including all 31 beats superimposed on top of each other. (b) First template T1, all 31 beat intervals are within the beat-to-beat interval physiological range between 167 ms to 25 ms (corresponding to a heart rate from 36 to 240 beats/minute). (c) Second template T2, 25 (80%) of the pulses have a cross-correlation coefficient greater than 0.8. (d) Third template T3, 24 (77%) of the pulses are within the amplitude bounds. (e) As the number of valid pulses is greater than 50%, the template shown in red is deemed to be valid.

For the initial stage, all the beats with a time interval outside the physiological range for beat-to-beat intervals from 25ms to 167ms (corresponding to a range of heart rates from 36 to 240 beats/minute) are rejected. All the beats within this range are aligned using their peaks as the reference points and the average of them is taken as the first template (T1) (see figure 6.29(b)).

During the second stage, a new template (T2) is computed by including only those beats with a cross-correlation coefficient between the beat and the template T1 greater than 0.8, (see figure 6.29(c)).

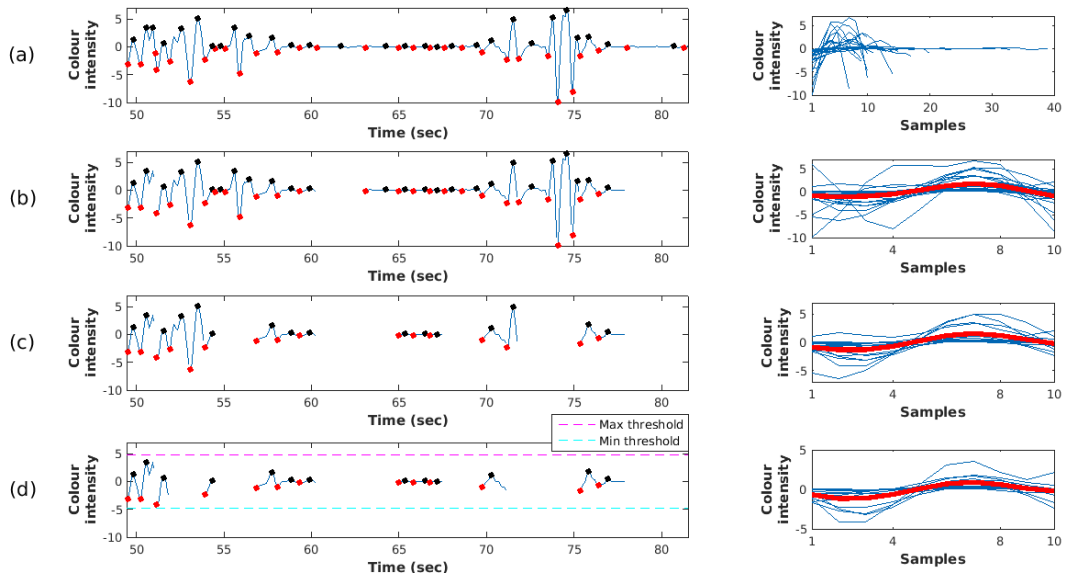


Figure 6.30: Computing an invalid template from a 30-second PPGi window. (a) Input PPGi time series including all 29 beats superimposed on top of each other. (b) First template T1, 26 beat intervals are within the beat-to-beat physiological range between 167 ms to 25 ms (corresponding to a heart rate from 36 to 240 beats/minute). (c) Second template T2, 16 (55%) of the pulses have a cross-correlation coefficient greater than 0.8. (d) Third template T3, only 13 (45%) of the pulses are within the amplitude bounds. As the number of valid pulses is less than 50%, the template is deemed to be invalid.

Finally, in the third stage a template (T3) is constructed by including all the beats with an amplitude less than three times the standard deviations computed from the valid beats in the previous stage (see figure 6.29(d)). This amplitude criterion helps to prevent the window template from including large oscillations due to motion artefacts in the PPGi signal.

If the number of beats that pass all three stages is greater than 50%, template T3 is taken as the window template, as shown in figure 6.29. Figure 6.30 shows a window for which the template is deemed to be invalid, as only approximately 45% of the beats pass the three-stage verification tests. If a template cannot be computed for the current window, the template from the previous window is used instead.

6.6.6 Multi-scale dynamic time warping

Once a valid template has been computed for a given window, an important step is to measure its similarity with each beat within the current window as a measure of quality. Dynamic Time Warping (DTW) is well suited for this type of pattern-matching problem [268]. DTW is a well-known algorithm for finding an optimal alignment between two time series that do not necessarily have the same length.

Dynamic Time Warping became popular for speech recognition to automatically cope with the time deformation which occurs with multiple utterances of the same sequence of words [268]. DTW also became popular for searching and comparing music information patterns [234].

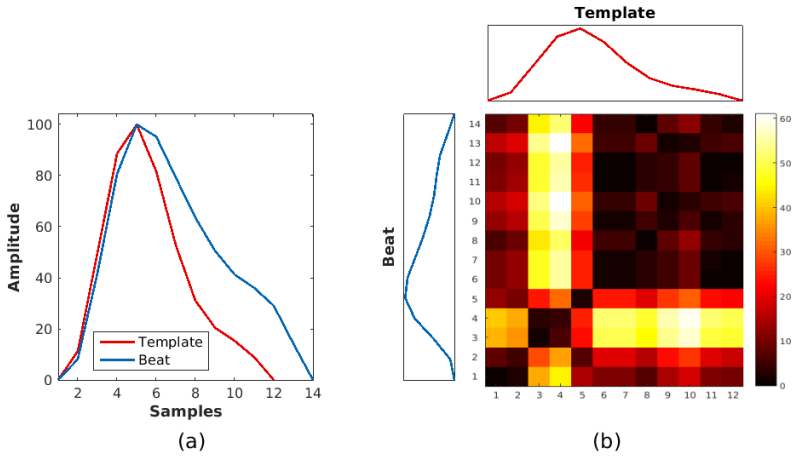


Figure 6.31: Computing the local distance matrix. (a) Normalised input beat (blue) and template (red). (b) The local cost matrix with darker regions representing areas where the beat and template are close, while brighter regions constitute areas where they are further apart.

DTW compares two time series, $X = x_1, x_2, ..., x_n$ of length N and $Y = y_1, y_2, ..., y_m$ of length M, representing the current beat to be tested and the window template respectively (see figure 6.31(a)). Given a feature space Υ defined for both x_n and y_m, a local cost measure is needed to compare both time series. The cost measure, also known as the local distance measure, is a function defined by:

$$c(x, y) : \Upsilon \times \Upsilon \rightarrow \Re \geq 0 \tag{6.18}$$

Typically, the function $c(x, y)$ is defined so that is has a small value (has a low cost) when x and y are similar to each other, and takes on a large value (higher cost) when x and y are different. Evaluating the local cost measure for each pair of elements of the time series X (the current beat to be tested) and Y(the window template), the cost matrix $C \in \Re^{N \times M}$ is obtained as shown in figure 6.31(b), where the sequence X (the beat) is plotted on the vertical axis and Y(the template) on the horizontal axis. The goal of DTW is to find an alignment path (also called the warping path) between X and Yhaving minimal overall cost.

The total cost $c_p(X, Y)$ of any warping path p of length L between the current PPGi beat being analysed X and the window template $Y$is given by:

$$c_p(X, Y) = \sum_{l=1}^{L} c(x_{ml}, y_{ml}) \tag{6.19}$$

An optimal warping path between X and $Y$is a warping path p^* with minimal total cost among all possible warping paths. Therefore, the DTW distance $DTW(X, Y)$ is defined as :

$$DTW(X,Y) = c_{p^*}(X,Y) \\ = min\{\, c_p(X,Y) \mid p\ is\ an\ (N,M)\ warping\ path\ \} \tag{6.20}$$

To compute the optimal path p^*, dynamic programming is used as it reduces the computational complexity to test every possible warping path. An accumulated cost matrix $D(n,m)$ is computed from the local cost matrix $C(X,Y)$ defined as:

$$D(n,m) = DTW(X(1:n), Y(1:m)) \tag{6.21}$$

The optimal minimal cost path runs through a valley of low cost values (darker areas) within the accumulated cost matrix $D(n,m)$, as shown in figure 6.32(b). The time alignment for the current PPGi beat and the window template is shown in figure 6.32(c).

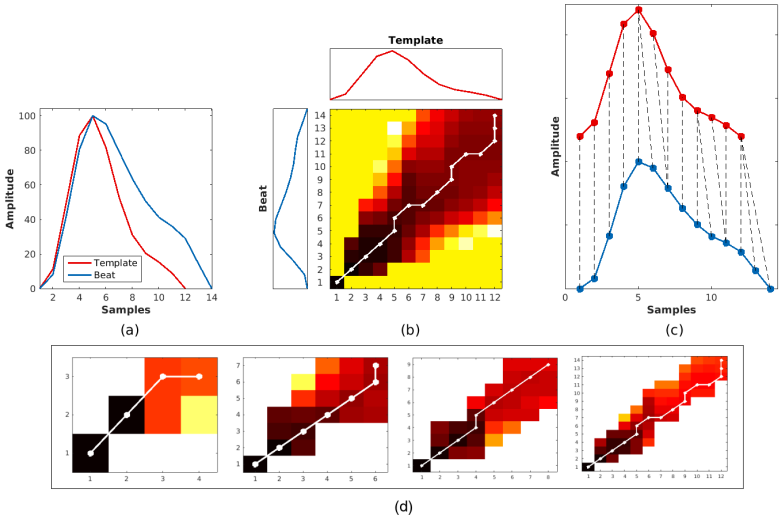


Figure 6.32: Time alignment for multi-scale Dynamic Time Warping. (a) Normalised input beat (blue) and template (red). (b) The accumulated distance matrix with darker regions present areas where the beat and template are close, while brighter regions constitute areas where they are further apart; the optimal minimum path is shown in white. (c) DTW time alignment between the beat (blue) and template (red). (d) Finding the minimum path in the iterative multi-scale Dynamic Time Warping algorithm.

Even though the classic DTW algorithm makes use of dynamic programming, it still requires all possible local distance measures to be computed for each pair of elements of the time series X and Y, as shown in the accumulated distance matrix (figure 6.32(b)) where the yellow and red cells are still computed even through the optimal minimal path does not traverse through them.

Given that all PPGi waveforms from all the regions of interest have to be analysed for their quality, an improvement on the classical DTW algorithm is proposed, based on the work done by Fitriani [15] and Salvador [281].

Firstly, to reduce the effects of short-term upward or downward trends in the PPGi signal, the current beat and template are rotated, as explained in the next section, and their amplitude normalised. Secondly, to reduce the number of cells for which to compute the accumulated

distance matrix $D(n, m)$, time alignment is done on downsampled or coarsened versions of the time series X and Y at different scales based on a piece-wise linear approximation method. At each scale, the lengths N and M are decreased and the optimal path found. The optimal path is then projected onto the next scale and the path search is only done within boundaries from the previous scale (see figure 6.32(d)).

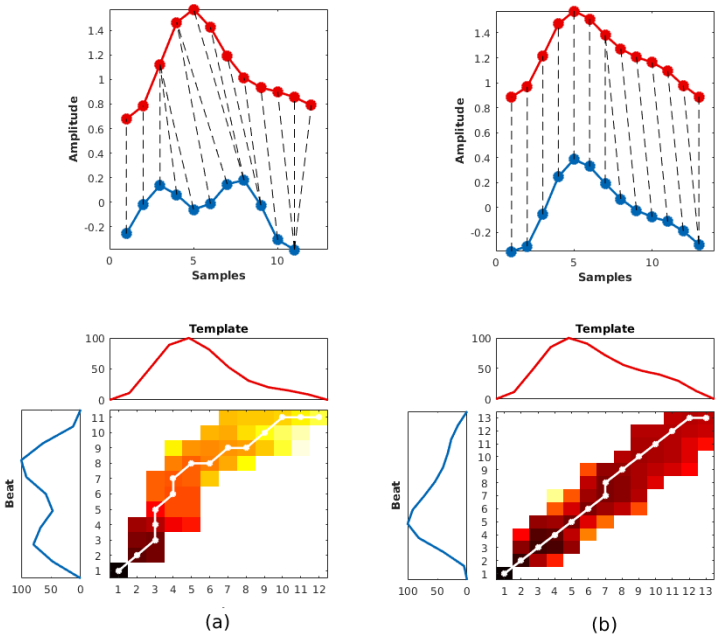


Figure 6.33: Time alignment and accumulated cost matrix examples for (a) poor and (b) good quality beats. Darker regions in the accumulated cost matrices represent areas where the beat and template are close while brighter regions constitute areas where they are further apart. The optimal minimum path is shown in white.

For each DTW scale, instead of the commonly used Manhattan distance, a local cost matrix is computed based on comparing the beat and template slopes. The accumulated cost matrix $D(n, m)$ is computed with specialised local path constraints to prevent distortion of the PPGi waveform by the alignment of one point from the beat to multiple consecutive points of the template.

Finally, after the overall minimal distance is found from all the DTW scales, a threshold is applied to decide if the beat is of good or poor quality. Figure 6.33 shows the time alignment for both cases. The following sections describe these algorithms in more detail.

Beat pre-processing

As shown in figure 6.34, the PPGi waveforms present variations in their amplitudes and trends. Moreover, the input beat and the window template have different time lengths, are rotated with a different angle, and usually have different amplitude ranges (as seen in figures 6.35(a) and 6.35(d)). A direct comparison would yield sub-optimal results, causing valid beats to be considered to be of poor quality.

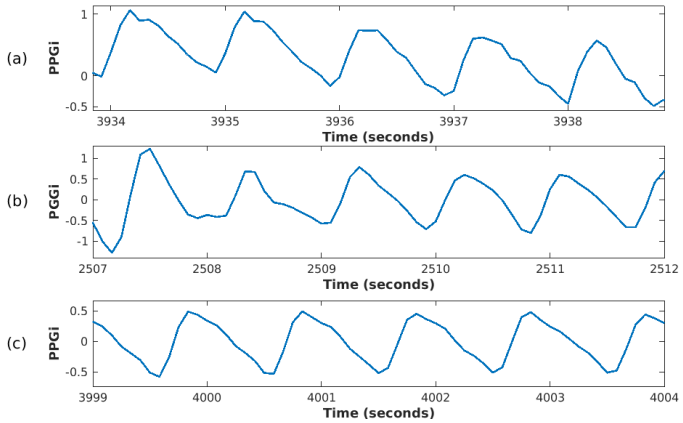


Figure 6.34: Three 5-second segments of a typical PPGi signal: (a) slow downward trend, (b) amplitude difference between beats during small movements, (c) regular PPGi from quiet periods.

Pre-processing is done in two stages: rotation and normalisation. Typical good-quality PPGi beats begin and end with similar amplitude values (see figure 6.34(c)). Therefore, the input beat and the window template are rotated so that their start and end values match, as seen in figure 6.35(b) and 6.35(e). The amplitudes of the PPGi beat and the window template are then normalised so as to lie within the range from 0 to 100, as seen in figure 6.35(c) and 6.35(f).

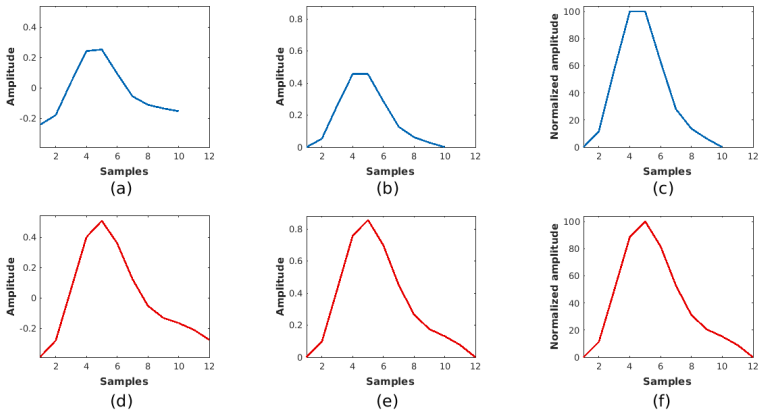


Figure 6.35: Pre-processing steps for Dynamic Time Warping: (a) Input beat, (b) rotated beat, (c) normalised beat, (d) input template, (e) rotated template, (f) normalised template.

Piece-wise linear approximation

The proposed multi-scale Dynamic Time Warping (DTW) algorithm works with different versions of the input beats and the window template, with different levels of coarseness in order to speed up the process. This is achieved using a piece-wise linear approximation (PLA) method in which the downsampling is non-linear in order to preserve as much of the shape of the beat as possible. The method segments the PPGi beat with a set of continuous lines for a given residual approximation error. Since the input beats are normalised, the target error represents a percentage of the amplitude instead of an absolute value.

Given an original PPGi beat with the point values $\{(x_1, y_1), (x_1, y_1), ..., (x_n, y_n)\}$ of length n, where x represents the discrete time points and y represents the amplitude values, the approximation is achieved by defining a distance function for every point in the PPGi beat to the segment line under consideration.

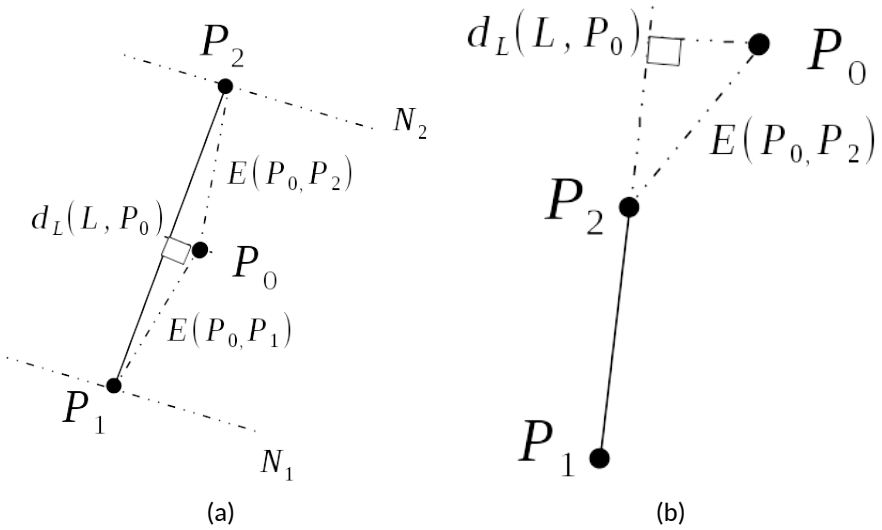


Figure 6.36: Comparison of distance measurements between a point and a line segment: (a) When the point P_0 is between the two normals N_1 and N_2, each Euclidean distance E is large; the algorithm breaks the line segment into two, causing the DTW algorithm to work with a less coarse version of the beat. Here the distance $d_L(L, P_0)$ is adequate. (b) The distance $d_L(L, P_0)$ from the point P_0 to the line segment $L(P_1, P_2)$ is small and inadequate; the algorithm extends the line segment from P_1 to P_0, and the shape of the peak is lost; here the Euclidean distance is adequate and is selected instead.

The distance between successive points in the PPGi grows large when the slope is at its steepest following the onset of each PPGi beat and its peak may be lost if the downsampling is not adjusted locally; conversely, the distance rapidly decreases after the peak of the PPGi. Therefore, the correct choice of the distant metric for every set of two points during each PPGi beat is essential for the optimal performance of the piece-wise linear approximation method.

The Euclidean distance is a common metric used to define the distance between two points $P_i(x_i, y_i)$ and $P_j(x_j, y_j)$ in the plane. It is given by the equation:

$$E(P_i, P_j) = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \tag{6.22}$$

The equation of a line L in the plane is given by:

$$L = ax + by + c = 0 \tag{6.23}$$

where a, b and c are real and known constants. Both a and b are non-zero quantities. The distance from a point $P_0(x_0, y_0)$ to the line L is :

$$d_L(L, P_0(x_0, y_0)) = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}} \tag{6.24}$$

A line $L(P_1, P_2)$ defined by two points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ is given by:

$$\begin{aligned} y &= \frac{y_2 - y_1}{x_2 - x_1}(x - x_1) + y_1 \\ L(P_1, P_2) &= (y_2 - y_1)x - (x_2 - x_1)y + x_2 y_1 - x_1 y_2 = 0 \end{aligned} \tag{6.25}$$

Substituting equation (6.25) into equation (6.23), the distance from a point $P_0(x_0, y_0)$ to the line $L(P_1, P_2)$ defined by two points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ can be expressed as:

$$d_L(L(P_1, P_2), P_0) = \frac{|(x_2 - x_1)(x_0 - y_1) - (y_2 - y_1)(x_0 - x_1)|}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}} \tag{6.26}$$

Given a point P_0 under analysis, figure 6.36(a) shows that choosing the Euclidean distance is inadequate as the algorithm breaks the line segment into two, causing the DTW algorithm to work with a less coarse version of the waveform and increasing computation time. The distance $d_L(L, P_0)$ from P_0 to the line segment $L(P_1, P_2)$ is therefore used instead in this case.

On the other hand, figure 6.36(b) shows a typical situation close to the peak of the PPGi pulse. This time, the distance $d_L(L, P_0)$ from the point P_0 to the line segment $L(P_1, P_2)$ is small, causing the algorithm to extend the line segment from P_1 to P_0, with the consequence that the shape of the PPGi peak is lost. Hence, the Euclidean distance $E(P_0, P_2)$ is used instead in this case.

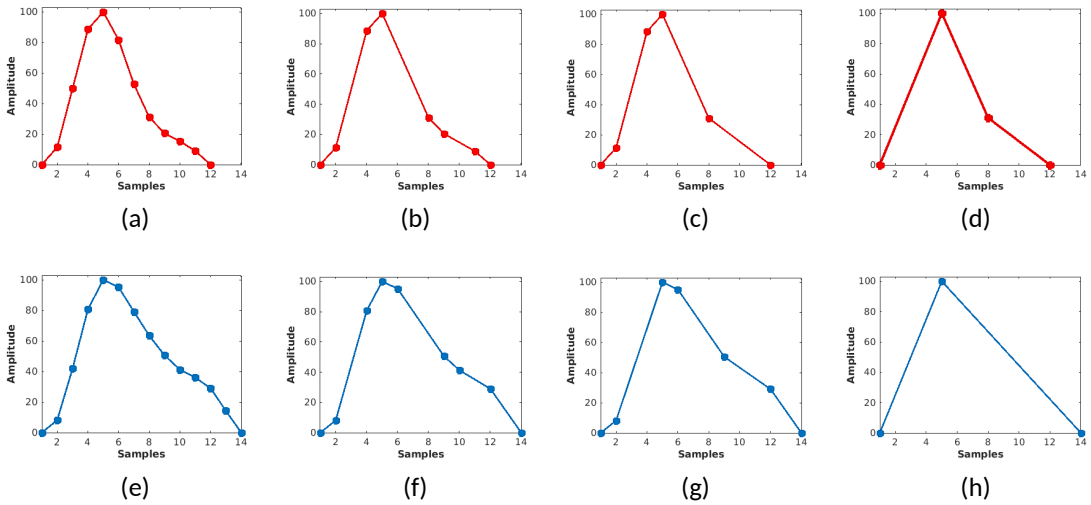


Figure 6.37: Comparison of piece-wise linear approximation for four different error thresholds. Template with errors of : (a) 0%, (b) 0.25%, (c) 0.5% and (d) 1%. The input waveform with errors of: (e) 0%, (f) 0.25%, (g) 0.5% and (h) 1%

Based on Koski et al [167], equation (6.27) defines the proposed distance metric d_M. The Euclidean distance is a valid measure only when the point P_0 under analysis is outside the two normals at the end points of the line segment defined by $L(P_1, P_2)$; otherwise, the distance from P_0 to the line segment is used.

$$d_M(P_1, P_0, P_2) = \begin{cases} E(P_0, P_1) & \text{, if } E(P_0, P_1)^2 + E(P_2, P_1)^2 < E(P_2, P_0)^2 \\ E(P_2, P_0) & \text{, if } E(P_2, P_1)^2 + E(P_2, P_0)^2 < E(P_0, P_1)^2 \\ d_L(L(P_1, P_2), P_0) & \text{, } otherwise \end{cases} \quad (6.27)$$

The approximation method is an iterative loop, scanning the waveform from the beginning and extending the current line until the point where the distance d_M crosses a given error threshold. When this error threshold is reached, the method starts a new line. Four error thresholds are used, with values of 1%, 0.5%, 0.25% and 0% of the normalised beat amplitude, as shown in figure 6.37.

Local path constraints

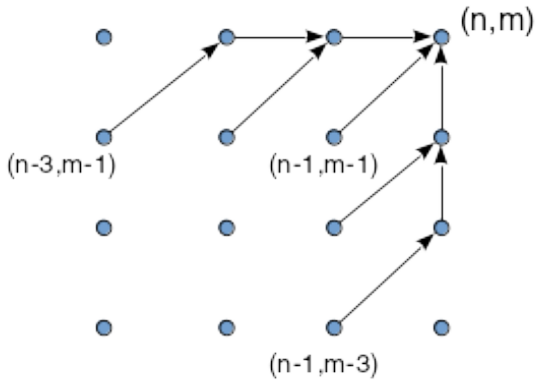


Figure 6.38: Local path constraints for DTW. As the typical PPGi waveform consists of a single upslope followed by a single downslope, the weights help favour the vertical paths in the cost matrix.

The accumulated cost matrix $D(n, m)$ in equation (6.21) is computed with specialised local path constraints (see figure 6.38) to prevent the distortion of the PPGi waveform by the alignment of one point from the beat to several data points of the template, or to prevent long vertical or horizontal paths. The accumulated matrix is defined by equation (6.28).

$$D(n, m) = \begin{cases} D(n-1, m-1) + (w_b(n) + w_t(m))c(n, m) \\ D(n-2, m-1) + (w_b(n-1) + w_t(m))c(n-1, m) + w_b(n)c(n, m) \\ D(n-1, m-2) + (w_b(n) + w_t(m-1))c(n, m-1) + w_t(m)c(n, m) \\ D(n-3, m-1) + (w_b(n-2) + w_t(m))c(n-2, m) + \\ \qquad + w_b(n-1)c(n-1, m) + w_b(n)c(n, m) \\ D(n-1, m-3) + (w_b(n) + w_t(m-2))c(n, m-2) + \\ \qquad + w_t(m-1)c(n, m-1) + w_t(m)c(n, m) \end{cases} \quad (6.28)$$

where w_t and w_b represent weights from the slopes for the template and beat respectively. As the typical PPGi waveform consists of a single upslope followed by a single downslope, the weights help favour the vertical paths in the cost matrix.

Path projection

After several coarse versions of the beat and the window template are created, the Dynamic Time Warping algorithm is run for each individual scale with the local path constraints presented in equation (6.28). Starting at the low scale (error threshold 1%), an initial optimal path is computed (see figure 6.39(a)). A mask of the optimal path is created for the next scale (figure 6.39(b)) and a new optimal path is computed analysing only the masked cells of the cost matrix in the new scale (figure 6.39(c)). This process is continued until the scale with error threshold of 0% is reached (figure 6.39(g)).

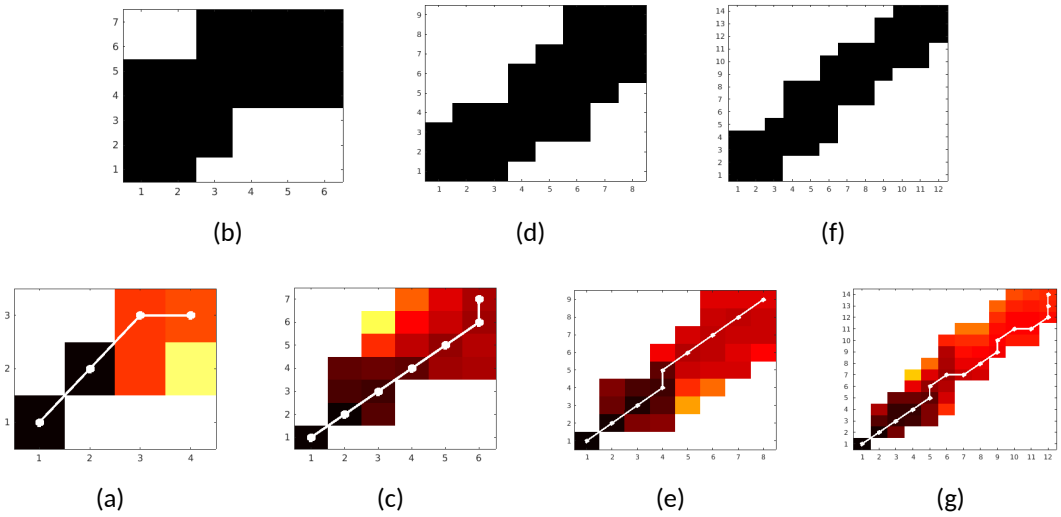


Figure 6.39: Optimal path computed by the iterative multi-scale dynamic time warping algorithm using error thresholds for the piece-wise linear approximation of (a) 1%, (c) 0.5%, (e) 0.25% and (g) 0%. The masks between iterations of each error threshold are shown in (b), (d) and (f). The horizontal axis in (a), (c), (e) and (g) corresponds to the number of red dots on the template from figure 6.37(d), (c), (b) and (a) respectively. The vertical axis in (a), (c), (e) and (g) corresponds to the number of blue dots on figure 6.37(h), (g), (f) and (e) respectively; representing the current PPGi beat.

6.6.7 Beat quality assessment

Deciding if ultimately a beat is of good or poor quality involves a simple decision rule based on the values of five Signal quality indices (SQIs): does a change point occur during the beat (see section 6.6.1)?; does the beat occur during periods of motion as computed by the activity index described in section 6.4.3?; is the beat time interval outside the range corresponding to heart rates between 36 and 240 beats/minute?; is the beat amplitude outside the ranges defined by the window (when computing the window template)?; and finally, is the beat clipped?.

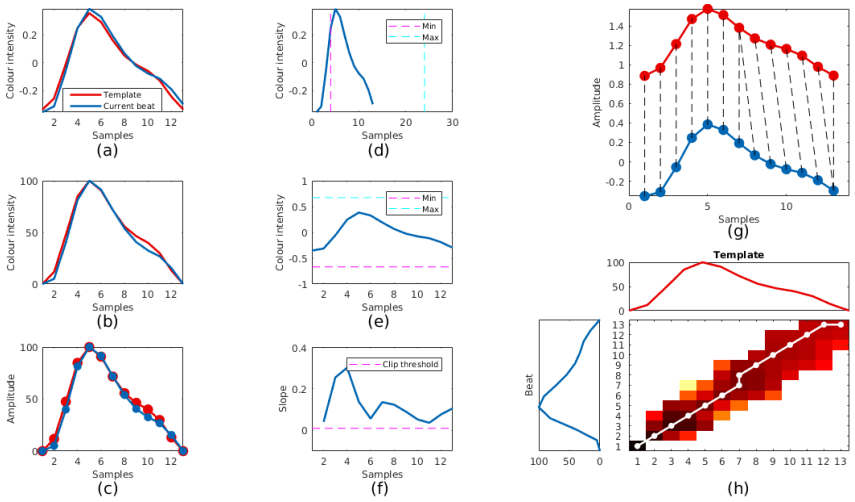


Figure 6.40: Beat quality metrics for a good-quality beat from figure 6.29. The resultant combined SQI metric for the beat is 1. (a) input template and beat, (b) rotated and normalised template and beat, (c) piece-wise linear approximation, (d) check if the beat is within physiological bounds for heart rate (36 to 240 $\mathrm{beats/min}$), (e) check if the beat is within the amplitude thresholds for the current window, (f) clipping detection, (g) Dynamic Time Warping showing the minimum path with a cost of 3.58, (h) Accumulated cost matrix showing the minimum path.

If any of the above criteria are met, the beat is rejected and no further processing is carried out. Otherwise, if the DTW minimum distance is below a given threshold, the beat is considered to be of good quality. Figure 6.40 shows how the metrics are used to identify good-quality beats. Figure 6.41 is an example of a poor-quality beat. Even though the beat is within the heart rate physiological range, the aligned beat is very different from the window template.

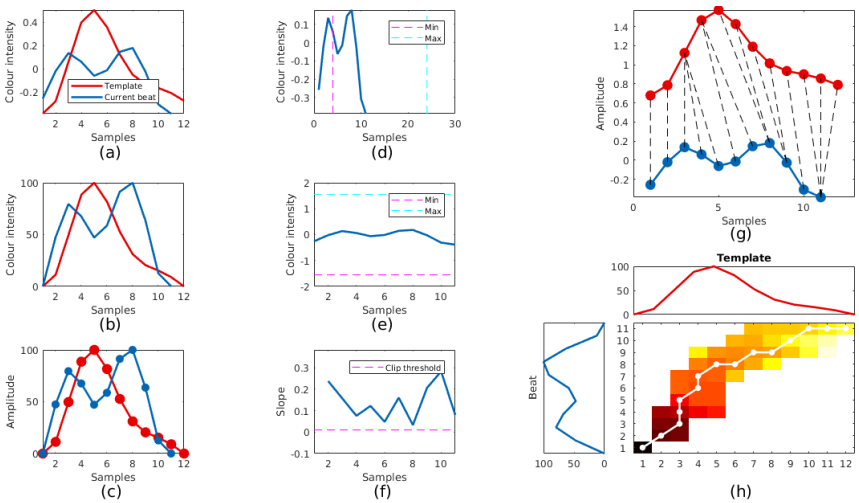


Figure 6.41: Beat quality metrics for a poor-quality beat for figure 6.29. The resultant combined SQI metric for the beat is 0. (a) input template and beat, (b) rotated and normalised template and beat, (c) piece-wise linear approximation, (d) check if the beat is within physiological bounds for heart rate (36 to 240 $\mathrm{beats/min}$), (e) check if the beat is within the amplitude thresholds for the current window, (f) clipping detection, (g) Dynamic Time Warping showing the minimum path with a cost of 28.10, (h) Accumulated cost matrix showing the minimum path.

6.6.8 Summary

Figure 6.42 presents the beat-by-beat quality analysis for a 5-minute PPGi signal for which all the proposed quality metrics described in the previous section were computed. The FFT spectrum is also shown in the right column (see section 6.7.2 below), to display how the heart rate frequency is visible as only the beats of good quality are included. The proposed quality metrics can correctly identify time periods for which the input camera signal in figure 6.42(a) is affected by several sources of noise, and therefore a reliable heart rate estimate cannot be computed. The resulting PPGi signal in figure 6.42(l) contains only the good quality pulses. The peak of the FFT spectrum is 1.12 Hz, corresponding to a heart rate of approximately 67.2 $\mathrm{beats/min}$, which is in good agreement to the patient's average reference heart rate of 66.7 $\mathrm{beats/min}$ during the corresponding 5-minute window.

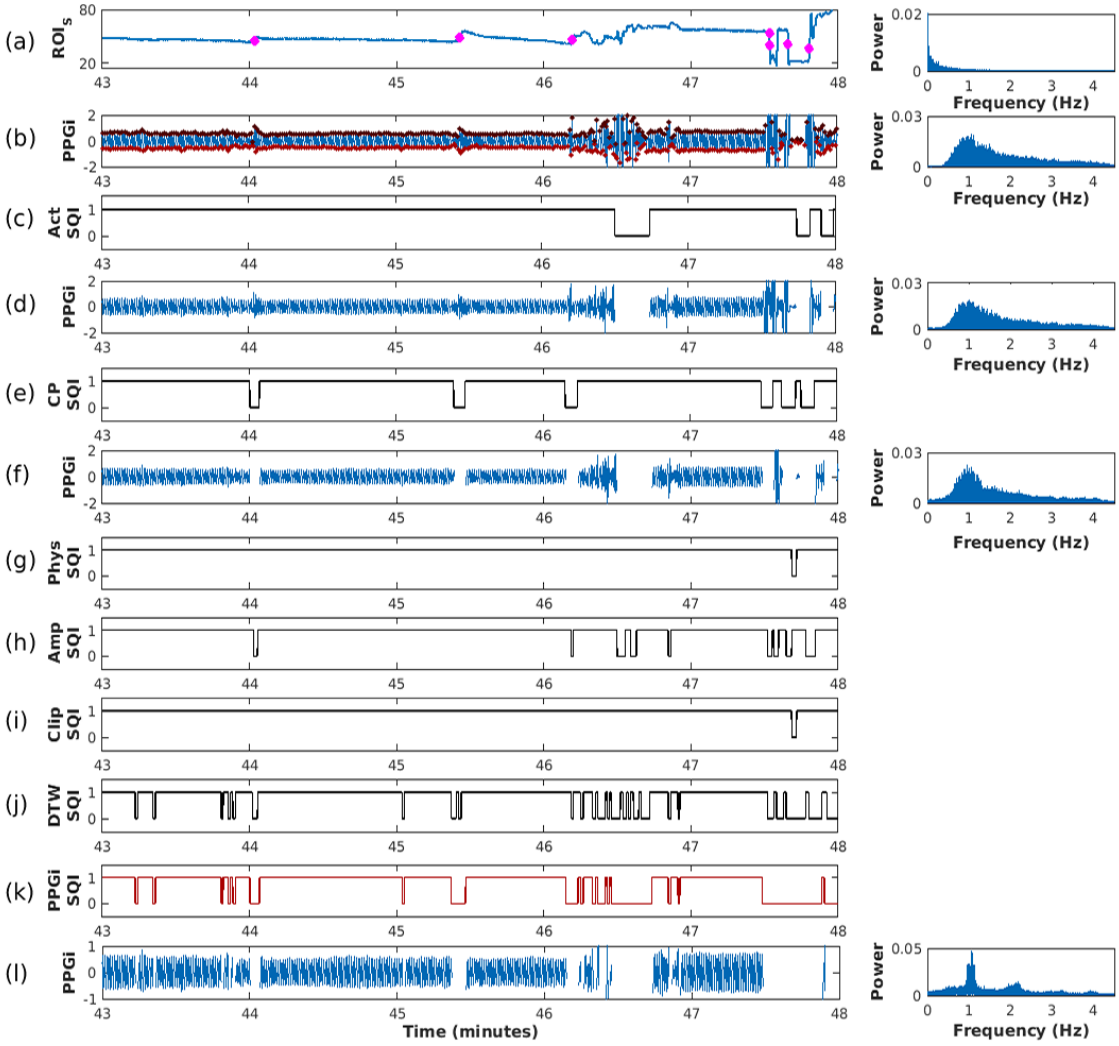


Figure 6.42: Beat quality assessment for a 5-minute window. (a) Input ROI_S with the change points marked in magenta, (b) the pre-processed PPGi signals with beat onsets and peaks marked in red and dark purple respectively, (c) activity index, (d) activity applied to the PPGi signal, (e) change point detection, (f) change point applied to the PPGi signal, (g) beats outside heart rate physiological range, (h) beats outside window amplitude range, (i) clipped beats, (j) DTW quality metric, (k) resulting PPGi SQI, (l) valid beats for the window

6.7 Heart rate estimation methods

As mentioned in section 6.5.2, to estimate heart rate from the video camera, a PPGi waveform is extracted for each of the three colour channels from all of the subject regions of interest (ROI_S). Subsequently, the quality of each input PPGi signal is individually analysed with the methods described in the previous sections. Each waveform is then subdivided into 15-second windows. Longer windows will lead to additional delays in a real-time implementation and decrease the percentage of data for which the heart rate can be estimated as patient activity will cause more windows to be rejected. The separation between consecutive windows was chosen to be 1 second to provide a large amount of overlap between windows. Therefore, heart rate estimates are reported every second.

The heart rate for each window from every input PPGi signal is computed individually using four algorithms described in this section: beat counting, Fast Fourier Transform (FFT), the angle of the dominant pole from an autoregressive model (AR) in the pole-zero plot, and autoregressive model with pole cancellation (ARk). Data fusion methods, described in the next section, select the optimal heart rate for the patient.

To help illustrate the estimation methodology, and to serve as a baseline comparison to the data fusion methods, this section describes heart rate estimation using the PPGi waveform extracted only from the green channel from a single region of interest.

6.7.1 Beat counting

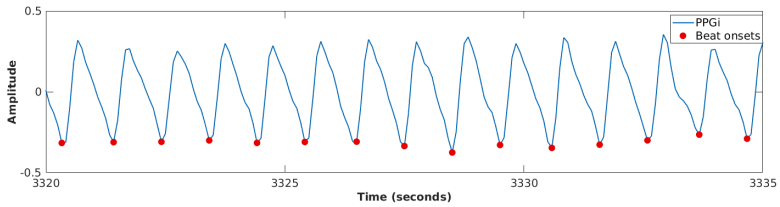


Figure 6.43: Heart rate estimation for a good quality 15-second time series using the beat count method. There are 14 beat pulses spanning a time length of 14.3 seconds, corresponding to a heart rate of 58.6 beats per minute.

Once the beat onsets are located and the quality of the signal is computed, counting the beats is a straightforward process. The heart rate estimate is taken as the number of pulses in the window, divided by the time interval from the first onset to the last onset. Figure 6.43 shows an example of the beat counting process for a good-quality 15-second window. The heart rate for this time period is estimated to be 58.6 $\mathrm{beats/min}$.

Counting the beats in a PPGi signal can also be done by counting the peaks . Figure 6.44 compares the two beat-counting approaches for a typical 4-hour dialysis session, for which the PPGi signal was extracted from a single region of interest located on the patient's forehead. Both methods produce heart rate estimates with a mean absolute error of 0.87 $\mathrm{beats/min}$ for approximately 40% (1:36 hours) of the total length of the video recording (4:01 hours).

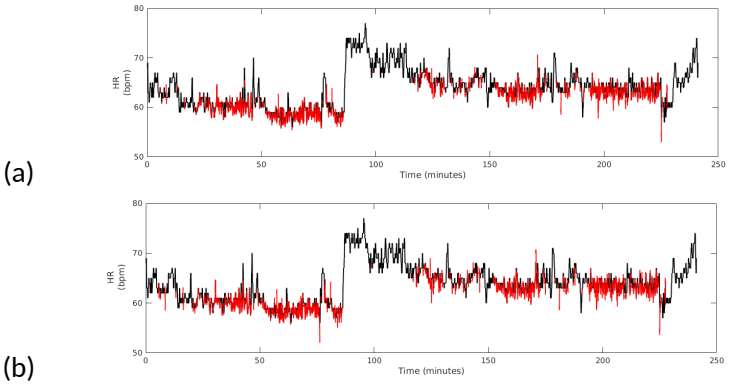


Figure 6.44: Heart rate estimation using the beat counting method from a single region of interest, located on the patient's forehead, from a typical 4-hour dialysis session. The reference heart rate produced by the finger pulse oximeter is plotted in black, whereas the camera estimates are plotted in red. (a) Counting the beat onsets, 40.20% (1:36 hours) of the data was estimated with a mean absolute error (MAE) of 0.87 $\mathrm{beats/min}$. (b) Counting the peaks, a comparable proportion of the data is estimated (40.27% corresponding to 1:37 hours) with a similar MAE of 0.87 $\mathrm{beats/min}$.

6.7.2 Fast Fourier Transform (FFT)

For the Fast Fourier Transform (FFT) method, the frequency of the maximum peak in the spectrum is taken to be the heart rate value for the window. A Hamming window is applied to the input PPGi window to minimise the effect of spectral leakage.

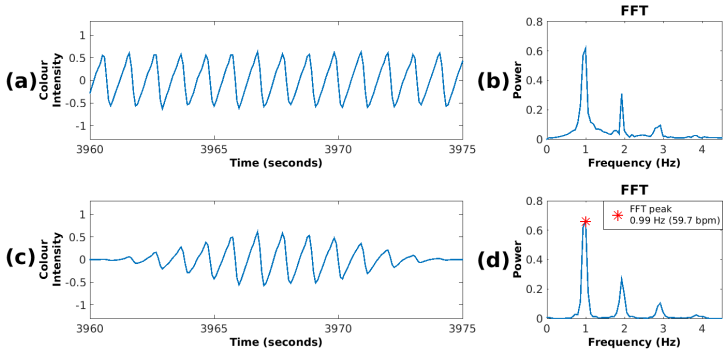


Figure 6.45: Heart rate estimation for a 15-second good-quality time series from a subject region of interest (ROI_S) using the FFT method. The maximum peak at around 1 Hz corresponds to a heart rate of 60 beats per minute. (a) Input time series; (b) the frequency components of the input time series; (c) a Hamming window applied to the input time series; (d) the resulting FFT spectrum, from which the heart rate estimate is computed.

Figure 6.45 shows an example of computing a heart rate estimate using the FFT method for a 15-second good-quality PPGi window extracted from a region of interest from the patient's forehead. The large peak close to 1 Hz corresponds to a heart rate of about 60 beats per minute.

The maximum peak in the FFT can also be prone to artefacts during periods of poor quality, as shown in figure 6.46. The heart rate estimate from the video camera data is computed as 48 beats per minute and the reference heart rate estimate reported by the pulse oximeter is 72 beats per minute.

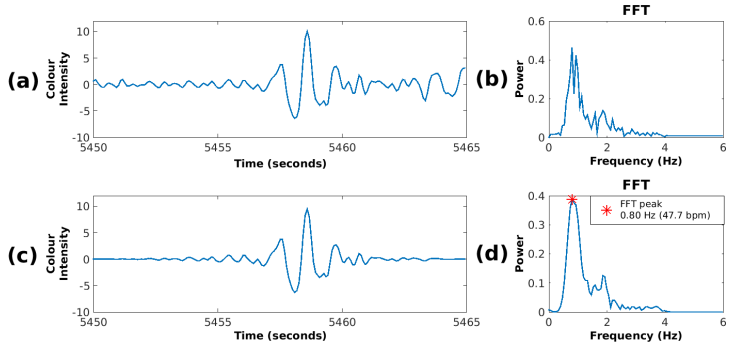


Figure 6.46: Heart rate estimation for a 15-second poor-quality time series from the subject region of interest (ROI_S) using the FFT method. The maximum peak at around 0.8 Hz corresponds to a heart rate of 48 beats per minute, while the reference heart rate reported by the pulse oximeter was 72 beats per minute. (a) Input time series; (b) the frequency components of the input time series; (c) a Hamming window applied to the input time series; (d) the resulting FFT spectrum, from which the heart rate estimate is computed.

Figure 6.47 shows the heart rate estimates using the FFT method for a typical 4-hour dialysis session for a PPGi signal extracted from a single region of interest located on the patient's forehead. The mean absolute error is 0.87 beats/min for approximately 40% (1:36 hours) of the total length of the video recording (4:01 hours).

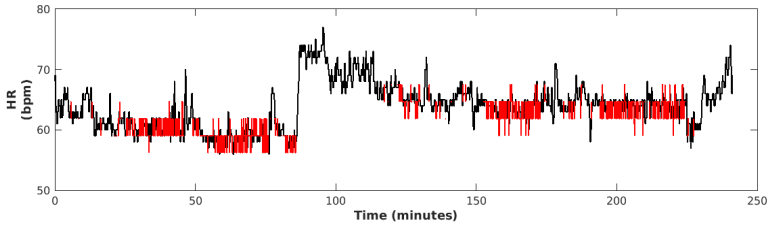


Figure 6.47: Heart rate estimation using the FFT method from a single region of interest, located on the patient's forehead, from a typical 4-hour dialysis session. The reference heart rate produced by the finger pulse oximeter is plotted in black, whereas the camera estimates are plotted in red. 40.20% (1:36 hours) of the data is estimated with a mean absolute error (MAE) of 1.02 beats/min.

The FFT algorithm used to estimate heart rate, computes the Discrete Fourier Transform (DFT) for each 15-second window of the PPGi signal. As can be seen in figure 6.47, FFT-based algorithms are affected by quantisation errors given by the frequency resolution (fr). The frequency resolution of a Discrete Fourier Transform is given by :

$$fr = \frac{f_s}{NFFT} \tag{6.29}$$

where f_s is the sampling frequency of the PPGi signal. $NFFT$ is the number of samples used to compute the FFT, given by the next power of two greater than the window sample size N:

$$2^{NFFT} \geq N \tag{6.30}$$

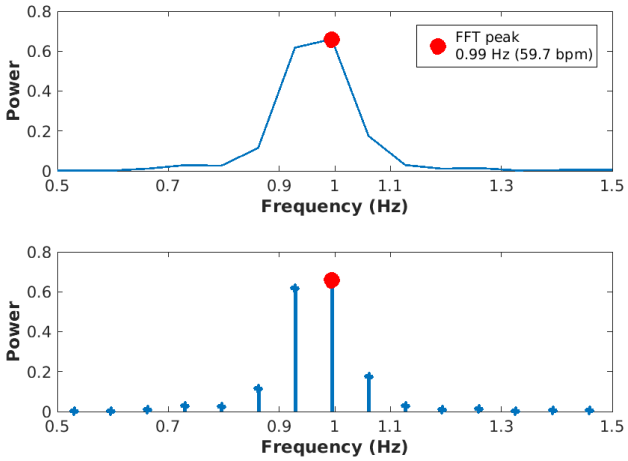


Figure 6.48: Frequency resolution for the Discrete Fourier Transform. (a) The FFT spectrum is typically plotted as connected lines. (b) Same FFT spectrum plotted as stem lines with a frequency resolution of 0.047 Hz, corresponding approximately to 2.81 $\mathrm{beats/min}$. The heart rate estimate of 59.7 $\mathrm{beats/min}$ could have been computed from a patient with a true heart rate between 58.3 $\mathrm{beats/min}$ to 61.1 $\mathrm{beats/min}$.

Given that the video data was recorded at a frame rate of 12 Hz, and the window length for heart rate estimation is 15 seconds, the frequency resolution is 0.047 Hz. Therefore, the each FFT bin represents approximately 2.81 beats per minute. This quantisation effect in the FFT can be seen in figure 6.48. The heart rate estimate of 59.7 $\mathrm{beats/min}$, corresponding to the peak of the FFT at 0.99 Hz, could have been computed from a patient with a true heart rate between 58.3 $\mathrm{beats/min}$ to 61.1 $\mathrm{beats/min}$ ($HR \pm fr/2$).

6.7.3 AR dominant pole

An autoregressive model (AR) is a type of random process that attempts to predict the output of a system based on its previous outputs, and can be specified for a sample time series x_t as:

$$x_t = c + \sum_{i=1}^{p} a_i x_{t-i} + \varepsilon_t \tag{6.31}$$

where :

- c is a constant (often omitted for simplicity).
- p is the model order.
- $a_1, ..., a_p$ are the model parameter coefficients.
- ε_t is a white noise Gaussian process.

The current value of x_t, is a linear combination of the p previous values of x_t and the current value of ε_t. The transfer function in the Z-domain relating the input to the output is given by:

$$H(Z) = \frac{1}{\sum_{k=1}^{p} a_k Z^{-k}} = \frac{Z^p}{(Z - Z_1)...(Z - Z_p)} = \frac{Z^p}{\vec{P}_1 \vec{P}_2 ... \vec{P}_p} \tag{6.32}$$

The denominator in equation (6.32) can be factored into p terms, where $\vec{P}_1\vec{P}_2...\vec{P}_p$ are vectors extending from any point Z in the complex plane to each of the p poles $Z_1, Z_2, ..., Z_p$ of $H(Z)$. As $H(Z)$ has no finite zeros, autoregressive models can be viewed as the output of an all-pole infinite impulse response filter whose input is white noise. A more detailed discussion of autoregressive models is presented in Appendix A.

After computing the autoregressive model for the PPGi window, the pole with the highest magnitude (largest radius in the zero-pole diagram in figure 6.49(e)) is identified and its frequency (angle on the pole-zero diagram) is used as the heart rate estimate. Figure 6.49 shows an example of the results from the AR dominant pole method obtained for a 15-second good-quality time series from the subject region of interest. The largest peak around 0.96 Hz indicates the dominant frequency corresponding to a heart rate of approximately 58 beats per minute.

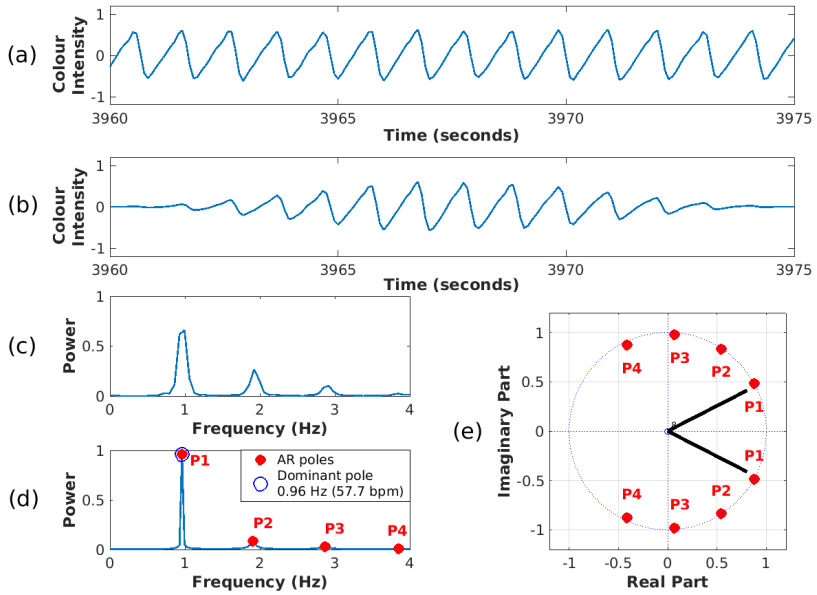


Figure 6.49: Computing the dominant pole for a typical 15-second good-quality window, AR model order 8, and 12 frames-per-second video, (a) Input time series; (b) a Hamming window applied to the input time series, (c) and (d) a comparison between the FFT and AR frequency components. The pole P_1 at about 0.96 Hz indicates the dominant frequency corresponding to a heart rate of approximately 58 beats per minute; (e) the Autoregressive Model zero-pole plot showing the 4 poles $P_1 to P_4$. The P_1 pole has the highest magnitude at 0.99.

Figure 6.50 shows the heart rate estimates using the AR dominant pole method for a typical 4-hour dialysis session for a PPGi signal extracted from a single region of interest located on the patient's forehead. The mean absolute error is 0.85 $\mathrm{beats/min}$ for approximately 35.28% (1:25 hours) of the total length of the video recording (4:01 hours).

Although autoregressive models are not affected by the same quantisation errors as the previous Fast Fourier Transform algorithm, the frequency associated with the angle of the dominant pole of an autoregressive model can often be of the harmonics of the fundamental frequency from the underlying PPGi time series. This dominant pole fluctuation causes a reduction of the proportion of time heart rate can be estimated in figure 6.50.

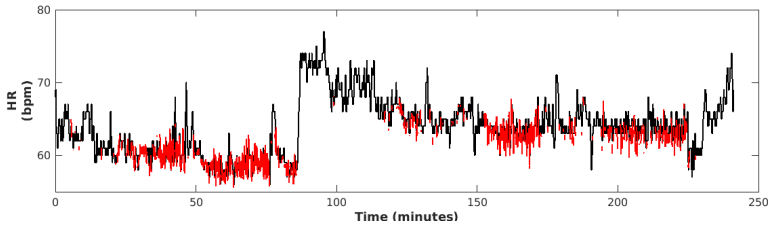


Figure 6.50: Heart rate estimation using the AR dominant pole method from a single region of interest, located on the patient's forehead, from a typical 4-hour dialysis session. The reference heart rate produced by the finger pulse oximeter is plotted in black, whereas the camera estimates are plotted in red. 35.28% (1:25 hours) of the data is estimated with a mean absolute error (MAE) of 0.85 $\mathrm{beats/min}$.

6.7.4 AR pole cancellation (ARk)

A challenge during video recordings in a real hospital setting is the ambient light interference from artificial lights (such as fluorescent lights) not only found during night time recordings but often within daylight hours as well [65]. The modulation of fluorescent lamps powered by alternating current is well known [340], resulting in flicker at twice the frequency of electrical supply (50 Hz in Europe results in a flicker at 100 Hz). Although the 100 Hz frequency of intensity variation is much higher than the heart rate being measured by the camera, in practice the intensity variations are aliased down to much lower frequencies because of the sampling process.

The images in the clinical study analysed in this thesis were sampled at 12 Hz, which is much lower than the sampling frequency which would be required (200 Hz) to avoid aliasing of the 100 Hz lights. Aliasing components (artefacts) are often found at frequencies such as 4 Hz and 2 Hz. However, it is not possible to predict exactly what frequencies will result from this aliasing process, and it is not effective simply to filter at specific frequencies, as the filters would need to be re-tuned in each setting to track the aliasing artefacts.

The PPGi signal obtained by the algorithms discussed previously is extracted from regions of interest on the patient's face. As before, these are called the subject regions of interest (ROI_S). To suppress the ambient light interference, a reference signal is also extracted from a region of interest not on the patient's body. This signal is called background region of interest (ROI_R).

Because the two regions of interest are both imaged by the same camera at the same time intervals, ambient light interference or aliasing artefacts will be present in both ROI_R and ROI_S. However, the signal from the subject region of interest will additionally have poles corresponding to spectral components representing the PPGi signal.

An autoregressive model of order 8 is computed for both ROI_S and ROI_R. The non-physiological frequency components can be identified as the poles which are common to both the autoregressive models. From equation (6.32), it should be possible to subtract the non-physiological frequency components in ROI_S from ROI_R by cancelling them from the denominator in equation (6.32) for the AR model for ROI_S. Poles are regarded as present in both the subject and reference regions of interest if they are within two angular degrees of each other. The new transfer function $H'(z)$, without these cancelled poles in the denominator, will now represent the transfer function for

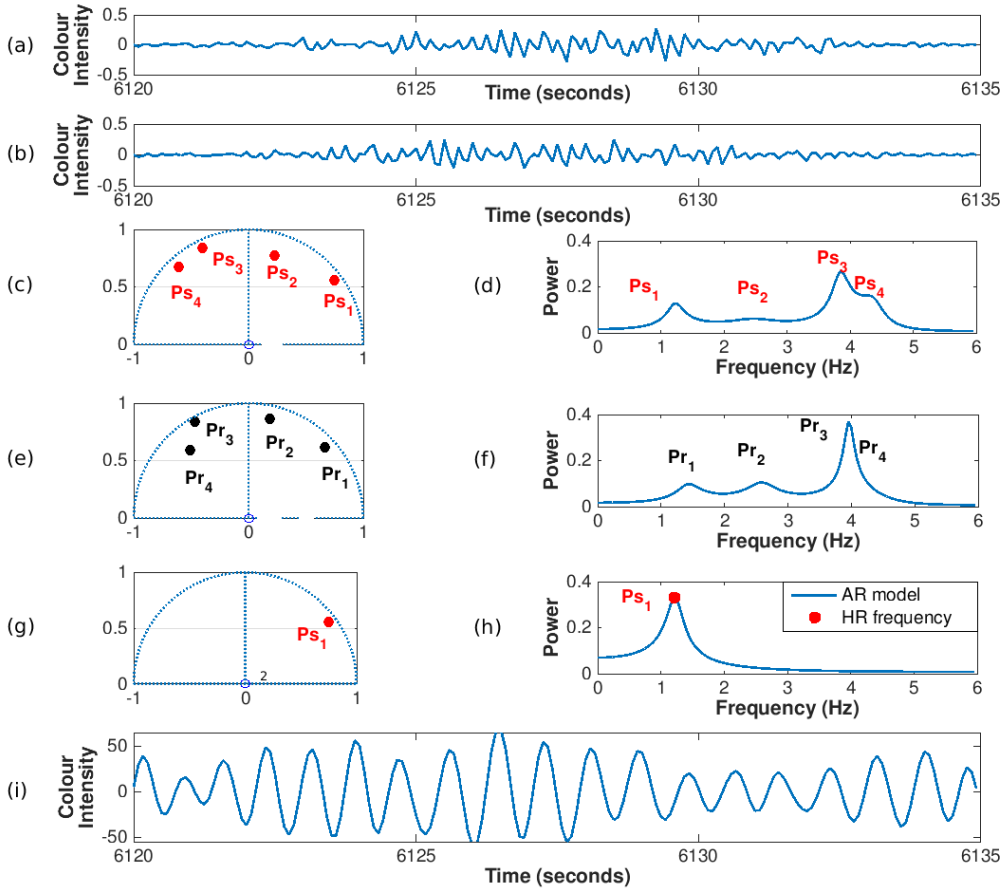


Figure 6.51: Autoregressive model pole cancellation process for a 15-second window from the (a) subject region of interest (ROI_S) and (b) background region of interest (ROI_R). (c) Frequency components and (d) zero-pole plot for ROI_S. (e) Frequency components and (f) zero-pole plot for ROI_R. (g) Similar poles from ROI_R in ROI_S are cancelled creating a new autoregressive model ARk and (h) the power spectrum is computed from the remaining pole. The heart rate estimate is taken as the peak of the reconstructed power spectrum at 1.2 Hz, corresponding to a heart rate of approximately 72 $\mathrm{beats/min}$. (i) From the remaining pole, a signal can be generated from a white noise input source.

the light intensity from the region of interest (from the subject's face) without the ambient light interference. This new autoregressive model is denoted by ARk. The pole cancellation process shown in figure 6.51 for a 15-second window.

After the non-physiological frequency components have been cancelled in ARk, the peak in the reconstructed power spectrum (more details presented in Appendix A) from the autoregressive model ARk is computed and its frequency is used as the heart rate estimate for the window (see figure 6.51(h)).

Figure 6.52 shows the heart rate estimates using the Ark method for a typical 4-hour dialysis session for a PPGi signal extracted from a single region of interest located on the patient's forehead. The mean absolute error is 1.09 $\mathrm{beats/min}$ for approximately 47.8% (1:55 hours) of the total length of the video recording (4:01 hours).

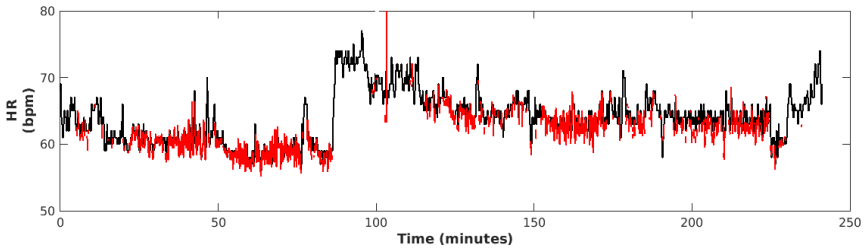


Figure 6.52: Heart rate estimation using the ARk method from a single region of interest, located on the patient's forehead, from a typical 4-hour dialysis session. The reference heart rate produced by the finger pulse oximeter is plotted in black, whereas the camera estimates are plotted in red. 47.8% (1:55 hours) of the data is estimated with a mean absolute error (MAE) of 1.09 $\mathrm{beats/min}$.

6.7.5 Summary

The methods presented in this section, estimate heart rate from a single region of interest. Although the estimation error is low (typically within 1 $\mathrm{beats/min}$) for the example record shown in figure 6.53, the proportion of time the estimates are valid is limited. This restriction is due to the fact that while the patient moves and interacts with the clinical staff during the typical 4 hours of a dialysis session, artefacts corrupt the single region of interest selected. More robust methods of combining all the regions of interest are presented in the following sections.

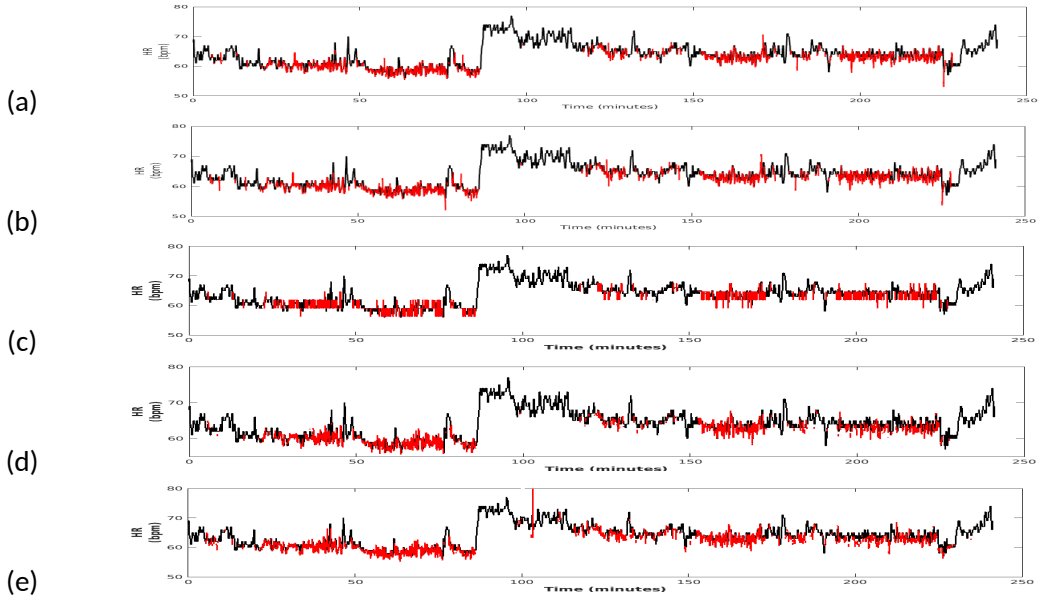


Figure 6.53: Summary of heart rate estimates using the proposed methods on a single region of interest, located on the patient's forehead, from a typical 4-hour dialysis session. The reference heart rate produced by the finger pulse oximeter is plotted in black, whereas the camera estimates are plotted in red. (a) Counting the beat onsets estimates 40.20% (1:36 hours) of the data with a mean absolute error (MAE) of 0.87 $\mathrm{beats/min}$. (b) Counting the beat peaks estimates 40.27% (1:37 hours) with a similar MAE of 0.87 $\mathrm{beats/min}$. (c) 40.20% (1:36 hours) of the data is estimated with the FFT method with a MAE of 1.02 $\mathrm{beats/min}$. (d) The AR dominant pole method estimates 35.28% (1:25 hours) with a MAE of 0.85 $\mathrm{beats/min}$. (e) 47.8% (1:55 hours) of the data is estimated with the pole cancellation method (ARk) with a MAE of 1.09 $\mathrm{beats/min}$.

6.8 Data fusion

Once the heart rate estimates are computed for each colour channel from each region of interest, the overall heart rate estimate for each of the methods in the previous section is computed using three data fusion techniques: *Simple data fusion*, for which the heart rate estimate is computed as the median of all the regions of interest that have an associated signal quality above a given threshold; *1D Kalman*, for which the overall heart rate estimate is computed using a single Kalman filter; *ND Kalman*, for which the estimates for every colour channel from each region of interest are tracked with an individual Kalman filter and then the output from each of the filters is combined to compute the overall heart rate estimate for the current window.

6.8.1 Simple data fusion

The first data fusion method computes the heart rate estimate for a given window, labelled HR_{window}, as simply the median of all the heart rate estimates from each colour channel from the regions of interest that have an associated signal quality index greater than 0.8, as defined by the following equation:

$$HR_{window} = median(HR_{ROI_{S,i}}) \ , \ \forall sqi(ROI_{S,i}) > 0.8 \\ , i = \{1, 2, ..., n\} \tag{6.33}$$

Figure 6.54 shows the results for a sample 4-hour dialysis recording. Taking into account the three colour channels from regions of interest on the patient's face, the simple data fusion technique improves the estimation when compared with the use of a single colour channel and a single region of interest (as in section 6.7). Simple data fusion applied to any of the methods gives heart rate estimates for around 70% of the time of the sample record shown in figure 6.54, with a mean absolute error close to 1 $\mathrm{beats/min}$. The best method with simple data fusion is the autoregressive model with pole cancellation (ARk), for which 78.62% (3:09 hours) of the data is estimated with a mean absolute error of 0.97 $\mathrm{beats/min}$.

6.8.2 Single Kalman filter

Computing heart rate estimates using the simple data fusion technique, as described in the previous section, is a clear improvement over the estimation of heart rate using a single region of interest. Nevertheless, simple data fusion is still affected by errors when the majority of the regions of interest produce incorrect heart rate estimates. These errors are usually caused by the patient's speech patterns, frowning or other facial changes, resulting in variations of the heart rate estimates unrelated to physiology.

A Kalman filter [154] is used to track the heart rate estimates output by the simple data fusion method, adjusting the estimates based on the signal quality of the data window. When the quality is low, the Kalman filter estimate is trusted more; when the quality of the data window is high, the heart rate estimate from the simple data fusion method is weighted more. This method is called "Kalman 1D".

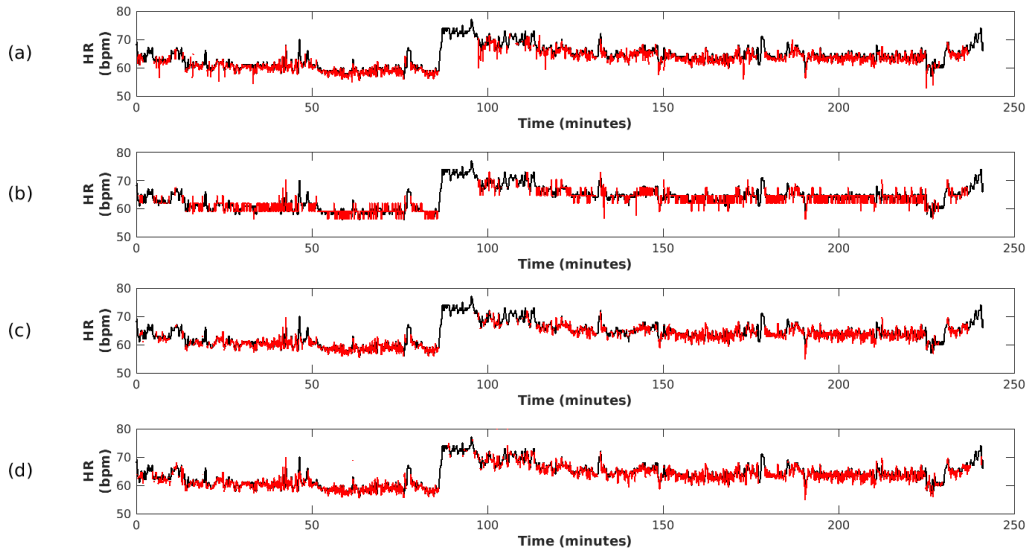


Figure 6.54: Summary of heart rate estimates using the proposed methods with the simple data fusion technique, from a typical 4-hour dialysis session. The reference heart rate produced by the finger pulse oximeter is plotted in black, whereas the camera estimates are plotted in red. (a) Counting the beat onsets estimates 71.35% (2:52 hours) of the data with a mean absolute error (MAE) of 0.85 $\mathrm{beats/min}$. (b) 71.35% (2:52 hours) of the data is estimated with the FFT method with a MAE of 1.05 $\mathrm{beats/min}$. (c) The AR dominant pole method estimates 68.51% (2:45 hours) with a mean absolute error (MAE) of 0.79 $\mathrm{beats/min}$. (d) 78.62% (3:09 hours) of the data is estimated with the pole cancellation method (ARk) with a mean absolute error of 0.97 $\mathrm{beats/min}$.

The Kalman filter may be considered a specific instance of a recursive Bayesian filter for the case where the probability densities on the state are Gaussian [300, 354]. The Kalman filter is popular in other fields such as robotics, being used for problems in tracking, localisation and navigation [300]. The Kalman filter is a recursive linear estimator which sequentially calculates an estimate for a continuous state $x(t)$ that evolves over time t on the basis of periodic observations $z(t)$ of the state. The Kalman filter uses an explicit statistical model not only for how the state evolves over time, but also for how the observations are related to the model parameters.

Although the generic Kalman filter model defines the evolution of a continuous-time system with continuous observations, for the estimation of heart rate from the video camera, the Kalman filter in this section is implemented for a discrete-time scenario, where $t_k = k$. The state x, representing the heart rate for a discrete-time process, is governed by the equation:

$$x(k) = F(k)x(k-1) + B(k)u(k) + G(k)v(k) \tag{6.34}$$

where $u(k)$ is a known control input, $v(k)$ is a random variable describing the uncertainty in the process model or evolution of the state. $F(k)$, $B(k)$ and $G(k)$ describe the contributions of the previous state (at time $k-1$), controls and process noise to state transition. An observation (measurement) model is defined as:

$$z(k) = H(k)x(k) + D(k)w(k) \tag{6.35}$$

where $w(k)$ is a random variable describing the uncertainty of the observation. $H(k)$ and $D(k)$ describe the contribution of the state and measurement noise to the observation. As stated earlier, an important assumption of the Kalman filter is that the random variables $v(k)$ and $w(k)$, representing the process and measurement noise respectively, are all Gaussian and zero-mean:

$$E[v(k)] = E[w(k)] = 0\ ,\ \forall k \tag{6.36}$$

with known covariances Q and R given by:

$$E[v(i)v^T(j)] = \delta_{ij}Q(i)\ ,\ E[w(i)w^T(j)] = \delta_{ij}R(i) \tag{6.37}$$

The estimate of the state at a time k given all information up to time k is written as $\hat{x}(k|k)$. The estimate of the state at a time k given only information up to time $k-1$, also called the prediction, is written as $\hat{x}(k|k-1)$

The Kalman filter operates in two stages, prediction and time update, also known as the predictor and corrector equations respectively. A prediction $\hat{x}(k|k-1)$ of the state at time k and its covariance $P(k|k-1)$ are computed by:

$$\hat{x}(k|k-1) = F(k)\hat{x}(k-1|k-1) + B(k)u(k) \tag{6.38}$$

$$P(k|k-1) = F(k)P(k-1|k-1)F^T(k) + G(k)Q(k)G^T(k) \tag{6.39}$$

After an observation $z(k)$ is made at time k, the state estimate and estimate covariance are updated according to:

$$\hat{x}(k|k) = \hat{x}(k|k-1) + K(k)[z(k) - H(k)\hat{x}(k|k-1)] \tag{6.40}$$

$$P(k|k) = P(k|k-1) - K(k)S(k)K^T(k) \tag{6.41}$$

where $K(k)$ is the Kalman gain given by:

$$K(k) = \frac{P(k|k-1)H(k)}{S(k)} \tag{6.42}$$

and $S(k)$ is the innovation covariance, computed according to:

$$S(k) = R(k) + H(k)P(k|k-1)H^T(k) \tag{6.43}$$

The difference between the observation $z(k)$ and the predicted observation $H(k)\hat{x}(k|k-1)$ in equation (6.40) is called the residual or innovation $r(k)$:

$$r(k) = z(k) - H(k)\hat{x}(k|k-1) \tag{6.44}$$

Given that a single state (the heart rate estimate) needs to be tracked, and following Tarassenko et al [323] and Li [186], the Kalman filter is reduced according to the following conditions:

- No control input is assumed, therefore $B(k) = 0$.
- The heart rate at time k is assumed to be equal to the heart rate at time $k-1$, therefore $F(k) = 1$.
- The constant heart rate assumption is only valid during times for which the patient is quiet. However, it is also known that the heart rate varies from beat to beat (phenomenon of heart rate variability [203]) and this is modelled by assigning a value of 0.1 to the process noise variance Q, similar to [186]. When the patient is moving, the heart rate is expected to change. This is modelled by assigning a value of 0.2 to the process noise variance Q during times of motion (identified by the activity index in section 6.4.3).
- Given that the process noise is modelled using the two different values of Q, a value of 1 is assigned to $G(k)$.
- A value of 1 is assigned to $H(k)$.

The variance of the measurement noise R is modelled in a similar way to [186]. In order for the variance to increase towards infinity when the signal quality tends to zero, $R(k)$ is defined as follows:

$$R(k) = R(k) * exp(1/SQI_{window}^2 - 1) \tag{6.45}$$

where SQI_{window} is the signal quality index computed for the data window. Note that the variance grows to infinity if the signal quality goes to zero.

Therefore the prediction equations 6.38 and 6.39 are modified as:

$$\hat{x}(k|k-1) = \hat{x}(k-1|k-1) \tag{6.46}$$

$$P(k|k-1) = P(k-1|k-1)F^T(k) + Q(k) \tag{6.47}$$

The measurement update equations are now given by:

$$\hat{x}(k|k) = \hat{x}(k|k-1) + K(k)[z(k) - \hat{x}(k|k-1)] \tag{6.48}$$

$$P(k|k) = P(k|k-1) - K(k)S(k)K^T(k) \tag{6.49}$$

$$K(k) = \frac{P(k|k-1)}{S(k)} \tag{6.50}$$

$$S(k) = R(k) + P(k|k-1) \tag{6.51}$$

$$r(k) = z(k) + \hat{x}(k|k-1) \tag{6.52}$$

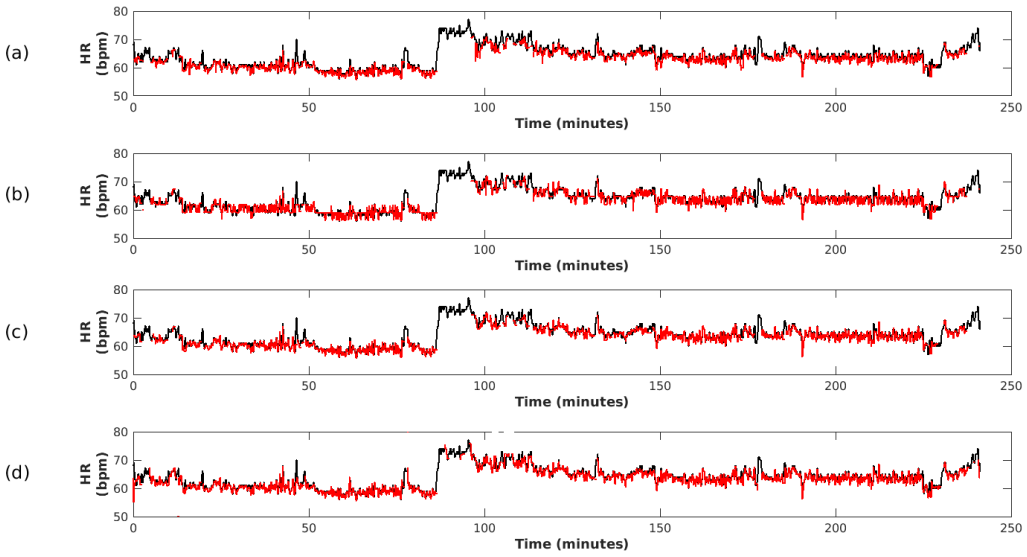


Figure 6.55: Summary of heart rate estimates using the proposed methods with a single Kalman filter, from a typical 4-hour dialysis session. The reference heart rate produced by the finger pulse oximeter is plotted in black, whereas the camera estimates are plotted in red. (a) Counting the beat onsets estimates 88.94% (3:34 hours) of the data with a mean absolute error (MAE) of 0.80 $\mathrm{beats/min}$. (b) 88.94% (3:34 hours) of the data is estimated with the FFT method with a MAE of 0.86 $\mathrm{beats/min}$. (c) The AR dominant pole method estimates 87.78% (3:34 hours) with a mean absolute error (MAE) of 0.68 $\mathrm{beats/min}$. (d) 95.47% (3:50 hours) of the data is estimated with the pole cancellation method (ARk) with a mean absolute error of 1.48 $\mathrm{beats/min}$.

Figure 6.55 shows the results for the sample 4-hour dialysis recording using the Kalman 1D data fusion method described above for the four heart rate estimation methods presented in section 6.7. The results are an improvement on the simple data fusion technique described in the previous section with all methods now able to estimate heart rate for more than 85% of the time of the sample record shown in figure 6.55. The best method is again the autoregressive model with pole cancellation (ARk) for which 95.47% (3:50 hours) of the data is estimated with a mean absolute error of 1.48 $\mathrm{beats/min}$.

6.8.3 Multiple Kalman filters

Section 6.8.2 described methods to compute the heart rate estimates using a single Kalman filter. With the methods described in this section, each colour channel from every region of interest is tracked with an individual Kalman filter, producing one estimate per ROI and per colour channel. The output from each Kalman filter is then combined following the proposed methods of Tarassenko et al [323] and Li et al [186]. This method is called "Kalman ND". The overall heart rate estimate for the current window HR_{window} is defined as:

$$HR_{window} = \sum_{l=1}^{n} \left(\frac{\prod_{j=1,j\neq l}^{n} \sigma_j^2}{\sum_{i=1}^{n} (\prod_{j=1,j\neq i}^{n} \sigma_j^2)} * HR_{ROI_{S,l}} \right) \ , \ l = 1, 2, ..., n \tag{6.53}$$

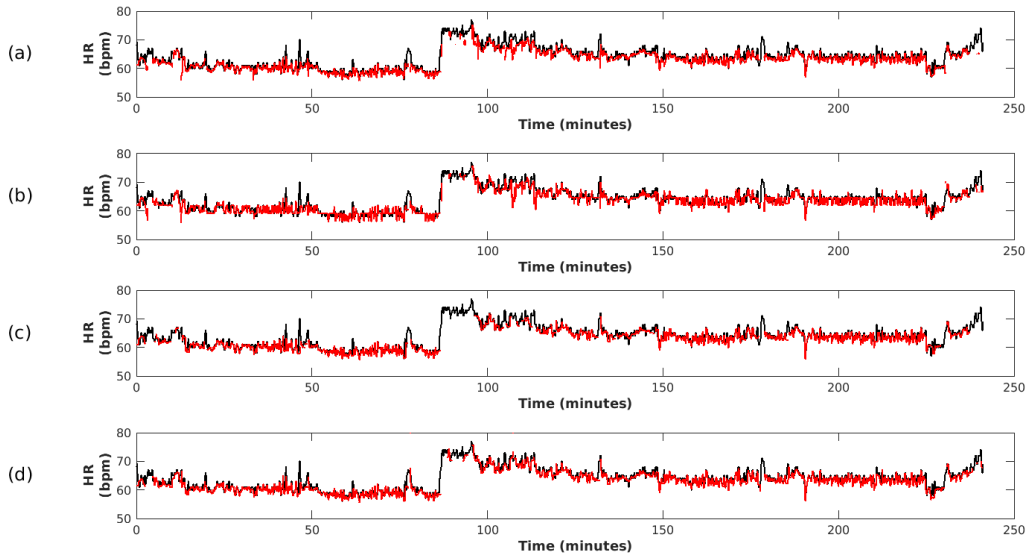


Figure 6.56: Summary of heart rate estimates using the proposed methods with multiple Kalman filters, from a typical 4-hour dialysis session. The reference heart rate produced by the finger pulse oximeter is plotted in black, whereas the camera estimates are plotted in red. (a) Counting the beat onsets estimates 94.26% (3:47 hours) of the data with a mean absolute error (MAE) of 0.92 $\mathrm{beats/min}$. (b) 94.26% (3:47 hours) of the data is estimated with the FFT method with a MAE of 0.91 $\mathrm{beats/min}$. (c) The AR dominant pole method estimates 87.74% (3:31 hours) with a mean absolute error (MAE) of 0.76 $\mathrm{beats/min}$. (d) 95.18% (3:49 hours) of the data is estimated with the pole cancellation method (ARk) with a mean absolute error of 1.24 $\mathrm{beats/min}$.

where $HR_{ROI_{S,l}}$ represents the heart rate estimate for the region of interest l. The residual σ is weighted by the SQI and is given by:

$$\sigma_i^2 = \left(\frac{r_i}{SQI_i} \right)^2 \tag{6.54}$$

where r_i is the Kalman filter residual for the region of interest i and SQI_i is the signal quality index for the data window for the region of interest i.

Figure 6.56 shows the results for the sample 4-hour dialysis recording using the Kalman ND data fusion technique for the four heart rate estimation methods presented in section 6.7. All methods are now able to estimate the heart rate for more than 90% of the time of the sample record shown in figure 6.56. The best method, as before, is the autoregressive model with pole cancellation (ARk), for which 95.18% (3:49 hours) of the data is estimated with a mean absolute error of 1.24 $\mathrm{beats/min}$.

6.8.4 Summary

The three data fusion techniques presented in this section improve the heart rate estimation process when compared with the methods using a single colour channel from a single region of interest, as described in section 6.7. The improvement comes from the ability of the data fusion methods to estimate heart rate accurately for a much greater proportion of the time.

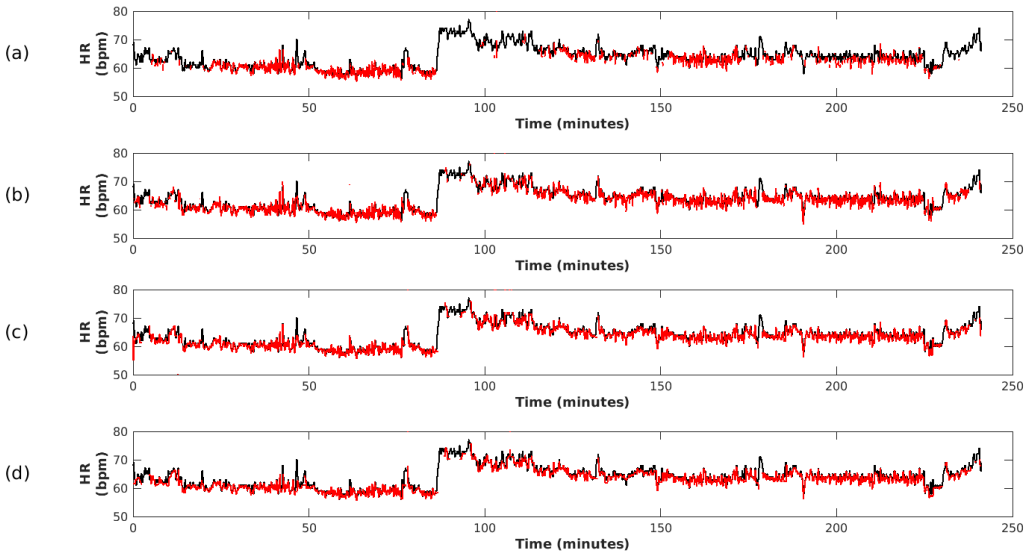


Figure 6.57: Summary of heart rate estimates using the ARk methods for all data fusion techniques, from a typical 4-hour dialysis session. The reference heart rate produced by the finger pulse oximeter is plotted in black, whereas the camera estimates are plotted in red. (a) For a single colour channel from a single region of interest, located on the patient's forehead, 47.8% (1:55 hours) of the data is estimated with a mean absolute error (MAE) of 1.09 $\mathrm{beats/min}$. (b) 78.62% (3:09 hours) of the data is estimated with the simple data fusion method with a MAE of 0.97 $\mathrm{beats/min}$. (c) 95.47% (3:50 hours) of the data is estimated with the single Kalman filter method with a MAE of 1.48 $\mathrm{beats/min}$. (d) 95.18% (3:49 hours) of the data is estimated with the multiple Kalman filters method with a MAE of 1.24 $\mathrm{beats/min}$.

Figure 6.57 shows the result of the the three heart rate data fusion algorithms and compares them with the simple method using a single channel and a single ROI for the same 4-hour dialysis recording as before, using the autoregressive model with pole cancellation (ARk) method as the benchmark algorithm. These results show that the proportion of time that the heart rate can be estimated for the record shown in figure 6.57 varies from 1:55 hours, for the "single-channel single-ROI" method, to 3:49 hours for the multi Kalman filter approach. The mean absolute error, when compared to the finger pulse oximeter, remain low, ranging from 0.97 $\mathrm{beats/min}$ to 1.24 $\mathrm{beats/min}$.

6.9 Results

6.9.1 Reference signals

According to table 5.4 in chapter 5, the number of video sessions recorded was 104, comprising a total video length of 369.1 hours. For 12 sessions, the recording from at least one pulse oximeter was interrupted due to patient discomfort, medical intervention, equipment malfunction or other external factors. 8 other sessions were discarded because the video camera recording equipment was interrupted by the clinical staff or the recording was not complete. Therefore, there were simultaneous recordings with both pulse oximeters and camera for only 84 sessions, consisting of a total recording time of 304.1 hours.

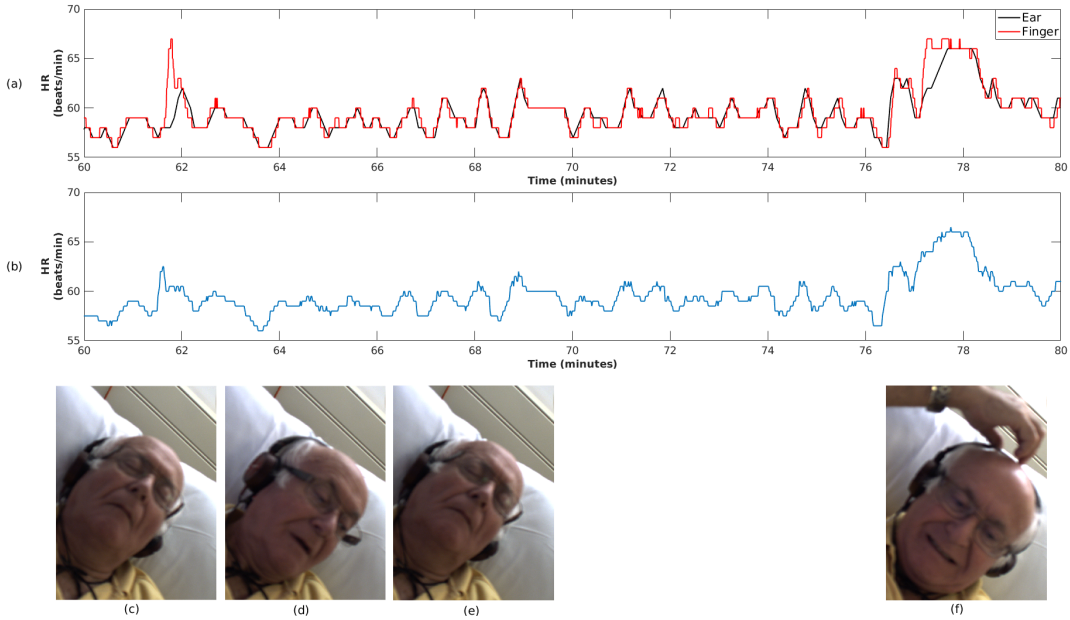


Figure 6.58: Comparison of the heart rate estimates provided by the pulse oximeters for a 20-minute section. (a) Heart rate values as reported by the ear pulse oximeter (in black) and the finger pulse oximeter (in red) for a sample 20-minute window from a typical dialysis session. (b) Reference heart rate computed as the mean of the estimates from each pulse oximeter. (c) Patient quietly sleeping around minute 61, (d) Short awakening period shortly before minute 62. (e) Patient sleeping during minute 63. (f) Patient actively moving around minute 77.

Figure 6.58(a) shows the comparison between the heart rate estimates produced by the two pulse oximeters for a 20-minute segment from a typical dialysis session. During this period, the patient was mostly sleeping (see figure 6.58(c) and 6.58(e)), hence the good agreement between the two estimates for most of the segment. Nevertheless, the difference between the estimates increases as the patient awakes for a short amount of time before minute 62 (see figure 6.58(d)) and during the period around minute 77 when the patient was actively moving (see figure 6.58(f)). Figure 6.58(b) shows the reference heart rate computed from the agreement between the estimates from the two pulse oximeters (see section 6.3).

6.9.2 Video analysis

The heart rate algorithms described in this chapter require video sessions during which the patient is stable and the location of the face is known. Following the criteria described in section 6.4.4 of this chapter, a further 23 sessions were rejected. Common reasons for rejecting a dialysis recording included periods of medical interventions or family visits during which the camera was moved away from the patient, video sessions with long time periods for which there was high patient activity, motion artefacts arising from the interaction between the patient and the clinical staff, or occurring when the patients were awake and engaging in conversation. Therefore, the resulting number of valid dialysis sessions is 61, comprising a total video length of 219.8 hours (approximately 60% of the total number of hours recorded originally), with a mean video length of 3.6 hours per dialysis session.

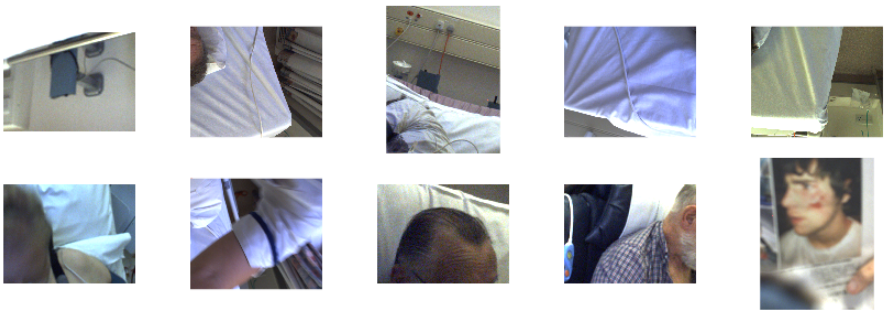


Figure 6.59: Examples of dialysis sessions that were rejected and therefore heart rate estimation was not carried out. Top row shows sessions where the camera was moved away from the patient to ease access for interventions. Bottom row shows sessions where the patient was only partially visible and the face could not be detected. The last image in the bottom row shows a face wrongly detected from the newspaper the patient was reading.

Figure 6.59 shows examples of images from selected sessions that were rejected by the video analysis algorithms. The top row represent sessions for which the camera was moved away from the patient to ease the access for clinical staff to perform a procedure, preserve the privacy of family members during visits or other personal circumstances reported by the patient to the nurse. The bottom row shows sessions for which the clinical staff were in the image frame during the video recording or the patient was only partially visible and therefore the face could not be detected. The last image on the right in the bottom row of figure 6.59 shows a snapshot from a dialysis session recorded from a male subject during which his face was wrongly detected from a picture on the back of the newspaper the patient was reading. Even though the wrong face was detected, heart rate was still successfully estimated form the video camera for short periods during which the patient motion was in synchrony with his heart beat.

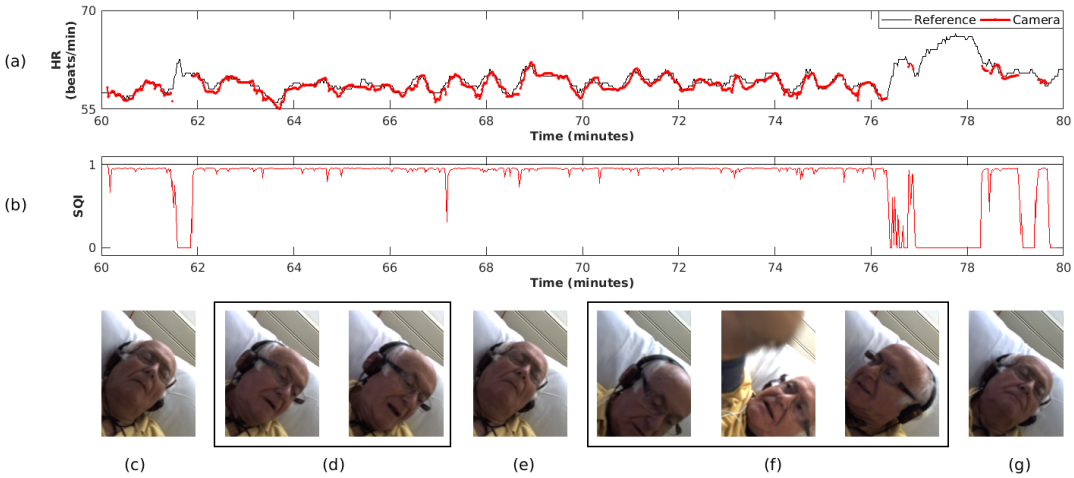


Figure 6.60: Heart rate estimation result for a sample 20-minute period from a typical 4-hour dialysis session. The reference heart rate is plotted in black, whereas the camera estimates are plotted in red. (a) The heart rate estimates. (b) Signal quality indices. Figures (c), (e) and (g) show snapshots from minutes 61, 66 and 79 respectively during which the patient was quietly sleeping. (d) Two sample images before minute 62 when the patient briefly woke up and move. (f) Three sample images between minutes 76 and 78 during which the patient was actively moving.

6.9.3 Heart rate estimation

Figure 6.60 shows the comparison between the reference heart rate estimates and the estimates from the video camera computed using the ARk algorithm and the Kalman ND data fusion method (see section 6.8.3) for the same 20-minute segment shown in figure 6.58. The camera estimates are in good agreement with the reference values for periods during which the patient was quietly sleeping, as shown in figures 6.60(c), 6.60(e) and 6.60(g) taken approximately at minute 61, 66 and 79 respectively. The activity algorithm correctly identified periods during which the patient was actively moving by decreasing the signal quality index and thereby preventing heart rate estimation, as shown in figure 6.60(d) for a period around minute 62 when the patient briefly woke up and in figure 6.60(f) for a period between minutes 76 and 78 when the patient was interacting with the clinical staff.

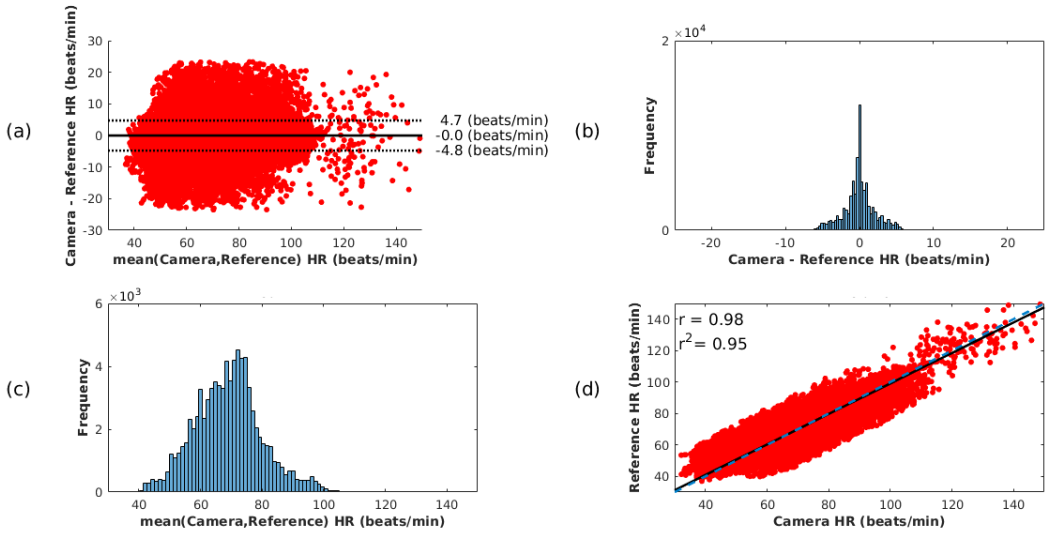


Figure 6.61: Agreement between the reference heart rate values (computed from the agreement between the two pulse oximeters) and the camera estimates (computed using the ARk algorithm and the Kalman ND data fusion method) for the valid dialysis video data, comprising a total recording time of approximately 219.8 hours. (a) The Bland-Altman plot presents no sensor bias. (b) The differences between the two pulse oximeters are normally distributed. (c) The distribution of the mean values show that most of the heart rate estimates are within the expected physiological range for adults. (d) The plot shows a strong correlation between the two devices, with a correlation coefficient of 0.98.

Figure 6.61 shows in more detail the agreement between the reference heart rate values and the video camera estimates (using the ARk algorithm and the Kalman ND data fusion method) for all the valid dialysis video data, comprising a total of recording time of 219.8 hours. The Bland-Altman plot, shown in figure 6.61(a), presents no sensor bias with narrow differences across the heart rate physiological range, implying a good agreement between the two measurements. The differences, shown in Figure 6.61(b), are normally distributed. The distribution of the mean values in figure 6.61(c), shows that the values are within the expected physiological range for adults. The values from the two devices have a positive correlation coefficient of 0.98, as shown in figure 6.61(d).

Table 6.2 presents the overall heat rate estimation results for all the methods and data fusion techniques analysed in this chapter. The mean absolute error, mean absolute deviation and the proportion of estimated time are reported for all the valid dialysis sessions, comprising a total recording time of 219.8 hours. The errors are measured comparing the heart rate estimates from the camera against the ground truth heart rate, computed from the mean of the two estimates from the pulse oximeters, as described in section 6.3.

Table 6.2: Heart rate estimation results for the all the methods and data fusion techniques, compared to the ground truth reference heart rate (computed from the mean of the two estimates from the pulse oximeters). Total time 219.8 hours.

Data fusion	Method	MAE (beats/min)	MAD (beats/min)	Time (%)
Simple	Beat count	2.1	2.2	42.3
	FFT	2.3	2.1	42.3
	AR dom pole	4.1	3.2	38.5
	ARk	2.6	1.9	51.7
Kalman 1D	Beat count	2.3	2.9	61.1
	FFT	2.2	3.1	61.1
	AR dom pole	3.8	3.1	56.4
	ARk	2.9	2.1	63.4
Kalman ND	Beat count	2.8	3.1	63.2
	FFT	3.4	4.1	63.2
	AR dom pole	4.1	4.2	59.6
	ARk	2.8	2.6	65.3

6.10 Discussion

6.10.1 Heart rate frequency maps

The estimation framework presented so far takes into account only the PPGi signal from the regions of interest on the patient's face as located by the face detection algorithms. The individual measurements are then combined to compute a single heart rate estimate. Another method to select regions of interest is to look for frequency content in the physiological range for heart rate in the video camera data. This requires a method for detecting periodic information in signals and a signal quality indicator indicative of the degree of confidence that the periodic information is valid.

Such a method can be based on an extension of autoregressive models, applying a threshold to the pole magnitude (the pole radius on the pole-zero plot as shown in figure 6.62) such that the spectral component represented by that pole is only regarded as valid if the pole magnitude exceeds a threshold. The frequency of the spectral component corresponding to a valid pole can be calculated from the phase angle of the pole. If the signal to be analysed is expected to have the cardiac frequency as the dominant frequency (as shown in figure 6.62), then the magnitude of the dominant pole is an indicator of how strong the cardiac frequency component is in the signal.

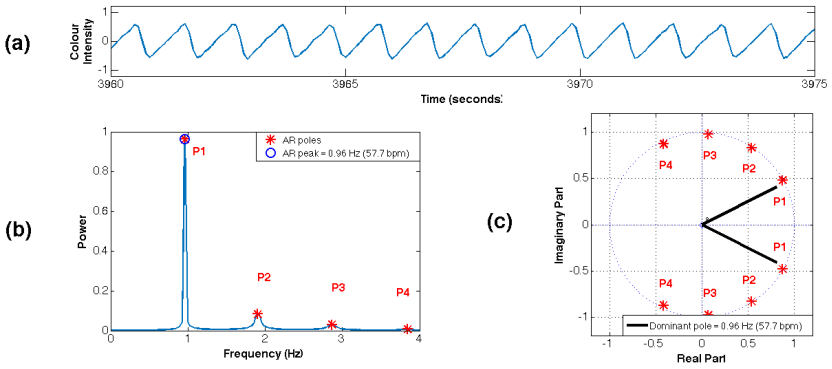


Figure 6.62: Computing the dominant pole for a typical PPGi window: (a) Input time series. (b) AR frequency components; the peak of the AR spectrum has a frequency of 0.96 Hz, corresponding to a heart rate of about 57.7 $\mathrm{beats/min}$. (c) the Autoregressive Model zero-pole plot shows the dominant pole P_1 with a magnitude of 0.99

The input image from a given colour channel of the video camera (typically the green channel) is this time divided into multiple small regions of interest, each typically of size 25x25 pixels rather than the 150x150 pixels regions used up until now on this chapter, as the aim of the work described in this section is to compute video maps with the greatest spatial resolution possible. For each of the new small ROI, the PPGi signal can be extracted by the methods presented in this chapter. An AR model of fixed order (typically 8) is applied to each PPGi signal, producing a set of poles, each with a given radius up to a maximum value of 1.0 (strength of the pole) and an angle (frequency corresponding to this pole). In each region of interest, the dominant pole is identified.

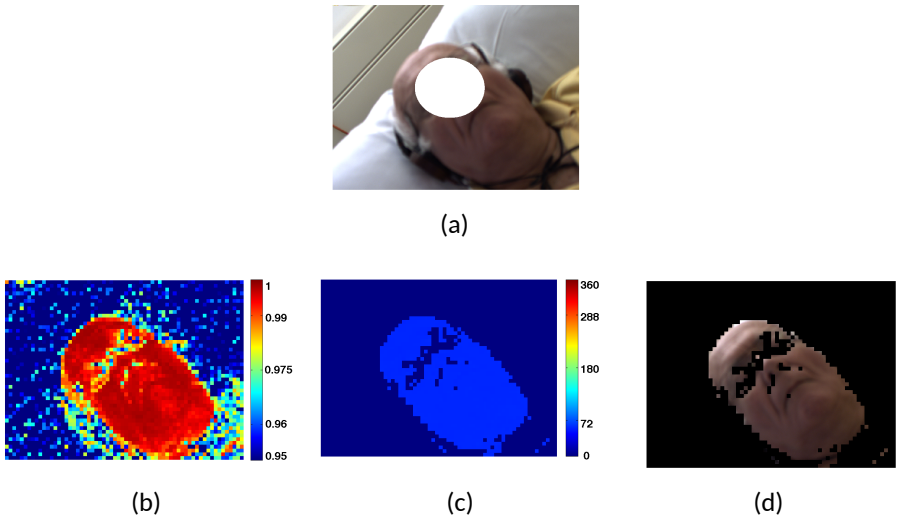


Figure 6.63: Heart rate maps for a typical 15-second video window using 25x25 pixel regions of interest. (a), The original image frame. (b) colour coding of the magnitude of the dominant pole, from 0.95 (minimum value-blue) to 1.0 (maximum value-red). The heart rate can potentially be computed from most of the face, apart from the eyes (the subject is wearing glasses), the nose and the mouth. The periodic component for heart rate is strongest on the patient's cheeks and forehead, corroborating other anatomical studies. (c) The actual heart rate values for each of the regions, (d) Mask of the pixels for which heart rate can be estimated from the image, unsurprisingly where skin is visible.

Typical results are shown in figure 6.62 which analyses the same 15 seconds of the input signals as in figure 6.49, However, in this case, the size of the subject regions of interest (ROI_S) is 25x25 pixels. Figure 6.63 shows how the radius and angle of the dominant pole, for each 25 x 25 region, are used to derive colour maps of heart rate. The first row in the figure shows the original input image. Figure 6.63(b) is a colour coding of the magnitude of the dominant pole, from 0.95 (minimum value blue) to 1.0 (maximum value red) for each 25 x 25 region. This is effectively a Signal Quality Index provided directly by the AR model, which shows in each case where the heart rate information can be estimated most accurately. It is clear from figure 6.63 that the most accurate estimation (according to the value of the dominant pole) is from most of the face, apart from the eyes (the subject is wearing glasses), the nose and the mouth. The periodic component for heart rate is strongest on the patient's cheeks and forehead, corroborating anatomical studies of the human face ([206]) where it is stated that there is a high skin blood flow in both of these regions. The actual values for the heart rate for each 25 x 25 pixel regions can be seen in Figure 6.63(b), where the values are colour-coded.

6.10.2 Continuous heart rate monitoring

As discussed in previous chapters, the appropriate monitoring of heart rate is vital to the physiologic assessment of the patient's health status. In an Intensive Care Unit (ICU) environment, heart rate is monitored continuously by using the electrocardiogram (ECG) as the gold standard method, where heart rate is typically measured as frequently as on a second-by-second basis [278].

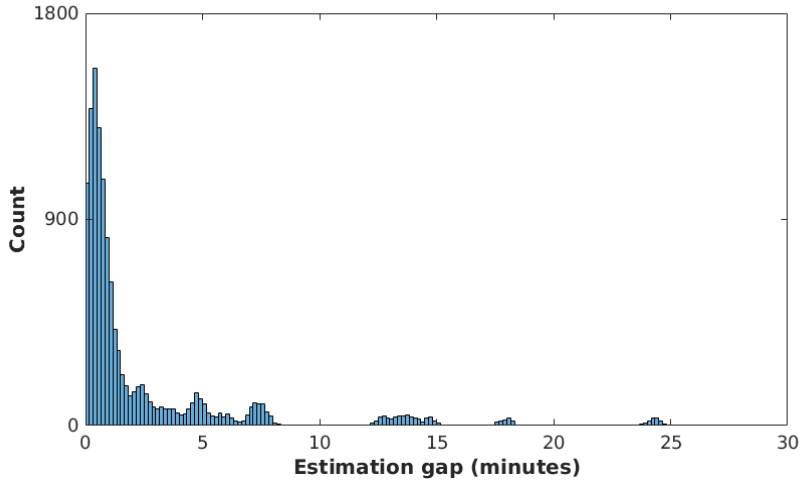


Figure 6.64: Gaps in heart rate estimation using the Kalman ND data fusion method. Most of the periods for which the camera could not produce a heart rate estimate were under 10 minutes.

Heart rate is often measured manually by nurses or other clinical staff, this is less frequently on other hospital wards. Evans et al [88] analysed a large set of publications raising concerns that there is limited guidance on the frequency that vital signs should be monitored, concluding that much of the information is based on surveys of nurses, clinical practice reports and expert opinions. For example, Schumacher et al [288] found that the frequency of the recording of vital

signs in post-operative patients varied between 15, 30 and 60 minutes, and as far as measuring vital signs once every 4 hours. Botti et al [35] interviewed nurses and found that vital signs were measured every 30 minutes for the first 4 hours following discharge from the recovery unit. More over, 86% of nurses have been reported to underestimate heart rate measurements in patients with atrial fibrillation [307].

Figure 6.64 shows the distribution of the time gaps between heart rate estimates from the camera for the entire valid dialysis data set, comprising a total recording time of 219.8 hours. Although heart rate estimates are not provided for every second of the entire video data, most of the periods for which the camera could not produce a reliable heart rate value were under 10 minutes, with a small subset of patients with longer gaps close to 15 and 25 minutes. This opens the possibility to monitor heart rate with a clinically relevant frequency.

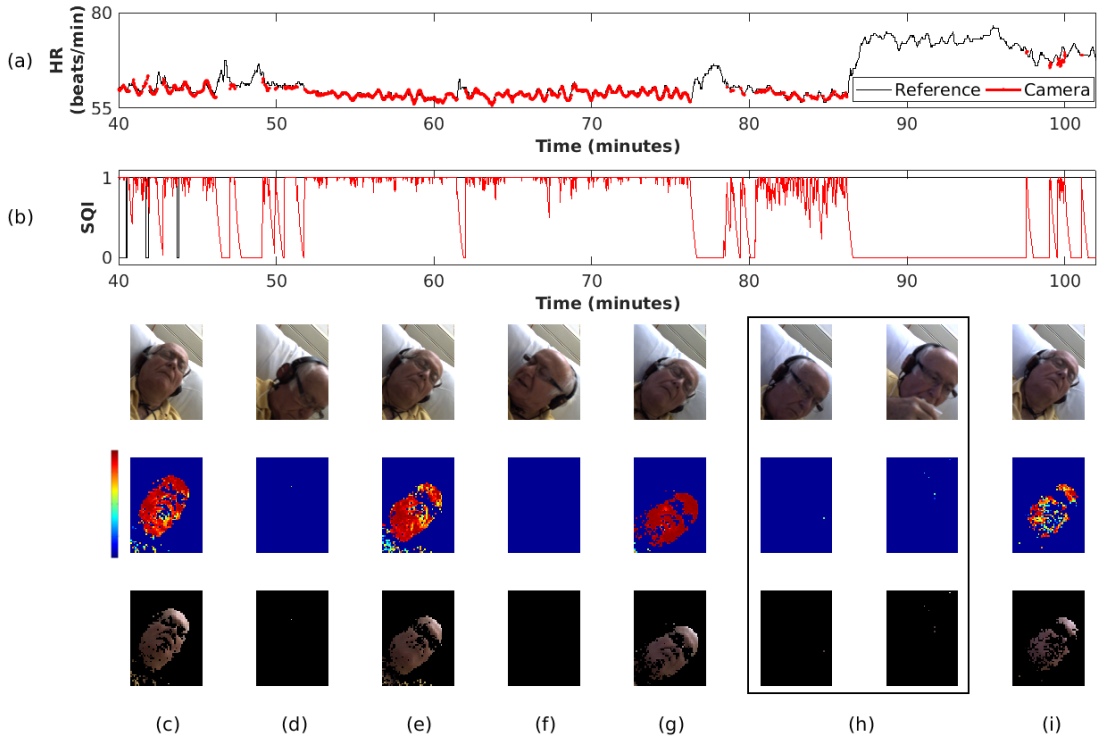


Figure 6.65: Heart rate estimation result for a sample 60-minute period from a typical dialysis session. Data for the reference heart rate is plotted in black, whereas the camera estimates are plotted in red. (a) Heart rate estimates comparison. (b) Signal quality indices. Columns (c) through (i) show sample snapshots from selected time periods: the top row shows the original video frames; the middle row shows the colour-coded heart rate frequency maps (red values represent areas with strong cardiac frequency and blue values represent areas where heart rate cannot be estimated); the bottom row shows the skin masks extracted from the original image frame using the frequency maps. (c) Patient quietly sleeping during minute 44. (d) During minute 48, the patient briefly awoke and changed his posture. (e) From about minute 50 to approximately minute 75, the patient was sleeping with only short periods of movements. (f) During minute 77, the patient was awake and moving. (g) The patient went back to sleep from minute 80 to approximately 86. (h) Two snapshots from the periods 86 to 99, during which the patient was actively moving, including having some tea. (i) The patient sleeping around minute 99.7.

The reasons for the camera not providing a reliable heart rate estimate are largely due to patient motion and the interaction with the clinical staff. Figure 6.65 shows the comparison between the reference heart rate estimates (computed from the agreement between the two pulse oximeters) and the camera estimates for a 60-minute segment from a typical dialysis session. The camera estimates are in good agreement with the reference values for most of the time periods during which there is minimal patient motion (see figure 6.65(c), 6.65(e), 6.65(g) and 6.65(i)). For these stable periods, the colour-coded heart rate frequency maps with red values represent areas with strong cardiac frequency from which heart rate was estimated. Periods during which the patient is actively moving, and therefore heart rate cannot be reliably estimated, are shown in figure 6.65(d), 6.65(f) and 6.65(h); in the latter, the patient is seen having some tea.

6.11 Conclusion

The accuracy of heart rate estimation from the video camera, in periods during which the subject is stable, is comparable to that of the two standard pulse oximeters used in this study, one attached to the ear lobe and the other to the patient's finger. The camera estimates using the best method (based on a multi-channel Kalman filter with an autoregressive model and pole cancellation) show a strong positive correlation and good agreement compared to the reference heart rate, with minimal bias for approximately 65% of the total recording time.

As opposed to current published literature, where a single colour channel is used to estimate heart rate (usually the green channel as described in chapter 4), our proposed algorithms analyse all three colour channels from multiple regions of interest and using different data fusion methods compute the overall heart rate for the patient.

The proposed approach of heart rate physiological maps from multiple regions of interest has several advantages compared to other methods where the face is analysed as a whole (see chapter 4). The detection of subjects in a video image can be performed, simply from the fact that their heart rate is a periodic signal derived from skin regions (shown in figure 6.63). This could be used to replace the face detection and tracking algorithms discussed and implemented as described in earlier sections of this chapter, or to augment the algorithms with independent information based purely on physiology. Analysing multiple regions of interests can also lead to more localised physiological studies, such as measuring the propagation of pulsatile arterial blood across the body (see chapter 3) and the potential to measure other vital signals such as pulse transit time (PTT) and blood pressure [246].

As most of the video recordings from the clinical data set were recorded in similar conditions (patients' upper torso at a similar distance from the camera), a common threshold for the activity index described in section 6.4 was computed. Further analysis will be needed if the video recording scenario changed, in which case an adaptive method will be required.

Chapter 7

Reference respiratory rate

7.1 Introduction

Respiratory rate is one of the core vital signs (with heart rate, blood pressure and temperature) that are recommended to be recorded at least daily as a standard measure for monitoring patients on acute hospital wards [114]. An abnormal respiratory rate has been shown to be an important predictor of serious events such as cardiac arrest or admission to an intensive care unit (ICU) [76]. However, in recent studies, it has been found that even though when the patient's primary problem is of a respiratory nature, respiratory rate in particular is often not recorded [126, 129, 211].

Several technologies have been developed that enable the monitoring of respiratory rate. During anaesthesia, capnography offers a reliable means of measuring pulmonary perfusion and respiratory rate through an optical sensor that measures the amount of carbon dioxide (CO_2) in the exhaled air present at each phase of the respiratory cycle. [267, 355]. Capnography is also a recommended technique for monitoring of respiratory depression during mechanical ventilation [45, 344].

Impedance pneumography (IP) is a commonly used technique in the intensive care unit (ICU), cardiac/coronary care unit (CCU) or a high dependency unit (HDU) because it can be measured simultaneously with the electrocardiogram (ECG). IP measures changes in the electrical impedance of the patient's thorax caused by respiration by injecting a high-frequency current through the ECG electrodes [249].

On the general wards of a hospital, a manual procedure is invariably used to record respiratory rate by counting chest wall movements during the regular nurse observations. It is recommended that the number of chest wall movements should be counted over a minute or two 30-second intervals, but this represents a significant investment in nursing time on the ward. As a result, accurate rates may only be recorded for as little as 37% of the time [30, 129]. Another technique is to use a thoracic expansion band that is worn around the chest. The elastic band measures the volume changes on the chest wall as it expands or contracts during the respiratory cycle.

This chapter begins by analysing the validity of the respiratory rate estimates provided by the manufacturer's equipment in section 7.2, it postulates that new reference respiratory rate values need to be computed as a gold standard measure to be compared with the camera estimates.

The overall methodology to compute the new gold standard reference respiratory rates is described in section 7.3. The extraction of a respiration signal using the ECG is covered in section 7.4. The estimation of respiratory rate using the ECG and chest belt is proposed in section 7.5 and results presented in section 7.6. Conclusions and recommendations are finally discussed in section 7.7.

7.2 Respiratory rate provided by Equivital LifeMonitor

As described in section 5.4, Hidalgo's Equivital LifeMonitor (see figure 5.3) is a wearable ambulatory monitoring system for collecting multiple physiological parameters. It provides, amongst other signals, two simultaneous respiratory rate estimates, the first one computed from the thoracic chest belt, and the second using the ECG.

7.2.1 The reference source signals

Chest belt

The chest belt from Hidalgo's LifeMonitor, shown in figure 7.1(a), is a harness that the patient wears on the upper torso. It contains an elastic thoracic expansion band that measures the patient's chest wall expansion and contraction as the patient breathes in and out. From the analysis of these movements, it derives a respiration signal at a sampling rate of 25.6 Hz (see figure 7.1(b)). Respiratory rate estimates are computed every 5 seconds from the respiration signal, as shown in figure 7.1(c).

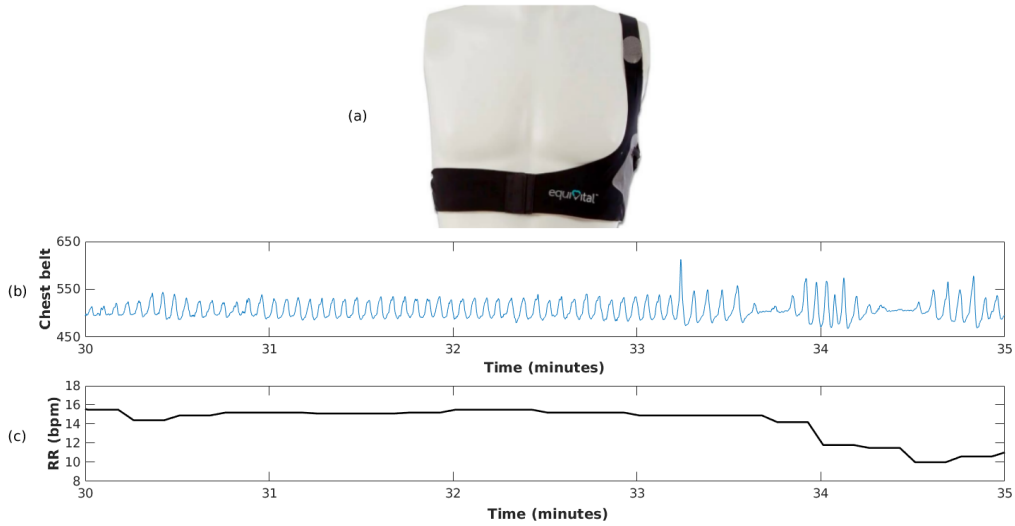


Figure 7.1: Respiratory rate estimates from Hidalgo's chest belt. (a) The Equivital EQ02 LifeMonitor chest belt/harness, (b) 5-minute sample respiration signal from the thoracic expansion belt. (c) The computed respiratory rate estimates.

The chest belt features a system of sensors embedded in a wearable textile fabric manufactured in a limited number of sizes to be able to cover most patients' body sizes. To record high-quality data, the belt has to be tightly fitted around the patient's torso, positioned in line with the bottom of the pectoral muscles. A shoulder strap provides additional support and improves mobility [191].

The electrocardiogram (ECG)

The electrocardiogram (ECG) is a recording of the electrical activity on the skin surface produced by the heart muscle depolarising as it goes through each cardiac cycle. In clinical diagnosis, it is common to record 12 channels (referred as leads) of ECG with 10 electrodes placed at standard locations on the thorax. The morphology of the ECG varies according to electrode placement; nevertheless common features labelled as PQRST by Einthoven [86] are always present in normal ECG and are shown in figure 7.2. More details of the underlying phenomena that give rise to these features and a description of a standard 12-lead ECG can be found in Clifford [64].

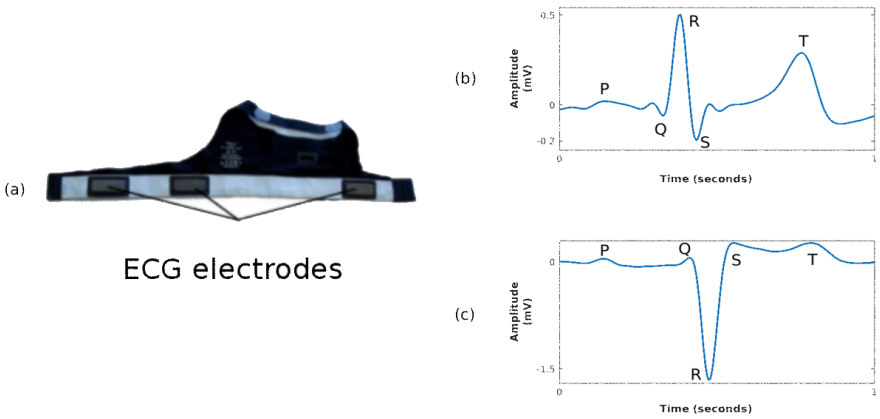


Figure 7.2: ECG recording by Hidalgo's LifeMonitor during a single cardiac cycle from the electrodes place at two locations and the PQRST features labelled accordingly. (a) The Equivital EQ02 LifeMonitor chest belt/harness. (b) The ECG output during a single cardiac cycle from the electrodes placed on the left side of chest (near the heart), resembling the "V6" lead from the 12-lead ECG standard. (c) The ECG output from the electrodes placed away from the chest, producing an inverted trace resembling the "V1" ECG lead.

With the ECG electrodes embedded inside the chest belt as shown in figure 7.2(a), Hidalgo's LifeMonitor records two channels of ECG (at a sampling rate of 256 Hz). Figure 7.2(b) shows the output during a single cardiac cycle from the electrodes placed on the left side of the chest (near the heart), resembling the standard "V6" lead from the 12-lead ECG; figure 7.2(c) presents the recording from the electrodes placed away from the heart on the right side of the chest, resembling the standard "V1" lead. The PQRST features are labelled accordingly.

The ECG electrodes are housed inside the same fabric as the thoracic expansion band, from where the respiration signal is recorded. Therefore, the quality of the ECG waveform also relies on the proper fit of the chest harness.

Another major factor that can distort the recordings is patient activity. Two five-minute recordings from one of the ECG channels are shown in figure 7.3. During a quiet period where the patient was sleeping, the ECG signal (see figure 7.3(a)) is clean and the respiratory estimates (see figure 7.3(b)) computed from the Hidalgo kit are within physiological range. In contrast, when the patient is active and interacting with the clinical staff, the ECG signal (see figure 7.3(c)) is noisier; its amplitude values are clipped and the respiratory rate estimates produced by the manufacturer (see figure 7.3(d)) are outside the expected physiological range.

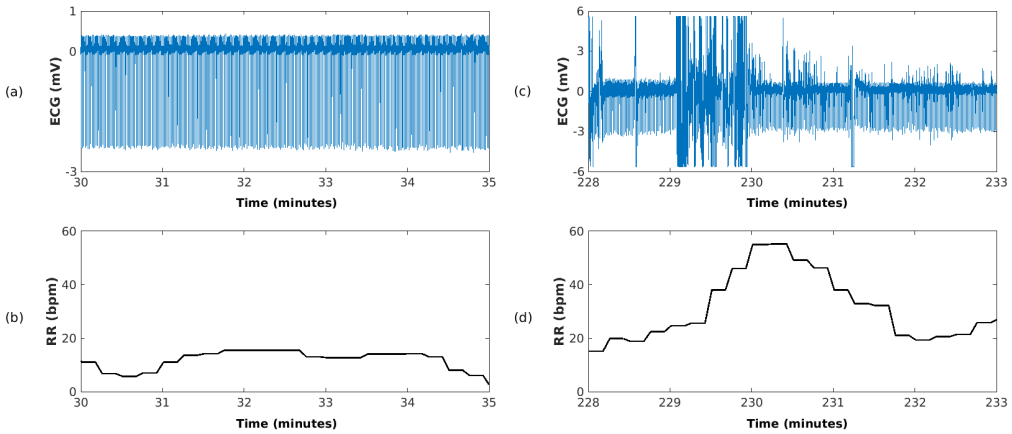


Figure 7.3: Respiratory rate estimates from electrocardiogram (ECG). (a) 5-minute sample ECG waveform from a period during which the patient was quietly sleeping, (b) the corresponding respiratory rate estimates. (c) 5-minute recording from a period during which the patient was active, (d) the reported respiratory rate estimates.

7.2.2 Reference respiratory rate comparison

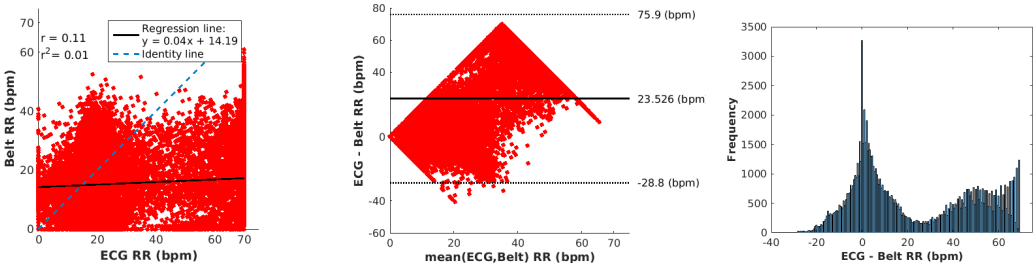
One of the primary goals of this thesis is to be able to compare the respiratory rate estimates from the camera to reference values provided by medical-grade equipment. Hence, the validity of the two respiratory rate estimates provided by the Hidalgo LifeMonitor needs to be established. This chapter focuses on the same subset of data as was used for heart rate estimation (see chapter 6). The total dataset consists of 84 video sessions, comprising a total recording length of 304.1 hours.

Comparing the overall dataset

Figure 7.4 presents an assessment of the overall dataset, including all the values provided by the manufacturer, without any processing applied to the source signals (ECG and chest belt). From the correlation plot in figure 7.4(a), a correlation coefficient of 0.11 shows that no positive linear association exists between the two respiratory rate estimates; the recording limits for respiratory rate established by the manufacturer as 0 to 70 breaths per minute [124] can also be noted.

The Bland-Altman plot, shown in figure 7.4(b), presents a large bias of more than 23 breaths per minute indicating that the two respiratory rate estimates are in disagreement; the respiratory rate limits imposed by the manufacturer can again be clearly seen. Figure 7.4(c) plots the distribution of the differences between the estimates from the two source signals, showing that the difference is not normally distributed.

When comparing the respiratory rate estimates provided by the manufacturer using both methods, the mean absolute error is 26.4 breaths per minute ($\mathrm{breaths/min}$) with a mean absolute deviation of 23.9 $\mathrm{breaths/min}$. Most of the errors in figure 7.4 originate from extreme values outside a patient's physiological range; this large discrepancy can be partly explained by the regular activity of the patient during the haemodialysis session (more details in section 6.4) or can be related to inaccuracies in the recording equipment.



(a) (b) (c)

Figure 7.4: The agreement between the two respiratory rate estimates provided by the Hidalgo LifeMonitor, computed from the ECG and chest belt for the overall data set (304.1 hours). (a) The correlation plot shows no correlation between the estimates, with a correlation coefficient of 0.11. (b) The Bland-Altman plot presents a positive bias of 23.5 $\mathrm{breaths/min}$. (c) The distribution of differences between the two source signals is not normally distributed.

Figure 7.5 shows the two respiratory rates estimates for a typical 4-hour dialysis recording; around minute 100, the patient was speaking on the phone; and around minute 176 (zoomed-in boxes), the patient was actively interacting with the clinical staff. At around minute 230, the nurse proceeded to gradually disconnect all the equipment as the dialysis session ended.

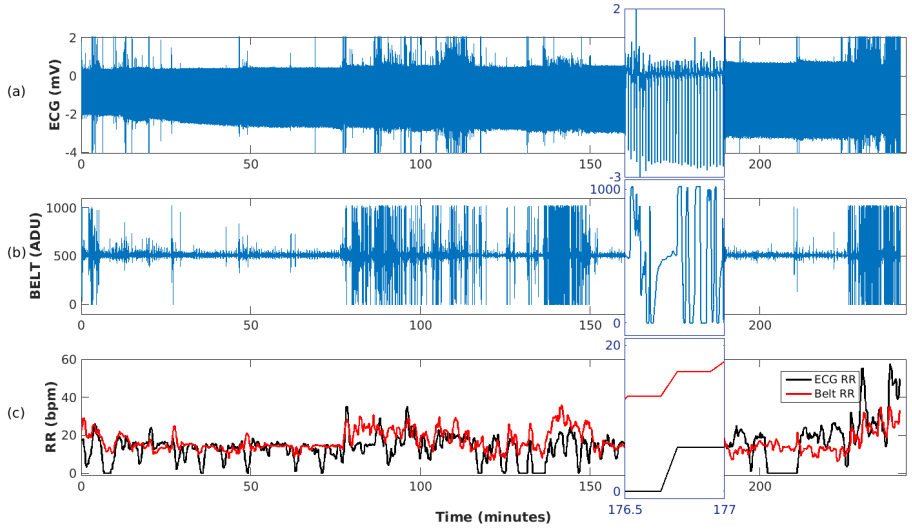
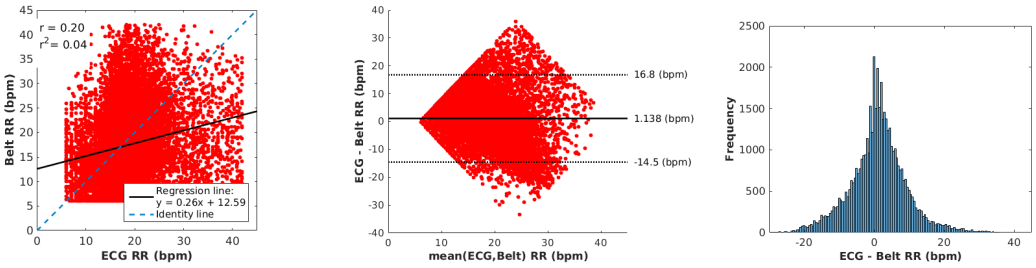


Figure 7.5: Comparison of the two respiratory rate estimates provided by Hidalgo's LifeMonitor for a typical dialysis session. (a) The ECG waveform, (b) the respiration signal from the chest belt, (c) the respiratory rate estimates from the ECG in black and chest belt in red. The zoomed-in boxes around minute 176 represent a large discrepancy between the two measurements at a time when the patient was actively interacting with the clinical staff.

Limiting to a physiological range

When limiting the respiratory rates estimates to a physiological range of 6 to 42 breaths per minute, the mean absolute error decreases to 6.0 $\mathrm{breaths/min}$ with a mean absolute deviation of 5.4 $\mathrm{breaths/min}$. Although the error decreases, the correlation coefficient is still low at 0.20, showing no significant correlation (see figure 7.6(a)).



(a) (b) (c)

Figure 7.6: The agreement between the respiratory rate estimates provided by Hidalgo's LifeMonitor using the ECG and chest belt waveforms for the overall data set (336.4 hours) when the estimated values are limited to the physiological range of 6 - 42 breaths/minute. (a) The correlation plot shows no correlation between the estimates, with a correlation coefficient of 0.20. (b) The Bland-Altman plot presents a positive bias of 1.14 $\mathrm{breaths/min}$. (c) The distribution of differences between the estimated values and the reference.

The Bland-Altman plot in figure 7.6(b) presents a positive bias of the estimates from the chest belt compared to the estimates computed from the ECG and a large spread indicating lack of agreement; the new physiological limits can also be clearly seen in the diamond-shaped plot. The distribution of the differences is now normally distributed (see figure 7.6(c)), as most of the previous errors in figure 7.4(c) were outside the physiological range (6 to 42 breaths per minute).

Table 7.1 summarises the error for the two respiratory rate estimates provided by the Hidalgo LifeMonitor. As already seen, the error is large when all values are taken into account. When restricting the values to a physiological range from 6 to 42 breaths per minute (corresponding to a frequency range from 0.1 Hz to 0.7 Hz), the error decreases but the estimates are only available for 31% of the time (approximately 94.3 hours).

Table 7.1: Mean absolute error and mean absolute deviation for the respiratory rate estimates provided by the Hidalgo LifeMonitor computed from the chest belt and the ECG for the overall dataset under analysis (336.4 hours). The valid physiological range for respiratory rate is taken to be from 6 to 42 $\mathrm{breaths/min}$.

Range of values compared	MAE ($\mathrm{breaths/min}$)	MAD ($\mathrm{breaths/min}$)	Time hours (%)
All values	26.4	23.9	304.1 (100%)
Valid physiological range	6.0	5.4	94.3 (31%)

7.2.3 Conclusion

Given the analysis presented in this section, the respiratory rate estimates provided by the Hidalgo LifeMonitor are not suitable as reference values to be compared with the respiratory rate estimates computed from the video camera. Therefore, new gold standard reference values need to be computed using state-of-the-art methods for analysing and fusing the two source signals: the ECG and the chest belt.

7.3 Overview of the estimation process

The overall methodology for the estimation of the reference respiratory rate is shown in figure 7.7. From the two main source signals, the ECG and the chest belt, the latter already records a respiration signal from which a respiratory rate can be computed. Therefore, the process starts with the extraction of a respiration signal from the ECG. Well known algorithms, commonly used in the research community and validated against publicly available physiological databases, are used to detect and assess the quality of each ECG beat; subsequently, a respiration signal is extracted using the ECG-derived respiration (EDR) algorithm.

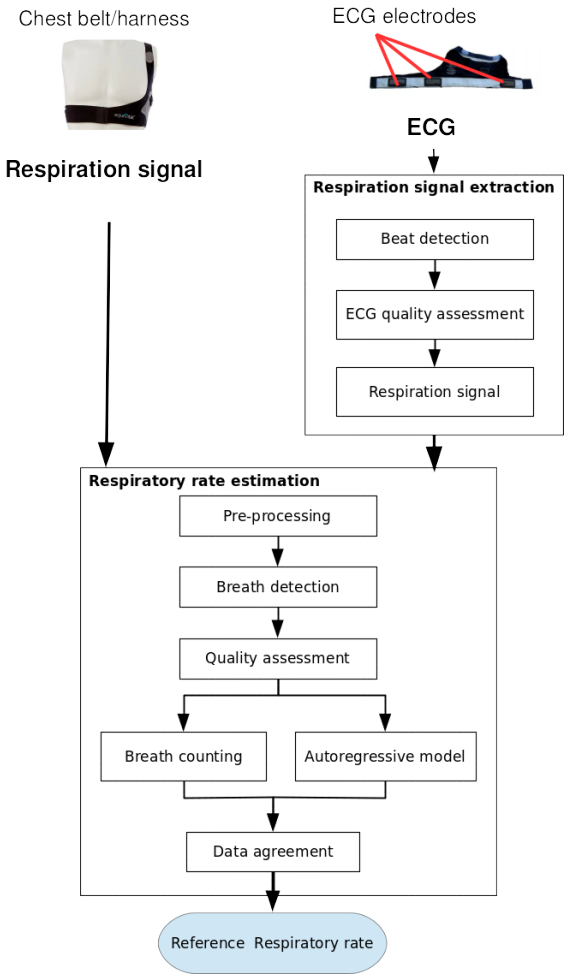


Figure 7.7: Overview of the respiratory rate estimation process.

The two respiration signals (from the ECG and chest belt) are pre-processed with algorithms to detect breath peaks and assess their quality. Respiratory rate is estimated using two methods: a time-domain technique by simply counting the number of breaths over a time period, and a frequency-domain technique based on autoregressive models. The proposed respiratory rate estimation methods are applied to sliding windows of 30 seconds with a separation between consecutive windows of 5 seconds. The final respiratory rate is computed as the median from the estimates for which the data is of good quality and do not vary by more than 2 $\mathrm{breaths/min}$.

7.4 Respiration signal extraction from ECG

7.4.1 Overview

The process of extracting a respiration signal from the ECG starts by simultaneously running two beat detectors, known to be sensitive to different types of noise [102]. Subsequently, in a similar manner as [186], a signal quality index is computed using different criteria for every ECG beat to prevent computing respiratory rate for periods when the ECG is noisy. Finally, the respiration signal is extracted from the two ECG channels by applying the ECG-derived respiration (EDR) algorithm.

7.4.2 Beat detection

Digital filter based

This first QRS detector is an adaptation of the algorithms described by Behar et al [29], Johnson et al [150] and Pan and Tompkins [252]. The ECG waveform is affected by various sources of noise, including muscle movement, power line interference (50 Hz in Europe), and baseline wander. Additionally it is known that most of the QRS energy is concentrated around the 5 Hz to 15 Hz band [64, 112, 325]. Therefore, the ECG waveform is initially run through a cascade of low-pass and high-pass filters with cut-off frequencies of 15 Hz and 5 Hz respectively. This process can be seen in figure 7.8 for a 5-second ECG waveform.

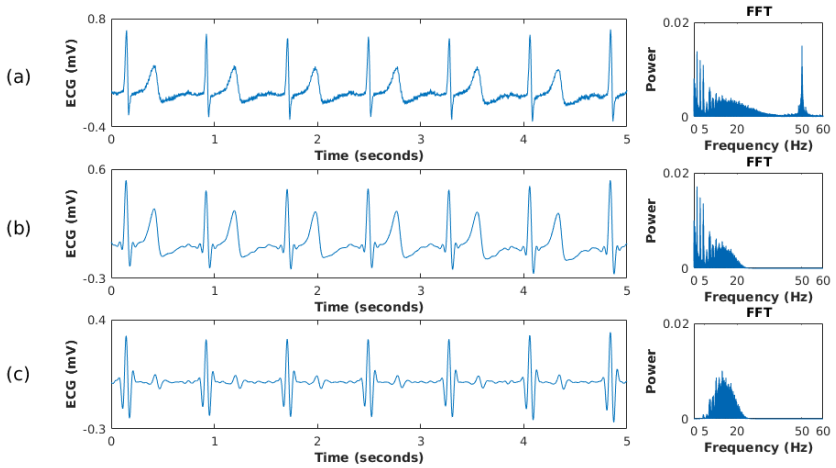


Figure 7.8: Digital filtering of the ECG waveform to enhance the QRS complex energy using a cascade of low-pass and high-pass filters with cut-off frequencies of 15 Hz and 5 Hz respectively. (a) Input ECG waveform, (b) low-pass filter, (c) high-pass filter

After band-pass filtering, four pre-processing steps are performed: derivative, squaring, integration and smoothing. The derivative of the ECG signal is computed to enhance the R peak slope. If $X = x_1, x_2, ..., x_n$ represents the filtered signal, the derivative $y(i)$ is computed as a five-point central difference according to [105] and defined by the equation:

$$y(i) = \frac{x(i-2) - 8x(i-1) + 8x(i+1) - x(i+2)}{12} \tag{7.1}$$

The time series is then squared to make all data points positive and emphasise the QRS complex. This is given by the equation:

$$S(i) = y(i)^2 \tag{7.2}$$

To improve the detection of other ECG features such as the Q and S onsets, moving-window integration is applied to the data. Similarly to [150], the length of the window is taken as the maximum QRS period of 150 ms, corresponding to 38 samples for the sampling rate of 256 Hz of the Hidalgo LifeMonitor. A 3-point median filter is subsequently applied to smooth the ECG data and reduce some artefacts. The pre-processing steps can be seen in figure 7.9 for a 15-second ECG waveform recorded while the patient was quietly sleeping.

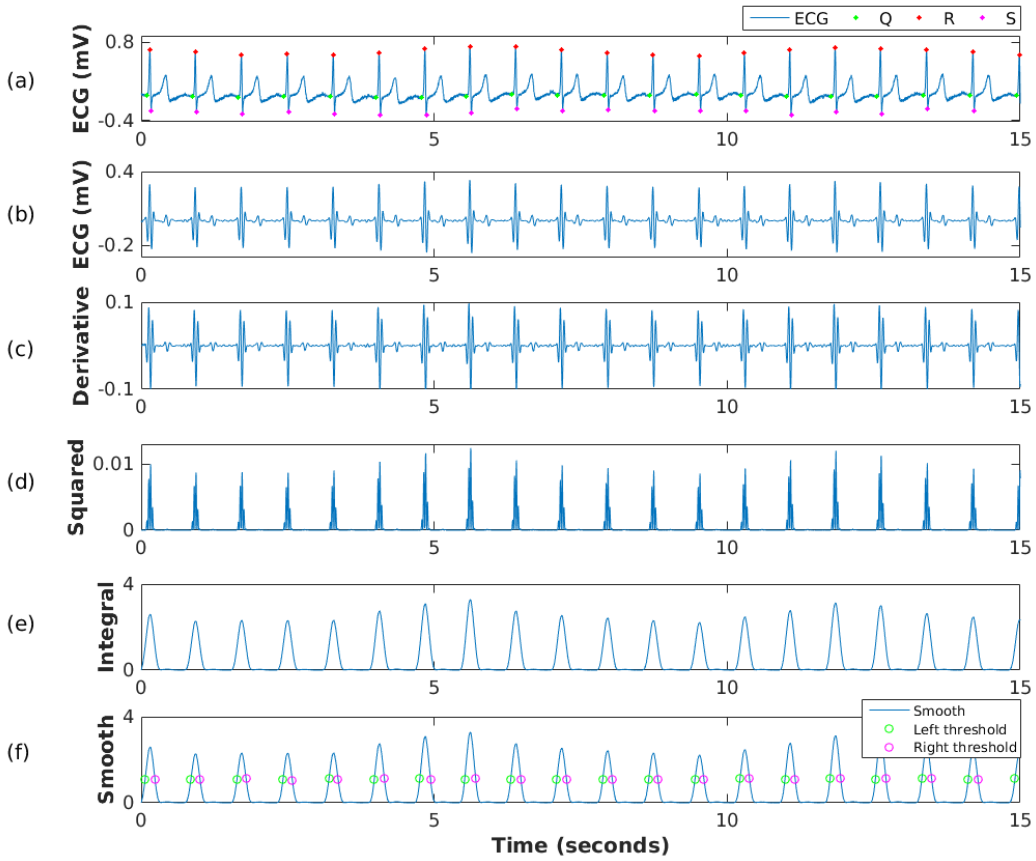


Figure 7.9: Processing steps for the QRS detector based on digital filtering of a 15-second ECG waveform. (a) Input ECG waveform with the Q,R and S peaks detected; (b) band-pass filtered, (c) derivative, (d) squared, (e) integration, (f) smoothing, showing the left and right thresholds for the QRS complexes for each potential beat.

After the ECG waveform has been pre-processed, R peaks are found by computing local maxima for values above two amplitude thresholds. A first threshold T_1 is computed as three times the mean of the smoothed ECG during a learning period w of 120 seconds at the start of each recording, given by the equation:

$$T_1 = 3 * \frac{1}{w} \sum_{i=1}^{w} ECG_{smooth}(i) \tag{7.3}$$

A list of all the ECG values greater than the threshold T_1 are identified. If there are gaps longer than two times the median gap length, a new threshold T_2 is computed as $T_2 = T_1/2$ and the new values greater than T_2 are added to the list. From all the values that are greater than any of the two thresholds, two indices are found to the left and right of each continuous segment corresponding to each potential ECG beat, as shown in figure 7.9(f). The R peaks are found by computing the local maxima from each potential beat (or minima for inverted ECG leads). To reduce the possibility of multiple contiguous beat detections, a refractory period of 250 ms (corresponding to 64 samples) is introduced after a beat is detected. The Q and S peaks are computed by finding the local minima (or maxima for inverted ECG leads) to the left and right of the R peak with a window of length 150 ms (corresponding to 38 samples) centred on the R peak. Figure 7.9(a) shows the input source ECG waveform and the subsequent steps for QRS feature detection.

Length transform based

The second QRS detector is an adaptation of an algorithm by Zong et al [375]. It consists of three stages: First, the ECG signal is digitally filtered, a curve length transform is then applied, and finally decision rules are applied to detect the QRS peaks.

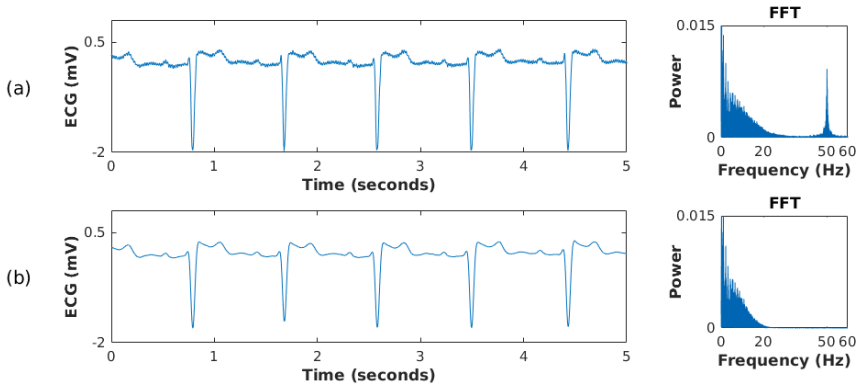


Figure 7.10: Digital filtering of the ECG waveform to enhance the QRS complex with a low-pass filter with a cut-off frequency of 15 Hz. (a) Input ECG waveform; (b) low-pass filter.

The digital filter stage starts by applying a low-pass filter with a cut-off frequency of 15 Hz to the input ECG waveform to reduce the effects of high-frequency components and enhance the QRS complex frequency information. In contrast to the previous detector, a high-pass filter is not applied because the length transform described below substantially suppresses low-frequency components. The digital filtering process can be seen in figure 7.10 for a 5-second ECG waveform.

After low-pass filtering the ECG signal, a curve length transform is computed over a window with a length equal to the maximum QRS period of 150 ms, corresponding to 38 samples. If $X = x_1, x_2, ..., x_n$ represents the filtered signal, the curve length transform L is defined by the equation:

$$L(w,i) = \sum_{k=i-w}^{i} \sqrt{C + (\Delta x(k))^2} \tag{7.4}$$

where $\Delta x(k) = x(k) - x(k-1)$, C is the sample interval and w is the length of the window. A 3-point median filter is subsequently applied to reduce some artefacts in the ECG. The pre-processing steps can be seen in figure 7.11 for a 15-second ECG waveform recorded while the patient was quietly sleeping.

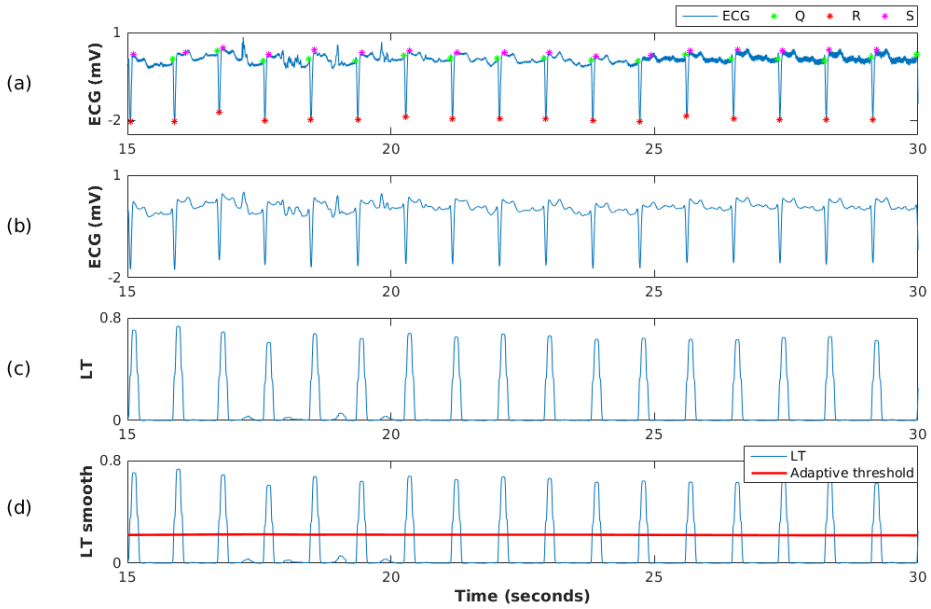


Figure 7.11: Processing steps for the QRS detector based on the length transform method for 15 seconds of an ECG waveform (a) Input ECG waveform with the Q,R and S peak detected; (b) low-pass filtered; (c) length transform (LT) signal; (d) smoothed LT waveform in blue with the adaptive amplitude threshold in red.

The decision rules to detect the QRS peaks consist of two steps: Defining a threshold for the length transform signal to find potential crossing points, and searching locally to detect the ECG peaks. The threshold is initialised as the mean value over a 120-second learning period and then later is dynamically adjusted based on the amplitude values of the length transform signal, as seen in figure 7.11(d).

When a sample of the length transform signal crosses the set threshold, the algorithm searches backwards and forwards for a minimum L_{min} and a maximum L_{max} respectively over a window of 125 ms. From these values, the R peak is found by computing the local maximum (or minimum for inverted ECG leads) in the original source ECG signal between the $[L_{min}, L_{max}]$ sample location. The Q and S peaks are computed by finding the local minima (or maxima for inverted ECG leads) around the location of the L_{min} and L_{max} samples. To reduce the possibility of multiple contiguous beat detections, a refractory period of 250 ms (corresponding to 64 samples) is introduced after a beat is detected. Figure 7.11(a) shows the input source ECG waveform and the subsequent steps for QRS feature detection.

7.4.3 Beat quality assessment

The quality of the ECG is computed in a similar manner as [186]. Four metrics are used: comparison of multiple beat detection algorithms on each of the two ECG channels (bSQI), comparison of the same beat detection algorithm on all of the leads (iSQI), evaluation of the kurtosis (kSQI), and computing the spectral distribution on each ECG lead (sSQI). All the quality metrics are applied to a 10-second sliding window centred on each beat (as computed by the "Digital filter based" beat detector). The four quality metrics are later combined to compute the final quality index of each beat from each of the ECG channels using a heuristic criterion similar to [150, 186].

Multiple beat detection on a single lead (bSQI)

Friesen et al [102] compared several beat detectors and showed that different algorithms are sensitive to different types of noise. The two beat detectors described in the previous sections achieved high sensitivity and positive predictivity on publicly available databases such as the MIT/BIH arrhythmia database [117, 375]; however the positive predictivity of the length transform algorithm decreases when the signal-to-noise ratio is low [186].

Oefinger [247] proposed that the quality of an ECG signal can be measured by the ratio of the number of beats detected synchronously (within a time interval) by two beat detectors to all the beats detected by any of the algorithms over a time window. The quality index for the kth beat over a 10-second time window w is defined as:

$$bSQI(k) = \frac{N_{matched}(k,w)}{N_{all}(k,w)} \tag{7.5}$$

where $N_{matched}$ is the number of beats that both algorithms agree for the window w, N_{all} is the total number of beats detected by both algorithms over the same window. Two beat detections are in agreement if they are within 150 ms as recommended by the Association for the Advancement of Medical Instrumentation (AAMI) [14].

Beat detection on a multiple leads (iSQI)

When running multiple beat detectors on the same ECG lead, bSQI can be low not only due to the inherent noise in the signal, but also due to errors in the algorithm, failing to detect a correct beat, or registering extra beats. There are several causes for these errors, including periods when the beat detector based on the length transform incorrectly marks a high amplitude T wave as an R peak, whereas the other beat detector correctly finds the R peak (see figure 7.12), decreasing the overall bSQI value for the sliding windows.

By running the digital filter based beat detector over the two available ECG leads, another measure of quality can be obtained as the ratio of the number of beats detected synchronously in both ECG channels to all the beats detected on any of the channels. Thus, the quality index for the kth beat over a 10-second time window w is defined as:

$$iSQI(k) = \frac{N_{matched}(k,w)}{N_{all}(k,w)} \tag{7.6}$$

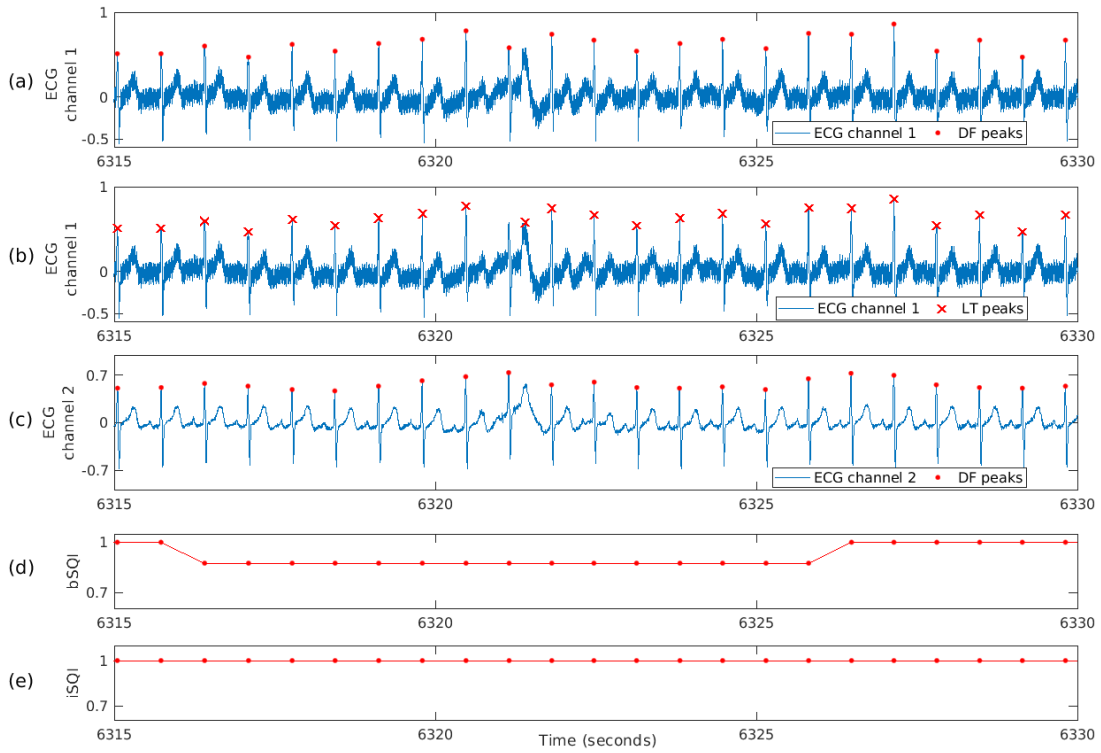


Figure 7.12: Comparison of the bSQI and iSQI metrics for 15 seconds of the two-channel ECG recorded by the Hidalgo LifeMonitor. (a) Digital filtering beat detector (DF) and (b) length transform beat detector (LT) applied to the first ECG channel. (c) digital filtering beat detector applied to the second ECG channel. (d) Due to a high amplitude T wave, the LT detector incorrectly marks the R peak, decreasing the overall bSQI value. (e) The multi channel metric iSQI correctly marks the ECG as a good-quality segment.

where $N_{matched}$ is the number of beats that the algorithm agrees on both leads for the window w, N_{all} is the total number of beats detected by the algorithm on the two ECG leads over the same window. Similarly to bSQI, two beat detections are in agreement if they are within 150 ms. Figure 7.12 shows the advantages of the multi-channel iSQI metric over the bSQI metric for a short ECG segment.

Kurtosis (kSQI)

From the central limit theorem, the distribution of a large number of independent, uncorrelated and identically distributed random variables will be approximately Gaussian [270]. On the other hand, correlated signals such as the ECG tend to have non-Gaussian distributions. There are several methods to verify if a signal is Gaussian distributed, a simple measure being kurtosis. Kurtosis is the degree of peakedness of a distribution and is defined as a normalised form of the fourth central moment of a distribution. For a discrete time series $x = x_1, x_2, ..., x_n$ of length N, the kurtosis can be estimated as:

$$\hat{K} = \frac{1}{N} \frac{\sum_{i=1}^{N} (x(i) - \mu)^4}{\sigma^4} \tag{7.7}$$

where μ is the mean and σ is the standard deviation of the time series. Gaussian distributions have a kurtosis of 3, whereas good-quality ECG has a kurtosis greater than 5 [122]. Other sources of noise in the ECG such as muscle artefacts, baseline wander and power-line interference have a kurtosis lower than 5 [64], as can be seen for the 10-second ECG segment shown in figure 7.13. Thus, the quality index for the kth beat over a 10-second time window w is defined as:

$$kSQI(k) = \begin{cases} 1 & \text{, if } \hat{K}(w) > 5 \\ 0 & \text{, } otherwise \end{cases} \tag{7.8}$$

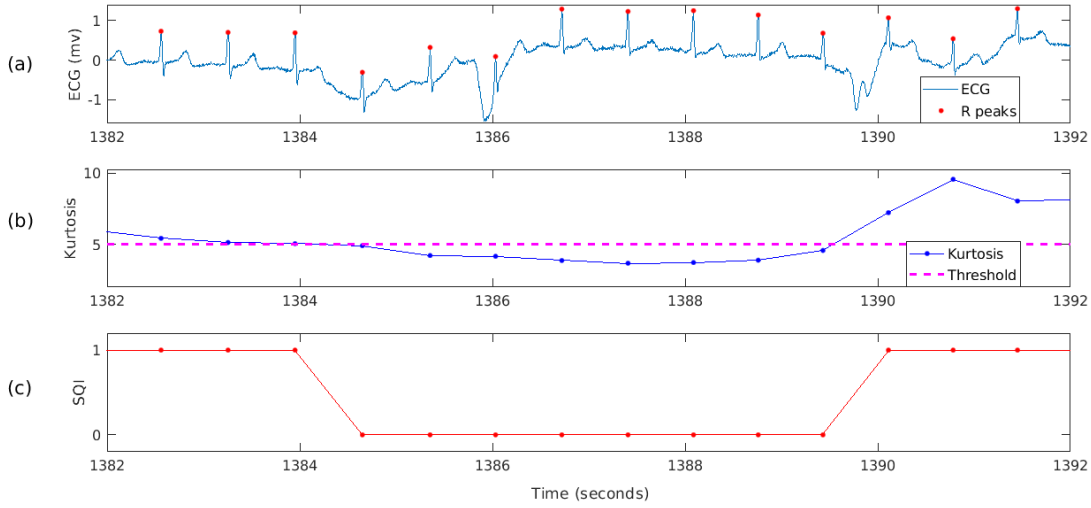


Figure 7.13: The kurtosis quality metric (kSQI) applied to a 10-second ECG signal. (a) The ECG waveform with the R peaks labelled, (b) kurtosis for each 10-second sliding window centred on each beat. (e) The computed kSQI values correctly mark the ECG as a poor quality segment.

Spectral distribution (sSQI)

According to Murphy [235] and Zong [375], the QRS energy is mainly concentrated in the 5 Hz to 15 Hz frequency range. The ratio of the power spectral density (PSD) in this range, when compared to the PSD for the overall signal, provides another measure of quality. Thus, the spectral distribution ratio (SDR) for the kth beat over a 10-second time window w is defined as:

$$SDR(k) = \frac{\int_{f=5}^{f=15} P(k,w)df}{\int_{f=5}^{f=40} P(k,w)df} \tag{7.9}$$

where $P$is the sum of the power over the specified frequency range. Low SDR values indicate muscle artefact, whereas high SDR values can be caused by electrode motion as can be seen in figure 7.14. Therefore, the quality index for the kth beat over a 10-second time window w is defined as:

$$sSQI(k) = \begin{cases} 1 & \text{, if } SDR \geq 0.5 \text{ and } SDR \leq 0.8 \\ 0 & \text{, } otherwise \end{cases} \tag{7.10}$$

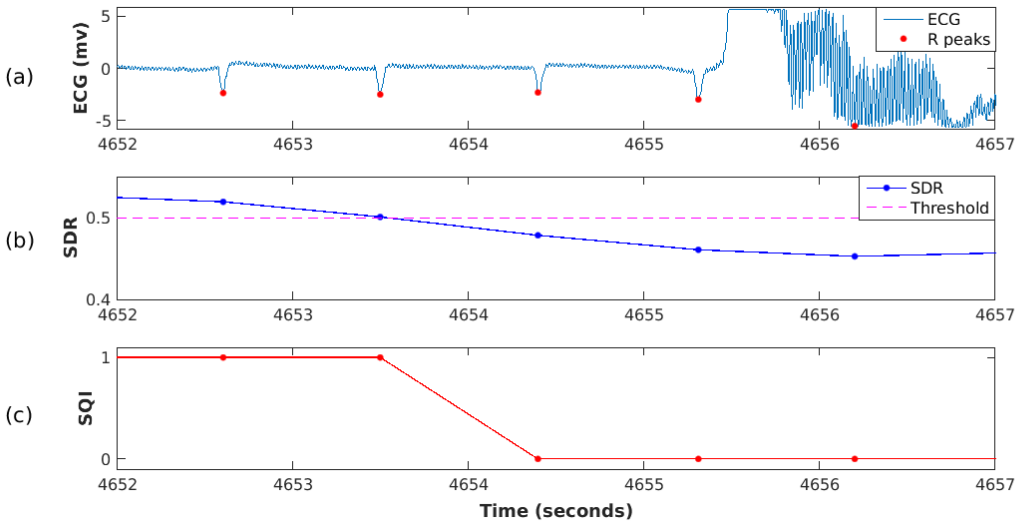


Figure 7.14: Quality metric based on spectral distribution (sSQI) applied to a 5-second ECG signal. (a) The ECG waveform with the R peaks labelled, (b) spectral distribution ratio (SDR) for each 10-second running windows centred on each beat. (e) The computed sSQI values correctly mark the ECG as a poor quality segment.

Combining the SQI metrics

Similar criteria to [375] are used to combine the four quality metrics described above into one measure given by the following equation:

$$ECGSQI(k) = \begin{cases} min(bSQI(k), iSQI(k)) & \text{, if } kSQI(k) = 1 \text{ and } sSQI(k) = 1 \\ bSQI(k) & \text{, if } kSQI(k) = 1 \text{ and } sSQI(k) = 0 \\ min(bSQI(k), iSQI(k)) * n & \text{, if } kSQI(k) = 0 \text{ and } sSQI(k) = 1 \\ bSQI(k) * n & \text{, if } kSQI(k) = 0 \text{ and } sSQI(k) = 0 \end{cases} \tag{7.11}$$

where $n = 0.7$ is a coefficient that adjusts the final SQI in the presence of statistical noise. When both kSQI and sSQI indicate the ECG segment to be of good quality, the minimum of bSQI and iSQI is used as a quality measure; however when sSQI is low, only bSQI can be used since noise outside the QRS frequency range is often present in all channels. Similar criteria are used when kSQI varies. Figure 7.15 shows the four quality metrics computed for a 60-second ECG signal, with the combined quality measure correctly identifying periods for which the ECG is of poor quality.

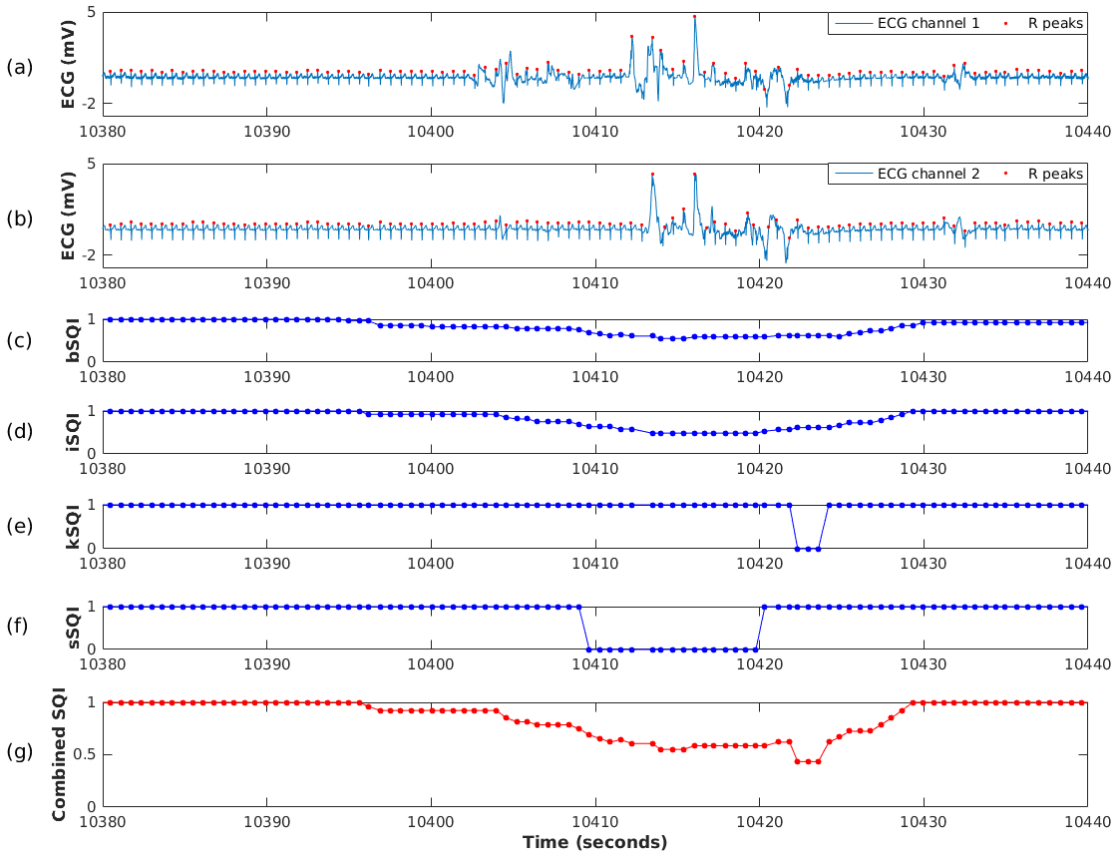


Figure 7.15: Signal quality assessment for 60 seconds of a two-channel ECG record: (a) First input ECG waveform channel with the detected R peaks plotted in red, (b) second input ECG waveform channel with the detected R peaks plotted in red. The individual beat quality measures: (c) bSQI, (d) iSQI, (e) kSQI, (f) sSQI. (g) The combined SQI.

7.4.4 Respiration signal extraction

After the quality of the each ECG lead has been computed, a respiration signal is extracted using three well known types of algorithms: Respiratory sinus arrhythmia (RSA), R-Peak amplitude (RPA) and ECG-derived respiration (EDR). Although the three types of algorithms are reported in the literature, EDR is shown in the next sections to be more appropriate due to the demographics of the patients undergoing haemodialysis treatment. All the algorithms process each beat of the ECG signals and assign the computed value to the index of the respective R peak. As such, they produce unevenly sampled time series. Therefore, these time series are linearly re-sampled to 4 Hz to analyse their frequency content in the later stages.

Respiratory sinus arrhythmia (RSA)

Respiratory sinus arrhythmia (RSA) is a modulation of the heart rate over the course of a respiratory cycle. The R-R interval decreases and increases during inspiration and expiration respectively. If the time series $t = t_1, t_2, ..., t_n$ represents the time vector for an ECG signal of length N, we define a sample RSA index for the kth beat using the following equation:

$$RSA(k) = t_k - t_{k-1} \tag{7.12}$$

Some studies have shown that RSA may play an active physiological role, improving the efficiency of pulmonary gas exchange by matching the timing of alveolar ventilation and its perfusion, thereby suppressing unnecessary heartbeats during expiration [366]. RSA depends on the regulation of the autonomic nervous system. It is most prominent in infants and is attenuated as humans age [123, 132]. RSA can also be affected by diabetes and cardiovascular diseases [133, 338]. These factors can potentially explain the poor quality of the RSA signal in our study, since the population of patients undergoing haemodialysis treatment is mainly comprised of the elderly and the unwell. Figure 7.16(c) and 7.17(c) show the RSA signals extracted for two 10-minutes ECG signals from two different patients.

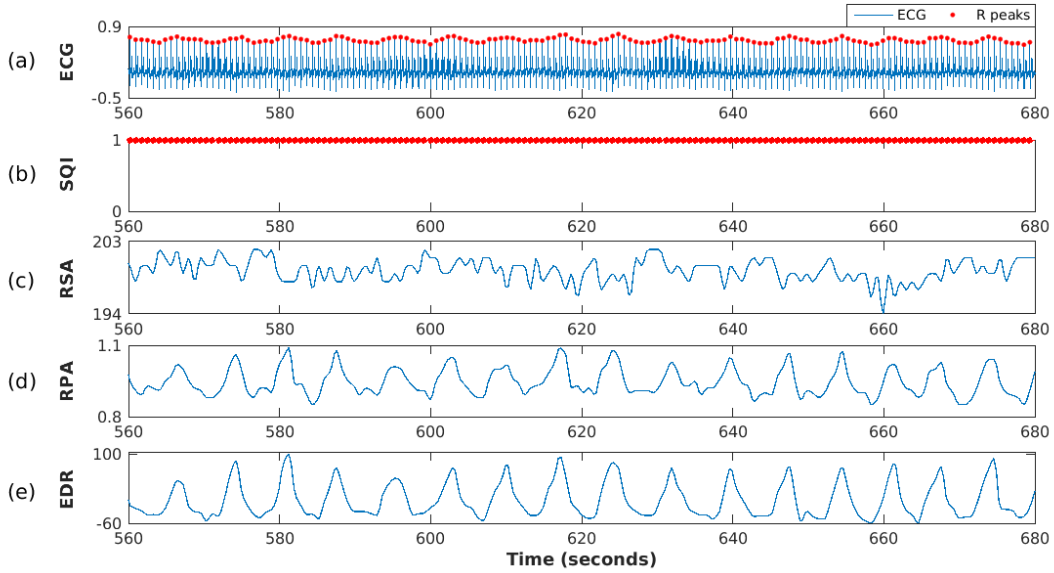


Figure 7.16: Respiration signal extraction comparison for a 120-second good-quality ECG signal. (a) Input ECG waveform with the detected R peaks shown in red, (b) SQI for each beat in the ECG waveform. The respiration signal extracted using: (c) respiratory sinus arrhythmia (RSA), (d) R-peak amplitude (RPA) and (e) ECG-derived respiration (EDR).

R-Peak amplitude (RPA)

The QRS amplitude is modulated as a result of the mechanical movement of the heart. This phenomenon is not affected by the autonomic nervous system, thus is less dependent on the age of the patient [230].

The R-Peak amplitude (RPA) method is computed by measuring the amplitude distance from the R to the S peak. Even though RPA is less affected by age than RSA, it is still of poor quality due to the higher noise when considering a single ECG channel and the sample accuracy of the beat detector. Figure 7.16(d) and 7.17(d) show the RPA signals extracted for two 10-minutes ECG signals from two different patients.

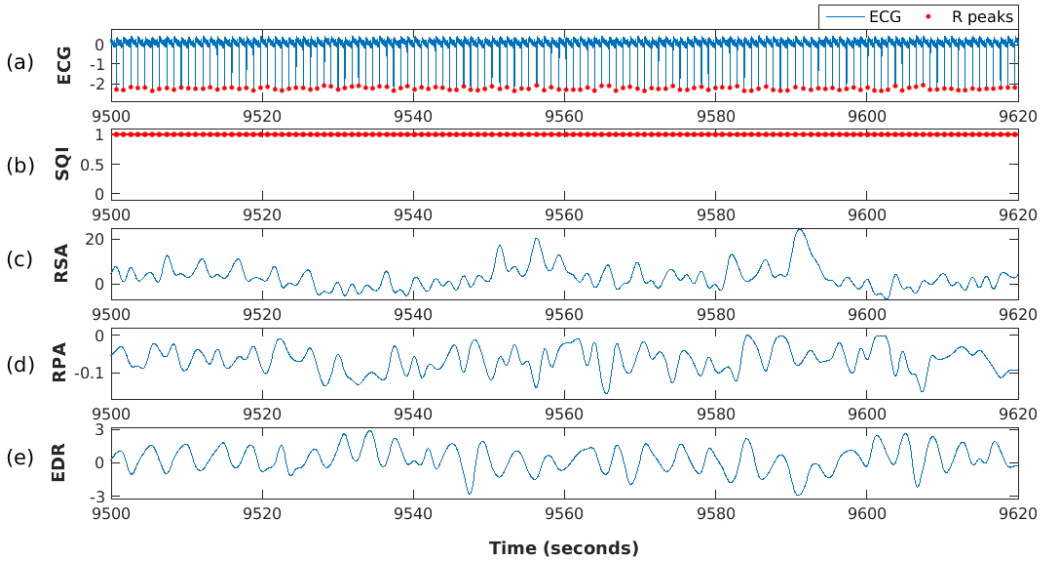


Figure 7.17: Extracting the respiratory signal from a 120-second good-quality ECG waveform. (a) Input ECG waveform with the detected R peaks shown in red; (b) SQI for each beat in the ECG waveform; (c) using the respiratory sinus arrhythmia (RSA) method; (d) using the R-peak amplitude (RPA) method; (e) using the EDR method.

ECG-derived respiration (EDR)

The ECG electrodes placed on the surface of the chest are affected by their relative motion with respect to the heart, and by changes in the electrical impedance of the thoracic cavity. The expansion and contraction of the chest during the respiratory cycle result in motion of the chest electrodes, producing a modulation in the direction of the mean cardiac electrical axis of the dipole model [64].

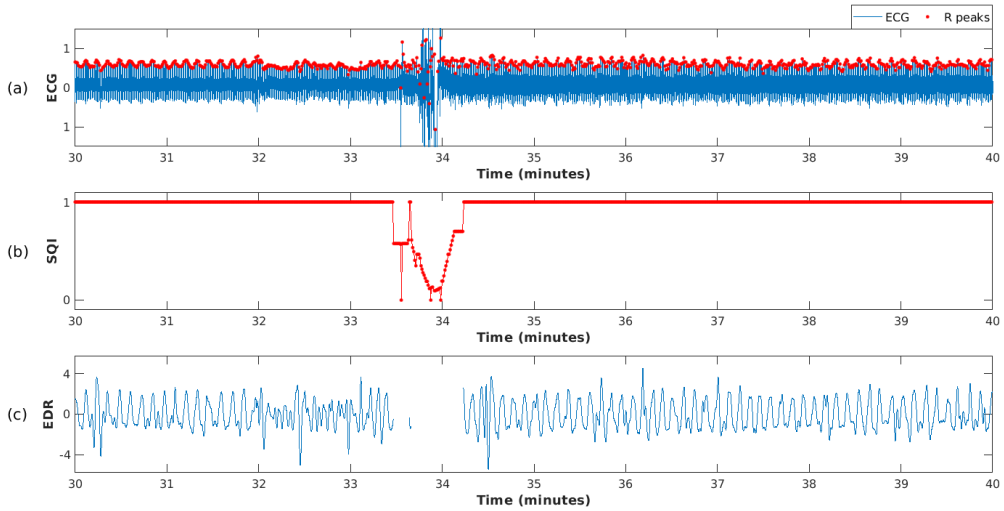


Figure 7.18: Extracting the respiratory signal from a 5-minute ECG waveform using the EDR method. (a) Input ECG waveform with the detected R peaks shown in red; (b) SQI for each beat in the ECG waveform; (c) respiration signal using the EDR method.

Moody et al [230] proposed a more robust algorithm to extract a respiration signal taking advantage of multiple channels from the ECG. An adaptation of the algorithm is used in this section. Initially, the baseline wander is subtracted by following the work of Brown et al [41]. Subsequently, the area of each QRS complex in each of the two leads is computed over a fixed window; the length of the window is equal to the median of all the PQ junctions to the J-point determined over a 120-second learning period. Given that the length of the window is fixed, the area is proportional to the mean amplitude of the signal. Finally, the respiration signal is obtained by the arctangent of the ratio of the areas of the two leads.

Figures 7.16 and 7.17 compare the three methods for extracting the respiration signal for different good quality ECG segments. It can be seen that the RSA method results in a poor signal with little or no respiratory rate information. The EDR method provides a good-quality signal from which respiratory rate can be computed.

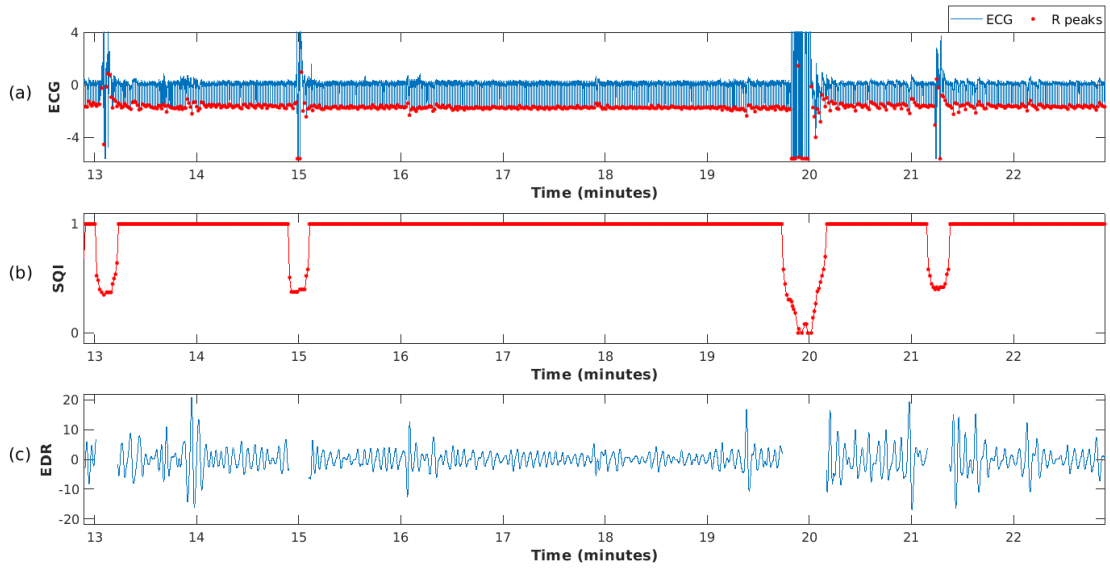


Figure 7.19: Extracting the respiratory signal from a 5-minute inverted ECG waveform using the EDR method. (a) Input ECG waveform with the detected R peaks shown in red; (b) SQI for each beat in the ECG waveform; (c) respiration signal using the EDR method.

Figures 7.18 and 7.19 show the performance of the beat detectors, signal quality metrics algorithms and the extraction of the respiration signal using the EDR method for two 10-minute ECG signals from two different patients. The algorithms correctly rule out periods when the ECG is noisy, and produce good-quality respiration signals from which respiratory rate can be computed.

7.4.5 Conclusion

A combination of algorithms to analyse the quality of the ECG channels is used. By using the two ECG channels recorded with the LifeMonitor used in the dialysis studies, the respiration signal extracted using the ECG-derived respiration technique can be used to compute respiratory rate estimates.

7.5 Respiratory rate estimation

7.5.1 Overview

Once a respiration signal has been extracted from the ECG using the techniques described above and if the respiration signal from the chest belt is available, each signal is pre-processed in three steps: Firstly, a band-pass filter is applied to reduce the effects of frequency content outside the respiratory range. Subsequently, the location of each breath peak is computed using a peak and valley algorithm. Finally, the quality of the respiration signal is assessed to discard periods for which the signal is noisy.

Respiratory rate is estimated from each respiration signal by using two methods applied to 30-second sliding windows: a time-domain technique by simply counting the number of breaths during the window, and a frequency-domain technique based on autoregressive models. The separation between consecutive windows is 5 seconds. The final respiratory rate is computed as the median from the estimates that do not vary by more than 2 breaths per minute and for periods where both respiration signals are of good quality.

7.5.2 Preprocessing

Digital filtering

The ECG electrodes and the chest belt are housed in the same chest harness that the patient wears and so are affected by various sources of noise, including muscle noise and movement. The physiological range for respiratory rate is taken to be from 6 to 42 breaths per minute, therefore a cascade of low-pass and high-pass filters with cut-off frequencies of 0.7 Hz and 0.1 Hz respectively are applied to the respiration signal. This process can be seen in figure 7.20 for a 1-minute respiration waveform from the chest belt.

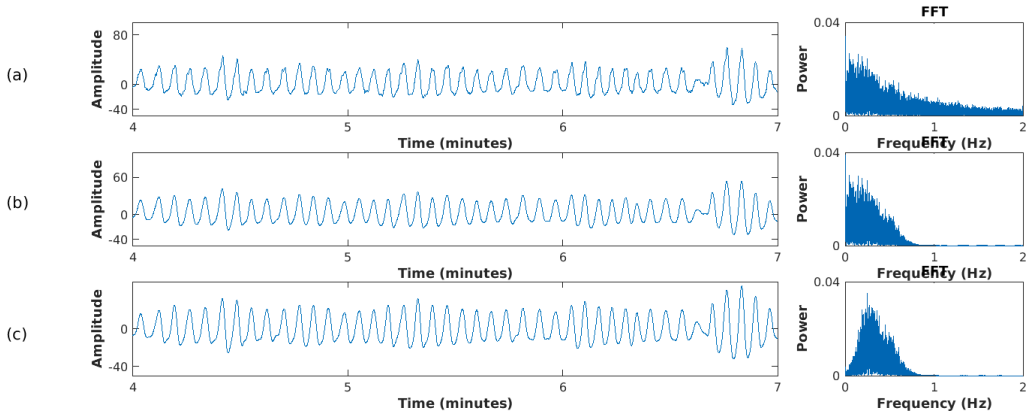


Figure 7.20: Digital filtering of the respiration waveform from the chest belt. (a) Input respiration signal, (b) low-pass filter with cut-off frequency of 0.7 Hz, (c) high-pass filter with cut-off frequency of 0.1 Hz

Breath detection

Breath detection is based on methods to find peaks and valleys in respiration waveforms by Ruangsuwana et al [277] and Lu et al [197]. Initially, a mean average curve (MAC) is computed as the mean over a running window, the length of the window being determined as the median of the respiratory period (computed as the peak of the FFT in each window) over a learning period of 2 minutes. Intersection points between the MAC curve and the respiration signal are found from all the values in the respiration signal above the MAC curve. The "up intercepts" points correspond to inspiration and the "down intercepts" points correspond to expiration in the respiratory cycle, as can be seen in figure 7.21 (a). The peak of the breath pulse is found by computing the local maximum between each up and down intercept points. Breath troughs are found by simply computing the local minimum between two peaks. The breath peak is valid only if its amplitude is at least 20% of the minimum breath amplitude (as computed as the median amplitude over the learning period of 2 minutes). A refractory period corresponding to the minimum breath interval (1.43 seconds) is added to prevent incorrectly detected consecutive peaks. Figure 7.21 shows the process of detecting the peaks and troughs from a 5-minute respiration signal extracted from the ECG using the EDR method.

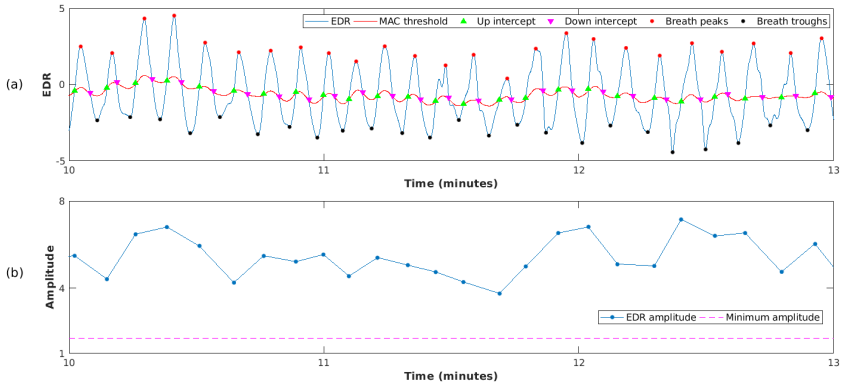


Figure 7.21: Finding the peaks and troughs from a 3-minute respiration signal extracted from the ECG using the EDR method. (a) The breath peaks (red circle) are found by computing the local maxima between the "up" and "down" intercept points (green and magenta triangles respectively) above the MAC signal (in red curve). Breath troughs are found by computing the local minima between peaks. (b) Breath pulse amplitude in blue with the minimum valid amplitude shown in magenta.

Quality assessment

The respiration signal is inherently noisy. As such, for the respiratory rate estimates to be reliable, a means for assessing the quality of the waveform is needed. Based on the work of Nemati et al [237], a quality index is computed by extracting a set of spectral features for each 30-second window and computing the spectral purity of the signal. Hjorth proposed the computation of moments as an approach to spectral feature extraction [125]. The i^{th} order spectral moment is defined by:

$$\bar{w}_i = \int_{-\pi}^{\pi} \omega^i S_x(e^{j\omega}) d\omega \tag{7.13}$$

where $S_x(e^{j\omega})$ is the power spectrum of the signal as a function of angular frequency $\omega = 2\pi f$, f being the sampling frequency. The odd-numbered moments are zero because the power spectrum is a symmetric function. The descriptors proposed by Hjorth (activity, mobility and complexity) are closely related to the even-numbered spectral moments [125]. An important property is that they can be efficiently computed in the time domain without having to compute the power spectrum. For a discrete time series $x(n) = x_1, x_2, ..., x_n$, Leif and Laguna [308] state that the spectral moments can be computed by the following equation:

$$\hat{\bar{w}}_i \approx \frac{2\pi}{N} \sum_{n=0}^{N-1} \left(x^{(i/2)}(n)\right)^2 \quad , \quad i = 0, 2, 4 \tag{7.14}$$

In ideal conditions, the respiration signal is a sinusoidal waveform with a period corresponding to the respiratory rate of the patient. Therefore, the extent to which the signal can be described by a single frequency is a measure of its quality. Barlow [26, 111] proposed the use of the "spectral purity index" (SPI) as a heuristic parameter designed to reflect the bandwidth of a signal , defined by the following equation:

$$\Gamma_{SPI}(n) = \frac{\bar{w}_2^2(n)}{\bar{w}_0(n)\bar{w}_4(n)} \tag{7.15}$$

The SPI index tends to unity for a noise-free sinusoidal signal, and decreases to zero as the bandwidth of the signal increases. Therefore, the signal quality of the respiration signal is computed as the spectral purity index for each 30-second running window. Figure 7.22 shows the quality index computed for a 10-minute respiration signal from the chest belt.

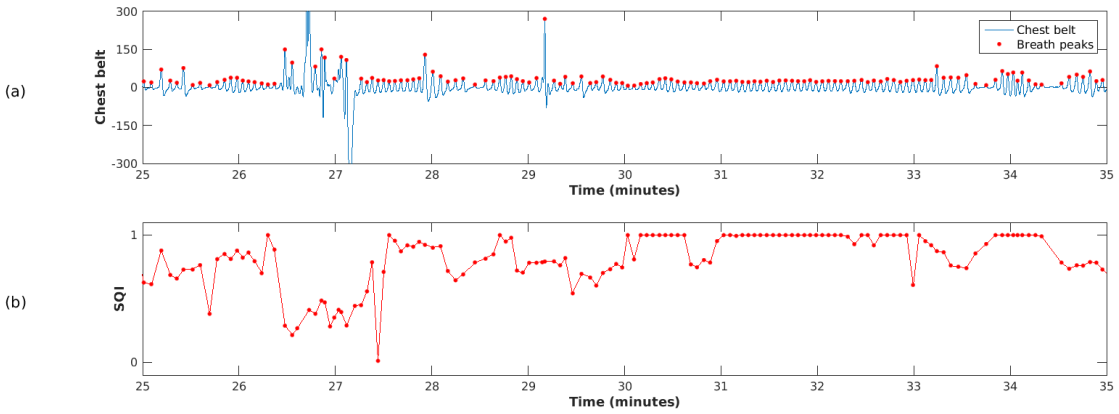


Figure 7.22: Spectral purity index for a 10-minute respiration signal from the chest belt. (a) Input chest belt signal with the breath peaks detected in red, (b) the spectral purity index.

7.5.3 Estimation methods

Breath counting

The breath counting method is computed as simply the median of all the breath intervals within each 30-second window. Figure 7.23 shows the result of computing the respiratory rate for a 10-minute respiration signal extracted from the ECG using the EDR method.

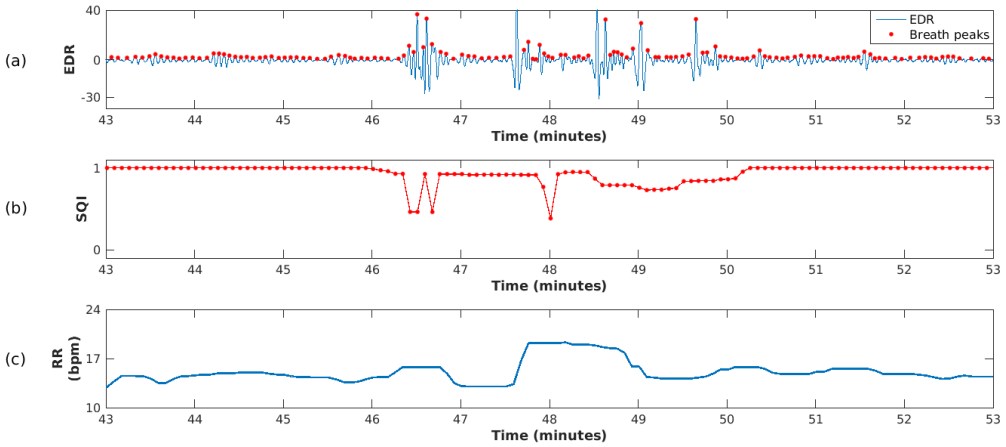


Figure 7.23: Respiratory rate estimation using the breath counting method for a 10-minute respiration signal extracted from the ECG using the EDR algorithm. (a) EDR signal with the breath peaks shown in red, (b) Signal quality index, (c) estimated respiratory rate.

Autoregressive model

Following a combination of Mason et al [209], Fleming [94] and Nemati et al [237], the respiratory rate for each 30-second window is computed using a four-stage process. Initially, an autoregressive model of order 8 is fitted to the input signal. Secondly, all poles that lie outside the physiological range for respiratory rate, taken to be 0.1 Hz to 0.7 Hz (corresponding to 6 to 42 breaths per minute respectively), are excluded. Subsequently, poles with a magnitude of at least 95% of the highest magnitude pole are kept. Finally, the respiratory rate for the window is taken to be the frequency of the pole with the smallest angle. Figure 7.24 shows the result of computing the respiratory rate using the Autoregressive Model method for a 10-minute respiration signal from the chest belt for the same period as figure 7.23.

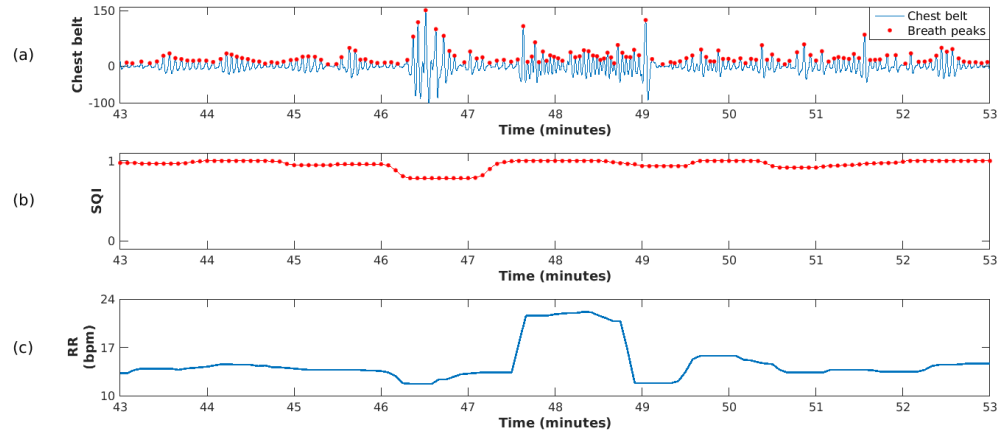


Figure 7.24: Respiratory rate estimation using the Autoregressive Model method for a 10-minute respiration signal extracted from the chest belt for the same period as figure 7.23. (a) Chest belt respiration signal with the breath peaks shown in red, (b) Signal quality index, (c) estimated respiratory rate.

7.5.4 Combined respiratory rate

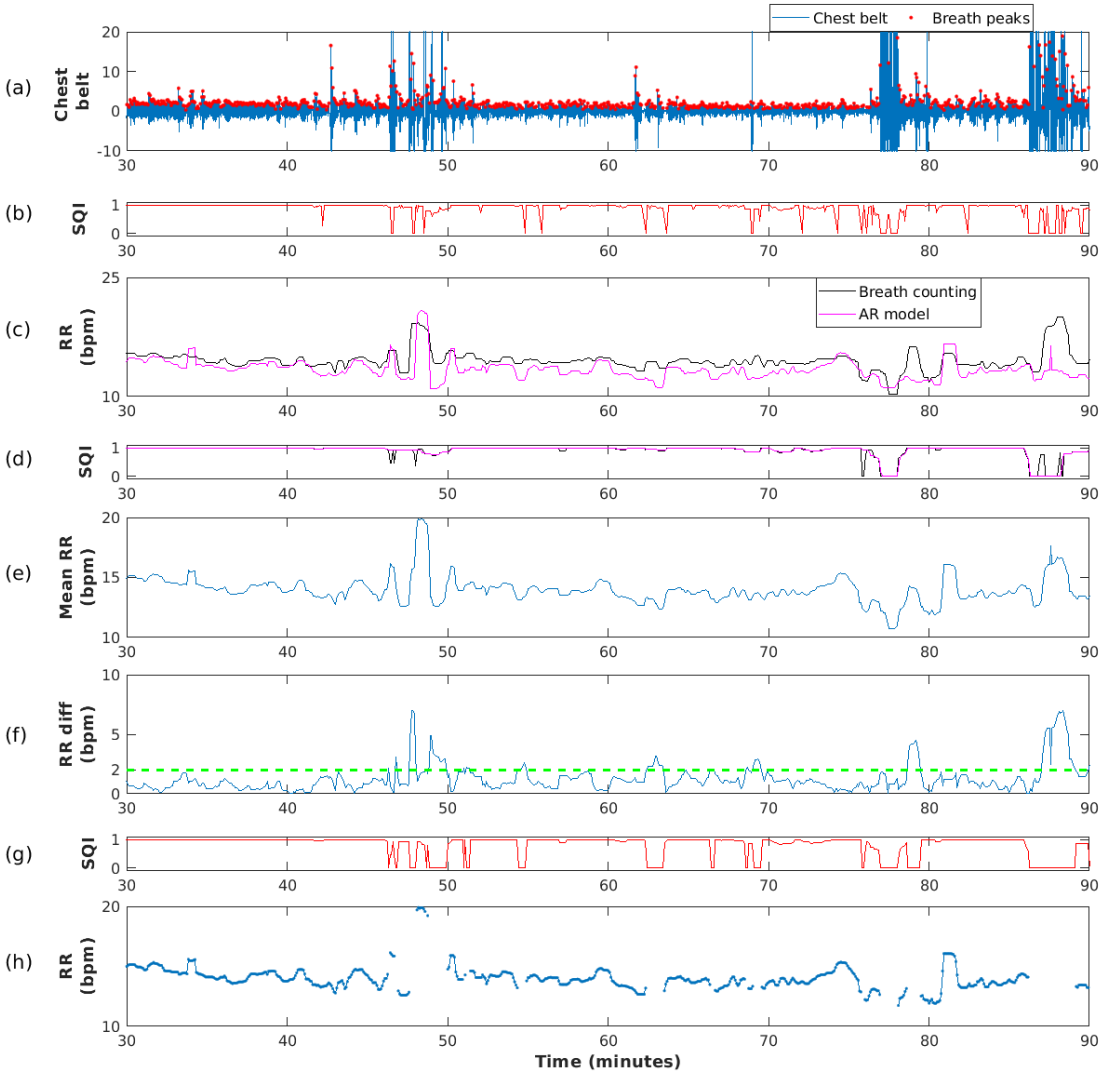


Figure 7.25: Respiratory rate estimation from 10-minute respiration signal from the chest belt for the same period as figure 7.23. (a) Input respiration signal from the chest belt with breath peaks shown in red; (b) SQI for the chest belt signal. (c) The respiratory rate estimates for the breath counting method in black and the AR model in magenta (d) The SQI for the breath counting method estimates in black and the AR model in magenta (e) The median respiratory rate estimates (f) The difference between the breath counting and AR model methods (g) The combined SQI for the respiratory rate estimates (h) The estimates for valid periods.

When no reference gold standard exists or when the reference data is of poor quality, new values are often computed as averages from only the estimates that are within a particular range. For respiratory rate, this range is usually taken as 2 or 3 breaths per minute, see for example Smith [306] or Pimentel et al [258]. Therefore, to combine the estimates computed from the two algorithms described in the previous section, the median of all the values is taken only from estimates that do not differ by more than 2 breaths per minute. Figure 7.25 shows the comparison for a 10-minute respiration signal.

7.6 Results

Table 7.2 compares the respiratory rate estimates from the ECG and chest belt as computed using the Autoregressive Model algorithm for the entire dataset when video was recorded (see chapter 6). From 304.1 hours, 32.3% (101.4 hours) of the estimates were within 2 breaths per minute, with a mean absolute error of 0.6 $\mathrm{breaths/min}$ and a mean absolute deviation of 0.5 $\mathrm{breaths/min}$.

Table 7.2: Respiratory rate estimation comparison between the ECG and chest belt respiration signals for the overall dataset, comprising 304.1 hours.

MAE ($\mathrm{breaths/min}$)	MAD ($\mathrm{breaths/min}$)	Time hours (%)
0.6	0.5	101.4 (32.3%)

Figure 7.26 shows a positive correlation of 0.96 with minimum bias in the Bland-Altman plot of 0.1 $\mathrm{breaths/min}$. The distribution of differences between the two estimates, shown in figure 7.26(c) is normally distributed.

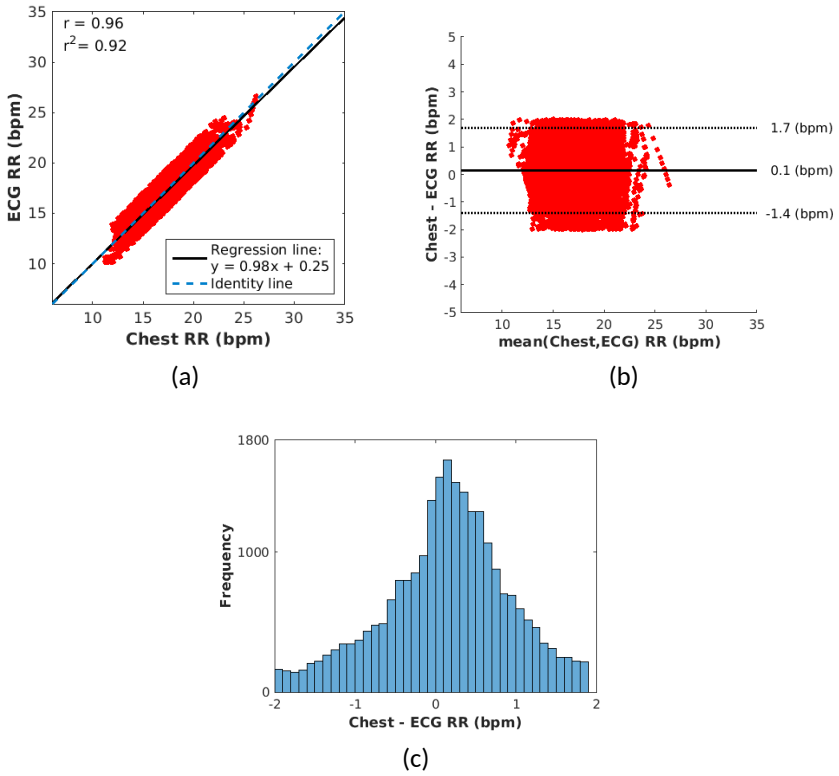


Figure 7.26: The agreement between the respiratory rate estimates computed from the ECG and chest belt for the overall data set (304.1 hours) when the difference between the estimated values is limited to 2 breaths/minute. (a) The correlation plot shows a strong positive correlation between the estimates, with a correlation coefficient of 0.96. (b) The Bland-Altman plot presents a bias close to zero. (c) The distribution of the differences between the estimated values and the reference.

7.7 Conclusion

Respiration signals are inherently noisy. For our dataset, there are several factors that can explain the source of disagreement between the respiratory rate estimates. The chest harness was manufactured in a limited number of sizes which were not always appropriate for the patients, as some of them were overweight and a proper fit was not possible. From the video recordings it can be seen in multiple instances, that patients, because of lying on their back for several hours undergoing an invasive procedure, experience fatigue leading to sudden movements that can disturb the location of the chest harness. The patient population for this study was comprised of the elderly, hence recording devices attached to the body were less tolerable, with several patients reporting discomfort to the nursing staff.

There were cases for which the required intervention by the clinical staff affected the normal functioning of the devices. On some occasions, there were equipment malfunction events that caused the manufacturer to upgrade their internal software or hardware.

The new re-derived respiratory rate estimates are consistent with values in the published literature. They were found to be of poor quality for between 37% to 61% of the time [149, 177]. These new gold standard estimates will now be used for comparison with the respiratory rate estimates from the video camera described in the next chapter.

Chapter 8

Respiratory rate estimation using a video camera

8.1 Introduction

This chapter discusses the algorithms and methodology to estimate respiratory rate from the video camera used in the clinical study for this thesis. The work carried on for heart rate estimation described in chapter 6 is expanded to process the video data and find suitable time periods during which respiratory rate can be estimated reliably. The respiratory rate estimates computed from the video camera are compared with the reference values, computed from the ECG and chest belt as described in chapter 7.

Section 8.2 summarises the overall estimation methodology. The strategy to select regions of interest is described in section 8.3. Algorithms for the extraction of respiration signals from the regions of interest are described in sections 8.4 and 8.5. Section 8.6 presents the methods for respiratory rate estimation and data fusion. The comparison of the estimated respiratory rate with the reference values is presented in section 8.7. The chapter finishes with the discussion and conclusion in section 8.9.

8.2 Overview of the estimation process

The overall methodology for respiratory rate estimation is shown in figure 8.1. As with heart rate estimation, the process starts by analysing the input video to identify time periods for which the location of the patient is known, discarding periods for which there is high activity, motion artefacts arising from the regular interaction between the patient and the clinical staff, or the patient is awake and engaging in a conversation.

Once stable periods are identified from the video recordings, regions of interest (ROI) from which respiratory rate can be estimated need to be selected on each video frame. To this purpose, several regions of interest are placed on two main areas of the image: The first group of ROI are located on the skin areas of patient's face and second group of ROI are placed on the patient's upper torso.

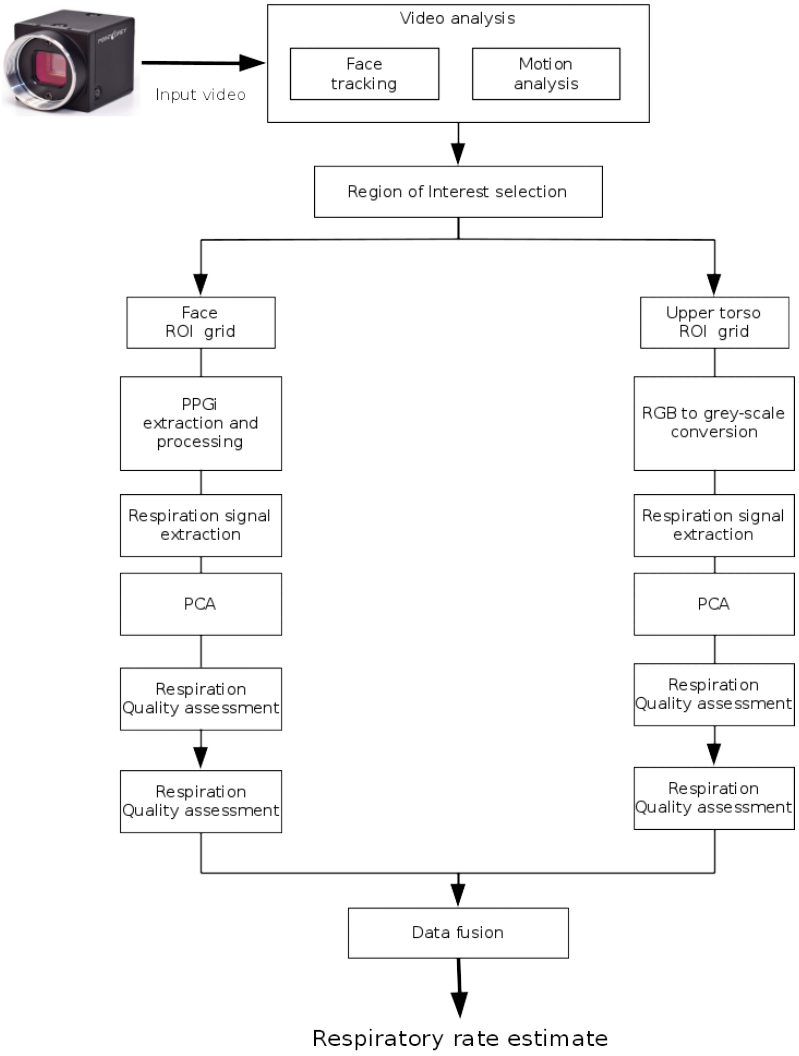


Figure 8.1: Overview of the respiratory rate estimation process.

PPGi signals are extracted, using the algorithms for heart rate estimation described in chapter 6, from the regions of interest on the patient's face for each of the three camera colour channels (red, green and blue colour respectively). Respiration signals are extracted from all the PPGi waveforms using the Respiratory-Induced Amplitude Variation (RIAV) algorithm. Respiration signals from the patient's upper torso are also extracted using grey-scale versions of the image frames.

The number of respiration signals to process is reduced using Principal Component Analysis. The quality of each principal component is assessed and respiratory rate is subsequently estimated for each signal using an Autoregressive Model.

To compute the overall respiratory rate estimate, the estimates from the regions of interest are combined using a data fusion technique based on Kalman filtering with a weighting determined by the signal quality of the respiration signal. The estimates from the camera are then compared with the reference respiratory rate computed as described in chapter 7.

8.3 Regions of interest for respiratory rate estimation

8.3.1 Physiological video maps

The physiological video maps described in section 6.10.1 and shown in the second row of figure 8.2, allow the development of a signal quality indicator indicative of the degree of confidence that the video data has frequency content in the heart rate physiological range. By colour coding the magnitude of the dominant pole in a 15-second window from 0.95 (minimum value blue) to 1.0 (maximum value red) for regions of interest (typically of 25×25 pixels in size) across the image frame, it can be seen in figure 8.2(b) that heart rate can be estimated using regions of interest on most of the patient's face, apart from the eyes (the subject is wearing glasses), the nose and the mouth.

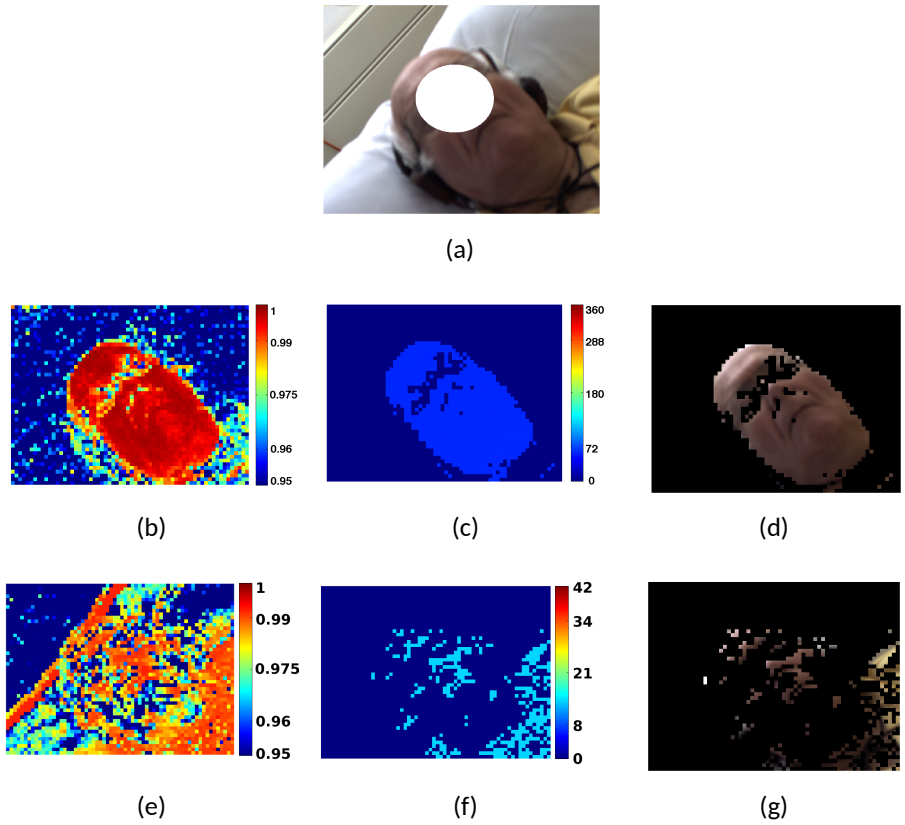


Figure 8.2: Physiological video maps for heart rate and respiratory rate from a typical 15-second video window using 25x25 pixel regions of interest. (a) The original image frame. By colour coding the magnitude of the dominant pole from an Autoregressive Model of order 8 between 0.95 (blue) and 1.0 (red), (b) heart rate can potentially be computed from most of the face, apart from the eyes (the subject is wearing glasses), the nose and the mouth. The periodic component for heart rate is strongest on the patient's cheeks and forehead, corroborating other anatomical studies. (c) The actual heart rate values for each of the regions, (d) mask of the pixel where heart rate can be estimated from the image, unsurprisingly where skin is visible. (e) Respiratory rate can potentially be computed not only from the patient's face but from other body areas such as chest. (f) The actual respiratory rate values for each of the regions, (g) mask of the pixel where respiratory rate can be estimated from the image.

The periodic component for heart rate is strongest on the patient's cheeks and forehead, corroborating anatomical studies of the human face [206] where it is stated that there is a high skin blood flow in both of these regions. The actual values for the heart rate for each 25×25 pixel region can be seen in Figure 6.63(b). Figure 8.2(d) shows the mask of the pixels where heart rate can be estimated from the image, unsurprisingly these correspond to where skin is visible.

The proposed technique can be expanded to visually analyse the image for frequency content in the respiratory rate physiological range, assumed to be between 6 and 42 breaths per minute. The third row of figure 8.2 shows the results of applying the physiological video maps algorithm by dividing every video frame into multiple small regions of interest (each typically 25 x 25 pixels) and analysing the image depending the frequency content in the respiratory rate physiological range. As opposed to heart rate, respiratory rate can be estimated not only from the subject's face, but also from the chest area, due to the patient's motion as a result of breathing.

8.3.2 Regions of interest selection

As shown in the previous section, for a robust estimation of respiratory rate, regions of interest corresponding to areas on the patient's skin and upper torso have to be considered. The nature of these signals differ in principle, as shown in figure 8.3. The signals extracted from the patient's skin are PPGi pulsatile time series, mainly caused by colour changes due to blood flow (figure 8.3(d) and 8.3(e)). The signals extracted from the upper torso (such as the thorax) are mainly affected by subject motion, as shown in figure 8.3(f).

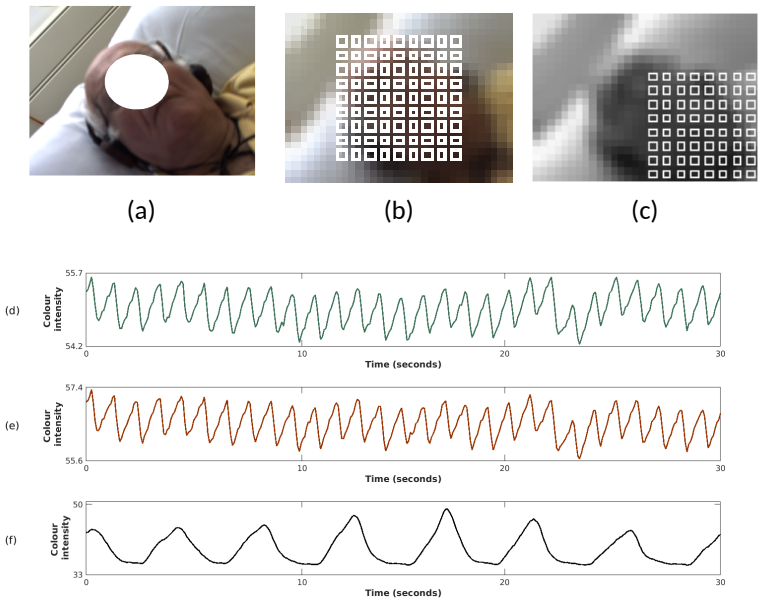


Figure 8.3: Respiratory rate maps for a typical 15-second video window using 25x25 pixel regions of interest. (a), The original image frame. (b) the 9x9 grid of all the regions of interest on the patient's face. (c) the 9x9 grid of all the regions of interest on the patient's chest. Sample waveforms extracted using (d) the green channel from a region of interest on the patient's face, (e) the red channel from a region of interest on the patient's face, (e) the grey-scale intensity variations from a 25x25 region of interest in on the patient's chest as seen in (c).

8.4 Respiration signal extraction from the face

8.4.1 Overview

The signals from the regions of interest on the patient's face correspond to the PPGi pulsatile time series extracted and analysed for heart rate estimation as described in chapter 6. Following the analysis for estimating heart rate from the video camera, PPGi signals with a low signal quality index are discarded. Similarly, the algorithms for face detection, tracking and patient's activity are used to discard time periods for which there is high activity, motion or other artefacts arising from the regular interaction between the patient and the clinical staff.

Once suitable time periods for vital signal estimation have been identified, respiratory signals are extracted from the relevant PPGi waveforms following the Respiratory-Induced Amplitude Variation (RIAV) algorithm, as described in section 3.6.8 of chapter 3. Given that there are large number of regions of interest on the patient's face, Principal Component Analysis (PCA) is applied to reduce the number of respiration signals to process. The quality of each respiration signal is analysed so that only two PCA components are kept.

8.4.2 Data windowing

Each PPGi signal is subdivided into 30-second windows with an overlap of 5-seconds between them. The following sections describe the signal processing applied to each individual waveform, analysing the PPGi data stream for each region of interest independently.

8.4.3 Respiratory-Induced Amplitude Variation (RIAV)

Similarly to the effect on the contact photoplethysmogram waveform as measured by a pulse oximeter, Respiratory-Induced Amplitude Variation (RIAV) is the modulation of the amplitude of the PPGi signal recorded by the camera due to respiration. RIAV is thought to be caused by a decrease in stroke volume from the left ventricle during changes in intra-thoracic pressure in the inspiratory cycle.

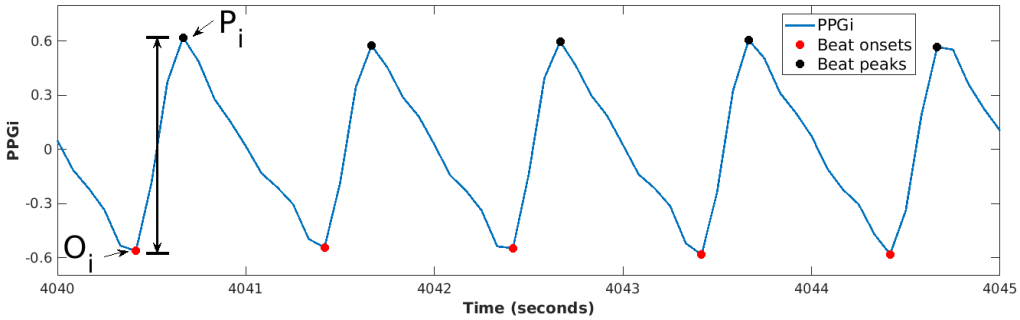


Figure 8.4: Measuring the beat amplitude from a typical PPGi signal. The beat amplitude is extracted by computing the difference between the corresponding peak "P" and onset "O" for every PPGi pulse. The peaks and onsets are detected following the adaptive Boxed Sum Slope Function (uBSSF) algorithm described in section 6.6.3 in chapter 6.

The peaks and onsets for every PPGi signal from the regions of interest on the patient's face are detected following the adaptive Boxed Sum Slope Function (uBSSF) algorithm described in section 6.6.3 in chapter 6. Subsequently, a respiration waveform is extracted by computing the amplitude of the PPGi signal, defined as the difference between the peaks and onsets for every PPGi pulse (see figure 8.4) and given by the following equation:

$$A_i = P_i - O_i \quad , \forall i = 1, 2, ..., N \tag{8.1}$$

where "A" is the beat amplitude, "P" and "O'" are the peak and onset of the i^{th} pulse respectively, and "N" is the number of beats. The resulting waveform is then resampled to 4 Hz as shown in figure 8.5.

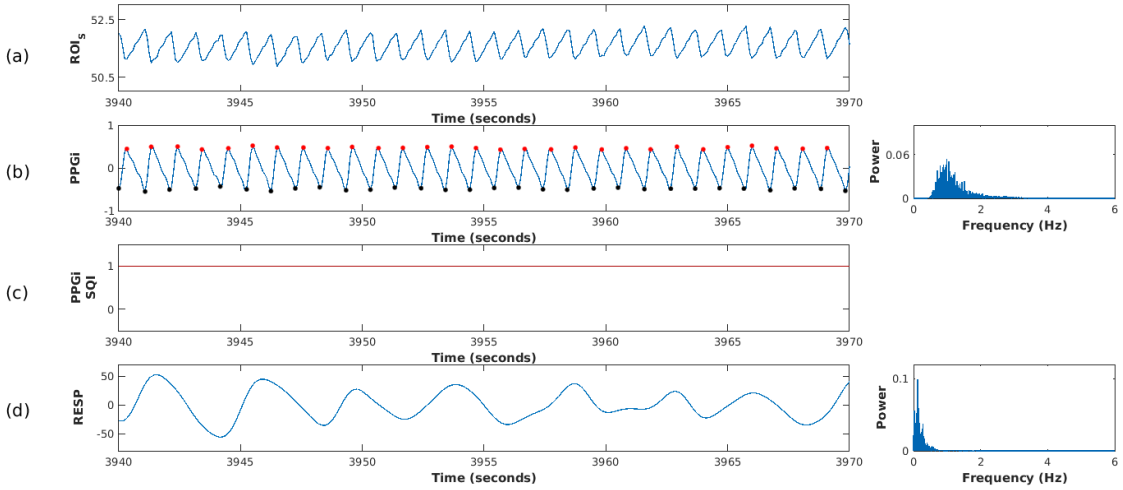


Figure 8.5: Extracting a respiration signal from a typical 30-second PPGi window. (a) Time series from the green channel from the region of interest. (b) PPGi signal processed with peaks and troughs detected. (c) Signal quality index for the PPGi waveform. (d) Resulting respiration signal extracted with the Respiratory-Induced Amplitude Variation (RIAV) algorithm.

8.4.4 Principal Component Analysis (PCA)

The number of respiration signals computed from every PPGi signal extracted from the regions of interest on a patient's face is large. As shown in figure 8.3(b), up to 243 PPGi waveforms are extracted for the three colour channels from each of the 81 regions (9x9 grid). As discussed in chapter 3, the breathing pattern of the patient is global, therefore, the data from all these regions of interest potentially contain similar respiratory frequency information. Therefore, the need to process every PPGi signal independently to estimate respiratory rate is neglected, reducing processing computing requirements and increasing the performance of the algorithms in a real-time environment. To reduce the number of signals to process, Principal Component Analysis (PCA) is applied to all the respiration signals computed from every PPGi signal, as described in the previous section .

Principal Component Analysis is a common technique used to identify and remove linear correlations among multi-dimensional data as an aid to lower its dimensions, to visualise, explore and to analyse the data with minimal loss of information; hence PCA is considered a dimensionality reduction algorithm [170] . The reduction is obtained by computing a new linear combination of the input variables, also known as projecting the data to a new set of dimensions, for which the projection error is minimised and the maximum amount of the variance is preserved [137]. These new linear combinations are termed the principal components. The most common use of PCA is as the first step in analysing large multivariate data sets [272], typical applications include signal de-noising, blind source separation and image data compression.

The PCA algorithms starts by defining an input dataset X containing the list of observations represented by the following matrix:

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & ... & x_{1,n} \\ x_{2,1} & x_{2,2} & ... & x_{2,n} \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ x_{m,1} & x_{m,2} & ... & x_{m,n} \end{bmatrix} \tag{8.2}$$

where the m rows represent the input variables (the 243 PPGi waveforms extracted from the three colour channels from each ROI), and the n columns represent the sample observations (the 120 samples from the 30-second window of each respiration signal sampled at 4 Hz).

Before computing the principal components, the mean of each input variable (the rows in equation (8.2)) is subtracted respectively from each row in the data set [136]. For each input variable, the mean is given by:

$$\mu_{j,i} = \frac{1}{n}\sum_{i=1}^{n} x_{j,i} \quad , \forall j = 1, 2, ..., m \tag{8.3}$$

The mean is subtracted from each respiration signal using equation (8.4). To simply notation, for the rest of the section we will use x to denote the zero mean input data set.

$$\hat{x}_{j,i} = x_{j,i} - \mu_{j,i} \quad , \forall j = 1, 2, ..., m \tag{8.4}$$

When the input data set contains values with different ranges or different units, some PCA algorithms also include a feature scaling step, whereby the input data is scaled to a common range, typically the standard deviation [1]. Since all the PPGi signals come from the same video camera and have the same value range (each pixel has been digitised with the same bit depth), this step is not needed for the camera data.

PCA computes a linear transformation of the input data X into a new matrix Y(of size $m \times n$), so that for a $m \times m$ matrix P:

$$Y = PX \tag{8.5}$$

The above equation projects the input data set X into the columns of P. Geometrically, the matrix P is a rotation and stretch which transforms X into Y. Thus, the rows of P effectively are the new basis for representing the columns of X. By assuming linearity, the problems is simplified to find the appropriate changes of basis that best describes X. The row vectors of P are the principal components of X.

The choice of the basis vectors P that best describe the input data set X depend on two factors: noise and redundancy [299]. For most data analysis techniques to work, noise in the input data must be low. There is no absolute scale to measure noise, rather a typical measure is signal-to-noise ratio (SNR), or the ratio of the variances between the signal and noise:

$$SNR = \frac{\sigma^2_{signal}}{\sigma^2_{noise}} \tag{8.6}$$

A high SNR indicates high precision data, while a low SNR indicates data contaminated by noise. In the case of the video camera data, a high SNR in the respiration signal extracted from a given 30-second PPGi window under analysis denotes that the signal is periodic, potentially modulated by the respiratory rate of the patient. A low SNR denotes the PPGi signal might be contaminated with movement artefacts or other sources of noise.

The quality of the PPGi signal will affect the ability to extract a valid respiration signal from it, and therefore the ability to estimate respiratory rate from the camera. Deciding if ultimately the PPGi signal is of good or poor quality is performed using the Beat-by-beat quality assessment based on Dynamic Time Warping (DTW) described in section 6.6 in chapter 6.

Redundancy in the input data set is more apparent in the video camera data, as the regions of interest across the patient's face potentially contain similar respiratory frequency information. Therefore, PCA de-correlates the original input data by finding the directions in which the variance is maximised. The variance of random variable Z of size n with mean μ is given by:

$$\sigma^2_Z = \frac{1}{n}\sum_{i=1}^{n}(Z_i - \mu)^2 \tag{8.7}$$

If the random variable Z is zero-mean, the variance can be written as:

$$\sigma^2_Z = \frac{1}{n}ZZ^T \tag{8.8}$$

where Z^T is the transpose of Z. Equation (8.8) can be generalised to measure how two variables Z and U change together, called the covariance, defined as:

$$\sigma^2_{ZU} = \frac{1}{n}ZU^T \tag{8.9}$$

The input data set X from the video camera data (see equation (8.2)) contains n columns, the observations during a given 30-second window; and m rows, the number of PPGi signals to process. The matrix X can be written in terms of its m row vectors, each of length n:

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & ... & x_{1,n} \\ x_{2,1} & x_{2,2} & ... & x_{2,n} \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ x_{m,1} & x_{m,2} & ... & x_{m,n} \end{bmatrix} = \begin{bmatrix} \mathrm{x}_1 \\ \mathrm{x}_2 \\ . \\ . \\ . \\ \mathrm{x}_m \end{bmatrix} \tag{8.10}$$

where $X \in \Re^{m \times n}$ and $\mathrm{x}_i^T \in \Re^n$. Equation (8.9) can be written to compute the variances between the rows x_i of the input matrix X as:

$$C_X = \frac{1}{n} X X^T = \frac{1}{n} \begin{bmatrix} \mathrm{x}_1\mathrm{x}_1^T & \mathrm{x}_1\mathrm{x}_2^T & ... & \mathrm{x}_1\mathrm{x}_m^T \\ \mathrm{x}_2\mathrm{x}_1^T & \mathrm{x}_2\mathrm{x}_2^T & ... & \mathrm{x}_2\mathrm{x}_m^T \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ \mathrm{x}_m\mathrm{x}_1^T & \mathrm{x}_m\mathrm{x}_2^T & ... & \mathrm{x}_m\mathrm{x}_m^T \end{bmatrix} \tag{8.11}$$

where $C_X \in \Re^{m \times m}$. The matrix C_X contains the variances of the m input variables in the diagonal entries, and all the possible covariances between the input variables on the off-diagonal entries. The matrix is called the "Covariance matrix". Covariance can be considered to be a measure of how well correlated two variables are.

PCA linearly transform the input matrix X into a new matrix Y following equation (8.5). To reduce redundancy, and therefore to achieve dimensionality reduction, PCA has the assumption that the variables in the transformed matrix Y should be as uncorrelated as possible. If the covariance matrix C_Y of the transformed data Y is known, the assumption is equivalent to minimise the covariances between the transformed variables (to minimise the off-diagonal entries in C_Y). Conversely, since PCA assumes the directions with the largest variances are the principal components, this implies to maximise the diagonal entries in C_Y. Ultimately, PCA requires the matrix P in equation (8.5) to be orthonormal.

PCA algorithm works by first selecting a normalised direction of the m-dimensional input space for which the variance in X is maximum, this constitutes the first row of the matrix P in equation (8.5). Subsequently, it selects another direction, perpendicular to all previous selected directions, for which the variance is maximum and are saved in the maxtrix P. This process continues until all m directions are processed. The resulting ordered vector inputs in P are called the "principal components".

The advantage of this simplified process is that efficient algorithms based on linear algebra decomposition can be used to find the principal components, such as Single Value Decomposition (SVD). More details of Singular Value Decomposition can be found in [299, 331].

Figure 8.6 shows the first 5 PCA components from a typical 30-second window; the first two PCA components explain more than 70% of the total variance of the data, whereas components 3 to 5 explain less than 5%.

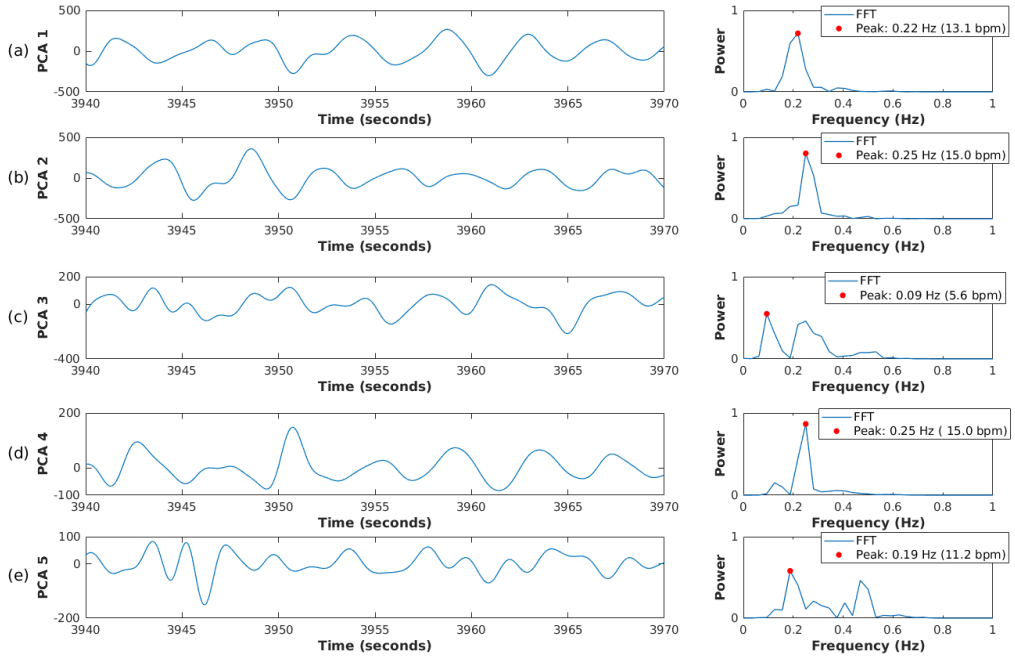


Figure 8.6: The first 5 principal components for typical 30-second respiration windows from (a) and (b) combined explain more than 70% of the total variance of the data, with (c) to (e) contributing less than 5%.

8.4.5 Respiration signal quality assessment

In ideal conditions, the respiration signal is a sinusoidal waveform with a period corresponding to the respiratory rate of the patient. Therefore, the extent to which the signal can be described by a single frequency is a measure of its quality. Barlow [26, 111] proposed the use of the "spectral purity index" (SPI) as a heuristic parameter designed to reflect the bandwidth of a signal , defined by the following equation:

$$\Gamma_{SPI}(n) = \frac{\bar{w}_2^2(n)}{\bar{w}_0(n)\bar{w}_4(n)} \tag{8.12}$$

where $\bar{w}_i$ is the i^{th} order spectral moment defined by:

$$\bar{w}_i = \int_{-\pi}^{\pi} \omega^i S_x(e^{j\omega}) d\omega \tag{8.13}$$

$S_x(e^{j\omega})$ is the power spectrum of the signal as a function of angular frequency $\omega = 2\pi f$, f being the sampling frequency.

The SPI index tends to unity for a noise-free sinusoidal signal, and decreases to zero as the bandwidth of the signal increases. As shown in figure 8.6, some PCA components include sources other than respiration. To choose which PCA component to select for respiratory rate estimation, a decision rule based on a signal quality index is applied. The quality index for the ith PCA component over a 30-second time window k is defined as:

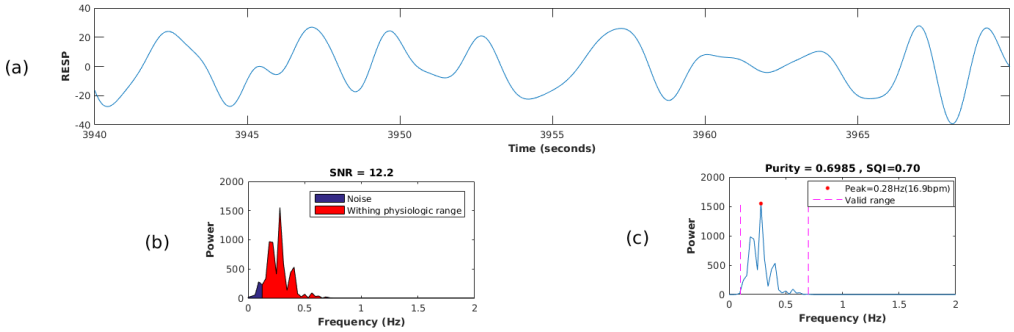


Figure 8.7: Quality assessment of a valid PCA component for a typical 30 second window: (a) Principal component input waveform, (b) the signal to noise ratio (SNR) is 12.2, (c) The peak of the FFT is at 0.28 Hz, corresponding to a respiratory rate of about 16.9 $\mathrm{breaths/min}$. This PCA component is valid with a spectral purity index of 0.69, therefore the PCA SQI is rounded to 0.7.

$$PCA_{SQI}(k_i) = \begin{cases} \Gamma_{SPI}(k_i) & \text{, if } 0.1 \text{ Hz} \leq FFT_{peak}(k_i) \leq 0.7 \text{ Hz and} \\ & PCA_{ex}(k_i) > 0.3 \text{ and} \\ & SNR(k_i) > 1 \\ 0 & \text{, otherwise} \end{cases} \tag{8.14}$$

where $\Gamma_{SPI}(k_i)$ is the "spectral purity index" as defined by equation (8.12), $FFT_{peak}(k_i)$ is the dominant frequency for the window (as determined by the peak of the FFT), $PCA_{ex}(k_i)$ is the fraction of the total variance explained by the PCA component and $SNR(k_i)$ is the signal-to-noise ratio. When the peak of the FFT is within the physiological range for respiratory rate (determined to be from 0.1 Hz to 0.7 Hz, corresponding to a range from 6 $\mathrm{breaths/min}$ to 42 $\mathrm{breaths/min}$), the PCA component explains more than 30% of the total variance and the signal-to-noise ratio is greater than 1.0, then the signal quality index for the given PCA component and window is equal to the spectral purity index. Otherwise, the signal quality index is equal to zero and the PCA component is discarded. Figure 8.7 shows a detail example for a valid PCA component.

For a typical 30-second window, figure 8.8 shows the two PCA components selected as being valid. The two components together account for more than 70% of the total variance of the data, have a high signal-to-noise ratio and the dominant frequency is within the physiological range for respiratory rate.

8.4.6 Conclusion

Extracting respiration signals from regions of interest on the patient's face using the methods so far described in this chapter presents little overhead, as much of the work is already performed when estimating heart rate. The signals from the large number of regions of interests are selected using the Principal Component Analysis algorithm, making respiratory rate estimation suitable for real-time implementation.

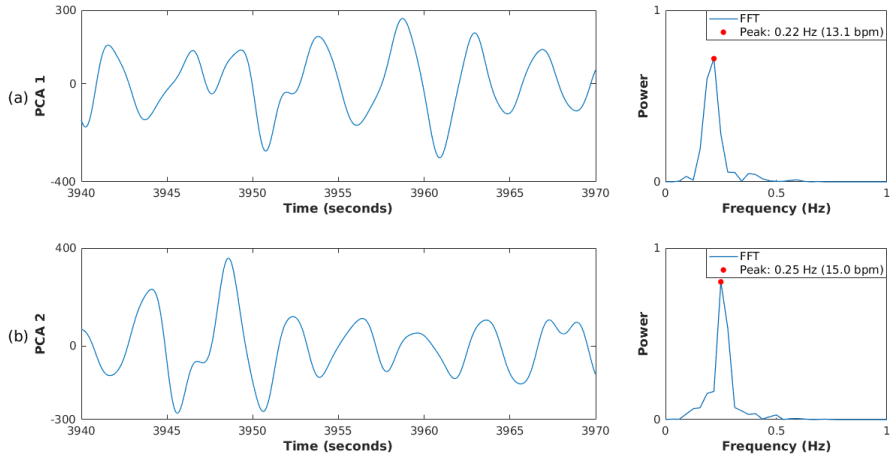


Figure 8.8: Valid PCA components for a typical 30-second window: (a) First PCA component, explaining 38% of the total variance of the data, with SQI of 0.61 and a dominant frequency of 0.22 Hz, corresponding to a respiratory rate of about 13.1 $\mathrm{breaths/min}$. (b) Second PCA component, explaining 36% of the total variance of the data, with SQI of 0.66 and a dominant frequency of 0.25 Hz, corresponding to a respiratory rate of about 15 $\mathrm{breaths/min}$.

8.5 Respiration signal extraction from the upper torso

8.5.1 Overview

Figures 8.2 and 8.3 showed examples for which respiratory rate can not only be estimated from regions of interest on the patient's face, but also from the motion of the upper torso, mainly the chest. The previous section presented algorithms to extract a respiration signal from the patient's face; this section describes algorithms and strategies to extract respiration signals from the subject's upper torso.

8.5.2 Data windowing

Each PPGi signal is subdivided into 30-second windows with an overlap of 5 seconds between them. The following sections describe the signal processing aplied to each individual window separately.

8.5.3 Respiration signal extraction

As the patient's upper torso is covered by clothing, the signal extracted does not depend on colour changes due to blood flow, but rather depends on the subject's motion patterns. Each colour image frame from the video data is converted to a grey-scale version using the following equation:

$$Gray_{i,j} = 0.2989 * R_{i,j} + 0.5870 * G_{i,j} + 0.1140 * B_{i,j} \tag{8.15}$$

where R,G and B are the colour intensity for the red, green and blue pixels respectively located at the row i and column j across the image frame. The mean over each region of interest is subsequently computed, as shown in figure 8.9 for a sample 5 regions of interest.

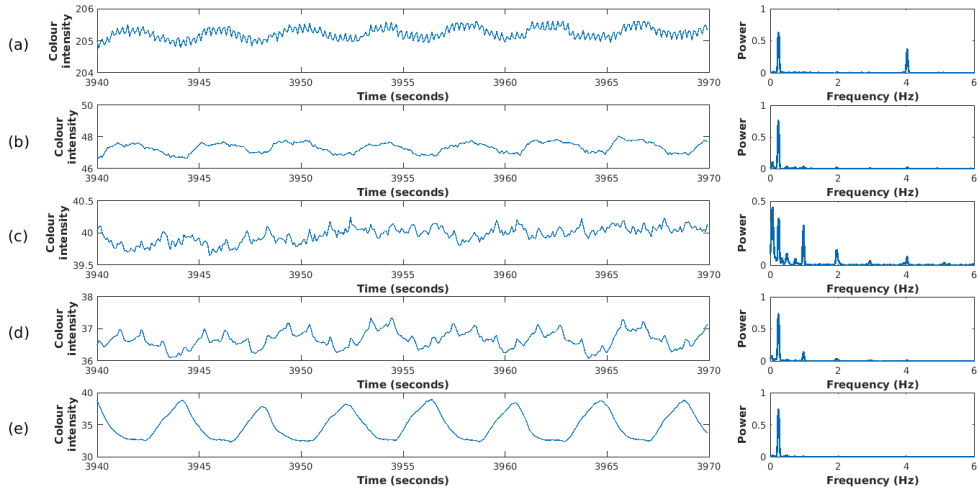


Figure 8.9: Extracting respiration signal from the patient's upper torso from a sample of 5 regions of interest for a typical 30-second window. Signals (a) to (d) are contaminated with sources other than respiration, whereas the pulsatile signal in (e) (located in the centre of the chest) is mainly caused by respiration.

8.5.4 Respiration signal quality assessment

As with the waveforms extracted from regions of interest on the patient's face, principal component analysis is applied to the respiration signals extracted from the patient's upper torso, reducing the number of waveforms to be further processed.

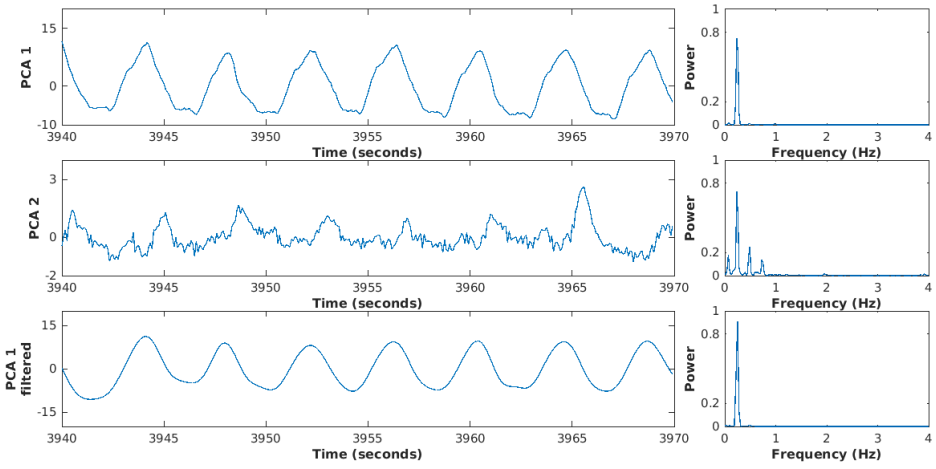


Figure 8.10: Assessing the quality of the first two PCA components for a typical 30-second window: (a) First PCA component, explaining 82.9% of the total variance, with a dominant frequency of 0.23 Hz corresponding to 14.06 breaths/min and a SQI of 0.81. (b) Second PCA component, explaining only 1.02% of the total variance, with a peak in the FFT at 0.25 Hz corresponding to 15 breaths/min and the SQI is 0.52. (c) First PCA component with a band-pass filtered applied cut-off frequencies of 0.1 Hz and 0.7 Hz, corresponding to a respiratory rate of 6 to 42 breaths/min

To assess the quality of each PCA component, the decision rule described in equation (8.14) is again applied. Finally, to improve the computation of the dominant frequency, each resulting valid PCA component is filtered with a 20th order band-pass filter with cut-off frequencies of 0.1 Hz and 0.7 Hz, corresponding to a respiratory rate of 6 to 42 breaths/min. Figure 8.10 shows the summary of the quality assessment for the resulting PCA components.

8.5.5 Conclusion

Extracting respiration signals from the patient's upper torso presents the advantage that most of the amplitude changes are due to respiration motion patterns and sources other than respiration can be easily discarded.

8.6 Respiratory rate estimation

To estimate respiratory rate (RR) from the extracted respiration signals, the input waveforms are subdivided in 30-second windows. Longer windows will lead to additional delays in a real-time implementation and decrease the percentage of data to be estimated as patient activity will cause more windows to be rejected. The separation between consecutive windows was chosen to be 5 seconds to provide a large amount of overlap between windows. Hence, respiratory rate estimates are reported every 5 seconds.

After all the respiration signals have been extracted, respiratory rate is estimated for each signal using an 8th-order Autoregressive Model. The "Multiple Kalman filters" algorithm described in section 6.8.3 of chapter 6 is applied to the following datasets:

- Using only the respiration signals from the patient's face.
- Using only the respiration signals from the patient's upper torso.
- Using all the respiration signals.

8.7 Results

According to the analysis of the reference data set discussed in chapter 7, the total time where the reference respiratory rate, as computed from the ECG and chest belt, agree within 2 breaths per minute is 101.4 hours.

Figure 8.11 shows in more detail the agreement between the reference respiratory rate values and the video camera estimates computed over the regions of interest on the patients' face for the valid dialysis video data, comprising a total recording time of approximately 101.4 hours. The Bland-Altman plot, shown in figure 8.11(a), presents minimal sensor bias with narrow differences across the respiratory rate physiological range, implying a good agreement between the two measurements. The distribution of the mean values in figure 8.11(c), shows that the values are within the expected physiological range for adults. The values from the two devices have a positive correlation coefficient of 0.85, as shown in figure 8.11(d).

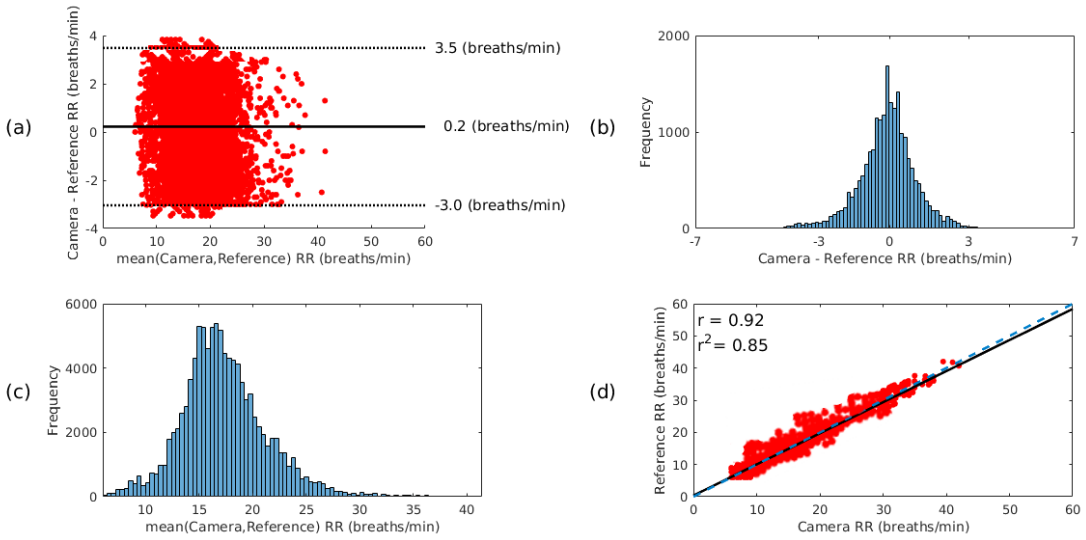


Figure 8.11: Agreement between the reference respiratory rate values (computed from the ECG and chest belt) and the camera estimates computed over the regions of interest on the patients' face for the valid dialysis video data, comprising a total recording time of approximately 101.4 hours. (a) The Bland-Altman plot presents minimal sensor bias. (b) The differences between the two estimates. (c) The distribution of the mean values show that most of the respiratory rate estimates are within the expected physiological range for adults. (d) The plot shows a strong correlation between the two devices, with a correlation coefficient of 0.85.

Table 8.1 compares the mean absolute error (MAE) and mean absolute deviation (MAD) of the three respiratory rate estimation strategies discussed on section 8.6 over the valid dialysis video data (total recording time of approximately 101.4 hours). The proportion of the time respiratory rate can be estimated from the camera is shown in the right column.

Table 8.1: Respiratory rate estimation comparison for the three ROI selection criteria for the overall valid dataset, comprising a total recording time of 101.4 hours.

ROI	MAE (breaths/min)	MAD (breaths/min)	Time (%)
Only face	2.2	2.1	60.3
Only upper torso	1.8	1.6	65.1
All combined	2.1	1.8	69.2

8.8 Discussion

Section 8.3.1 discussed the extension of the physiological video maps algorithm developed for heart rate to include a visual analysis of a signal quality metric indicative of the degree of confidence that the video data included frequency content in the respiratory rate physiological range. It was shown that respiratory rate could be estimated not only from regions of interest on the subject's face, but also from regions across the chest area, mainly the thorax and diaphragm.

The nature of the information extracted from the regions of interest on the face and chest differs in principle. The signals extracted from regions of interest on the patient's face were PPGi pulsatile time series, mainly caused by colour changes due to blood flow. The signals extracted from the upper torso (such as the thorax) were mainly affected by subject motion as a result of breathing.

Although the difference in accuracy is not large, as shown in table 8.1, the advantage to have two areas from which respiratory rate can be estimated increases the proportion of time the signal can be computed. In a typical hospital setting, large portions of the patient's skin are not always visible, therefore motion-based respiratory rate estimation is a viable alternative. Similarly, when the thorax of a patient is covered or is obstructed, respiratory rate can still be estimated from the analysis of the PPGi signals from skin areas such as the face.

One limitation of the analysis presented in this chapter is the poor quality of the reference data and the lack of a gold-standard reference respiratory rate to which the camera estimates can be compared. The reference respiratory rate, computed in chapter 7 from the ECG and chest belt, is consistent with other analysis presented in the published literature, which were found to be of poor quality for between 37% to 61% of the time [149, 177]. Therefore, the results obtained from the camera algorithms presented in section 8.7 of this chapter have to assessed in the context of the analysis discussed in chapter 7.

8.9 Conclusion

The algorithms for respiratory rate estimation presented in this chapter, build on top of the work carried out for heart rate estimation described in chapter 6. As a result, fusing the results from the PPGi waveforms from regions of interest on the patient's face with the motion-dependent signals from the subject's upper torso, the proportion of time that respiratory rate can be estimated is increased without degrading the accuracy.

As the large number of regions of interest to process is reduced using PCA, the algorithms are potentially suitable for real-time implementation, as described in chapter 10.

Chapter 9

Towards SpO_2 estimation using a video camera

9.1 Introduction

Oxygen is a chemical substance essential to the functioning of each cell in the human body and, therefore, necessary to sustain life. It is important to monitor if organs are receiving a sufficient supply of oxygen as it is being delivered to all the body parts. The measurement of blood oxygenation, also known as oxygen saturation (or peripheral oxygen saturation (SpO_2) when provided by a pulse oximeter), is an important indicator of a patient's health. Several techniques have been proposed for the monitoring of SpO_2 over the past few decades (see chapter 3). This chapter presents the methodology and algorithms that were developed for identifying changes in SpO_2 using the video camera described in section 5.4.2 and visible ambient light.

Section 9.2 summarises the overall process of identifying changes in SpO_2 from the video camera signals. Section 9.3 describes the analysis of the SpO_2 estimates provided by the two pulse oximeters used in the study and used as reference values to be compared with the camera-derived estimates. Sections 9.4 and 9.5 describe in detail the algorithms used to process the video recordings and estimate SpO_2 changes. Section 9.6 presents the peripheral oxygen saturation estimation results with respect to the two pulse oximeters used as reference devices. Section 9.7 discusses the results in the context of patients undergoing haemodialysis treatment. Finally, conclusions are presented in section 9.8.

9.2 Overview of the estimation process

The overall methodology for SpO_2 estimation from the camera is shown in figure 9.1. As our goal is to be unobtrusive and not affect regular patient care, the process begins by using the methods presented in chapter 6 to analyse the input video to identify time periods for which the location of the patient is known, discarding periods for which there is high activity, motion, artefacts arising from the regular interaction between the patient and the clinical staff, or the patient is awake and engaging in a conversation.

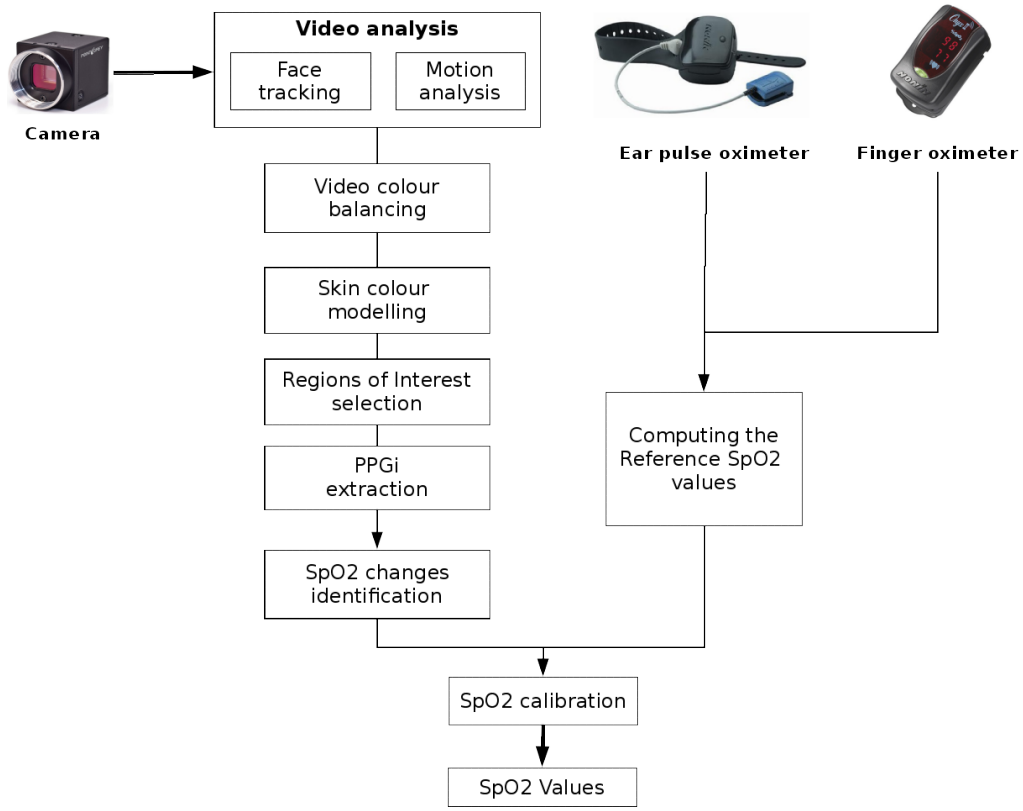


Figure 9.1: Overview of the process for identifying changes in peripheral oxygen saturation.

Once stable periods are identified from the video recordings, the images are colour-balanced to reduce the effects of changes in illumination in the scene. Following previous research on skin colour modelling, canonical skin tones are obtained by converting the colour-balanced images from RGB to Normalised Colour Coordinate (NCC) space. Subsequently, all the regions of interest on the patient's face (as detected by the face-tracking algorithms) are processed using the signal quality index methods used for heart rate analysis (see chapter 6). Once all regions of interest of good quality are identified, a representative PPGi signal is extracted from which changes in SpO_2 can be estimated using a modified version of the ratio of ratios method introduced in section 3.6.6.

As with the heart rate analysis described in chapter 6, a new "reference SpO_2" is computed as the mean from the estimates provided by two transmission-mode pulse oximeters (located on the finger and on the ear lobe) that do not vary by more than 2%.

From all the dialysis sessions recordings, eighteen 5-minute segments were manually selected, for which the reference SpO_2 decreased by more than 5%. For the time periods selected, the estimates from the pulse oximeters were used to calibrate the value obtained from the analysis of the video camera signals.

9.3 Reference SpO_2

9.3.1 Dual PPG monitoring during haemodialysis

The reference SpO_2 estimates for this study were provided by two transmission-mode pulse oximeters, one located on the finger and the other on the ear lobe (see chapter 5). Most of the studies reported in the literature use only one pulse oximeter as a reference device, recording the PPG waveform and SpO_2 estimates from a single body site. Using SpO_2 estimates from more than one pulse oximeter introduces several physiological and non-physiological factors to be considered, taking into account the haemodynamics of the patient population under study.

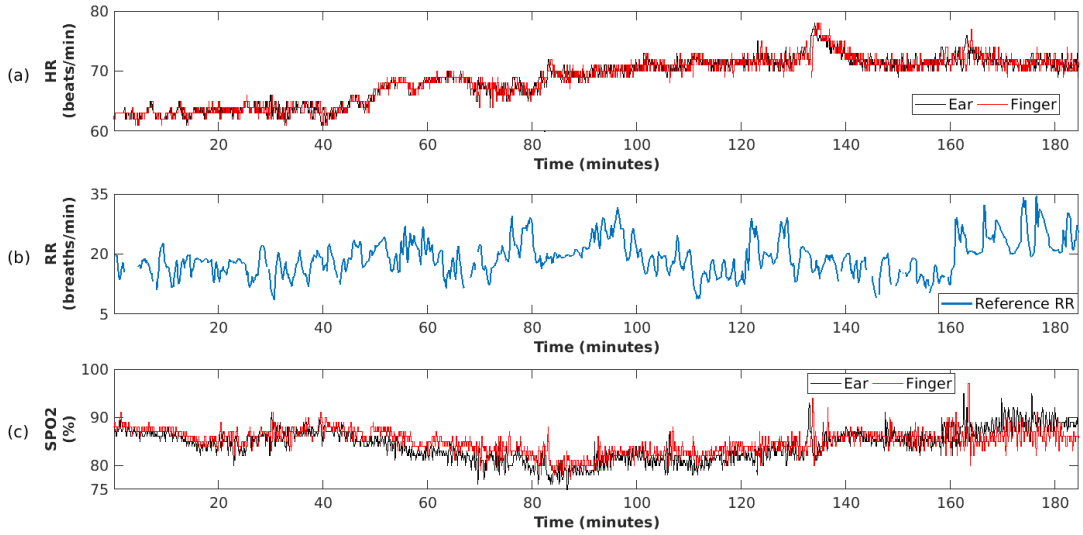


Figure 9.2: A dialysis session for a 54-year old male for which there is a good agreement between the heart rate and SpO_2 estimates from the two reference pulse oximeters connected to the ear lobe (black traces) and finger (red traces). (a) The heart rate, (b) the estimated respiratory rate computed using the ECG and Hidalgo chest belt/harness, (c) the SpO_2 estimates.

Figure 9.2 shows the recording of one dialysis session from a 54-year old male where there is a good agreement between the heart rate and SpO_2 estimates from the two pulse oximeters. For reference, the estimated respiratory rate computed using the ECG and Hidalgo chest belt/harness is also shown. During this session, the effect of low blood oxygen levels (minute 80 onwards) are seen to have an impact on the respiratory rate. In an effort to increase the amount of oxygen in the blood, breathing increases above the normal rate (typically around 12 to 16 breaths per minute for healthy individuals). Rates of 24 breaths per minute or even higher may occur. As such, feeling short of breath is generally one of the first and most obvious symptoms of low blood oxygenation. The heart rate also increases to help oxygen-containing blood circulate throughout the body in an attempt to meet the cells' needs [219].

The characteristics of peripheral circulation at various body sites can result in variations of the recorded PPG waveform and, therefore, the computed SpO_2 values. As shown in figure 9.3(d), the differences in timing can negatively affect the comparison between two SpO_2 estimates, as it is difficult to agree on a valid reference value particularly between minutes 23 and 25.

The accuracy of pulse oximeters has been reported to decrease when rapid changes in heart rate or oxygen saturation occur [200]. The discrepancy between the computed SpO_2 values from the two reference pulse oximeters is more apparent during low saturation periods (see figure 9.3(c)). The haemodialysis treatment also has an effect on the patient's cardiovascular status.

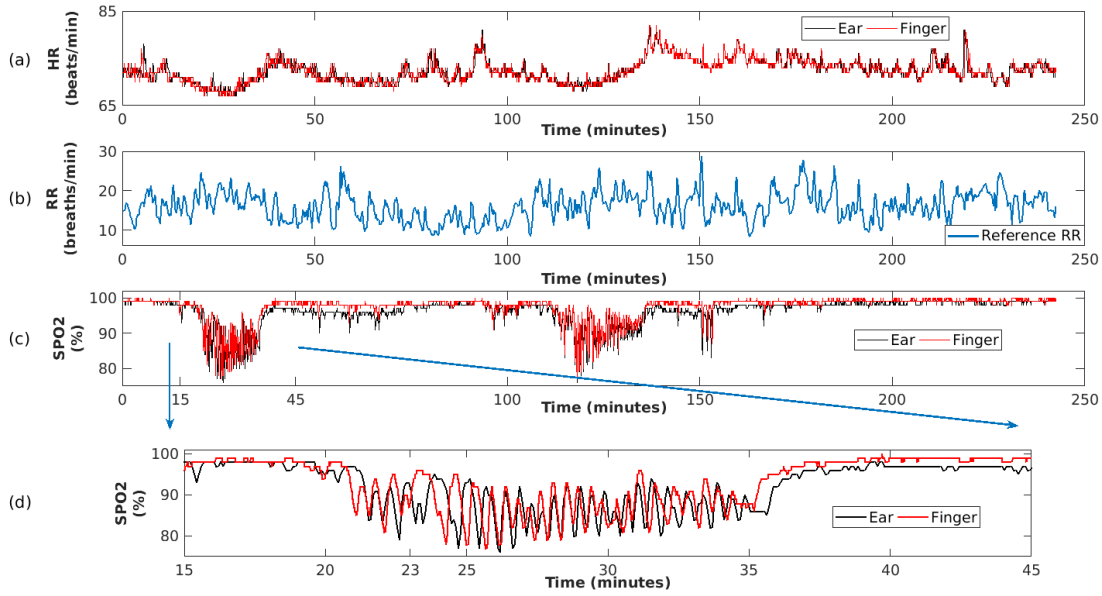


Figure 9.3: Recording from a 56-year old male with seven months on dialysis, diagnosed with Diabetic Nephropathy, Type 2 Diabetes and Hypertension. Although there is a good agreement for the SpO_2 estimates from the two reference pulse oximeters connected to the ear lobe (black traces) and finger (red traces), errors are apparent during low saturation periods. (a) The heart rate, (b) the estimated respiratory rate computed using the ECG and Hidalgo chest belt/harness, (c) the SpO_2 estimates. (d) A 30-minute extract from the SpO_2 estimates, showing the discrepancies between the two pulse oximeters.

Patients undergoing haemodialysis treatment are often diagnosed with other illnesses. In figure 9.3, the patient is a 56-year old male, with seven months under dialysis treatment prior to enrolling into the study, diagnosed with Diabetic Nephropathy, Type 2 Diabetes and Hypertension. Kidney disease amongst diabetics is common. Diabetes affects the arteries of the body and as the kidneys filter blood from many arteries, kidney problems are a particular risk for people with diabetes. Hypertension is a long-term medical condition in which the blood pressure in the arteries is persistently elevated. Long-term high blood pressure is a major risk factor for coronary artery disease, stroke, heart failure, peripheral vascular disease, vision loss, and chronic kidney disease [175, 218].

Haemodialysis is an invasive treatment in itself. During haemodialysis, blood is extracted from the patient via a surgically-formed vascular conduit (the Artereo-venous Fistula - AVF) at around 1000 to 1500 millilitres per minute, and pumped through a synthetic filter (the dialyser) integrated into the dialysis machine to be later returned into the patient's bloodstream (see figure 5.1 in chapter 5). Functioning as an interface between the patient's extracorporeal blood and the dialysis fluid (dialysate), the dialyser permits diffusion and ultrafiltration (UF).

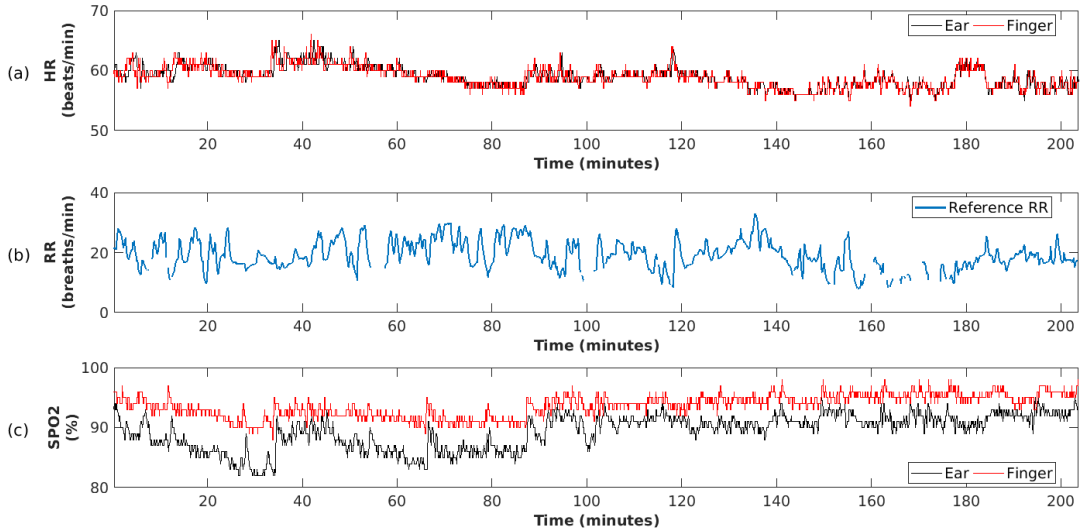


Figure 9.4: Recording from an 85-year old male with two and a half years on dialysis, diagnosed with Hypertension and given Valsartan and Amlodipine. Although the SpO_2 estimates from the two reference pulse oximeters connected to the ear lobe (black traces) and finger (red traces) follow a similar trend, the ear pulse oximeter underestimates the values, making it difficult to compute the gold standard SpO_2. (a) The heart rate, (b) the estimated respiratory rate computed using the ECG and Hidalgo chest belt/harness, (c) the SpO_2 estimates.

Arteriovenous fistulae are the gold standard for vascular access for haemodialysis and are formed from a surgical anastamosis between the venous system and the radial or brachial artery in the distal or proximal forearm, respectively. The anastomosis allows arterial blood to bypass the capillary bed and pass directly into the venous system. The increased venous pressure and flow promote thickening or arterialisation of the vascular wall [219]. The invasive nature of the fistula at the arterial-venous shunt could also have an effect on recordings of oxygen saturation during haemodialysis.

Hypovolaemia also occurs during haemodialysis. Hypovolaemia is a decrease in the volume of blood in the body, which can be due to clinical treatments such as dialysis, or during blood donation. Dialysis patients undergo pure fluid removal by ultrafiltration from the blood space, resulting in haemoconcentration. The transition from hypervolaemia, due to fluid overload in the dialysis patient before treatment, to relative hypovolaemia during ultrafiltration can have deteriorating effects on the cardiovascular status of patients. Tissue perfusion with oxygen and nutrients is dependent on an adequate blood pressure to overcome the opposing forces to blood flow during blood volume reduction [106, 222].

Patients on haemodialysis are given medications with several side effects. Figure 9.4 shows the recording for an 85-year old male with two and a half years on dialysis treatment, diagnosed with hypertension and antineutrophil cytoplasmic antibody (ANCA) Vasculitis, a type of autoimmune swelling caused by autoantibodies that result in the inflammation of blood vessels. The patient was administered Valsartan and Amlodipine, two commonly used medications to slow the development of end-stage renal disease.

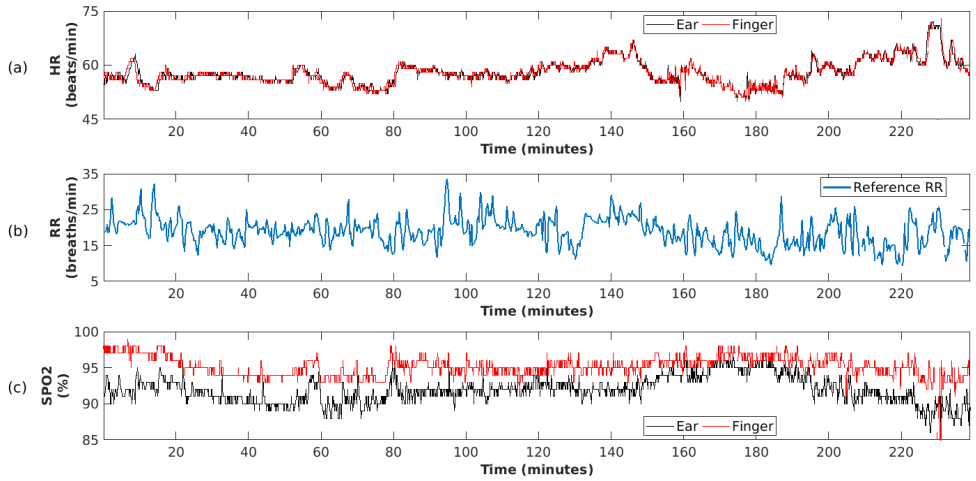


Figure 9.5: Recording from a 75-year old female with two months on dialysis, diagnosed with diabetes and hypertension. The SpO_2 estimates from the pulse oximeter connected to the ear lobe (black traces) underestimates the values provided by the finger pulse oximeter (red traces) at the beginning of the session. There is a gradual correction as the session continues, but the offset gets larger as the session ends, making it difficult to compute the gold standard SpO_2. (a) The heart rate, (b) the estimated respiratory rate computed using the ECG and Hidalgo chest belt/harness, (c) the SpO_2 estimates.

Valsartan blocks the actions of angiotensin II, a peptide hormone that causes vasoconstriction and a subsequent increase in blood pressure [157]. Amlodipine, a calcium channel blocker, is used in the management of hypertension to dilate blood vessels and improve blood flow. Calcium stimulates the actions of cells. Calcium channels are found within the heart muscle, blood vessels, and other cells. When the channels open, calcium flows into the cell, causing them to contract. Figure 9.4, shows that very often the ear pulse oximeter (black traces) underestimates oxygen saturation values when compared with the finger pulse oximeter (red traces), making it difficult to compute the true underlying oxygen saturation for the patient.

The skin is arguably the largest organ in the body; as such it possesses a remarkable range of functions and adaptations that include waterproofing and forming a barrier from infection, a site of vasoconstriction during orthostasis and a heat exchanger with the environment. The latter role arises from its position on the surface of the body and its capacious volume for blood flow allowing rapid heat loss [44]. Occasionally, little shunts can open up in the blood vessels of the skin that bypass the capillaries and send arterial blood directly into the venous system. It is thought that this is part of the cooling/warming mechanism in humans, and it is likely that there are more of these shunts in the blood vessels of the earlobe than the finger [55]. This shunting of arterial blood into the veins can then cause some pulsatility in the venous system and trick the pulse oximeter into thinking that some of this poorly-saturated venous blood is in fact arterial, as pulse oximeters are designed to take advantage of the pulsatile nature of arterial blood flow (see chapter 3). In these cases, the SpO_2 reading is derived from light interacting partly with arterial blood and partly with venous blood. The result leads to SpO_2 values which are lower than the 'true' SpO_2 of arterial blood [55, 219].

Female hormones also play a role, which might explain the differences shown in figure 9.5, a dialysis session from a 75-year old female diagnosed with diabetes and hypertension. The recording starts with a clear offset for the ear pulse oximeter, which gives values lower than those of the finger pulse oximeter. This offset gradually corrects itself as the session progress, but gets larger at the end of the session. Note that, in figures 9.4 and 9.5, although the ear pulse oximeter underestimates the SpO_2 values, the heart rate provided by the two pulse oximeters are in good agreement, corroborating the notion that the probes were correctly attached to the patient.

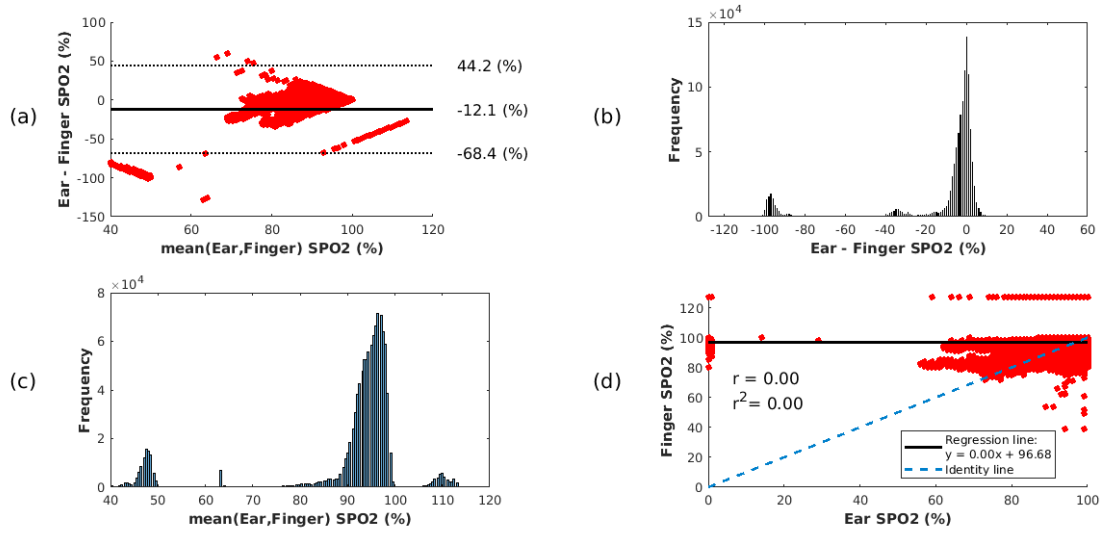


Figure 9.6: Agreement between the original SpO_2 values provided by the two reference pulse oximeters for the entire dialysis dataset (304.1 hours). (a) The Bland-Altman plot presents a large error, with the ear pulse oximeter underestimating SpO_2. (b) The distribution of the differences between the two pulse oximeters. (c) The distribution of the mean values shows that some of the SpO_2 estimates reported are outside physiological range (typically taken between 70% and 100%). (d) The plot shows no correlation between the two devices.

Section 6.3 in chapter 6 discussed other non-physiological factors that can affect the timing and values recorded from the two pulse oximeters. SpO_2 estimates are reported at different time intervals: the Hidalgo kit provides one peripheral oxygen saturation measurement every 5 seconds, corresponding to a sampling frequency of 0.2 Hz; in contrast, the finger pulse oximeter provides three SpO_2 estimates every second, corresponding to a sampling frequency of 3 Hz. Moreover, according to the finger pulse oximeter manufacturer, each reported SpO_2 estimate is the result of averaging values over the duration of 8 heart beats. Documentation was not available on the algorithms or window size used for reporting SpO_2 estimates from the ear pulse oximeter. It is important to also note that the SpO_2 estimates from the two pulse oximeters are reported as integers, with the rounding strategy not documented by the manufacturers.

The agreement between the original values as reported by the two pulse oximeters is presented in figure 9.6 for the entire data set (corresponding to approximately to 304.1 hours). The two devices show no correlation, large errors and a bias of 12.1% for the ear pulse oximeter with respect to the finger pulse oximeter. Figure 9.6(c) shows that some of the SpO_2 estimates reported are outside physiological range, typically taken between 70% and 100%.

A direct comparison, on a sample by sample basis, of the SpO_2 estimates between the camera and each of the two pulse oximeters can therefore potentially be affected by factors not only caused by physiology but also by the recording set-up and the different signal processing techniques used by the manufacturers.

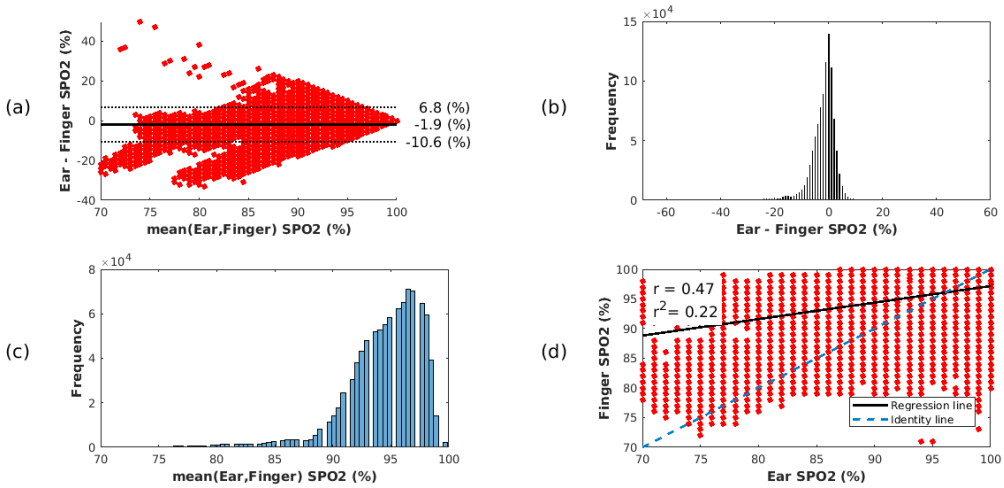


Figure 9.7: Agreement between the original SpO_2 values provided by the two reference pulse oximeters for the entire dialysis dataset (304.1 hours). (a) The Bland-Altman plot presents a sensor bias, with the ear pulse oximeter underestimating SpO_2. (b) The distribution of the differences between the two pulse oximeters. (c) The distribution of the mean values shows that most of the SpO_2 estimates reported are above 90%. (d) The plot shows a poor correlation between the two devices with a correlation coefficient of 0.47.

Figure 9.7 shows the agreement between the SpO_2 estimates reported by the two pulse oximeters when the values are restricted to a physiological range between 70% and 100%. The devices are still poorly correlated with large errors and a bias of 1.9% for the ear pulse oximeter with respect to the finger pulse oximeter. Although large errors exists between the SpO_2 estimates from the two pulse oximeters, the mean absolute error (MAE) is 3.2% with a mean absolute deviation (MAD) of 3.6%. Therefore, as with the heart rate analysis described in chapter 6, a new reference SpO_2 can be computed from the time periods when the reported values by the two pulse oximeters are in good agreement. This new "ground truth" reference SpO_2 is used to evaluate the estimates derived from the camera.

9.3.2 Computing the reference SpO_2

To compute the new reference SpO_2, a three-stage algorithm is proposed. The technique is similar to the procedure used in computing the reference heart rate described in section 6.3.3 of chapter 6. Initially, as the two pulse oximeters provide values at different time intervals, the SpO_2 estimates from the two devices are re-sampled to 1 Hz. Secondly, the two signals are time aligned by the delay found for the reference heart rate. Finally, the new reference SpO_2 is computed, on a second-by-second basis, as the mean of the estimates from each pulse oximeter that do not vary by more than 2%. This process is shown in figure 9.8 for a 10-minute segment from a typical dialysis session.

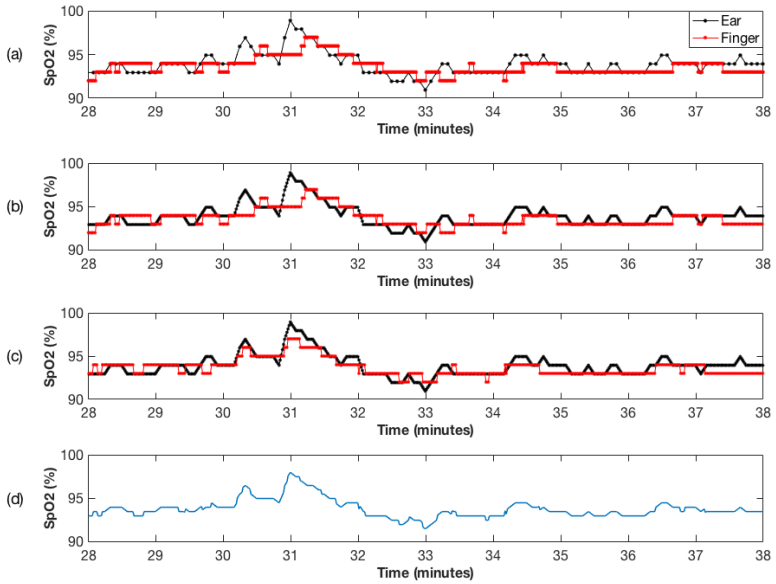


Figure 9.8: Four-stage process for computing the reference SpO_2 values from the two pulse oximeters for a 10-minute segment from a typical dialysis session. (a) Original time series for which the SpO_2 from the ear pulse oximeter (in black) is sampled at 0.2 Hz and the SpO_2 from the finger pulse oximeter (in red) is sampled at 3 Hz. (b) Both input signals are resampled to 1 Hz, (c) time-aligned signals after the finger pulse oximeter was found to be delayed by approximately 20 seconds with respect to the ear pulse oximeter, (e) Finally, the reference SpO_2 is computed as the average of the values provided by the two pulse oximeters.

9.3.3 Results

According to table 5.4 in chapter 5, the number of video sessions recorded was 104, comprising a total video length of 369.1 hours. For 12 sessions, the recording from at least one pulse oximeter was interrupted due to patient discomfort, medical intervention, equipment malfunction or other external factors. 8 other sessions were discarded because the video camera recording equipment was interrupted by the clinical staff or the recording was not complete. Therefore, there was simultaneous recordings with both pulse oximeters and camera for only 84 sessions, consisting of a total recording time of 304.1 hours.

Table 9.1 shows the error between the SpO_2 estimates from the two pulse oximeters for the original values provided by the manufacturers, values restricted to a physiological range and for the post-processed values used for computing the reference SpO_2. For the original values (see figure 9.6), corresponding to a total recording time of approximately 304.1 hours, estimates present a large mean absolute error (MAE) of 13.3% with a mean absolute deviation (MAD) of 28.2% for 304.1 hours. When the SpO_2 estimates are restricted to a physiological range between 70% and 100% (see figure 9.7), the estimates have a mean absolute error of 3.2% with a mean absolute deviation of 3.6% for 87.1% of the time, corresponding to approximately 264.9 hours. This is because, for approximately 12.9% of the time (39.4 hours), any of the two pulse oximeters flagged their measurements as errors. From the dataset, when computing the reference SpO_2, only 49.8% (approximately 151.5 hours) of SpO_2 values differ by no more than 2% with a mean absolute error of 1.0% and a mean absolute deviation of 0.8%.

Table 9.1: Comparison of the SpO_2 estimates provided by the two reference pulse oximeters placed on the ear lobe and the finger for the overall patient population, comprising a total recording length of 304.1 hours.

Data	MAE (%)	MAD (%)	Time hours (%)
Original values, no processing	13.3	28.2	304.1 (100%)
Original values, good quality values	3.2	3.6	264.9 (87.1%)
Reference SpO_2 computation	1.0	0.8	151.5 (49.8%)

Figure 9.9 shows in more detail the agreement of the SpO_2 estimates from the two pulse oximeters for the values used in computing the reference SpO_2. The Bland-Altman plot has a bias close to zero with narrow differences across the SpO_2 range, showing a good agreement between the two measurements. The distribution of the mean SpO_2 values is presented in figure 9.9(c), showing that most of the patients have a peripheral oxygen saturation above 90%. The values from the two devices are now highly correlated with a positive correlation coefficient of 0.91 (compared with figure 9.7(d)).

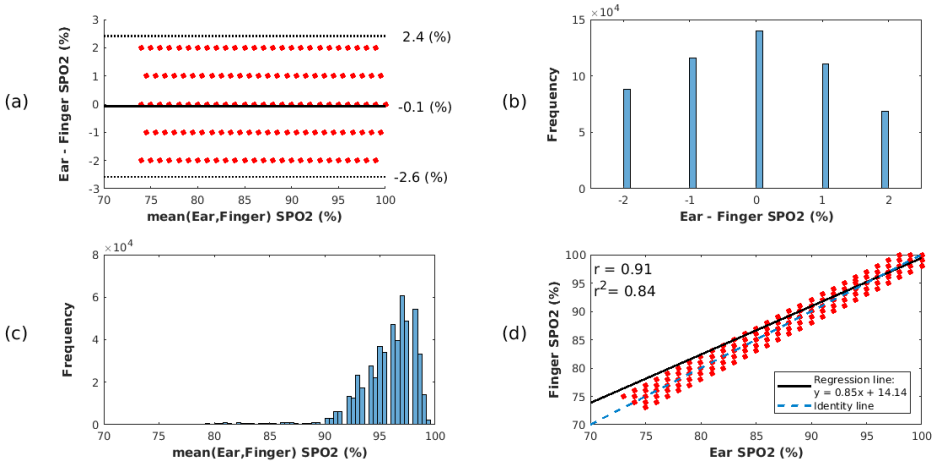


Figure 9.9: Agreement between the SpO_2 values provided by the two pulse oximeters used to compute the reference SpO_2 for the entire dialysis dataset, comprising a total recording time of approximately 304.1 hrs. (a) The Bland-Altman plot presents a bias close to zero. (b) The distribution of the differences between the two pulse oximeters. (c) The distribution of the mean values show that most of the SpO_2 estimates reported are above 90%. (d) The plot shows a good correlation between the two devices with a correlation coefficient of 0.91.

9.3.4 Conclusion

Given the analysis described above, when the SpO_2 estimates from the two pulse oximeters are aligned in time, the measurements are within 2% of each other for 49.8% of the time (151.5 hours). In those cases, the two time series are highly correlated. Therefore, the proposed reference SpO_2, computed from the mean of the estimates from the two pulse oximeters, can be used to calibrate and compare the estimates derived from the camera, for those 151.5 hours of data.

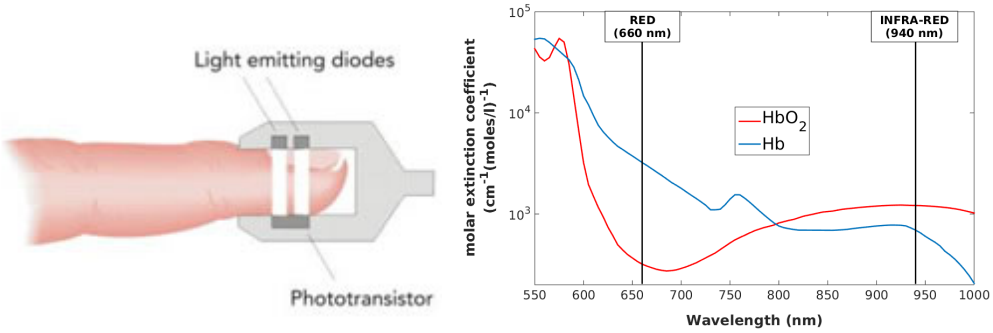
9.4 Skin colour analysis

Unlike the methods presented in previous chapters that mainly find a frequency component to estimate heart rate or respiratory rate, identifying changes in SpO_2 depends on the colour reproduction of the video camera sensor from skin regions. One major challenge when recording video is that digital cameras are not able to distinguish between colour changes in an object and colour shifts caused by varying illumination in uncontrolled environments, such as in a hospital. Colour changes can be due to several factors such as varying light levels, the temporary presence of shadows during the intervention of the clinical staff, or the varying light colour from changes in the spectral power distribution of light sources (daylight, fluorescent or incandescent).

It is important, therefore, to properly balance the colour reproduction of the video recordings and to choose the right colour channels to be used for identifying changes in SpO_2. The choice of colour channels can be based on the traditional wavelengths used in pulse oximetry.

9.4.1 The choice of wavelengths in a pulse oximeter

A pulse oximeter traditionally uses two LEDs and a photo-diode placed on opposite sides of a vascular bed such the finger tip, shown in figure 9.10(a) (copied from figure 3.10(a) for clarity). The red (R) and infra-red (IR) channels are often chosen, typically at 660 nm and 940 nm wavelengths respectively (see figure 9.10(b), copied from figure 3.7) as the absorption spectra are different at those wavelengths for deoxygenated haemoglobin (Hb) and oxyhaemoglobin (HbO_2). Oxygen saturation is then derived from the ac and dc signals recorded from the red and IR channels. The LEDs of a pulse oximeter have very tight wavelengths specifications, usually with a spectral bandwidth of $\pm 3\ \mathrm{nm}$, as variations can lead to high inaccuracies in SpO_2 estimation [81, 353].



(a) (b)

Figure 9.10: Typical wavelengths chosen in pulse oximetry. (a) Two LEDs and a photodiode are usually placed on opposite sides of a vascular bed such the finger tip (Source: Parin Dedhia, Cornell). (b) Oxyhaemoglobin (HbO_2) and deoxygenated haemoglobin (Hb) exhibit markedly different absorption characteristics in the red and infra-red parts of the spectrum (Source: Nonin).

Although wavelengths in the red and infra-red range are typically chosen in pulse oximetry, other wavelengths are also used, such as in the sensor proposed by Lochner et al [192] that uses organic light-emitting diodes (OLED) at the green (532 nm) and red (626 nm) wavelengths (see figure 9.11).

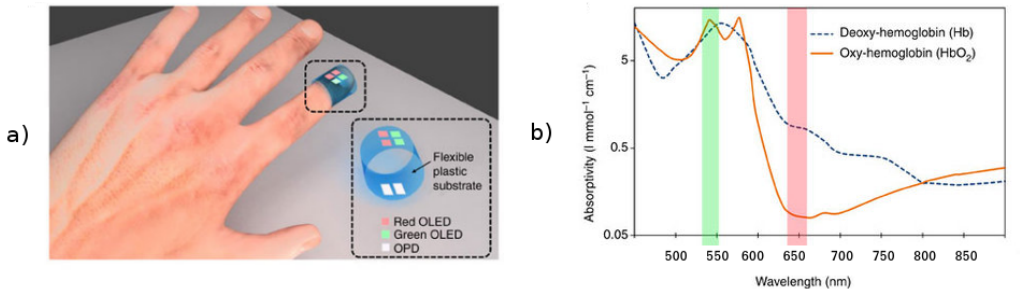


Figure 9.11: HbO_2 and Hb also exhibit different absorption characteristics in the green and red parts of the spectrum. (a) Pulse oximetry sensor composed of two organic LEDs. (b) HbO_2 (orange solid line) and Hb (blue dashed line) absorption spectra at the green (532 nm) and red (626 nm) wavelengths of interest (Source: Lochner et al [192]).)

9.4.2 Colour reproduction in digital cameras

The imaging sensor in a video camera with a Bayer colour filter array, such as the one used for the clinical study in this thesis, reproduces the scene by filtering the light in the visible spectrum (400 nm - 700 nm) with three components which are typically red (R), green (G) and blue (B), as shown in figure 9.12 (copied from figure 4.1 for clarity). The resulting recorded colour signal contains light from the illumination sources, the light reflected from a surface or a combination of both in a similar manner to the human vision system [187, 363].

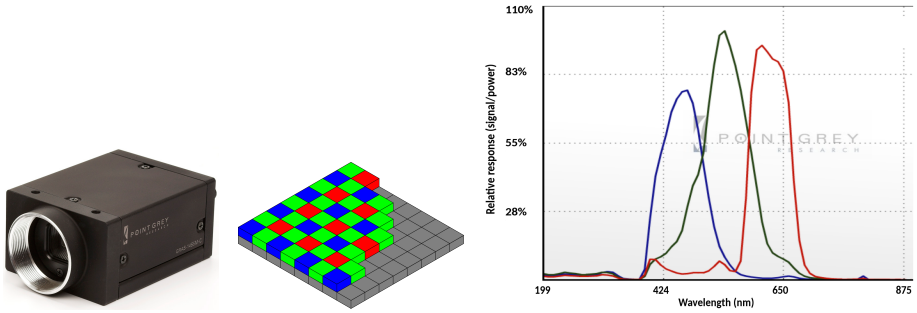
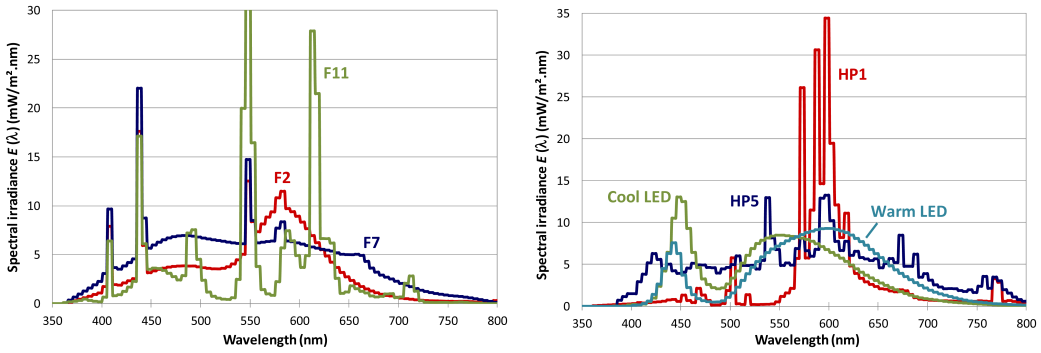


Figure 9.12: The video camera used in the study: (a) Grasshopper2 camera with the Sony ICX625 progressive scan CCD sensor. (b) The Bayer filter pattern used in the sensor: 50% green, 25% red and 25% blue. (c) Response curve for the three colour channels (Source: PointGrey).

Unlike the photo-diode used in a pulse oximeter, colour cameras cannot directly measure the spectra of colour signals because spectral accuracy is sacrificed for spatial resolution [98]. Since the spectral data for a single point in the scene is described with three values (called tristimulus values), the (R,G,B) values are only an approximation of the true incoming colour signal spectra. There exists another major problem caused by this spectral data compression: colour samples with different reflectances can become metameric, so that different combinations of light across the wavelength range can produce an equivalent sensor response. This is more evident when the same colour object is recorded as two different colours when affected by different illumination sources, or when two different colours cannot be discriminated under the effects of other illuminants [363].



(a) (b)

Figure 9.13: The spectral irradiance for different common light sources: (a) Fluorescent illuminants F2, F7 and F11. (b) Illuminant HP1 from a high pressure sodium lamp, illuminant HP5 for a metal halide lamp, and a standard cool and warm LED spectrum for typical LED lights. (Source: Minnaert et al [226])

Minnaert et al [226] analysed the spectral response for several common indoor illuminants. The spectral irradiance of three fluorescent lighting sources is shown in Figure 9.13(a): F2, a cool white fluorescent lamp with a correlated colour temperature (CCT) of 4230 K, mostly used in office buildings; F7, a broad-band fluorescent lamp with CCT = 6500 K, representing average daylight; and F11, a tri-band fluorescent lamp with CCT = 4000 K, typically used in warehouse lighting. Figure 9.13(b) shows the spectral irradiance for other typical light sources found indoors.

The output of a colour camera can be modelled with the following equation [187]:

$$D = \frac{\int \eta_D(\lambda) I_p(\lambda) S(\lambda)\, d\lambda}{\int \eta_D(\lambda) I_c(\lambda)\, d\lambda} \tag{9.1}$$

where D is the red, green or blue response, λ is the light wavelength, η is the response of a spectral filter, I_p is the spectral power distribution (SPD) of the prevailing illumination, I_c is the spectral power distribution of the calibration illumination, and S is the spectral reflectance of the surface. The nominator in equation (9.1) describes the resulting recorded image as depending only on the main illumination, object reflectance and camera sensitivity. The denominator models the white balance, adjusting the gains of the camera so that the response for white is equal in every channel. For example, the response of a white pixel in a 8-bit colour camera is adjusted to have a value of 255 for all three RGB channels.

This dependence on white balance and correct colour reproduction is particularly challenging in equation (9.1). When the prevailing and calibration illumination are the same, then the output image is called a canonical or calibrated image, and colours are canonical colours. However, when the prevailing and calibration illumination differ, the output image is called a non-canonical image and its modelling is more problematic. Martinkaupp et al [208] proposed a normalisation procedure, which assumes that the prevailing illumination is originally I_{np} , its normalisation factor is the constant factor f_p , and the calibration illumination is I_{nc} , which is normalised by the constant factor f_c . If we insert these variables into equation (9.1), we can derive the following output response for the red channel (and similarly for the other two channels):

$$
\begin{aligned}
R &= \frac{\int \eta_R(\lambda) I_p(\lambda) S(\lambda)\, d\lambda}{\int \eta_R(\lambda) I_c(\lambda)\, d\lambda} \\
&= \frac{\int \eta_R(\lambda) \frac{I_{np}(\lambda)}{f_p} S(\lambda)\, d\lambda}{\int \eta_R(\lambda) \frac{I_{nc}(\lambda)}{f_c}\, d\lambda} \\
&= \frac{f_c}{f_p} \frac{\int \eta_R(\lambda) I_{np}(\lambda) S(\lambda)\, d\lambda}{\int \eta_R(\lambda) I_{nc}(\lambda)\, d\lambda}
\end{aligned}
\tag{9.2}
$$

The ratio f_c/f_p is close to unity only when the illumination conditions are the same. Different choices for normalisation methods may produce different results [187, 208].

9.4.3 Colour balancing

Changes in illumination during the video recordings can have a detrimental effect on the skin chromaticities perceived by a video camera and therefore result in a poor performance of methods for skin modelling [207]. Fluorescent light sources can have strong variations in their red, green or blue components and change the colour appearance of the face, as shown in figure 9.14. Therefore, there is a need to balance the colour reproduction from the raw video recordings across all the dialysis sessions.



(a) (b) (c) (d)

Figure 9.14: The changes in illumination conditions during the video recordings can have a detrimental effect on the skin colour perceived by a video camera: (a) recording during a dark winter morning, (b) fluorescent light source with a strong blue component, (c) regular summer morning, (d) fluorescent light source with a strong green and red component

Each video frame was colour balanced by following the generalised diagonal transformation model [364], expressed in the following equation:

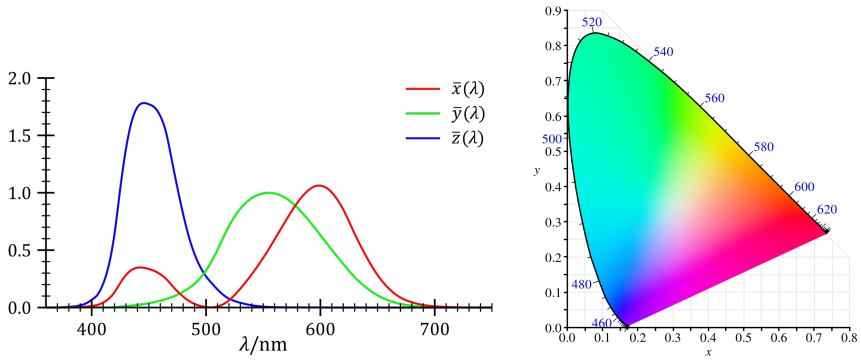
$$
RGB_{out} = F_{out} \times D \times F_{in} \times RGB_{in} \tag{9.3}
$$

where RGB_{in} is a set of 3×1 camera input RGB values for a given pixel, F_{in} is 3×3 matrix transforming RGB to an intermediate colour space, D is 3×3 diagonal matrix whose values vary with illuminant and effectively performs the colour balancing, F_{out} is 3×3 matrix transforming the intermediate colour space value back into RGB colour space, and RGB_{out} is a set of 3×1 colour balanced RGB values. The matrix D can be computed from:

$$D = \begin{bmatrix} \frac{R_{display}}{R_{scene}} & 0 & 0 \\ 0 & \frac{G_{display}}{G_{scene}} & 0 \\ 0 & 0 & \frac{B_{display}}{B_{scene}} \end{bmatrix} \tag{9.4}$$

where RGB_{scene} are the RGB colour values of the original non-colour balanced image, in the intermediate colour space, for a point that should be white. Similarly, $RGB_{display}$ are the values of the display white point of the intermediate colour space. The correct choice of diagonal parameters guarantees that the scene white point is mapped into the display white point [364].

The International Commission on Illumination (CIE) XYZ colour space has been used as an intermediate colour space for colour balancing [343, 364]. An important attribute of the XYZ colour space is that it is device independent and, as such, is typically used to report the spectral response of samples measured by a colorimeter or a spectrophotometer [343]. The XYZ "tristimulus" values describe any colour that can be perceived by an average human observer and are assigned to the red, green and blue curves respectively by colour matching functions [330].



(a) (b)

Figure 9.15: The CIE XYZ colour space: (a) The standard observer colour matching functions, (b) Colour space chromaticity diagram (Source: Minnaert et al [226])

The XYZ's colour matching functions, shown in figure 9.15(a), $\overline{x}(\lambda)$, $\overline{y}(\lambda)$ and $\overline{z}(\lambda)$ are the numerical descriptions of the chromatic response of the observer. They can be thought of as the spectral sensitivity curves of three linear light detectors yielding the tristimulus values X, Y and Z. Collectively, these three functions are known as the CIE standard observer [119]. The tristimulus values for each colour are given in terms of the standard observer by:

$$\begin{aligned} X &= \int_{380}^{780} S(\lambda)\overline{x}(\lambda)\, d\lambda \\ Y &= \int_{380}^{780} S(\lambda)\overline{y}(\lambda)\, d\lambda \\ Z &= \int_{380}^{780} S(\lambda)\overline{z}(\lambda)\, d\lambda \end{aligned} \tag{9.5}$$

where S is the spectral radiance and λ is the wavelength of the equivalent monochromatic light (measured in nanometers). To show the colour spectrum visible to humans, the "CIE xy Chromaticity" diagram is often used. The XYZ tristimulus values specify a colour's lightness, hue and saturation; whereas the *chromaticity* specifies the hue and saturation only. The chromaticity values x, y, z are computed using the formulas in equation (9.6).

$$\begin{aligned} x &= \frac{X}{X+Y+Z} \\ y &= \frac{Y}{X+Y+Z} \\ z &= \frac{Z}{X+Y+Z} \end{aligned} \tag{9.6}$$

As the chromaticity values all add to unity ($x + y + z = 1$), the xy diagram is used to display the colour gamut of a device, shown in figure 9.15(b).

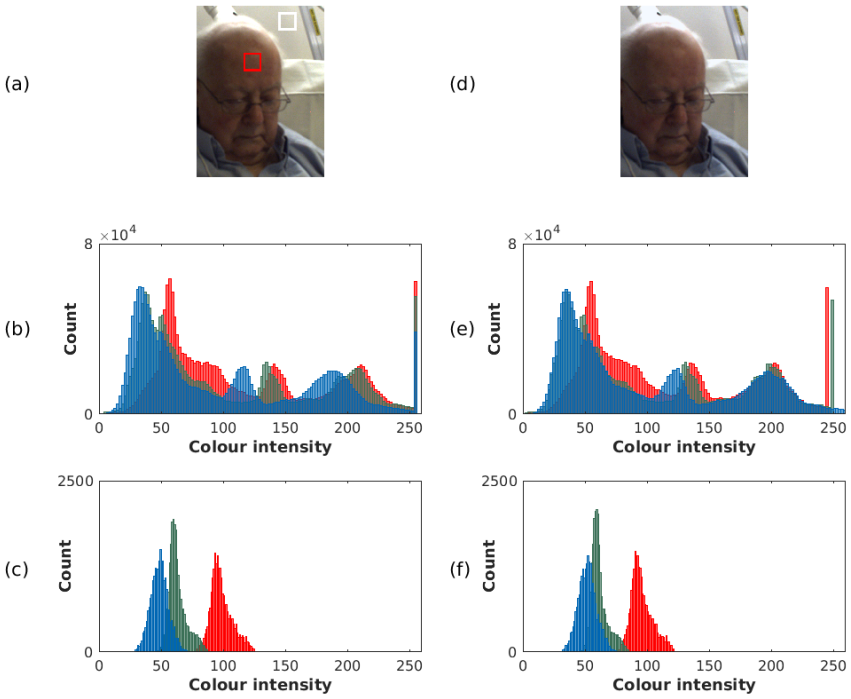


Figure 9.16: Colour balancing for a video recording with small blue hues. (a) Original RGB image with the white point selected in the white square and the sample skin region in the red square, (b) histogram for the whole original RGB (c) histogram for the skin region. (d) Colour balanced RGB image (e) Histogram of the colour balanced image, (f) histogram for the skin region.

Most of the video recordings from the dialysis clinical data set were recorded from a similar distance. The reference white point for matrix D in equation (9.4) was therefore manually chosen from the wall behind the patient, seen as the white squares in figures 9.16(a) and 9.17(a). This choice of the reference white point has the advantage of it not being in contact with the patient. Other potential reference points (such as the pillow and bed) can contain frequency components in synchrony with the heart rate or respiratory rate as demonstrated by the physiology maps in section 6.10.1.

Examples of the results of the colour balancing process can be seen in figures 9.16 and 9.17 for scenes affected by different colour shifts by the ambient illuminants, from small blue hues (figure 9.16) to strong green and red components that made the scene appear yellow (figure 9.17).

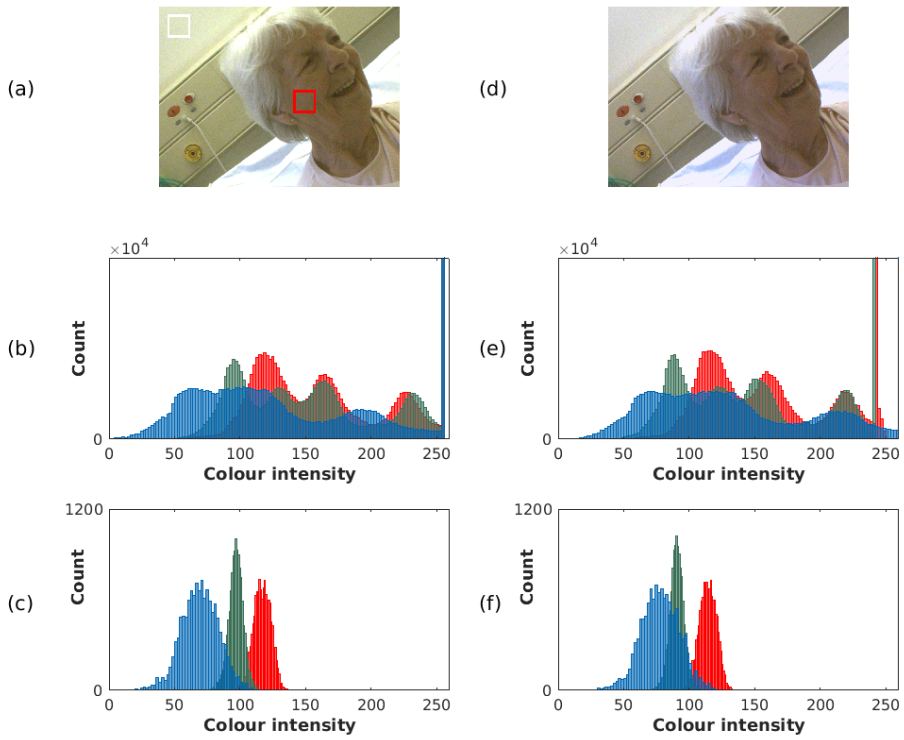


Figure 9.17: The result of the colour balancing algorithms for a video recording with strong green and red components that made the scene appear yellow. (a) Original RGB image with the white point selected in the white square and the sample skin region in the red square, (b) histogram for the whole original RGB (c) histogram for the skin region. (d) Colour balanced RGB image (e) Histogram of the colour balanced image, (f) histogram for the skin region.

9.4.4 Colour spaces for skin

The colour balancing methods discussed in the previous section reduce the colour cast distortions produced by the different illuminants present during video recording. Nevertheless, the colours of objects perceived by a colour camera are also dependent on the changing lightness (how dark or light the scene is) and conditions that have a strong effect on the intensities of the recorded colours.

The appropriate choice of a colour space representation for modelling the skin colour changes under different illumination conditions is critical when developing robust techniques against illumination changes. Therefore, the colour-balanced RGB images from the video recordings were converted to the Normalised Colour Coordinate (NCC) space as it has been shown to be among the most usable colour spaces for skin chromaticity modelling [187, 207]. In the Normalised Colour Coordinate colour space, the chromaticity values (noted as r, g, b lowercase) are computed using the following equations:

$$
\begin{aligned}
r &= \frac{R}{R+G+B} \\
g &= \frac{G}{R+G+B} \\
b &= \frac{B}{R+G+B}
\end{aligned}
\tag{9.7}
$$

Figure 9.18 shows the result of converting an image that was colour-balanced with the algorithms described in the previous section to the NCC colour space. It is important to note how the skin areas are more visible in the NCC colour space compared to the regular RGB.

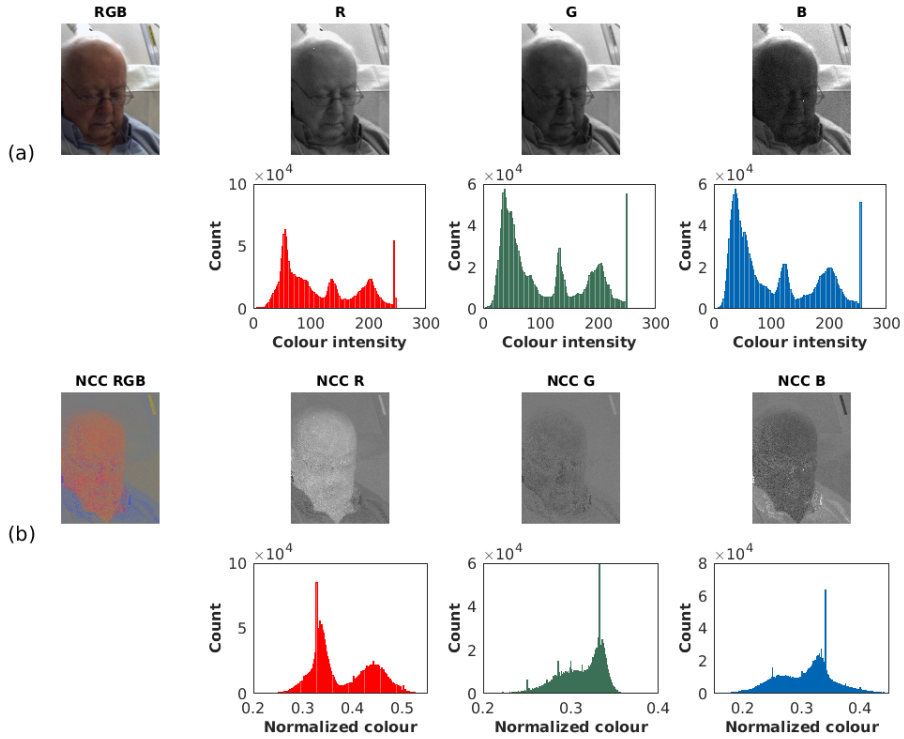


Figure 9.18: Colour balancing the images: (a) Colour balanced RGB image with the corresponding histograms. (b) Conversion to Normalised Colour Coordinates (NCC) and the corresponding histograms for each channel.

As the NCC values in equation (9.7) all add to unity ($r + g + b = 1$), the chromaticity NCC_r NCC_g diagram is often used to display the skin locus of an image. Figure 9.19(b) shows the chromaticity diagram for the original camera RGB input image before colour balancing, whereas figure 9.19(f) shows the chromaticity diagram after colour balancing and final colour space conversion.

The advantage of using the NCC colour space for extracting a pulsatile signal to subsequently estimate SpO_2 can be seen in figure 9.19 for a 30-second extraction of the mean values from a region of interest on the patient's forehead (red squares in the figures). The changing lightness of the scene is present in the original input RGB values from the camera, where as the illumination effects are reduced after colour balancing and colour space conversion, improving the extraction of a PPGi time series.

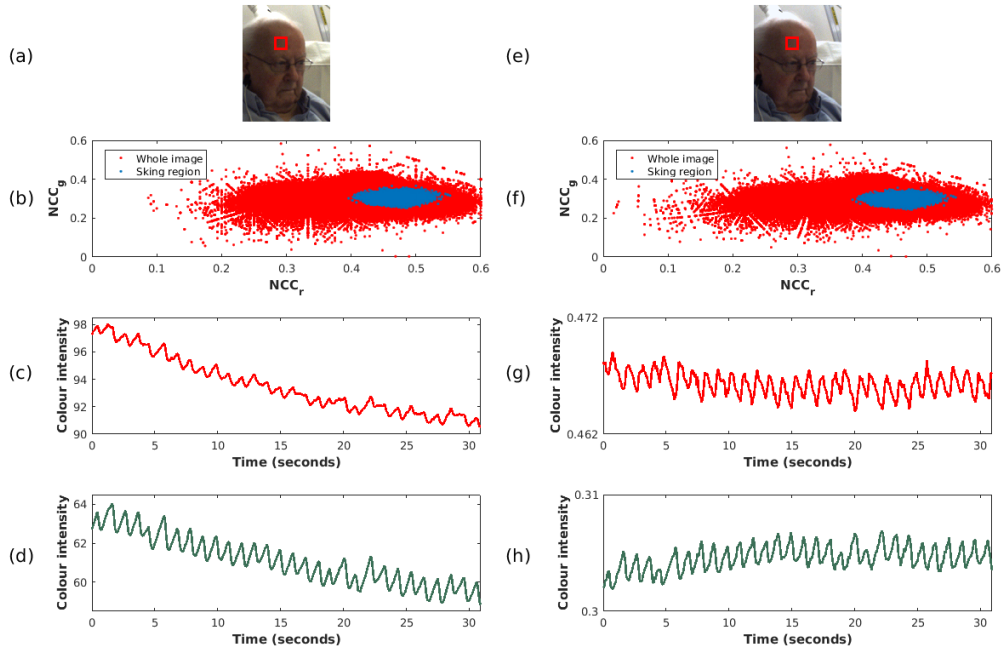


Figure 9.19: Comparison of the extraction of a 30-second pulsatile signal from a single region of interest (red square on the patient's forehead) between the original video (left column) and after colour processing (right column). (a) Original RGB image output from camera; (b) chromaticity diagram; the red colour represents the chromaticity for the whole image, whereas the blue is computed from the skin areas only. The pulsatile time series for the original red and green channels are shown in (c) and (d) respectively. (e) The image after colour processing with (f) the chromaticity diagram. The pulsatile time series for the NCC_r and NCC_g channels are shown in (g) and (h) respectively.

9.5 Identifying changes in peripheral oxygen saturation

The process of identifying changes in peripheral oxygen saturation relies on the work developed for heart rate estimation in chapter 6 to analyse the input video. Time periods for which the location of the patient is known are identified, discarding periods for which there is high activity, motion, artefacts arising from the regular interaction between the patient and the clinical staff, or the patient is awake and engaging in a conversation.

Once stable periods are identified from the video recordings, the colour-balanced video, processed by the algorithms described in section 9.4, is subdivided into 15-second windows. Longer windows will lead to additional delays in a real-time implementation and decrease the percentage of data for which the SpO_2 can be estimated as patient activity will cause more windows to be rejected. The separation between consecutive windows was chosen to be 1 second to provide a large amount of overlap between windows.

Each 15-second time window is processed as follows: A single PPGi signal is extracted from several regions of interest (ROI) positioned on the patient's face for which the signal quality is high. Subsequently, a ratio of the normalised signal amplitudes, simply labelled "R", is computed from the red and green colour channels. As a last step, the ratio "R" is calibrated with the reference SpO_2 estimates to make the comparison of absolute oxygen saturation values possible.

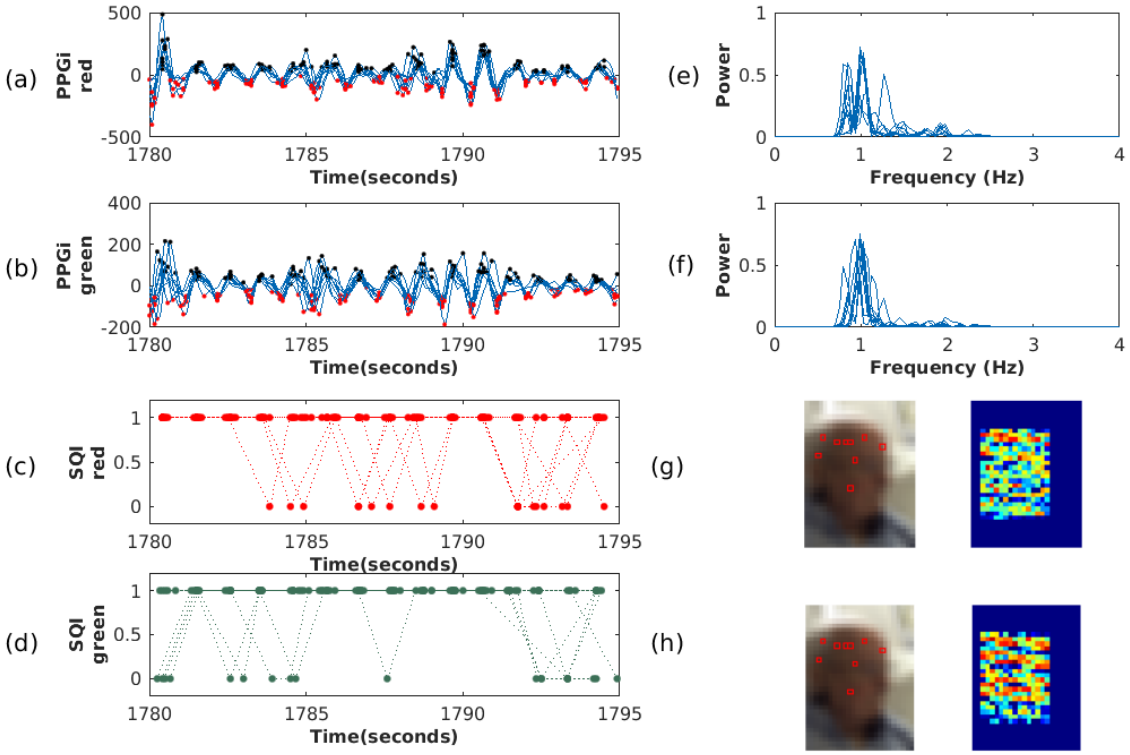


Figure 9.20: PPGi signal extraction for a 15-second window from high-quality regions of interest. The time aligned PPGi signals are shown in (a) and (b) for the red and green channels respectively. The quality of the signals is shown in (c) and (d). (e) and (f) show the frequency content for the red and green channels respectively. (g) The location of the high-quality ROIs in the image frame for the red channels together with the physiological video maps. Similarly, (h) shows where the high-quality ROIs for the green channels are located.

9.5.1 PPGi signal extraction

Using the output of the face detection and tracking algorithms described in previous chapters, a grid of multiple regions of interest (of size 150x150 pixels) is placed on the entirety of the patient's face. A PPGi signal is extracted from each region of interest for the red and green colour channels for each 15-second window. The quality of each PPGi signal is analysed using the algorithms described in chapter 6 for heart rate estimation and only the regions of interest for which the ratio of number of high-quality beats (with SQI = 1) and total number of beats is above 0.8 for both colour channels are selected. The location of the high-quality regions of interest for a sample 15-second window is shown as red squares in figures 9.20(g) and 9.20(h) for the red and green channels respectively. As expected, the areas of high-quality data are mostly located on the patient's forehead and cheeks.

Each PPGi signal from all the high-quality regions of interest is time-aligned by subtracting the sample delay as computed from the lag for which the normalised cross-correlation has the largest absolute value. A sample of the time-aligned signals is shown in figures 9.20(a) and 9.20(b) for the red and green channels respectively with their respective beat quality indices and frequency content.

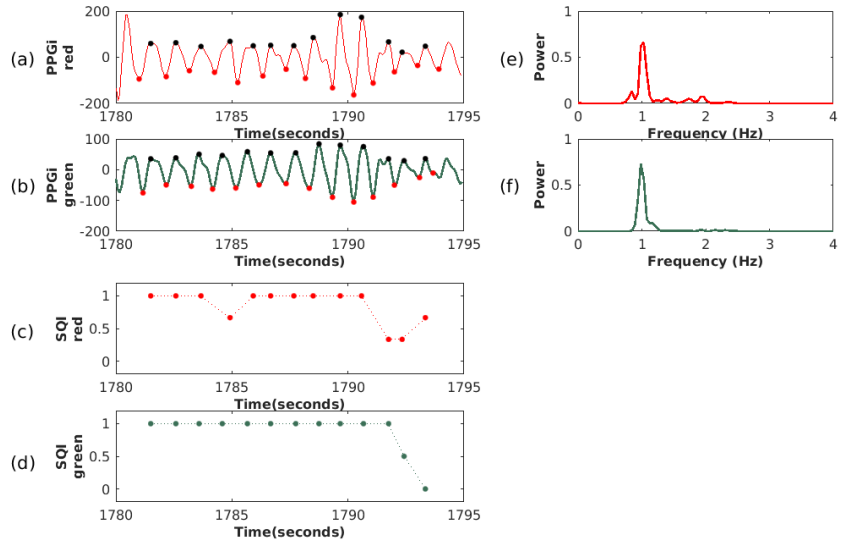


Figure 9.21: PPGi template computed for a 15-second window using the red and green colour channels, shown in (a) and (b) respectively. The quality of the signals is shown in (c) and (d). (e) and (f) show the frequency content for the red and green channels respectively.

A single PPGi signal, the template, is subsequently computed as the mean of all the time-aligned PPGi signals. Peaks and troughs are computed using the same algorithms as for heart rate estimation. Figure 9.21 shows the computed PPGi signals for the same 15-second window as in figure 9.20.

9.5.2 Computing the ratio of normalised signal amplitudes

Using the peaks and troughs detected in each PPGi signal template from the two colour channels, the means for the AC and DC components are computed from all the beats for which the SQI, as computed by the algorithms for heart rate estimation, is equal to 1.0. Figure 9.22 shows the result of computing the mean AC and DC components for a 15-second PPGi signal.

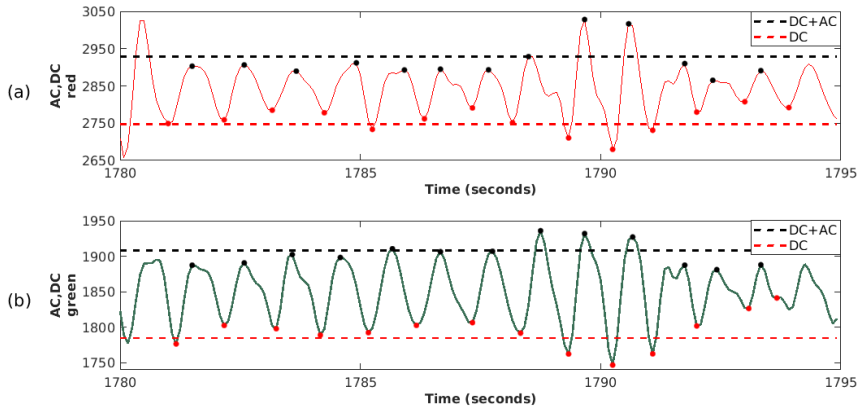


Figure 9.22: Computing the AC and DC components for a 15-second PPGi signal for the red and green colour channels, shown in (a) and (b) respectively.

The theory of pulse oximetry was initially developed by adapting the theory of oximetry, which itself is based on the Beer-Lambert law [81]. Pulse oximetry manufacturers measure a ratio, simply called "R" of the transmitted (or reflected) light at the red and infra-red wavelengths which is proportional to arterial oxygen saturation (SaO_2). To compute the ratio "R", the AC amplitudes of the signals at both wavelengths are typically scaled or normalised [81, 253, 367]. A more detailed description has already been presented in chapter 3.

Following pulse oximetry theory, the signal amplitude for both red and green channels is normalised using the following equation:

$$A_\lambda = \frac{DC_\lambda + AC_\lambda}{DC_\lambda} \tag{9.8}$$

where λ represents either the red or green colour channel. The ratio "R" of these normalised signal amplitudes between the red and green video camera signals is computed by:

$$R = \frac{log(A_{red})}{log(A_{green})} \tag{9.9}$$

9.5.3 SpO_2 calibration

Pulse oximeters are usually calibrated based on real blood samples from healthy volunteers by using proprietary regression functions to obtain optimal values between the measured ratio "R" (between the red and infra-red wavelengths) and arterial oxygen saturation (SaO_2). The arterial oxygen saturation level which is measured and displayed by a pulse oximeter is called SpO_2 [156].

To determine whether an absolute value of SpO_2 can be obtained from the video camera data, we use a linear model similar to that proposed by [81, 315] between the ratio "R" computed from the camera signals and the reference SpO_2 provided by the pulse oximeters, as computed in section 9.3:

$$SpO2 = \lambda_0 + \lambda_1 R \tag{9.10}$$

Figure 9.23 shows the regression result for a 5-minute segment from a 54-year old male recorded while he was asleep during the dialysis treatment.

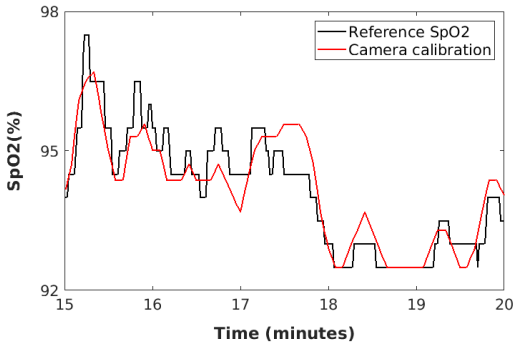


Figure 9.23: Calibration from the ratio of normalised signal amplitudes "R" computed from the camera (red trace) and the reference SpO_2 provided by the pulse oximeters (black trace).

9.6 Results

From the 104 video sessions recorded in total, twelve sessions were discarded because the recording from at least one pulse oximeter was interrupted due to patient discomfort, the patient required medical intervention or there was an equipment malfunction. 8 other sessions were discarded because there was malfunction of the camera recording equipment or the recording was not complete. A further 23 sessions were rejected using the video analysis algorithms described in section 6.4, ruling out video sessions with long periods for which there is high patient activity, motion artefacts arising from the interaction between the patient and the clinical staff, or the patients are awake and engaging in conversation (see chapter 6).

Section 9.3 described circumstances for which the SpO_2 estimates from the two reference pulse oximeters varied greatly depending on the patient's physiology and diagnoses. Therefore, sessions from patients diagnosed with diabetes or under certain medications (such as angiotensin II receptor antagonist (A2RB) or beta-blockers) were also discarded.

From the resulting dialysis sessions, 18 segments, each with a duration of 5 minutes, were manually selected using the following criteria: The SpO_2 estimates provided by the two pulse oximeters differed by less than 2%; the computed reference SpO_2 decreased by more than 5% across the SpO_2 physiological range (80% to 100%); patients were relatively still, as computed by the activity/motion algorithms described in section 6.4; and no sudden changes in illumination occurred, as computed by the algorithms in section 6.6.1.

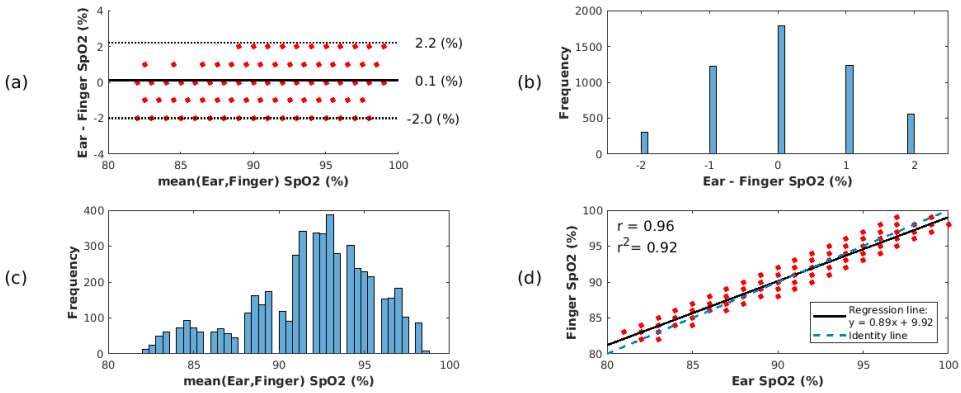


Figure 9.24: Agreement between the SpO_2 values provided by the two pulse oximeters for the eighteen 5-minute segments used to evaluate the camera-derived estimates, comprising a total recording time of 90 minutes. (a) The Bland-Altman plot presents a bias close to zero. (b) The distribution of the differences between the two pulse oximeters. (c) The distribution of the mean values across the SpO_2 physiological range from 80% to 100%. (d) The plot shows a good correlation between the two devices with a correlation coefficient of 0.96.

Figure 9.24 shows in more detail the agreement between the SpO_2 estimates from the two pulse oximeters for the eighteen 5-minute segments. The Bland-Altman plot shows a bias close to zero, with narrow differences across the SpO_2 physiological range, showing a good agreement between the two measurements. The values from the two devices are highly correlated with a positive correlation coefficient of 0.96.

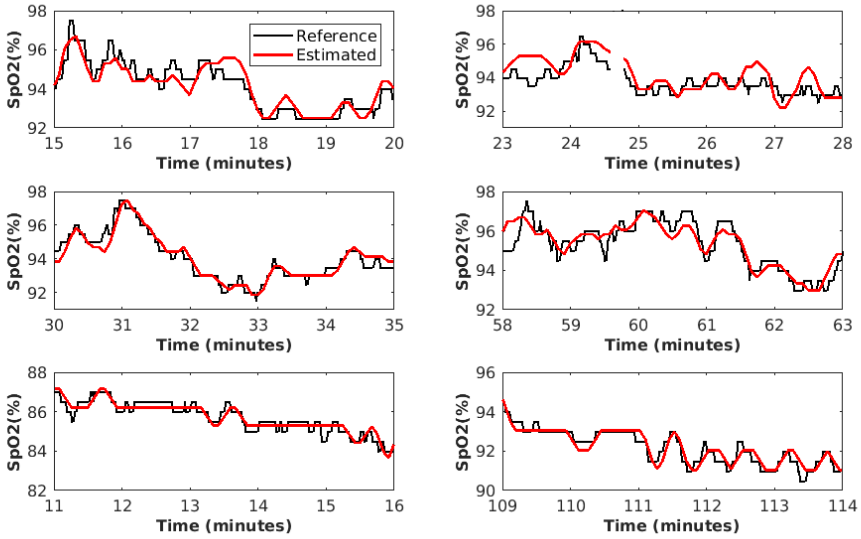


Figure 9.25: Comparison between the SpO_2 estimates from the camera (in red) and the reference SpO_2 values (in black) for six (out of the eighteen) 5-minute sample desaturation segments.

Using equation (9.10), each of the eighteen 5-minute segments was calibrated individually with the reference SpO_2 values computed from the two pulse oximeters. Figure 9.25 shows a sample of six 5-minute periods where a 5% decrease in the reference SpO_2 occurred. The reference SpO_2 values are shown in black, whereas the camera-derived estimates are shown in red.

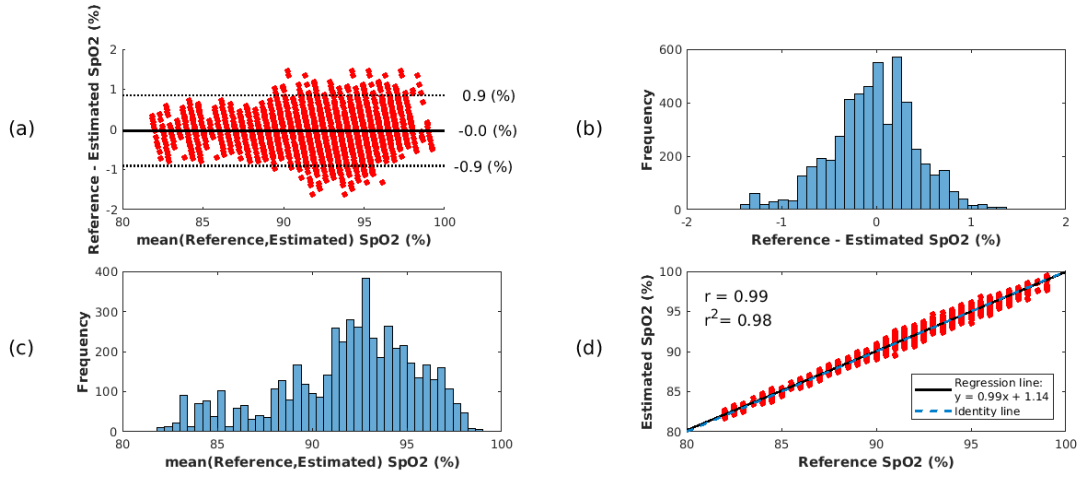


Figure 9.26: Agreement between the reference SpO_2 values and the camera-derived estimates for the eighteen 5-minute segments used to evaluate the camera-derived estimates, comprising a total recording time of 90 minutes. (a) The Bland-Altman plot presents no sensor bias. (b) The distribution of the differences between the camera-derived estimates and the reference SpO_2. (c) The distribution of the mean values across the SpO_2 physiological range from 80% to 100%. (d) The plot shows a good correlation between the camera-derived estimates and the reference values, with a correlation coefficient of 0.99.

Figure 9.26 shows the agreement between the SpO_2 estimates derived from the camera signals and the reference SpO_2 (as computed by the algorithms described in section 9.3) for the eighteen 5-minute segments. The Bland-Altman plot shows no sensor bias, with narrow differences across the SpO_2 physiological range, showing a good agreement between the two reference SpO_2 values and the camera-derived estimates. The two sets of values are highly correlated with a positive correlation coefficient of 0.99. Table 9.2 presents the errors between the two pulse oximeters and the camera estimates for all the selected 5-minute segments, comprising a total recording time of 90 minutes.

Table 9.2: Comparison of the SpO_2 estimates derived from the camera signals and the error between the pulse oximeters for eighteen 5-minute segments, comprising a total recording time of 90 minutes.

Signals	**MAE** (%)	**MAD** (%)
Ear vs Finger SpO_2	0.8	0.7
Camera vs Reference SpO_2	0.4	0.3

9.7 Discussion

The accuracy of the SpO_2 estimates derived from the video camera signals is comparable to that of the two standard pulse oximeters used in this study for the selected eighteen 5-minute segments, after individual calibration. The low error in the estimates is obtained by minimising the effects of external factors (such as patient motion, sudden changes in external illumination), ruling out sessions from patients with abnormal physiology, and by calibrating the ratio of ratios "R" (as computed by equation (9.9) from the red and green colour channels) with the SpO_2 estimates provided by the pulse oximeters for each 5-minute segment individually.

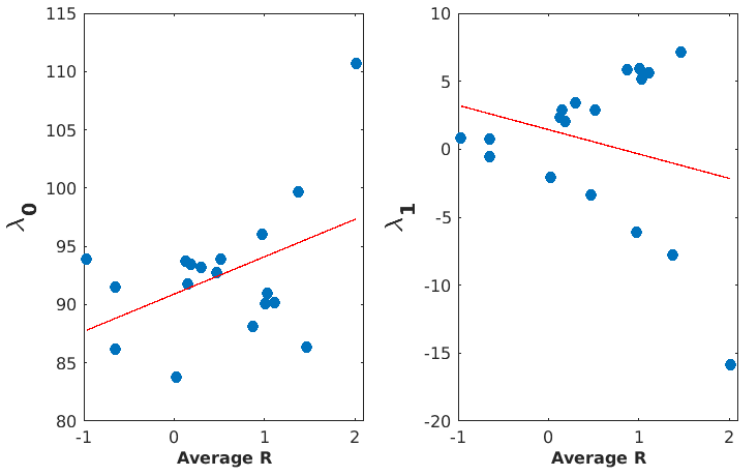


Figure 9.27: Relationship between the average ratio "R" from the normalised signal amplitudes from the red and green colour channels and the linear regression parameters λ_0 and λ_1 for all eighteen sessions. The line of best fit is shown in red.

The dependence of the linear regression parameters λ_0 and λ_1 in equation (9.10) for all the 18 sessions under study is shown in figure 9.27. Although no strong relationship exists between the average ratio "R" and the regression parameters for each session, two groups are clearly visible for λ_1, one having a positive and the other a negative relationship. Note that some values of R are negative because the colour pre-processing algorithms introduced in this chapter can result in negative values.

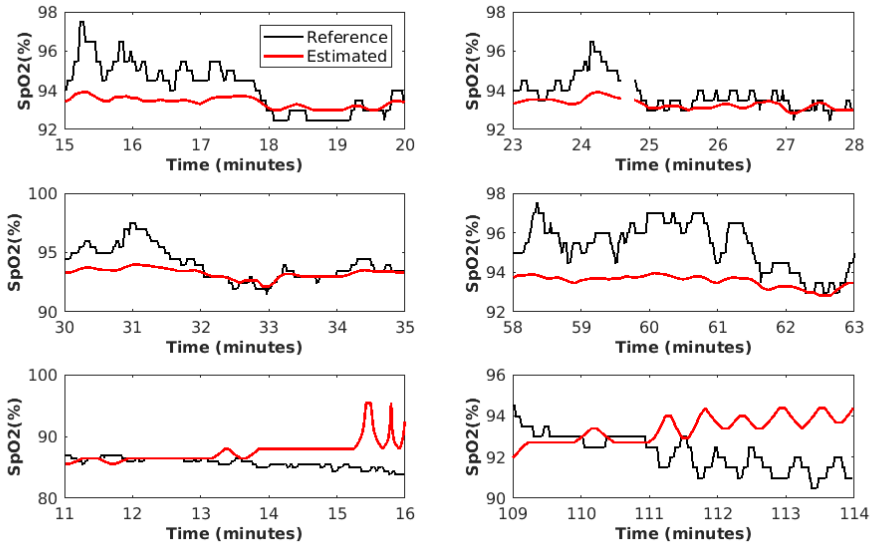


Figure 9.28: Comparison between the SpO_2 estimates from the camera (in red) and the reference SpO_2 values (in black) for six 5-minute sample desaturation segments. The linear regression parameters were computed from the remaining 12 sessions independently.

When calibrating the camera-derived SpO_2 using all the data from 12 sessions out of 18 (as opposed to calibrating each session individually) and evaluating on the remaining 6, two relationships are found, as shown in figure 9.28. For the 4 sessions at the top of the figure, the camera-derived SpO_2 estimates follow a similar trend as the reference SpO_2 values, although at a different scale. For the two sessions at the bottom of the figure, the camera-derived SpO_2 estimates follow a trend opposite to the trend in the reference values.

Figure 9.29 shows the relationship of the linear regression parameters λ_0 and λ_1 with the average ratio "R" for the sessions for which the camera-derived SpO_2 estimates follow a similar trend to that of the reference SpO_2 values. The strong relationships between the parameters can indicate the possibility of quantifying, to some degree of accuracy, changes in SpO_2 for periods of significant desaturation under the conditions of the dialysis study.

The issue of SpO_2 calibration for the entire video data set remains a challenge to be solved. Angle changes caused by subject motion, as well as changes in the spectrum of the incident illumination, will affect the amplitude of the AC signal for each of the two colour channels differently, resulting in inaccuracies in the SpO_2 estimates from the camera. Most clinical scenarios however, only require the detection of a decrease in SpO_2 values rather than an accurate knowledge of the absolute value of oxygen saturation at any one time.

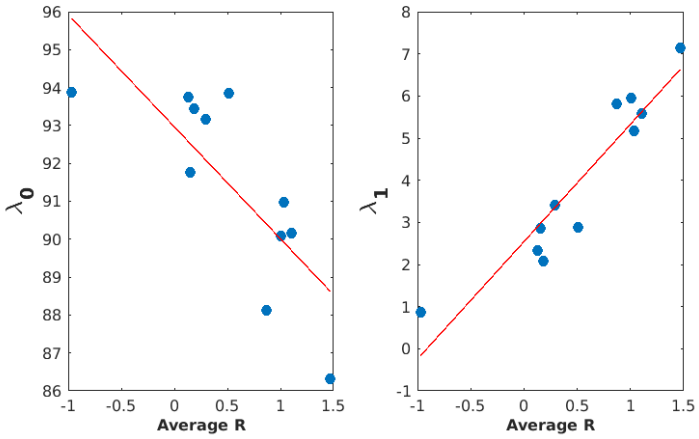


Figure 9.29: Relationship between the average ratio "R" from the normalised signal amplitudes from the red and green colour channels and the linear regression parameters λ_0 and λ_1 for sessions that the camera-derived SpO_2 follow a similar trend to the reference SpO_2. The line of best fit is shown in red.

9.8 Conclusion

In this chapter, we have described the results from algorithms for identifying changes in peripheral oxygen saturation using a video camera and ambient light for a small number of manually selected time segments for which a significant desaturation occurred. The SpO_2 algorithms presented in this chapter, build on the work carried for heart rate and respiratory rate estimation described in chapters 6 and 8 respectively.

The validation of any new method of deriving oxygen saturation is hampered by the difficulties associated with trying to obtain SpO_2 values over the most clinically relevant range, from 80% to 95%. In the dialysis population for this study, there are only a handful of cases of significant desaturation. Nevertheless, the results indicate that the proposed analysis of the video camera signals using two colour channels, and after colour processing, could provide an indication of oxygen desaturation, provided that there are no substantial changes in subject orientation or in scene illumination.

Chapter 10

Real-time implementation

10.1 Introduction

The algorithms for vital-sign estimation from video cameras presented in the previous chapters were developed using the MATLAB programming language and were designed to retrospectively analyse the video data recorded from the dialysis study. This chapter discusses the extensions added to the data acquisition software developed to record video and pulse oximetry data (see chapter 5) to implement the algorithms for heart rate and respiratory rate estimation proposed in chapters 6 and 8 in a real-time environment.

A real-time implementation of the proposed algorithms presents several challenges. The accuracy of the heart rate and respiratory rate algorithms depends not only on the ability to accurately locate the patient within the image frame, but also on the extraction several high-quality photoplethysmographic imaging (PPGi) signals from the different regions of interest across the patient's face and upper torso.

As discussed in chapter 4, a video camera's imaging pipeline includes several components, such as: the imaging sensor, lenses, analogue-to-digital conversion circuit, image post-processing algorithms, data format conversion and storage of the image frames in a typical video file. This pipeline is usually supplied by several hardware and software vendors, contains several proprietary algorithms and the file formats often use proprietary lossy compression codecs. These factors can corrupt the extracted PPGi signal. Therefore, a vital-sign monitoring system ideally requires the acquisition of the unprocessed raw image data with minimum alteration and the ability to control the imaging sensor recording capabilities such as sensor gain, shutter exposure time, dynamic range or brightness levels.

10.2 Software architecture

The vital-sign estimation software extends the data acquisition system described in chapter 5 adding modules for the real-time implementation of the image and signal processing algorithms for heart rate and respiratory rate estimation proposed in this thesis, as shown in figure 10.1. The user interface is extended to support not only workstations, but also mobile devices.

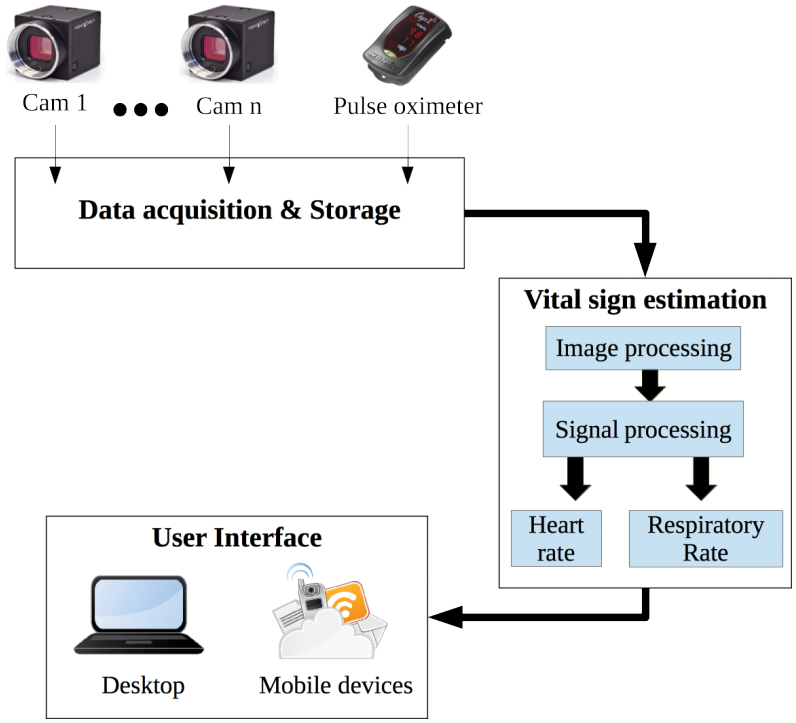


Figure 10.1: Overall software architecture for the real-time vital-sign estimation system.

The vital-sign estimation software was mainly written in portable the C/C++ programming languages, compliant with the latest "C11" and "C++14" ANSI standards. Similarly to the data acquisition system, the vital-sign estimation package was developed following the safety guidelines presented in section 5.4.2; so that the system, with minor modifications, may be submitted for medical software certification in the future.

10.2.1 Supported platforms

As the software was written in portable programming languages, following well-known accepted development processes and guidelines, the core of the modules is able to run on the main three desktop operating systems: Microsoft Windows 7; Apple OS X 10.10 or above; and several Linux distributions running kernels 4.2 or above, such as Fedora or Ubuntu. With some modifications to the code that interacts with the operating systems and cameras, most parts of the software are also able to run on mobile devices such as the iPhone and iPad running iOS 9 or above.

10.2.2 Programming languages and external libraries used

The software was built on top of some well-known and widely-used libraries. Table 10.1 lists the most relevant programs and libraries used. The core of the software infrastructure was developed using the C++ programming language. Code that directly interacts with hardware, accesses low-level system resources from the operating system or performs critical tasks such as memory allocation or thread scheduling, was developed using the C programming language.

Three major C/C++ compilers were used to validate the code: GNU GCC, LLVM/Clang and Microsoft Visual C++. The Qt framework was used as the main user interface library for interacting with the user and reporting the patient's physiological status on screen. On mobile systems, Apple's native Cocoa Touch user interface libraries were used instead. Rendering of the video frames on screen and time series plots was done directly on the GPU using the cross-platform OpenGL libraries.

Table 10.1: Compilers and third party dependency libraries required to build the real-time vital-sign estimation system

Software	**Linux**	**Microsoft Windows**	**Apple OS X**	**Apple iOS**
Operating system	Kernel 4.2	Windows 7	OS X 10.10/10.11	iOS 9
Build & Test	CMake 3.5			
Compiler	GNU GCC 5.4	Visual C++ 2015	Apple LLVM/Clang 7.3	
C++ infrastructure	Boost 1.58			
User interface	Qt framework 5.0			Apple native
Graphics rendering	OpenGL Core Profile 3.2			
Media	FFMpeg 1.2			
Imaging framework	OpenCV 2.4.11			
Graphics computing	Nvidia CUDA 7.0			Apple native
Threading	Intel Threading Building Blocks (TBB) 4.3			Apple native
Signal processing	Intel Math Kernel Library (MKL) Intel Integrated Performance Primitives (IPP)			Apple native

The image processing algorithms were developed using the data structures and facilities provided by the OpenCV framework. When performance was an issue, due to the large raw uncompressed images from the video cameras, the algorithms were extended to compute directly on the CPU using the Nvidia CUDA framework.

Similarly to the data acquisition software, the vital-sign estimation software requires multiple tasks to run in parallel. As such, it was built on top of the Intel Threading Building Blocks library, dynamically adapting to the different characteristics of modern multi-core CPUs.

The signal processing algorithms described in previous chapters were designed in MATLAB; for real-time use, they were ported and implemented using the C programming language on top of Intel's Math Kernel Library (MKL) and Integrated Performance Primitives (IPP).

10.3 Extensions to the data acquisition system

Due to the variety of hardware found on modern computers and mobile devices, the data acquisition system was extended to capture data not only from the reference medical devices used in the dialysis study, but also from several video cameras manufacturers and pulse oximeter models.

Custom user interfaces were developed to suit the needs of the nurses or other clinical staff who were in charge of the data collection process. Operating system drivers and a common hardware abstraction layer were also created to follow the design guidelines for a data collection system.

10.3.1 Pulse oximeter capture

Although the dialysis study described in chapter 5 used the Nonin pulse oximeter model 4100, the vital-sign estimation software was extended to communicate with Nonin pulse oximeters models 9560 and 3150. These pulse oximeters communicate with a PC or laptop using the serial port profile (SPP) over a Bluetooth connection. The advantage of these new pulse oximeters is that the estimated heart rate and SpO_2 are shown in a display on the device itself, helping in the rapid comparison with the estimates from the video camera.

10.3.2 Video capture

Video cameras supplied by Original Equipment Manufacturers (OEM) such as Point Grey, are often supplied with their own specific software and an Application Programming Interface (API) to capture and control the cameras. As one of our goals was to estimate vital signs using off-the-shelf cameras across multiple operating systems, new cross-platform routines were developed using the native operating system libraries for Linux, OS X and iOS to capture video from generic cameras such as external USB webcams or the embedded cameras in mobile devices.

The Video4Linux 2 (V4L2) video capture API was used to record video from generic cameras in Linux. The real-time estimation software requires the manual control of brightness, white balance, pixel format, compression and other more advanced camera settings. As such, only two models of webcams are supported: the Logitech HD Pro Webcam C910 and HD Pro Webcam C920.

Camera recording in Apple's OS X was developed using the AV Foundation framework. It provides native interface libraries to access, at a low level, audiovisual data and tools for the capture and playback of video.

Video capture in Microsoft Windows was developed using the cross-platform OpenCV library interface. Image capture properties such as exposure time, white balance, pixel format or compression were not available. Hence, the robustness and accuracy of vital-sign estimation are likely to be reduced.

10.4 User interface

Although the visualisation component of the data collection system was done in OpenGL, the Qt framework was chosen to develop the user interface (UI) elements that the research nurses used in the study, such as buttons for starting/stopping the collection process and customised windows to configure the devices. Qt is a cross-platform application and user interface framework developed in C++. It targets various desktop, embedded and mobile operating systems with a single code base.

10.4.1 Video rendering with OpenGL

OpenGL is a cross-platform application programming interface for rendering 2D and 3D computer graphics. It was originally developed by Silicon Graphics Inc. in the early 1990s. Currently, OpenGL is managed by the Khronos Group, a non-profit technology consortium.

The support for multiple platforms is of high importance for the real-time vital-sign estimation system. The libraries use the GPU for processing data, leaving the main CPU free for other important routines, such as saving the video frames to disk. The OpenGL 3.2 core profile was used as the reference standard for use with the real-time system.

10.5 Real-time vital-sign estimation

This section describes the design considerations and steps taken to implement a real-time vital-sign estimation system running on high-end workstations and laptops. Section 10.6 discusses the implementation of the software in mobile devices such as the iPhone or iPad.



Figure 10.2: Real-time vital-sign estimation running on an Apple Laptop. (a) heart rate reported by the Nonin pulse oximeter, (b) respiratory rate set by the artificial metronome, (c) progress bar representing an artificial metronome - subjects were asked to breathe at the displayed frequency, (d) heart rate and respiratory rate estimated from the pulse oximeter PPG, (e) heart rate and respiratory rate reported from processing the video camera data in real-time.

Figure 10.2 shows the main user interface of the vital-sign estimation application running on an Apple Laptop. The reference heart rate from the Nonin pulse oximeter is shown in figure 10.2(a). The reference respiratory rate, shown in figure 10.2(b), is given by the artificial metronome in figure 10.2(c). The artificial metronome is a blue bar moving from left to right at the bottom of the screen. The subject adjusts his or her respiratory rate to be in synchrony with the period set by the metronome. The estimated heart rate and respiratory rate from the pulse oximeter PPG are shown in figure 10.2(d). The estimated vital signs from the camera are shown in figure 10.2(e).

10.5.1 Image processing

OpenCV is an open source multi-platform computer vision library written in the C/C++ programming languages designed for computational efficiency with a strong focus on real-time applications [38]. OpenCV provides low-level image-processing infrastructure and basic high-level algorithms such as colour conversion and image filtering [266]. Of particular interest is that OpenCV includes a GPU interface for software written using Nvidia CUDA. CUDA is a parallel computing platform and programming model introduced in 2006 by NVIDIA. It is widely used in applications from physics, manufacturing, finance, and other computationally intense fields.

Traditionally, CPUs are optimised for fast single-thread execution cores, designed to execute one or two threads concurrently. Graphics processing units (GPU), on the other hand, are optimised for high multi-thread data-parallel processing. Transferring input data and intermediary computation from the CPU to GPU memory (and vice-versa) at high frame rates presents a problem. As the raw uncompressed images from the video cameras are large, the algorithms that process the incoming images (such as face detection or the extraction of PPGi from the multiple regions of interests) benefit from maintaining the data in GPU memory, so they are available for the subsequent processing steps. Keeping the data in the GPU memory also helps later processing stages for rendering the results on screen through OpenGL.

10.5.2 Signal processing algorithms

To migrate the signal processing algorithms described in chapters 6 and 8 from Matlab to a real-time environment required the use of two major libraries from Intel: Math Kernel Library (MKL) and Integrated Performance Primitives (IPP). These libraries contain a set of low-level routines providing basic vectorised and threaded Linear Algebra, Fast Fourier Transforms (FFT), Vector Math and Statistics functions.

Most of the vital-sign estimation algorithms were implemented using the C programming language on top of Intel's libraries. A small proportion of the code that performed some linear algebra computations required the use of the FORTRAN programming language for a faster interaction with the Basic Linear Algebra Subprograms (BLAS) in the MKL library.

10.6 Real-time vital-sign estimation on mobile devices

The software described in section 10.5 was developed for high-end workstations with a powerful graphics processing unit. Non-contact vital-sign estimation using mobile devices presents benefits that could disrupt the delivery of healthcare in developed and developing countries. Therefore, the real-time system was extended to support mobile devices from Apple such as the iPad and iPhone running on the iOS operating system, as seen in figure 10.3.

Apple development tools provided an easy pathway for porting the native code developed in C/C++ programming languages and OpenGL video rendering for the high-end workstation described in section 10.5. Porting to other platforms such as Google's Android operating system, will require further investigation and validation.

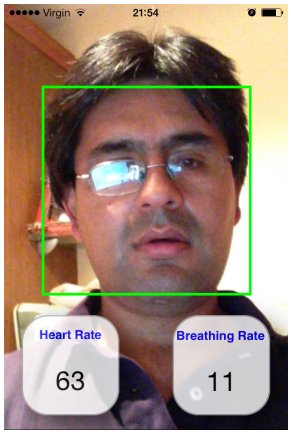


Figure 10.3: Real-time heart rate and respiratory rate estimation running on a mobile phone.

10.6.1 Image and signal processing

In the current mobile devices, there is no access to the hundreds of GPU cores available in a desktop system. Therefore, the real-time system developed for the Apple iOS operating system makes use of the Accelerate Framework for implementing the algorithms for heart rate and respiratory rate estimation described in chapters 6 and 8 respectively. The Accelerate framework is a set of libraries containing high-performance vector-accelerated libraries providing similar functionality to Intel's MKL and IPP libraries. As such, the difficulty of porting the image and signal processing algorithms to this framework was low.

10.7 Conclusion

This chapter described the design and implementation of real-time software for the estimation of vital-signs developed to run on high-end Linux workstations. The software was developed following the software life-cycle processes outlined in the "BS EN 62304:2006" British standard, with the goal that in the future, with some modification, the system can be submitted for medical software certification.

The use of software development guidelines and industry best practices was an important factor that allowed the system to acquire vital signals reliably. The standards and guidelines also helped to develop portable code that runs on different platforms and form factors such as workstations and mobile devices.

Non-contact real-time heart rate and respiratory rate estimation is possible using specialised video cameras in high-end workstations and modern laptops. As computing power increases and higher quality digital sensors become affordable, the potential to use off-the-shelf video cameras increases. In addition, the design and implementation of real-time software for heart rate and respiratory rate estimation for mobile devices, such as the Apple iPad and iPhone, will become easier to implement as their computing power increase and the performance of their cameras improve.

Chapter 11

Conclusions and future work

11.1 Introduction

Current monitoring systems available to track changes in the standard vital signs (heart rate, respiratory rate, peripheral oxygen saturation (SpO_2), blood pressure and temperature) in clinical practice involve sensors that require direct contact with the patients. Chapters 1 and 2 discussed the clinical need and advantages of non-contact vital-sign monitoring over these traditional methods together with description of some technologies being proposed by the research community.

Most of the current work in video-based non-contact vital-sign monitoring has so far been performed over short time periods of typically up to a couple of minutes per recording session, under tightly-controlled conditions with relatively still and healthy subjects. The research carried out for this thesis introduced technology to accurately monitor, unobtrusively without interfering with the day-to-day clinical care, the heart rate and respiratory rate of patients undergoing haemodialysis treatment in the Renal Unit of the Churchill Hospital in Oxford using a standard colour camera and ambient light. Changes in peripheral oxygen saturation were identified from the video camera as preliminary work towards the non-contact monitoring of SpO_2. A total of 104 dialysis sessions were recorded from 40 patients during the course of one year, comprising an overall video recording time of approximately 370 hours.

This chapter starts, in section 11.2, with a discussion of the challenges faced and solutions proposed during the design of the clinical study analysed in this thesis. Section 11.3 reviews the analysis of the reference physiological data provided by the medical instrumentation for comparison with the estimates derived from the video camera signals. A summary of the important results from the proposed video-based vital-sign estimation algorithms is presented in section 11.4 with a discussion of the major challenges and limitations found during the analysis of the recorded data. Section 11.5 discusses future work and presents new potential research avenues in the field of non-contact sensing. Conclusions and final remarks are presented in section 11.6.

11.2 Clinical study

One of the goals of this thesis was the development of vital-sign estimation algorithms for patients in a real-world hospital setting. The advantage of recording videos from patients undergoing haemodialysis treatment is that, in a relatively short amount of time (approximately 4 hours), these patients experience a wide range of physiological values (see chapter 5). This therefore helps to validate the proposed vital-sign estimation algorithms. Nevertheless, the clinical study presented several challenges to overcome, mainly with the requirements for the recording and storage of large quantities of video data, and the time synchronisation across all the recording equipment.

11.2.1 Video recording

At the time of the design of the clinical study, there was no recording system that provided the level of control needed of the video camera or satisfied the requirements of the data acquisition process outlined in chapters 6 through 9. To this end, major effort was devoted to the development of a new software package and the validation of a hardware system to record high volumes of video data. The system also allows the recording of data from other medical devices such as pulse oximeters (see chapter 5).

The data acquisition software that interacts with medical devices or the video camera was developed using high-performance programming languages and following international standards that address safety and best practices on critical systems across several industries. The feedback from the clinical staff also enabled the development of customised user interfaces to support not only the clinical study for this thesis, but the execution of other studies in our research group.

11.2.2 Data storage

Issues regarding the acquisition, storage and transfer of data were carefully considered in our study design. The high-resolution video camera used in the study recorded approximately 3 terabytes of data for each of the 104 dialysis sessions, requiring a total of more than 300 terabytes of storage space for a single copy of the recorded videos without considering backups.

The transfer of the data recorded from the hospital to the university's research centre for later analysis and the data integrity for the long-term archival of the database were major challenges. The data from each dialysis session was individually copied to external encrypted hard disks and physically transferred on a regular basis from the hospital ward to a research storage cluster within the university, where multiple backup copies of each session were stored. Transferring the data electronically would have been logistically impractical.

Patient privacy was a major concern, as our study was the first of its kind for the Oxford University Hospitals (OUH) National Health Service (NHS) Foundation Trust. Therefore, methods by which appropriately anonymised data was collected and stored securely were implemented. The study was compliant with the Data Protection Act and adhered to the National Health Service (NHS) code of confidentiality.

11.2.3 Time synchronisation

The clinical study recorded data from a heterogeneous group of devices and technologies, as described in section 5.4. To compare the output of the vital-sign estimation algorithms from the video camera with the estimates produced by the reference devices, reliable timestamps are necessary across all recording equipment. Custom software was developed using the network time protocol (NTP) [225] to automatically synchronise the data recorded from all the reference devices and the video camera at the beginning of each dialysis session.

11.3 Computing reference signals

The recording of patient data for a clinical study within a hospital presents many challenges, especially, the recording of the reference physiological signals (heart rate, respiratory rate and SpO_2 in the case of this thesis) against which the camera estimates are compared and evaluated.

Most of the studies reported in the literature use only one data source or reference device as the gold standard, for example the use of one pulse oximeter to record a single set of heart rate or SpO_2 values, without an extensive analysis of the quality and accuracy of the reference device signals. With this in mind, our clinical study was designed to have at least two reference data sources for heart rate, respiratory rate and SpO_2.

11.3.1 Dual PPG monitoring during haemodialysis

The reference heart rate and SpO_2 estimates were provided by two transmission-mode pulse oximeters, one located on the finger and other on the ear lobe (see chapter 5). Using the values reported from more than one pulse oximeter introduced some physiological and non-physiological factors to be considered. The characteristics of peripheral circulation at various body sites can result in variations of the recorded PPG waveform and therefore of the computed heart rate and SpO_2 values, as discussed in sections 6.3 and 9.3 respectively.

The haemodialysis population in our study was mostly comprised of the elderly, with an average age of 64.7 years (see table 5.4). Most of the patients were diagnosed with illnesses such as diabetes, hypertension and other autoimmune diseases. Patients were also given medications such as calcium-channel blockers to regulate blood pressure by relaxing and widening blood vessels in the arterial walls. Other medications included Amlodipine, an angiotensin II receptor antagonists used to help regulate blood pressure and heart rate to slow the progression of kidney disease. These, and other medications, potentially affected the patients' physiology, and as a consequence, the values recorded by the pulse oximeters.

Although the pulse oximeter sensors used in our study were provided by the same manufacturer (Nonin), there were other non-physiological factors that affected the timing and values of the two recorded PPG waveforms and the estimates derived from them. Differences arose from the recording set-up, processing context, versions of the installed software, different connection technologies or simply the rate at which the values were reported.

For twelve dialysis sessions out of the total of 104, the recording from at least one pulse oximeter was interrupted due to patient discomfort, medical intervention, to preserver the privacy of visiting family members, equipment malfunction or other external factors. Therefore, there was simultaneous recordings with both pulse oximeters for only 92 sessions, giving a total recording time of 304.1 hours.

The heart rate values, as provided by the pulse oximeters, occasionally presented large errors of more than 50 beats/min as shown in figure 6.4. From the 304.1 hours for which simultaneous pulse oximetry recordings were available, only 79.7% (approximately 242.7 hours) of the values differed by no more than 5 beats/min (see section 6.3), as recommended by the "ANSI/AAMI EC13:2002 Cardiac monitors, heart rate meters, and alarms" standard [13].

The pulse oximeter attached to the ear often underestimated the SpO_2 values with respect to the pulse oximeter attached to the finger. From the 304.1 hours of simultaneous recording time, only 49.8% (approximately 151.5 hours) of the values differed by no more than 2% (see section 9.3), as recommended by the published literature and the "ISO 80601 Particular requirements for basic safety and essential performance of pulse oximeter equipment" standard [143].

A direct comparison (on a sample-by-sample basis) of the heart rate or SpO_2 values between the video camera and each of the two pulse oximeters can potentially be affected not only by physiological factors but also by the recording set-up and internal processing context for each manufacturer. Therefore, new reference heart rate and SpO_2 estimates were computed from the average of the reported values from the two pulse oximeters during time periods for which there was good agreement (values that vary by no more than 5 beats/min for heart rate and 2% for SpO_2). These new "ground truth" reference heart rate and SpO_2 values were used to evaluate the performance of the estimates derived from the video camera presented in chapters 6 and 9 for heart rate and SpO_2 respectively.

11.3.2 Continuous monitoring of respiratory rate

As described in section 5.4, two reference respiratory rates were recorded using a harness worn by the patient around the thorax: the first one computed from an elastic thoracic band that measures the patient's chest wall expansion and contraction as the patient breathes in and out, and the second using the electrocardiogram (ECG).

The respiration signal recorded from the expansion and contraction of a thoracic chest band was inherently noisy. The respiratory rate estimates, as provided by the manufacturer, had errors of more than 20 breaths/min when compared to the estimates computed from the ECG (see figure 7.4). There were several factors that explained the sources of disagreement between the respiratory rate estimates (more details in section 6.4 and section 7.2). Most of the errors originated from values outside the adult physiological range, taken to be as 6 to 42 breaths/min. The chest harness was manufactured in a limited number of sizes and, therefore, a proper fit was not always possible for each patient. Motion artefacts changed the location of the chest harness, disturbing the measurements. There were cases that required medical intervention, affecting negatively the normal functioning of the devices.

Given these errors, a direct comparison between the camera-derived respiratory rate and the values provided by the manufacturer could be misleading. Therefore, two new respiratory rates were computed using state-of-the-art methods for analysing the two source signals: the ECG and the chest band. The final reference respiratory rate was computed as the mean from the new estimates from the ECG and chest band for which the data was of good quality and did not vary by more than 2 breaths/min, as recommended in the literature. From the 304.1 hours of simultaneous recording time, only 32.3% (approximately 101.4 hours) of the values differed by no more than 2 breaths/min (see section 7.2), consistent with the literature where valid respiratory rates have been found to be valid between 37% to 61% of the time [149, 177].

11.4 Non-contact vital-sign estimation summary

The main objective of the thesis was to investigate how accurately vital signs such as heart rate and respiratory rate can be estimated and changes in SpO_2 identified using a standard colour camera and ambient light. A second major objective was to obtain calibration data for vital-sign estimation algorithms from a real-world hospital setting.

11.4.1 Heart rate

The algorithms proposed in chapter 6 start by analysing the input video to identify the time periods for which the location of the patient is known, discarding periods for which there is high activity or which are affected by motion artefacts. As one of the main requirements of the clinical study was to be unobtrusive and not affect regular patient care, this process was particularly challenging.

As opposed to an intensive care unit (ICU), where patients are often sedated and resting, patients during haemodialysis treatment were usually awake and engaged in conversation with nurses, visiting family members or used media devices such as mobile phones. In several cases, the video camera was moved away from the patient to ease access for a medical intervention or to preserve the privacy of patients or their family who wished not to appear in the video frame. Some patients withdrew from the study due to personal reasons.

Following the criteria described in section 6.4.4, 23 recording sessions were rejected. The resulting number of dialysis sessions with valid video data and simultaneous recordings from both pulse oximeters was 61, comprising a total video length of 219.8 hours, approximately 60% of the total number of hours recorded originally.

Even though the face detection, tracking and activity algorithms described in section 6.4 can potentially be applied to a diverse set of environments, they were developed and trained to deal with the challenges imposed by videos recorded in the Renal Unit. Further research and validation is needed to obtain similar performance results if the video recording scenario changes, for example by recording in other hospital units or even at home.

As most of the video recordings from the clinical data set were recorded under similar conditions (patients' upper torso at a similar distance from the camera), a common threshold was often for the activity index algorithm to judge a time period as high activity and, therefore, to not produce a heart rate estimate. An adaptive approach could potentially improve the results.

Once stable periods were identified from the video recordings, a set of photoplethysmographic imaging (PPGi) signals were extracted from multiple regions of interest across the patient's face. Heart rate was estimated from each region of interest individually, using several methods to assess the quality of the PPGi signals. To compute the overall heart rate estimate, the estimates from each region of interest were combined using three data fusion techniques.

During video recording, modulation from the artificial lighting equipment (fluorescent lights or bed-side lamps) in the Renal Unit was identified, resulting in the light sources flickering at twice the frequency of the electrical supply (50 Hz in the UK). As the video in the clinical study was recorded at 12 Hz, aliasing components (artefacts) from these light sources were often found in the PPGi signals extracted from the video data. The exact frequencies from this aliasing process were difficult to predict, making the estimation of heart rate challenging. Because the ambient light interference affected the entire video frame, these aliased frequency components found in the PPGi signals extracted from regions of interest on the patient's face, were also present in the signals extracted from a region of interest manually selected from the background wall. Since the time course extracted from the background wall was not expected to contain frequency content relating to the patient's physiology, the ambient light interference was removed from the PPGi signals by cancelling the common poles from the autoregressive models fitted to both regions of interest (face and the background wall).

The pole cancellation method proved to be successful because of the ease of the manual selection of the background region of interest from the homogeneous recording configuration across all sessions. An automated method would require the segmentation of the background wall or other objects in the video and the proper selection of regions of interest depending on their frequency content.

The accuracy of heart rate estimation from the video camera, for periods during which the subject is stable, is comparable to that of the two standard pulse oximeters used in the clinical study. The camera estimates using the best method (based on a multi-channel Kalman filter applied to autoregressive models after pole cancellation) showed a strong positive correlation and good agreement compared to the reference heart rate, with minimal bias for approximately 65% of the valid recording time.

By applying a common threshold to the magnitude of the radius of the dominant pole of an autoregressive model, the proposed methods for heart rate estimation discussed in chapter 6 were used to construct maps of distribution of heart rate information across the entire image frame. This was effectively used as a signal quality metric to select the regions of interest that contain frequency content in the physiological range for heart rate in the video camera data and, hence, highlight areas from which heart rate can be estimated most accurately. The analysis presented in section 6.10.1 showed that the periodic components for heart rate are strongest on the patient's cheeks and forehead, corroborating published anatomical studies of the human face which indicate that there is a high skin blood flow in both of these regions.

11.4.2 Respiratory rate

The physiological video maps developed for heart rate in chapter 6 allowed for the development of a signal quality metric indicative of the degree of confidence that the video data included frequency content in the heart rate physiological range. The proposed technique was expanded in chapter 8 to visually analyse the video data for frequency content in the physiological range for respiratory rate. It was shown that, as opposed to heart rate, respiratory rate could be estimated not only from regions of interest on the subject's face, but also from regions across the chest area, mainly the thorax and diaphragm. The nature of the information extracted from these two areas differs in principle. The signals extracted from regions of interest on the patient's face were PPGi pulsatile time series, mainly caused by colour changes due to blood flow. The signals extracted from the upper torso (such as the thorax) were mainly affected by subject motion as a result of breathing.

Similarly to heart rate estimation, respiratory rate estimation also relies on the process of identifying suitable time periods for which the location of the patient is known and the effects of subject motion or interaction with the clinical staff are minimal. The algorithms for respiratory rate estimation were built on top of the work carried out for heart rate estimation described in chapter 6. Once stable periods were identified from the video recordings, two groups of regions of interest were selected: from the patient's face and from the patient's upper torso. The quality of the PPGi signals extracted from the patient's face was assessed and respiratory signals were extracted from each region of interest and for each of the three colour channels. Respiratory signals from the patient's upper torso were extracted using grey-scale versions of the image frames. Algorithms for data reduction were used due to the large number of regions of interest to process. Respiratory rate was estimated from each respiratory signal and data fusion strategies similar to heart rate processing were used to compute a final respiratory rate estimate for the patient.

By fusing the results from the PPGi waveforms from regions of interest on the patient's face with the motion-dependent signals from the subject's upper torso, the accuracy of respiratory rate estimation from the video camera, in periods during which the subject is stable, was comparable to that of the reference respiratory rate computed from the ECG and chest band. The camera estimates showed a strong positive correlation and good agreement compared to the reference respiratory rate, with minimal bias for approximately 70% of the valid recording time.

11.4.3 Towards SpO_2

Chapter 9 presented preliminary work to identify changes in peripheral oxygen saturation using a video camera and ambient light. The proposed methods for heart rate or respiratory rate estimation mainly find a frequency component in the recorded signals. Identifying changes in SpO_2 depends on the colour reproduction of the video camera sensor from areas for which the skin was visible. Unlike the sensor used in a pulse oximeter, colour cameras cannot directly measure the spectra of colour signals because spectral accuracy is sacrificed for spatial resolution. Since the spectral data for a single point in the scene is described with three values (red, green and blue), the recorded values are only an approximation of the true incoming colour signal spectra.

Different combinations of external illuminants can produce contrasting sensor responses: an object can be recorded as two different colours when affected by different illumination sources, or two different colours cannot be discriminated under the effects of other illuminants [363]. Colour changes in the Renal Unit were due to several factors such as varying light levels, the temporary presence of shadows during the intervention of the clinical staff, or the varying light colour from changes in the spectral power distribution of light sources (daylight, fluorescent or incandescent). Therefore, the algorithms discussed in chapter 9 proposed methods for colour-balancing the data recorded by the video camera and used colour space transformation to model the skin colour changes under different illumination conditions.

The validation of any new method of deriving oxygen saturation is hampered by the difficulties associated with trying to obtain SpO_2 values over the most clinically relevant range, from 80% to 95%. In the patient population from the dialysis study, there were only a handful of cases of significant desaturation, defined as a SpO_2 decrease of at least 5%. Moreover, Section 9.3 described circumstances for which the SpO_2 estimates from the two reference pulse oximeters varied greatly depending on the patient's physiology, medications and diagnoses. Therefore, to validate the proposed algorithms, 18 segments, each with a duration of 5 minutes, were manually selected from the dialysis sessions for which the SpO_2 estimates provided by the two pulse oximeters differed by less than 2%, the computed reference SpO_2 decreased by more than 5% across the SpO_2 physiological range (80% to 100%), minimal patient motion occurred and there were no sudden changes in illumination. The results presented in chapter 9 indicate that the proposed analysis of the video camera signals using two colour channels, and after colour processing, can provide an indication of oxygen desaturation during the selected periods under analysis.

11.4.4 Real-time implementation

Chapter 5 described the software development guidelines and industry best practices followed for the development of the data acquisition system used to record video and pulse oximetry data for the dialysis study. The standards and guidelines allowed the extension of the data acquisition system adding modules for the real-time implementation (see chapter 10) of the image and signal processing algorithms for heart rate and respiratory rate estimation proposed in chapters 6 and 8 respectively.

In the initial phase of the research carried on for this thesis, due to the complexity of the algorithms developed for the non-contact estimation of heart rate and respiratory rate, real-time analysis was only possible using specialised high-end video cameras attached to high-performance workstations. By the culmination of the thesis, as the computing power increased, higher quality digital sensors became affordable and advanced software optimisations were possible, the proposed algorithms are now able to run on the latest mobile devices such as the iPhone and iPad.

As the computing power of consumer-level workstations increases and the performance of off-the-shelf video cameras improves, the design and implementation of more advanced algorithms, such as the work described in the next section, will become ever possible,

11.5 Future work

This section discusses future work and potential research avenues in the field of non-contact sensing arising from the research carried out for this thesis.

11.5.1 Patient identification

Chapters 1 and 2 stated that non-contact vital-sign monitoring technologies provide several advantages over traditional methods. One advantage of a video-based solution is the potential of using a single video camera to simultaneously monitor the physiology from multiple patients. This configuration would improve user acceptability by reducing the burden of wearing continuous vital-sign monitoring hardware, reducing cabling and decreasing the costs of implementing and maintaining new technology.

The simultaneous monitoring of patients would require automated methods for the recognition of the patients from individual clues in the video recordings, such as the face. Face recognition is a task that humans perform routinely and effortlessly on a daily basis. The availability of computers and the continued increase in their processing power has created great interest in the imaging research community with several competing methods being proposed [187].

Face recognition also presents an important advantage over other biometric technologies for the identification of individuals, such as fingerprint or iris. Video can be recorded from a distance, rendering face analysis unobtrusive.

11.5.2 Body and skin segmentation

The methods for the non-contact estimation of heart rate, respiratory rate and changes in SpO_2 proposed in chapters 6 to 9, rely on the detection of a patient in the video frame and the selection of regions of interest from which the vital signs can be estimated. The performance of these two tasks is essential to the successful estimation of vital signs from the video camera.

Methods based on face detection that rely on the colour-reproduction of the camera sensor can potentially be less reliable in a hospital setting and present several challenges. Clinical staff constantly interact with the patients. Patients are awake, their posture and change in position are spontaneous. Many clinical care routine activities are undertaken several times a day, such as the recording of vital signs or administering medication.

Lighting conditions change not only throughout the day but also depend on the season of the year, with long and bright days during summer and short and dark days during winter. Artificial light sources, such as room and overhead lights, cause reflections on the skin and change how the video cameras record colour. Reflections or shadows of people walking across the room also affect the recordings.

The algorithms proposed in this thesis also rely on the selection of skin regions of interest on the patient's face and the correct identification of the patient's upper torso. Therefore, algorithms for robust skin and body part segmentation can potentially improve not only the results but also increase the amount of time that a patient can continuously be monitored by a video camera.

Recently, advanced methods such as multi-task convolutional neural networks (CNN) have been proposed for the detection of the presence of a patient and segmenting the patient's skin regions [52]. These methods can produce accurate results and be robust to changes in different skin tones, pose variations, lighting variations, routine interaction with clinical staff, mixed-race subjects and low-light settings.

11.5.3 Physiological video maps

The proposed approach of heart rate physiological maps from multiple regions of interest presented in chapters 6 and 8 allowed for the development of signal-quality metrics indicative of the degree of confidence that the video data included frequency content in the heart rate or respiratory rate physiological ranges. This approach can potentially be expanded to the detection of subjects in a video image, simply from the fact that their physiological information (such as their heart rate) is a periodic signal derived from skin regions. This could be used to replace the face detection and tracking algorithms in the imaging pipeline discussed in section 6.4, or to augment the algorithms with independent information based purely on physiology.

By continuously computing the physiological maps throughout the video sequence, multiple regions of interests could also lead to more localised physiological studies, such as measuring the propagation of pulsatile arterial blood across the body (see chapter 3) and the potential to measure other vital signals such as pulse transit time (PTT) and blood pressure [246].

11.5.4 Continuous non-contact vital-sign estimation

The algorithms proposed and the technology used in the clinical study analysed in this thesis were developed and evaluated to estimate vital signals from video signals recorded during the day time. Providing continuous estimates of vital signs from a video camera to the clinical staff throughout the entire 24 hours could potentially widen the acceptance of the video-based estimation technology. This could prove to be especially important during the night, where availability of clinical staff is reduced. This undoubtedly would require the design of new clinical studies, the evaluation of new infra-red cameras and other potential sources of illumination such as infra-red light-emitting diodes (LED), and the development of new algorithms to process the night-time video data.

11.6 Concluding remarks

Non-contact vital-sign monitoring provides several advantages over traditional methods because no subject participation is required to set the equipment up. It requires no skin preparation, causes no skin irritation, decreases the risk of infection, and has the potential to be seamlessly integrated into the patient's care in the hospital or clinical practice, and lifestyle at home.

Measuring multiple vital signs from a single non-contact sensor is also beneficial as it would improve user acceptability by reducing the burden of wearing continuous vital-sign monitoring hardware, reducing cabling and decreasing the costs of implementing and maintaining new technology.

Appendix A

Autoregressive Modelling

An autoregressive (AR) model is a type of random process that attempts to predict the output of a system based on a linear combination of its previous values [48]. Given a time series x_t, an autoregressive model can be defined as:

$$x_t = c + \sum_{i=1}^{p} a_i x_{t-i} + \varepsilon_t \tag{A.1}$$

Where :

- c is a constant (often omitted for simplicity).
- p is the model order.
- $a_1, ..., a_p$ are the model parameter coefficients.
- ε_t is a white noise Gaussian process with zero mean and variance σ^2.

The current value of x_t, is a linear combination of the p previous values of x_t and the current value of ε_t. The transfer function in the Z-domain relating the input to the output is given by:

$$H(z) = \frac{1}{\sum_{k=1}^{p} a_k Z^{-k}} = \frac{Z^p}{(Z - Z_1)...(Z - Z_p)} = \frac{Z^p}{\vec{P}_1 \vec{P}_2 ... \vec{P}_p} \tag{A.2}$$

The denominator in equation (A.2) can be factored into p terms, where $\vec{P}_1 \vec{P}_2 ... \vec{P}_p$ are vectors extending from any point Z in the complex plane to each of the p poles $Z_1, Z_2, ..., Z_p$ of $H(z)$. These poles, each of which is one of a complex-conjugate pari, are the roots of the denominator of $H(z)$. As $H(z)$ has no finite zeros, autoregressive models can be viewed as the output of an all-pole infinite impulse response filter whose input is white noise:

$$\varepsilon_t \rightarrow H(z) \rightarrow X_t \tag{A.3}$$

It is known that the power spectral density (PSD) of a linear filter, such as $P_x(Z)$, is related to the PSD of its input process, $P_\varepsilon(Z)$, by the following equation [158]:

$$P_x(Z) = H(Z) H^*(\frac{1}{Z^*}) P_\varepsilon(Z) \tag{A.4}$$

As ε_t, the signal driving the system, is a Gaussian white noise source with zero mean and variance σ^2, its PSD is therefore constant:

$$P_{\varepsilon}(x) = \sigma^2 \Delta t \tag{A.5}$$

Replacing equation (A.5) into equation (A.4), the power spectral density of the output is therefore given by:

$$P_x(Z) = H(Z)H^*(\frac{1}{Z^*})\sigma^2 \Delta t \tag{A.6}$$

To evaluate the frequency response, $z = exp(j2\pi f \Delta t)$ is substituted into $H(Z)$ over the range $-(1/2\Delta t) \leq f \leq (1/s\Delta t)$. This is equivalent to evaluating $H(Z)$ around the unit circle in the complex plane. The PSD is therefore given by:

$$PSD(f) = |H(e^{j2\pi f \Delta t})|^2 \sigma^2 \Delta t = \frac{\sigma^2 \Delta t}{|1 + \sum_{i=1}^{p} a_i e^{-j2\pi f \Delta t}|^2} \tag{A.7}$$

Each complex-conjugate pair of the poles in $H(Z)$ corresponds to one resonant frequency in $PSD(f)$. Therefore, a pth order AR model has p poles and r resonant frequencies, $f_1, f_2, ..., f_r$, where $r = p/2$ when p is even and $r = (p+1)/2$ when p is odd [48]. Figure A.1 shows an example of the Autoregressive Model of order 8 applied to the a 15-second green-channel time series from a region of interest on the subject's cheek.

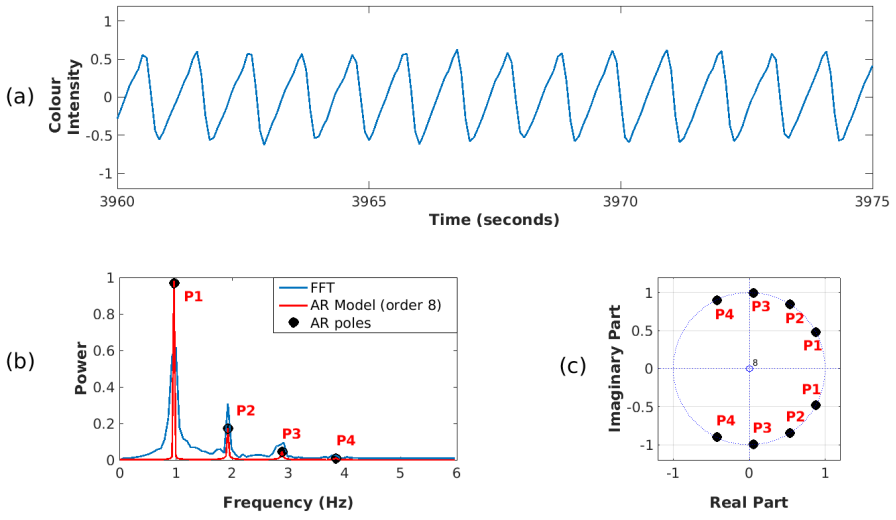


Figure A.1: Autoregressive Model computation for a typical 15-second window and model order 8 from a 12 frames-per-second video, (a) Input time series, (b) Comparison between the FFT and AR frequency components, (c) Autoregressive Model zero-pole plot showing the 4 poles P_1 to P_4

Appendix B

Bayesian change point detection

Change points, or step changes, are abrupt variations or discontinuities in a data sequence, as shown in the step change at around minute 11 in figure B.1(a) during which the video camera automatically modified the gain of the green colour channel to compensate for a change in the overall lighting conditions in the scene.

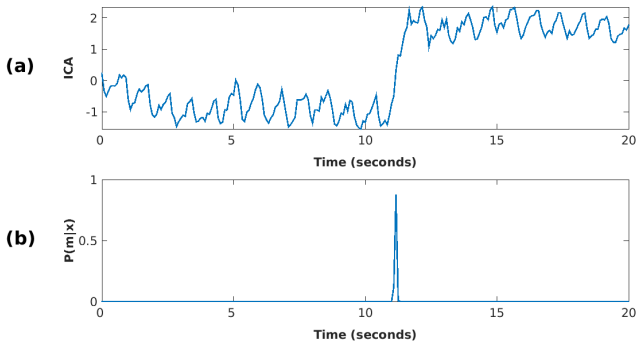


Figure B.1: Bayesian change point detection algorithm applied to a 20-second PPGi signal extracted from the green colour channel. (a) Input signal, (b) change point detected.

To estimate the location of a step change (see figure B.1(b)), a Bayesian change point detection algorithm based on [276] and [2] is used in the thesis. Given a data sequence x of N samples with Gaussian noise added:

$$x_i = \begin{cases} \mu_1 + \epsilon_i, & \text{if } i < m. \\ \mu_2 + \epsilon_i, & \text{otherwise.} \end{cases} \tag{B.1}$$

where the noise samples ϵ_i are assumed to be independent and m is the step change location. The likelihood of the data is given by the joint probability of the noise samples ϵ_i:

$$P(x|\{\mu_1 \mu_2 \sigma m\}) = \prod_{i=1}^{N} P(\epsilon_i) \tag{B.2}$$

where σ is the standard deviation of the Gaussian noise; μ_1, μ_2 and σm are the known time series parameters. The probability density function for the noise samples is defined by:

$$P(\epsilon) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\epsilon-\mu)^2}{2\sigma^2}} \tag{B.3}$$

Substituting equation (B.3) into B.2, the likelihood is given by:

$$P(x|\{\mu_1\mu_2\sigma m\}) = (2\pi\sigma^2)^{-\frac{N}{2}} exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{m}(x_i-\mu_1)^2 + \sum_{i=m+1}^{N}(x_i-\mu_2)^2\right)\right] \tag{B.4}$$

Given Bayes' theorem, the posterior probability density of the signal parameters $\mu_1\mu_2\sigma m$ given the data sequence x can be computed by:

$$P(\{\mu_1\mu_2\sigma m\}|x) = \frac{P(x|\{\mu_1\mu_2\sigma m\})P(\{\mu_1\mu_2\sigma m\})}{P(x)} \tag{B.5}$$

If it is assumed that all parameter values are equally likely a priori, then:

$$\begin{aligned} P(\mu_1) &= k_1 \\ P(\mu_2) &= k_2 \\ P(log(\sigma)) = k_3 &\Longrightarrow P(\sigma) \propto \frac{1}{\sigma} \end{aligned} \tag{B.6}$$

where k_1, k_2 and k_3 are constants. Substituting equation (B.6) in B.5, the posterior probability density of the signal parameters is modified to be the following expression:

$$P(\{\mu_1\mu_2\sigma m\}|x) \propto \frac{1}{\sigma} P(x|\{\mu_1\mu_2\sigma m\}) \tag{B.7}$$

By integrating over the parameters μ_1, μ_2 and σm , the posterior in equation (B.7) can be modified to account only for the parameter of interest ([31]), which is the location of the step change m:

$$P(m|x) = \int_0^{+\infty} d\sigma \int_{-\infty}^{+\infty} d\mu_1 \int_{-\infty}^{+\infty} d\mu_2 P(\{\mu_1\mu_2\sigma m\}|x) \tag{B.8}$$

According to [276], μ_1 and μ_2 in equation (B.8) can be computed using the following integral identity:

$$\int_{-\infty}^{+\infty} exp(-ax^2 - bx - c)dx = \sqrt{\frac{\pi}{a}} exp\left(\frac{b^2}{4a} - c\right) \tag{B.9}$$

Therefore, the first μ_1 is integrated as:

$$\begin{aligned} P(\{\mu_2\sigma m\}|x) &= \int_{-\infty}^{+\infty} P(\{\mu_1\mu_2\sigma m\}|x)d\mu_1 \\ &\propto \frac{(2\pi\sigma^2)^{-\frac{N}{2}}}{\sqrt{m}} exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{m} x_i^2 - \frac{1}{m}(\sum_{i=1}^{m} x_i)^2 + \sum_{i=m+1}^{N}(x_i-\mu_2)^2\right)\right] \end{aligned} \tag{B.10}$$

Similarly, μ_2 can be integrated as:

$$
\begin{aligned}
P(\{\sigma m\}|x) &= \int_{-\infty}^{+\infty} P(\{\mu_2 \sigma m\}|x) d\mu_2 \\
&\propto \frac{(2\pi\sigma^2)^{-\frac{N-1}{2}}}{\sqrt{m(N-m)}} exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{N} x_i^2 - \frac{1}{m}(\sum_{i=1}^{m} x_i)^2 - \frac{1}{N-m}(\sum_{i=m+1}^{N} x_i)^2\right)\right]
\end{aligned}
\tag{B.11}
$$

The parameter σ can be integrated using the gamma integral expression:

$$
\int_{-\infty}^{+\infty} x^{\alpha-1} exp(-Qx) dx = \frac{\Gamma(\alpha)}{Q^{\alpha}} \tag{B.12}
$$

Finally, the posterior probability density of the step change location m given the data sequence x can be computed by:

$$
\begin{aligned}
P(m|x) &= \int_{-\infty}^{+\infty} P(\{\sigma m\}|x) d\sigma \\
&\propto \frac{1}{\sqrt{m(N-m)}}\left[\sum_{i=1}^{N} x_i^2 - \frac{1}{m}(\sum_{i=1}^{m} x_i)^2 - \frac{1}{N-m}(\sum_{i=m+1}^{N} x_i)^2\right]^{-\frac{N-2}{2}}
\end{aligned}
\tag{B.13}
$$

Figure B.1 shows the Bayesian change point detection algorithm applied to a 10-minute window from the green-channel of the subject region of interest.

Bibliography

[1] Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010.

[2] Ryan Prescott Adams and David JC MacKay. Bayesian online changepoint detection. *arXiv Cornell University Library*, 2007.

[3] Paul S Addison, James N Watson, Michael L Mestek, and Roger S Mecca. Developing an algorithm for pulse oximetry derived respiratory rate (RRoxi): a healthy volunteer study. *Journal of clinical monitoring and computing*, 26(1):45–51, 2012.

[4] Geeta S Agashe, Joseph Coakley, and Paul D Mannheimer. Forehead pulse oximetry: Headband use helps alleviate false low readings likely related to venous pulsation artifact. *Anesthesiology*, 105:1111–1116, 2006.

[5] Farah Q AL-Khalidi, Reza Saatchi, Derek Burke, H Elphick, and Stephen Tan. Respiration rate monitoring methods: A review. *Pediatric pulmonology*, 46(6):523–529, 2011.

[6] Ben Allen, Anthony Brown, Katja Schwieger, Ernesto Zimmermann, Wasim Q Malik, David J Edwards, Laurent Ouvry, and Ian Oppermann. Ultra wideband: Applications, technology and future perspectives. *International workshop on convergent technologies*, 2005.

[7] John Allen. *Measurement and analysis of multi-site photoplethysmographic pulse waveforms in health and arterial disease*. PhD thesis, University of Newcastle upon Tyne, 2002.

[8] John Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological measurement*, 28(3), 2007.

[9] John Allen and Alan Murray. Similarity in bilateral photoplethysmographic peripheral pulse wave characteristics at the ears, thumbs and toes. *Physiological measurement*, 21(3):369, 2000.

[10] John Allen and Alan Murray. Effects of filtering on multisite photoplethysmography pulse waveform characteristics. In *Computers in Cardiology*, pages 485–488, 2004.

[11] John Allen, Klaus Overbeck, Gerard Stansby, and Alan Murray. Photoplethysmography assessments in cardiovascular disease. *Measurement and Control*, 39(3):80–83, 2006.

[12] Steven Allender, Peter Scarborough, Vito Peto, Mike Rayner, Jose Leal, Ramon Luengo-Fernandez, and Alastair Gray. *European cardiovascular disease statistics*. European Heart Network, 2008.

[13] ANSI. Cardiac monitors, heart rate meters, and alarms. Standard ANSI/AAMI EC13: 2002, American National Standards Institute - Association for the Advancement of Medical Instrumentation, 2002.

[14] ANSI. Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms. Standard ANSI/AAMI EC57:2012, American National Standards Institute - Association for the Advancement of Medical Instrumentation, 2012.

[15] Brian W Anthony et al. *Multiscale Dynamic Time and Space Warping*. PhD thesis, Massachusetts Institute of Technology, 2008.

[16] Nadarajah Aravindhan and David George Chisholm. Sulfhemoglobinemia presenting as pulse oximetry desaturation. *The Journal of the American Society of Anesthesiologists*, 93 (3):883–884, 2000.

[17] The Motor Industry Software Reliability Association. *MISRA-C:2004 Guidelines for the use of the C language in critical systems*. MISRA, second edition, 2004.

[18] The Motor Industry Software Reliability Association. *MISRA C++:2008 Guidelines for the use of the C++ language in critical systems*. MISRA, first edition, 2008.

[19] The Renal Association. The seventeenth annual report, Dec 2014. URL `https://www.renalreg.org/reports/2014`. [Online; accessed 14-May-2017].

[20] AE Aubert, L Welkenhuysen, J Montald, L De Wolf, H Geivers, J Minten, Hugo Kesteloot, and H Geest. Laser method for recording displacement of the heart and chest wall. *Journal of biomedical engineering*, 6(2):134–140, 1984.

[21] Terry M Baird and Michael R Neuman. Effect of infant position on breath amplitude measured by transthoracic impedance and strain gauges. *Pediatric pulmonology*, 10(1):52–56, 1991.

[22] Guha Balakrishnan, Fredo Durand, and John Guttag. Detecting pulse from head motions in video. *IEEE Conference on Computer Vision and Pattern Recognition*, Jun 2013.

[23] Steven J Barker. "motion-resistant" pulse oximetry: a comparison of new and old models. *Anesthesia & Analgesia*, 95(4):967–972, 2002.

[24] Steven J Barker and Kevin K Tremper. The effect of carbon monoxide inhalation on pulse oximetry and transcutaneous PO2. *Anesthesiology*, 66(5):677–679, 1987.

[25] Steven J Barker and Kevin K Tremper. Pulse oximetry: applications and limitations. *International anesthesiology clinics*, 25(3):155–175, 1987.

[26] John S Barlow. *The electroencephalogram: its patterns and origins*. MIT press, 1993.

[27] Terence W Barrett. History of ultrawideband (UWB) radar & communications: pioneers and innovators. In *Proceedings Progress in Electromagnetics Symposium*, pages 1–42, 2000.

[28] Bryce Bayer. Color imaging array, Jul 1976. US Patent 3,971,065.

[29] Joachim Behar, Alistair Johnson, Gari D Clifford, and Julien Oster. A comparison of single channel fetal ECG extraction methods. *Annals of biomedical engineering*, 42(6):1340–1353, 2014.

[30] Amirta Benning, Maisoon Ghaleb, Anu Suokas, Mary Dixon-Woods, Jeremy Dawson, Nick Barber, Bryony Dean Franklin, Alan Girling, Karla Hemming, Martin Carmalt, et al. Large scale organisational intervention to improve patient safety in four UK hospitals: mixed method evaluation. *British Medical Journal*, 342:d195, 2011.

[31] James O Berger, Brunero Liseo, and Robert L Wolpert. Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*, 14(1):1–28, 1999.

[32] J Martin Bland and Douglas G Altman. Measuring agreement in method comparison studies. *Statistical methods in medical research*, 8(2):135–160, 1999.

[33] J Martin Bland and Douglas G Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *International Journal of Nursing Studies*, 47(8): 931–936, 2010.

[34] C Bohr, K Hasselbalch, and A Krogh. Concerning a biologically important relationship–the influence of the carbon dioxide content of blood on its oxygen binding. *Skand. Arch. Physiol*, 16:402–412, 1904.

[35] Mari A Botti and Jennifer O Hunt. The routine of post anaesthetic observations. *Contemporary Nurse*, 3(2):52–57, 1994.

[36] Frédéric Bousefsaf, Choubeila Maaoui, and Alain Pruski. Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate. *Biomedical Signal Processing and Control*, 8(6):568–574, Nov 2013.

[37] WS Boyle and GE Smith. Charge-coupled devices-a new approach to mis device structures. *IEEE Spectrum*, 8(7):18–27, 1971.

[38] Gary Bradski and Adrian Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. O'reilly, 2008.

[39] David Briers, Donald D Duncan, Evan Hirst, Sean J Kirkpatrick, Marcus Larsson, Wiendelt Steenbergen, Tomas Stromberg, and Oliver B Thompson. Laser speckle contrast imaging: theoretical and practical limitations. *Journal of biomedical optics*, 18(6):066018–066018, 2013.

[40] Timothy D Brooks, David A Paulus, and William E Winkle. Infrared heat lamps interfere with pulse oximeters. *The Journal of the American Society of Anesthesiologists*, 61(5):630–630, 1984.

[41] Lewis F Brown and Shivaram P Arunachalam. Real-time TP knot algorithm for baseline wander noise removal from the electrocardiogram. In *Rocky Mountain Bioengineering Symposium*, 2009.

[42] A Buchs, Y Slovik, M Rapoport, Chaim Rosenfeld, B Khanokh, and Meir Nitzan. Right-left correlation of the sympathetically induced fluctuations of photoplethysmographic signal in diabetic and non-diabetic subjects. *Medical and Biological Engineering and Computing*, 43 (2):252–257, 2005.

[43] Robert G Buckley, Stephen E Aks, James L Eshom, Robert Rydman, Jeffrey Schaider, and Phillip Shayne. The pulse oximetry gap in carbon monoxide intoxication. *Annals of emergency medicine*, 24(2):252–255, 1994.

[44] Nigel T Cable. Unlocking the secrets of skin blood flow. *The Journal of physiology*, 572(3): 613–613, 2006.

[45] G Cacho, JL Pérez-Calle, A Barbado, JL Lledó, R Ojea, and CM Fernández-Rodríguez. Capnography is superior to pulse oximetry for the detection of respiratory depression during colonoscopy. *Revista Espanola de Enfermedades Digestivas*, 102(2):86, 2010.

[46] Bradley Carlson. Comparison of modern CCD and CMOS image sensor technologies and systems for low resolution imaging. In *Proceedings of IEEE Sensors*, volume 1, pages 171–176, 2002.

[47] CG Caro and JA Bloice. Contactless apnoea detector based on radar. *The Lancet*, 298(7731): 959–961, 1971.

[48] Shelley Cazares, Mary Moulden, Christopher Redman, and Lionel Tarassenko. Tracking poles with an autoregressive model: a confidence index for the analysis of the intrapartum cardiotocogram. *Medical engineering & physics*, 23(9):603–614, Nov 2001.

[49] William T Cecil, Kevin J Thorpe, Eugene E Fibuch, and Gerald F Tuohy. A clinical evaluation of the accuracy of the Nellcor N-100 and Ohmeda 3700 pulse oximeters. *Journal of Clinical Monitoring and Computing*, 4(1):31–36, 1987.

[50] Giovanni Cennini and Vincent Jeanne. Method of controlling a function of a device and system for detecting the presence of a living being, Sep 2010. WO Patent Application 2,010,100,593.

[51] Giovanni Cennini, Jeremie Arguel, Kaan Akşit, and Arno van Leest. Heart rate monitoring via remote photoplethysmography with motion artifacts reduction. *Optics express*, 18: 4867–4875, 2010.

[52] Sitthichok Chaichulee, Mauricio Villarroel, João Jorge, Carlos Arteta, Gabrielle Green, Kenny McCormick, Andrew Zisserman, and Lionel Tarassenko. Multi-task convolutional neural network for patient detection and skin segmentation in continuous non-contact vital sign monitoring. In *Proceedings of the 12th International IEEE Conference on Face & Gesture Recognition*, 2017.

[53] AVJ Challoner. Photoelectric plethysmography for estimating cutaneous blood flow. *Non-invasive physiological measurements*, 1:125–151, 1979.

[54] Edward D Chan, Michael M Chan, and Mallory M Chan. Pulse oximetry: understanding its basic principles facilitates appreciation of its limitations. *Respiratory medicine*, 107:789–99, 2013.

[55] Nisha Charkoudian. Skin blood flow in adult human thermoregulation: how it works, when it does not, and why. In *Mayo Clinic Proceedings*, volume 78, pages 603–612. Elsevier, 2003.

[56] Sergey Y Chekmenev, Aly A Farag, and Edward A Essock. Multiresolution approach for non-contact measurements of arterial pulse using thermal imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop*, 2006.

[57] Yu M Chi and Gert Cauwenberghs. Wireless non-contact eeg/ecg electrodes for body sensor networks. In *Body Sensor Networks (BSN), 2010 International Conference on*, pages 297–301. IEEE, 2010.

[58] Yu M Chi, Stephen R Deiss, and Gert Cauwenberghs. Non-contact low power EEG/ECG electrode for high density wearable biopotential sensor networks. In *Sixth IEEE International Workshop on Wearable and Implantable Body Sensor Networks*, pages 246–250, 2009.

[59] Yu Mike Chi, Tzyy-Ping Jung, and Gert Cauwenberghs. Dry-contact and noncontact biopotential electrodes: methodological review. *IEEE Reviews in Biomedical Engineering*, 3: 106–119, 2010.

[60] Eirini Christinaki, Giorgos Giannakakis, Franco Chiarugi, Matthew Pediaditis, Galateia Iatraki, Dimitris Manousos, Kostas Marias, and Manolis Tsiknakis. Comparison of blind source separation algorithms for optical heart rate monitoring. In *4th International Conference on Wireless Mobile Communication and Healthcare (Mobihealth)*, pages 339–342, 2014.

[61] DG Clayton, RK Webb, AC Ralston, D Duthie, and WB Runciman. A comparison of the performance of 20 pulse oximeters under conditions of poor perfusion. *Anaesthesia*, 46(1): 3–10, 1991.

[62] DG Clayton, RK Webb, AC Ralston, D Duthie, and WB Runciman. Pulse oximeter probes a comparison between finger, nose, ear and forehead probes under conditions of poor perfusion. *Anaesthesia*, 46(4):260–265, 1991.

[63] SF Cleary, F Nickless, LM Liu, and R Hoffman. Studies of exposure of rabbits to electromagnetic pulsed fields. *Bioelectromagnetics*, 1(3):345–352, 1980.

[64] Gari D Clifford, Francisco Azuaje, and Patrick McSharry. *Advanced methods and tools for ECG data analysis*. Engineering in medicine & biology series. Artech House, 2006.

[65] David Clifton, Mauricio Villarroel, and Lionel Tarassenko. Remote monitoring of vital signs, 2012. US Patent App. 14/240,291.

[66] David Clifton, Mauricio Villarroel, and Lionel Tarassenko. Analysing video images of a subject to identify spatial image areas which contain periodic intensity variations, 2014. WO Patent App. PCT/GB2014/050,211.

[67] David Clifton, Marco Pimentel, Alistair Johnson, Peter Charlton, Ahmar Shah, Alessandro Guazzi, and Lionel Tarassenko. Robust estimation of respiratory rate from pulse oximeters. *IEEE Transactions on Biomedical Engineering*, 2015.

[68] Julie-Ann Collins, Aram Rudenski, John Gibson, Luke Howard, and Ronan O'Driscoll. Relating oxygen partial pressure, saturation and content: the haemoglobin–oxygen dissociation curve. *Breathe*, 11(3):194, 2015.

[69] R.B. Colton. Radar in the U.S. Army. In *Proceedings of IRE*, volume 1, 1945.

[70] Julius H Comroe Jr and Stella Botelho. The unreliability of cyanosis in the recognition of arterial anoxemia. *The American journal of the medical sciences*, 214(1):1–6, 1947.

[71] B Conrad and P Schönle. Speech and respiration. *Archiv für Psychiatrie und Nervenkrankheiten*, 226(4):251–268, 1979.

[72] Ernest Cooke, Sally Bowcock, and Albert Smith. Photoplethysmography of the distal pulp in the assessment of the vasospastic hand. *Angiology*, 36(1):33–40, 1985.

[73] Lockheed Martin Corporation. *Joint strike fighter air vehicle c++ coding standards for the system development and demonstration program*. Number Document No. 2RDU00001 Rev C. Lockheed Martin, first edition, 2005.

[74] Charles J Coté, E Andrew Goldstein, William H Fuchsman, and David C Hoaglin. The effect of nail polish on pulse oximetry. *Anesthesia & Analgesia*, 67(7):683–686, 1988.

[75] Angela Coulter, Suzanne Parsons, Janet Askham, World Health Organization, World Health Organization, et al. *Where are the patients in decision-making about their own care?* World Health Organization Regional Office for Europe, 2008.

[76] Michelle A Cretikos, Rinaldo Bellomo, Ken Hillman, Jack Chen, Simon Finfer, and Arthas Flabouris. Respiratory rate: the neglected vital sign. *Medical Journal of Australia*, 188(11): 657, 2008.

[77] John Crowe and Damianos Damianou. The wavelength dependence of the photoplethysmogram and its implication to pulse oximetry. In *14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 6, pages 2423–2424, Oct 1992.

[78] John A D'Andrea, Brenda L Cobb, and John O de Lorge. Lack of behavioural effects in the rhesus monkey: high peak microwave pulses at 1.3 GHz. *Bioelectromagnetics*, 10(1):65–76, 1989.

[79] Gerard de Haan and Vincent Jeanne. Robust Pulse Rate From Chrominance-Based rPPG. *IEEE transactions on biomedical engineering*, 60:2878–86, 2013.

[80] Joost De Kock and Lionel Tarassenko. Pulse oximetry: Theoretical and experimental models. *Medical and Biological Engineering and Computing*, 31(3):291–300, 1993.

[81] Joost P De Kock. *Pulse oximetry: theoretical and experimental models*. PhD thesis, University of Oxford, 1991.

[82] Denise Felber Dietrich, Christian Schindler, Joel Schwartz, Jean-Claude Barthélémy, Jean-Marie Tschopp, Frédéric Roche, Arnold von Eckardstein, Otto Brändli, Philippe Leuenberger, Diane R Gold, et al. Heart rate variability in an ageing population and its association with lifestyle and cardiovascular risk factors: results of the SAPALDIA study. *Europace*, 8(7):521–529, 2006.

[83] Amy Droitcour. *Non-contact measurement of heart and respiration rates with a single-chip microwave Doppler radar*. PhD thesis, Stanford University, 2006.

[84] Amy Droitcour, Victor Lubecke, Jenshan Lin, and Olga Boric-Lubecke. A microwave radio for Doppler radar sensing of vital signs. In *IEEE MTT-S International Microwave Symposium Digest*, volume 1, pages 175–178, 2001.

[85] Zachary V Edmonds, William R Mower, Luis M Lovato, and Rosaelva Lomeli. The reliability of vital sign measurements. *Annals of emergency medicine*, 39(3):233–237, 2002.

[86] W Einthoven, G Fahr, and A De Waart. On the direction and manifest size of the variations of potential in the human heart and on the influence of the position of the heart on the form of the electrocardiogram. *American heart journal*, 40(2):163–211, 1950.

[87] Renars Erts, Janis Spigulis, Indulis Kukulis, and Maris Ozols. Bilateral photoplethysmography studies of the leg arterial stenosis. *Physiological measurement*, 26(5):865, 2005.

[88] David Evans, Brent Hodgkinson, and Judith Berry. Vital signs in hospital patients: a systematic review. *International journal of nursing studies*, 38(6):643–650, 2001.

[89] ML Evans and LA Geddes. An assessment of blood vessel vasoactivity using photoplethysmography. *Medical instrumentation*, 22(1):29–32, 1988.

[90] Mark Everingham, Josef Sivic, and Andrew Zisserman. "Hello! My name is... Buffy" – Automatic Naming of Characters in TV Video. In *Proceedings of the British Machine Vision Conference*, 2006.

[91] I Fine and A Weinreb. Multiple scattering effect in transmission pulse oximetry. *Medical and Biological Engineering and Computing*, 33(5):709–712, 1995.

[92] Susannah Fleming. *Measurement and fusion of non-invasive vital signs for routine triage of acute paediatric illness*. PhD thesis, University of Oxford, 2010.

[93] Susannah Fleming, Matthew Thompson, Richard Stevens, Carl Heneghan, Annette Plüddemann, Ian Maconochie, Lionel Tarassenko, and David Mant. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies. *The Lancet*, 377(9770):1011–1018, 2011.

[94] Susannah G Fleming and Lionel Tarassenko. A comparison of signal processing techniques for the extraction of breathing rate from the photoplethysmogram. *International Journal of Biological and Medical Sciences*, 2(4):232–6, 2007.

[95] Rich Fletcher and Jing Han. Low-cost differential front-end for Doppler radar vital sign monitoring. In *IEEE MTT-S International Microwave Symposium Digest*, pages 1325–1328, 2009.

[96] Mia Folke, L Cernerud, M Ekström, and Bertil Hök. Critical review of non-invasive respiratory monitoring in medical care. *Medical and Biological Engineering and Computing*, 41(4): 377–383, 2003.

[97] Jong Yong A Foo and Stephen J Wilson. A computational system to optimise noise rejection in photoplethysmography signals during motion or poor perfusion states. *Medical and Biological Engineering and Computing*, 44(1-2):140–145, 2006.

[98] Brand Fortner and Theodore E Meyer. *Number by colors: A guide to using color to understand technical data*. Springer Science & Business Media, 2012.

[99] CI Franks, BH Brown, and DM Johnston. Contactless respiration monitoring of infants. *Medical and biological engineering*, 14(3):306–312, 1976.

[100] CI Franks, JB Watson, BH Brown, and EF Foster. Respiratory patterns and risk of sudden unexpected death in infancy. *Archives of disease in childhood*, 55(8):595–599, 1980.

[101] Yoav Freund, Robert E Schapire, et al. Experiments with a new boosting algorithm. In *ICML*, volume 96, pages 148–156, 1996.

[102] Gary M Friesen, Thomas C Jannett, Manal Afify Jadallah, Stanford L Yates, Stephen R Quint, and H Troy Nagle. A comparison of the noise sensitivity of nine QRS detection algorithms. *IEEE Transactions on Biomedical Engineering*, 37(1):85–98, 1990.

[103] Marc Garbey, Arcangelo Merla, and I Pavlidis. Estimation of blood flow speed and vessel location from thermal video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR*, volume 1, 2004.

[104] Marc Garbey, Nanfei Sun, Arcangelo Merla, and Ioannis Pavlidis. Contact-free measurement of cardiac pulse based on the analysis of thermal imagery. *IEEE Transactions on Biomedical Engineering*, 54(8):1418 –1426, Aug 2007.

[105] Curtis F Gerald and Patrick O Wheatley. Applied numerical methods, 2004.

[106] Michael J Gesquiere, Aymen A Awad, David G Silverman, Robert G Stout, Denis H Jablonka, Tyler J Silverman, and Kirk H Shelley. Impact of withdrawal of 450 ml of blood on respiration-induced oscillations of the ear plethysmographic waveform. *Journal of clinical monitoring and computing*, 21(5):277–282, 2007.

[107] Davide Giavarina. Understanding Bland Altman analysis. *Biochemia medica*, 25(2):141–151, 2015.

[108] Thomas Gleixner and Douglas Niehaus. Hrtimers and beyond: Transforming the linux time subsystems. In *Linux Symposium*, volume 1, pages 333–346. Citeseer, 2006.

[109] Charlie Goldberg and Jan Thompson. A practical guide to clinical medicine, 2005. URL `https://meded.ucsd.edu/clinicalmed/vital.htm`. [Online; accessed 14-May-2017].

[110] EAG Goldie. A device for the continuous indication of oxygen saturation of circulating blood in man. *Journal of Scientific Instruments*, 19(2):23, 1942.

[111] Irina I Goncharova and John S Barlow. Changes in EEG mean frequency and spectral purity during spontaneous alpha blocking. *Electroencephalography and clinical neurophysiology*, 76(3):197–204, 1990.

[112] Henk G Goovaerts, Henk H Ros, Thomas J Van Den Akker, and Hans Schneider. A digital QRS detector based on the principle of contour limiting. *IEEE Transactions on Biomedical Engineering*, 2(BME-23):154–160, 1976.

[113] Edwin Gordy, David L Drabkin, et al. Spectrophotometric studies XVI. Determination of the oxygen saturation of blood by a simplified technique, applicable to standard equipment. *Journal of Biological Chemistry*, 227(1):285–299, 1957.

[114] Thomas R Gravelyn and John G Weg. Respiratory rate as an indicator of acute respiratory dysfunction. *Jama*, 244(10):1123–1125, 1980.

[115] K Grossmann, H Rolka, H Reinhardt, and HC Reissmann. A new bedside procedure for computerized analysis of pulse curves in the diagnosis of peripheral arterial diseases. *Zeitschrift fur Gefasskrankheiten*, 16(4):324–329, 1986.

[116] BSI Group. Medical device software - Software life-cycle processes. Standard BS EN 62304:2006, British Standards Institution, 2008.

[117] Patrick S Hamilton and Willis J Tompkins. Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database. *IEEE Transactions on Biomedical Engineering*, (12): 1157–1165, 1986.

[118] CD Hanning and JM Alexander-Williams. Pulse oximetry: a practical review. *British Medical Journal*, 311(7001):367, 1995.

[119] AC Harris and IL Weatherall. Objective evaluation of colour variation in the sand-burrowing beetle Chaerodes trachyscelides White (Coleoptera: Tenebrionidae) by instrumental determination of CIELAB values. *Journal of the Royal Society of New Zealand*, 20(3):253–259, 1990.

[120] Alastair Harrison and Paul Newman. TICSync: Knowing When Things Happened. In *Proceedings of IEEE International Conference on Robotics and Automation*, May 2011.

[121] Ken M Harrison. Cameras and CCD's. In *Astronomical Spectroscopy for Amateurs*, Patrick Moore's Practical Astronomy Series, pages 87–95. Springer New York, 2011.

[122] Taigang He, Gari Clifford, and Lionel Tarassenko. Application of independent component analysis in removing artefacts from the electrocardiogram. *Neural Computing & Applications*, 15(2):105–116, 2006.

[123] Joel B Hellman and Ralph W Stacy. Variation of respiratory sinus arrhythmia with age. *Journal of applied physiology*, 41(5):734–738, 1976.

[124] Hidalgo. Equivital EQ02 LifeMonitor. URL `http://www.equivital.co.uk/products/tnr/sense-and-transmit`. [Online; accessed 14-May-2017].

[125] Bo Hjorth. EEG analysis based on time domain properties. *Electroencephalography and clinical neurophysiology*, 29(3):306–310, 1970.

[126] Timothy J Hodgetts, Gary Kenward, Ioannis G Vlachonikolis, Susan Payne, and Nicolas Castle. The identification of risk factors for cardiac arrest and formulation of activation criteria to alert a medical emergency team. *Resuscitation*, 54(2):125–131, 2002.

[127] Randy Hoffman. *Measurements to determine potential interference to GPS receivers from ultrawideband transmission systems*. US Department of Commerce, National Telecommunications and Information Administration, 2001.

[128] T Hoffmann, B Eilebrecht, and S Leonhardt. Respiratory monitoring system on the basis of capacitive textile force sensors. *IEEE Sensors Journal*, 11(5):1112–1119, 2011.

[129] Jacqueline Hogan. Why don't nurses monitor the respiratory rates of patients? *British Journal of Nursing*, 15(9), 2006.

[130] HyunDae Hong and Martin D Fox. Noninvasive detection of cardiovascular pulsations by optical Doppler techniques. *Journal of biomedical optics*, 2(4):382–390, 1997.

[131] John R Howell, M Pinar Menguc, and Robert Siegel. *Thermal radiation heat transfer*. CRC press, 2010.

[132] WJ Hrushesky, D Fader, O Schmitt, and V Gilbertsen. The respiratory sinus arrhythmia: a measure of cardiac age. *Science*, 224(4652):1001–1004, 1984.

[133] Heikki V Huikuri, MJ Niemelä, Sirkku Ojala, Asko Rantala, MJ Ikäheimo, and KE Airaksinen. Circadian rhythms of frequency domain measures of heart rate variability in healthy subjects and patients with coronary artery disease. effects of arousal and upright posture. *Circulation*, 90(1):121–126, 1994.

[134] Kenneth Humphreys. *An investigation of remote non-contact photoplethysmography and pulse oximetry*. PhD thesis, National University of Ireland Maynooth, 2007.

[135] Kenneth Humphreys, Tomas Ward, and Charles Markham. Noncontact simultaneous dual wavelength photoplethysmography: A further step toward noncontact pulse oximetry. *Review of Scientific Instruments*, 78(4):044304–044306, Apr 2007.

[136] Aapo Hyvärinen, Jarmo Hurri, and Patrick O Hoyer. *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision.*, volume 39. Springer Science & Business Media, 2009.

[137] Lassi Hyvärinen. Principal component analysis. *Mathematical Modeling for Industrial Processes*, 19(6):82–104, 1970.

[138] Ilias Iakovidis and Octavian Purcarea. eHealth in Europe: from Vision to Reality. *Studies in health technology and informatics*, 134:163–168, 2008.

[139] IEC IEC. 62304: 2006 medical device software–software life cycle processes. *International Electrotechnical Commission, Geneva*, 2006.

[140] Nonin Medical Inc. Pulse Oximetry, . URL `http://www.nonin.com/PulseOximetry`. [Online; accessed 14-May-2017].

[141] Point Grey Research Inc. Grasshopper2 camera, . URL `http://ptgrey.com/grasshopper2-gige-cameras`. [Online; accessed 14-May-2017].

[142] Shuhei Ishida, Naruhiro Shiozawa, Yoshihisa Fujiwara, and Masaaki Makikawa. Electrocardiogram measurement during sleep with wearing clothes using capacitively coupled electrodes. In *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2647–2650, 2007.

[143] ISO. Medical electrical equipment. particular requirements for basic safety and essential performance of pulse oximeter equipment. Standard BSN EN ISO 80601-2-61:2011, British-Standard-Institution - International Organization for Standardization, Apr 2011.

[144] ITU Telecom. The world in 2009: ICT facts and figures, Oct 2009.

[145] SI Ivashov, VV Razevig, AP Sheyko, and IA Vasilyev. Detection of human breathing and heartbeat by remote radar. In *Progress in Electromagnetic Research Symposium*, pages 663–666, 2004.

[146] Carlos M Jarque and Anil K Bera. A test for normality of observations and regression residuals. *International Statistical Review*, pages 163–172, 1987.

[147] James R Jauchem, Ronald L Seaman, Heather M Lehnert, Satnam P Mathur, Kathy L Ryan, Melvin R Frei, and William D Hurt. Ultra-wideband electromagnetic pulses: Lack of effects on heart rate and blood pressure during two-minute exposures of rats. *Bioelectromagnetics*, 19(5):330–333, 1998.

[148] Vincent Jeanne, Frederik De Bruijin, Ruud Vlutters, Giovanni Cennini, Dmitri Chestakov, et al. Processing images of at least one living being, Sep 2010. WO Patent Application 2,010,100,594.

[149] Anders Johansson, Per Åke Öberg, and Gunnar Sedin. Monitoring of heart and respiratory rates in newborn infants using a new photoplethysmographic technique. *Journal of clinical monitoring and computing*, 15(7-8):461–467, 1999.

[150] Alistair EW Johnson, Joachim Behar, Fernando Andreotti, Gari D Clifford, and Julien Oster. R-peak estimation using multimodal lead switching. In *Computers in Cardiology*, pages 281–284, 2014.

[151] Deric P Jones. Medical electro-optics: measurements in the human microcirculation. *Physics in Technology*, 18(2):79, 1987.

[152] Amal Jubran. Pulse oximetry. *Intensive care medicine*, 30(11):2017–2020, 2004.

[153] David M Kagle, Christian M Alexander, Robert S Berko, M Giuffre, and Jeffrey B Gross. Evaluation of the Ohmeda 3700 pulse oximeter: steady-state and transient response characteristics. *Anesthesiology*, 66(3):376–380, 1987.

[154] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82(1):35–45, 1960.

[155] Walter Karlen, Srinivas Raman, J Mark Ansermino, and Guy A Dumont. Multiparameter respiratory rate estimation from the photoplethysmogram. *IEEE Transactions on Biomedical Engineering*, 60(7):1946–1953, 2013.

[156] S Käste, F Noller, S Falk, A Bukta, E Mayer, and D Miller. Volunteer study for sensor calibration. *HP Journal*, pages 1–3, 1997.

[157] Bertram G Katzung, Susan B Masters, Anthony J Trevor, et al. *Basic & clinical pharmacology*, volume 8. Lange Medical Books/McGraw-Hill New York, 2004.

[158] Steven M Kay and Stanley L Marple. Spectrum analysis—a modern perspective. *Proceedings of the IEEE*, 69(11):1380–1419, 1981.

[159] Joseph F Kelleher. Pulse oximetry. *Journal of clinical monitoring*, 5(1):37–62, 1989.

[160] M Kendall, A Stuart, and K Ord. *The advanced theory of statistics*, volume 1. Griffin, 6 edition, 1994.

[161] Raghbir Singh Khandpur. *Handbook of biomedical instrumentation*. Tata McGraw-Hill Education, 1987.

[162] Jong-min Kim and Hugh S Mathewson. Venous congestion affects arterial hemoglobin saturation measured by the pulse oximeter. *Anesthesiology*, 63(3):A174, 1985.

[163] Ko Keun Kim, Yong Kyu Lim, and Kwang Suk Park. The electrically noncontacting ECG measurement on the toilet seat using the capacitively-coupled insulated electrodes. In *26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 1, pages 2375–2378, 2004.

[164] Ko Keun Kim, Yong Gyu Lim, Jung Soo Kim, and Kwang Suk Park. Capacitive-Coupled ECG Measurement and Its Applications for Ubiquitous Healthcare. In *International Conference on Convergence Information Technology (ICCIT)*, pages 811–814, 2007.

[165] Bruce Koeppen and Bruce Stanton. *Berne and Levy Physiology*. Elsevier Health Sciences, 2008.

[166] Lingqin Kong, Yuejin Zhao, Liquan Dong, Yiyun Jian, Xiaoli Jin, Bing Li, and Yun Feng. Non-contact detection of oxygen saturation based on visible light imaging device using ambient light. *Optics Express*, 21:179–183, 2013.

[167] Antti Koski and Martti Juhola. Segmentation of digital signals based on estimated compression ratio. *IEEE Transactions on Biomedical Engineering*, 43(9):928–938, 1996.

[168] Alexei A Kostenko, Alexander I Nosich, and Irina A Tishchenko. Radar prehistory, Soviet side: three-coordinate L-band pulse radar developed in Ukraine in the late 30's. In *IEEE International Symposium Antennas and Propagation Society*, volume 4, pages 44–47, 2001.

[169] Kurt Kramer, James O Elam, George A Saxton, and William N Elam. Influence of oxygen saturation, erythrocyte concentration and optical depth upon the red and near-infrared light transmittance of whole blood. *American Journal of Physiology*, 165(1):229–246, 1951.

[170] Mark A Kramer. Nonlinear principal component analysis using autoassociative neural networks. *AIChE journal*, 37(2):233–243, 1991.

[171] Jure Kranjec, S Beguš, G Geršak, and J Drnovšek. Non-contact heart rate and heart rate variability measurements: A review. *Biomedical Signal Processing and Control*, 13:102–112, 2014.

[172] Bruce Krieger, Drew Feinerman, A Zaron, and F Bizousky. Continuous noninvasive monitoring of respiratory rate in critically ill patients. *CHEST Journal*, 90(5):632–634, 1986.

[173] Manjula Kurella Tamura, Kenneth E Covinsky, Glenn M Chertow, Kristine Yaffe, C Seth Landefeld, and Charles E McCulloch. Functional status of elderly adults before and after initiation of dialysis. *New England Journal of Medicine*, 361(16):1539–1547, 2009.

[174] S. Yunkap Kwankam. What e-Health can offer. *Bulletin of the World Health Organization*, 82:800 – 802, Oct 2004.

[175] Daniel T Lackland and Michael A Weber. Global burden of cardiovascular disease and stroke: hypertension at the core. *Canadian Journal of Cardiology*, 31(5):569–571, 2015.

[176] Donna L Lamping, Niculae Constantinovici, Paul Roderick, Charles Normand, Lynne Henderson, Susan Harris, Edwina Brown, Reinhold Gruen, and Christina Victor. Clinical outcomes, quality of life, and costs in the North Thames Dialysis Study of elderly people on dialysis: a prospective cohort study. *The Lancet*, 356(9241):1543–1550, 2000.

[177] Vagn Højkjær Larsen, Per-Henrik Christensen, Henrik Oxhøj, and Torben Brask. Impedance pneumography for long-term monitoring of respiration during sleep in adult males. *Clinical Physiology*, 4(4):333–342, 1984.

[178] Bok Y Lee. *Handbook of noninvasive diagnostic techniques in vascular surgery*. Appleton & Lange, US, 1981.

[179] June-Sok Lee, You-Young Jung, Byung-Soo Kim, and Sung-Jea Ko. An advanced video camera system with robust AF, AE, and AWB control. *IEEE Transactions on Consumer Electronics*, 47 (3):694 –699, Aug 2001.

[180] Kual-Zheng Lee, Pang-Chan Hung, and Luo-Wei Tsai. Contact-free heart rate measurement using a camera. *IEEE Canadian Conference Computer and Robot Vision*, pages 147–152, 2012.

[181] Thomas Martin Lehmann, Claudia Gonner, and Klaus Spitzer. Survey: Interpolation methods in medical image processing. *IEEE Transactions on Medical Imaging*, 18(11):1049–1075, 1999.

[182] Magdalena Lewandowska, Jacek Ruminski, Tomasz Kocejko, and Jedrzej Nowak. Measuring pulse rate with a webcam; a non-contact method for evaluating cardiac activity. *Federated Conference on Computer Science and Information Systems (FedCSIS)*, pages 405–410, 2011.

[183] Changzhi Li and Jenshan Lin. Complex signal demodulation and random body movement cancellation techniques for non-contact vital sign detection. In *IEEE MTT-S International Microwave Symposium Digest*, pages 567 –570, Jun 2008.

[184] Changzhi Li, Yanming Xiao, and Jenshan Lin. Robust overnight monitoring of human vital signs by a non-contact respiration and heartbeat detector. In *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2235 –2238, Sep 2006.

[185] Q Li and GD Clifford. Dynamic time warping and machine learning for signal quality assessment of pulsatile signals. *Physiological Measurement*, 33(9):1491, 2012.

[186] Qiao Li, Roger G Mark, and Gari D Clifford. Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a kalman filter. *Physiological measurement*, 29(1):15, 2008.

[187] Stan Z Li and Anil K Jain. *Handbook of face recognition*. springer, 2011.

[188] WS Lim, MM Van der Eerden, R Laing, WG Boersma, N Karalus, GI Town, SA Lewis, and JT Macfarlane. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax*, 58(5):377–382, 2003.

[189] Yong Gyu Lim, Ko Keun Kim, and Suk Park. ECG measurement on a chair without conductive contact. *IEEE Transactions on Biomedical Engineering*, 53(5):956–959, 2006.

[190] Dave Litwiller. CCD vs. CMOS. *Photonics Spectra*, 35(1):154–158, 2001.

[191] Yan Liu, Shai H Zhu, Guo H Wang, Fei Ye, and Peng Z Li. Validity and reliability of multiparameter physiological measurements recorded by the Equivital LifeMonitor during activities of various intensities. *Journal of occupational and environmental hygiene*, 10(2): 78–85, 2013.

[192] Claire M Lochner, Yasser Khan, Adrien Pierre, and Ana C Arias. All-organic optoelectronic sensor for pulse oximetry. *Nature communications*, 5, 2014.

[193] Alfredo Lopez and Philip C Richardson. Capacitive electrocardiographic and bioelectric electrodes. *IEEE Transactions on Biomedical Engineering*, (1):99–99, 1969.

[194] Stavros Loukogeorgakis, Rex Dawson, Nirree Phillips, Christopher N Martyn, and Stephen E Greenwald. Validation of a device to measure arterial pulse wave velocity by a photoplethysmographic method. *Physiological measurement*, 23(3):581, 2002.

[195] A Louw, C Cracco, C Cerf, A Harf, P Duvaldestin, F Lemaire, and L Brochard. Accuracy of pulse oximetry in the intensive care unit. *Intensive care medicine*, 27(10):1606–1613, 2001.

[196] Paris B Lovett, Jason M Buchwald, Kai Stürmann, and Polly Bijur. The vexatious vital: neither clinical measurements by nurses nor an electronic monitor provides accurate measurements of respiratory rate in triage. *Annals of emergency medicine*, 45(1):68–76, 2005.

[197] Wei Lu, Michelle M Nystrom, Parag J Parikh, David R Fooshee, James P Hubenschmidt, Jeffrey D Bradley, and Daniel A Low. A semi-automatic method for peak and valley detection in free-breathing respiratory waveforms. *Medical physics*, 33(10):3634–3636, 2006.

[198] Andrew Lumb. *Nunn's applied respiratory physiology*. Churchill Livingstone, Elsevier, 2010.

[199] Christen Lundsgaard, D Van Slyke, and Maude E Abbott. Cyanosis. *Canadian Medical Association Journal*, 13(8):601, 1923.

[200] N Lutter, E Kozma, C Raushmayer, and J Schuettler. Rapid changes in heart rate and oxygen saturation decrease the clinical performance of motion-resistant pulse oximeters. *Proceeding from the Society for Technology in Anesthesia Annual Meeting*, 2003.

[201] Y Maeda, M Sekine, T Tamura, A Moriya, T Suzuki, and K Kameyama. Comparison of reflected green light and infrared photoplethysmography. In *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2270–2272, 2008.

[202] Pierre Magnan. Detection of visible photons in CCD and CMOS: A comparative view. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 504(1):199–212, 2003.

[203] Marek Malik and A John Camm. Heart rate variability. *Clinical cardiology*, 13(8):570–576, 1990.

[204] Edward Marcantonio, Sylvia McKean, Michael Goldfinger, Sharon Kleefield, Mark Yurkofsky, and Troyen Brennan. Factors associated with unplanned hospital readmission among patients 65 years of age and older in a medicare managed care plan. *The American journal of medicine*, 107(1):13–17, 1999.

[205] George Mardirossian and Roland E Schneider. Limitations of pulse oximetry. *Anesthesia progress*, 39(6):194, 1992.

[206] Slaheddine Marrakchi and Howard I Maibach. Biophysical parameters of skin: map of human face, regional, and age-related differences. *Contact dermatitis*, 57(1):28–34, 2007.

[207] Birgitta Martinkauppi. *Face colour under varying illumination-analysis and application*. PhD thesis, University of Oulu, 2002.

[208] Birgitta Martinkauppi and Graham D Finlayson. Designing a simple 3-channel camera for skin detection. In *Color and Imaging Conference*, volume 2004, pages 150–156. Society for Imaging Science and Technology, 2004.

[209] Laura Mason and Lionel Tarassenko. Quantitative assessment of respiratory derivation algorithms. In *Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 2, pages 1998–2001, 2001.

[210] T Matsui, S Suzuki, K Ujikawa, T Usui, S Gotoh, M Sugamata, Z Badarch, and S Abe. Development of a non-contact screening system for rapid medical inspection at a quarantine depot using a laser Doppler blood-flow meter, microwave radar and infrared thermography. *Journal of Medical Engineering and Technology*, 33(5):403–409, 2009.

[211] Jackie McBride, Debbie Knight, Jo Piper, and Gary B Smith. Long-term effect of introducing an early warning score on respiratory rate charting on general wards. *Resuscitation*, 65(1): 41–44, 2005.

[212] Thomas E McEwan. Body monitoring and imaging apparatus and method, Nov 1996. US Patent 5,573,012.

[213] Thomas E McEwan. Body monitoring and imaging apparatus and method, Jun 1998. US Patent 5,766,208.

[214] Paul E McKenney, Dipankar Sarma, Ingo Molnar, and Suparna Bhattacharya. Extending RCU for realtime and embedded workloads. In *Ottawa Linux Symposium*, pages 123–138, 2006.

[215] David McMaid, Michael Drummond, and Marc Suhrcke. *How can European health systems support investment in and the implementation of population health strategies?* WHO Regional Office for Europe, 2008.

[216] Martina Meinke, Gerhard Müller, Jürgen Helfmann, and Moritz Friebel. Optical properties of platelets and blood plasma and their influence on the optical behavior of whole blood in the visible to near infrared wavelength range. *Journal of biomedical optics*, 12(1): 014024–014024, 2007.

[217] Yitzhak Mendelson and Joel C Kent. Variations in optical absorption spectra of adult and fetal haemoglobins and its effect on pulse oximetry. *IEEE Transactions on Biomedical Engineering*, 36(8):844–848, 1989.

[218] Shanthi Mendis, Pekka Puska, Bo Norrving, et al. *Global atlas on cardiovascular disease prevention and control*. World Health Organization, 2011.

[219] David Meredith. *Continuous Monitoring During Haemodialysis*. PhD thesis, University of Oxford, 2014.

[220] DJ Meredith, D Clifton, P Charlton, J Brooks, CW Pugh, and L Tarassenko. Photoplethysmographic derivation of respiratory rate: a review of relevant physiology. *Journal of medical engineering & technology*, 36(1):1–7, 2012.

[221] James H Merritt, Johnathan L Kiel, and William D Hurt. Considerations for human exposure standards for fast-rise-time high-peak-power electromagnetic pulses. *Aviation, space, and environmental medicine*, 66(6):586–589, 1995.

[222] Paul M Middleton, Gregory SH Chan, Emma O'Lone, Elizabeth Steel, Rebecca Carroll, Branko G Celler, and Nigel H Lovell. Spectral analysis of finger photoplethysmographic waveform variability in a model of mild to moderate haemorrhage. *Journal of clinical monitoring and computing*, 22(5):343, 2008.

[223] FC Mihm and Bruce D Halperin. Noninvasive detection of profound arterial desaturations using a pulse oximetry device. *Anesthesiology*, 2(1):85–87, 1985.

[224] Glenn A Millikan. The oximeter, an instrument for measuring continuously the oxygen saturation of arterial blood in man. *Review of scientific Instruments*, 13(10):434–444, 1942.

[225] David L Mills. Internet time synchronization: the network time protocol. *IEEE Transactions on Communications*, 39(10):1482–1493, 1991.

[226] Ben Minnaert and Peter Veelaert. A proposal for typical artificial light sources for the characterization of indoor photovoltaic applications. *Energies*, 7(3):1500–1516, 2014.

[227] John Gregory Mitchell and John Mitchell. *From telehealth to e-health: The unstoppable rise of e-health*. Commonwealth Department of Communications, Information Technology and the Arts, 1999.

[228] Jakob T Moller, Tom Pedersen, Lars S Rasmussen, Per F Jensen, Bente D Pedersen, Odd Ravlo, Niels H Rasmussen, Kurt Espersen, Nils W Johannessen, Jeffrey B Cooper, et al. Randomized evaluation of pulse oximetry in 20,802 patients; i. *Anesthesiology*, 78(3):436–444, 1993.

[229] George Moody. Spectral analysis of heart rate without resampling. *Proceedings of Computers in Cardiology Conference*, pages 715–718, 1993.

[230] George B Moody, Roger G Mark, Andrea Zoccola, and Sara Mantero. Derivation of respiratory signals from multi-lead ECGs. *Computers in cardiology*, 12(1985):113–116, 1985.

[231] Umberto Morbiducci, Lorenzo Scalise, Mirko De Melis, and Mauro Grigioni. Optical vibrocardiography: A novel tool for the optical monitoring of cardiac activity. *Annals of biomedical engineering*, 35(1):45–58, 2007.

[232] RW Morris, M Nairn, and TA Torda. A comparison of fifteen pulse oximeters. Part I: a clinical comparison; Part II: a test of performance under conditions of poor perfusion. *Anaesthesia and intensive care*, 17(1):62–73, 1989.

[233] Isar Mostafanezhad, Olga Boric-Lubecke, Victor Lubecke, and Danilo P Mandic. Application of empirical mode decomposition in removing fidgeting interference in Doppler radar life signs monitoring devices. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 340–343, 2009.

[234] Meinard Müller. *Information retrieval for music and motion*, volume 2. Springer, 2007.

[235] Vrudhula K Murthy, Thomas M Grove, George A Harvey, and L Julian Haywood. Clinical usefulness of ECG frequency spectrum analysis. In *Proceedings of the Annual Symposium on Computer Application in Medical Care*, page 610. American Medical Informatics Association, 1978.

[236] K Nakajima, T Tamura, and H Miike. Monitoring of heart and respiratory rates by photoplethysmography using a digital filtering technique. *Medical engineering & physics*, 18(5):365–372, 1996.

[237] Shamim Nemati, Atul Malhotra, and Gari D Clifford. Data fusion for improved respiration rate estimation. *EURASIP journal on advances in signal processing*, 2010:10, 2010.

[238] Nicolas Nesseler, Jean-Vincent Frénel, Yannick Mallédant, Jeff Morcet, Philippe Seguin, and Yoann Launey. Pulse oximetry and high-dose vasopressors: a comparison between forehead reflectance and finger transmission sensors. *Intensive Care Medicine*, 38:1718–1722, 2012.

[239] Bruce G Nickerson, Catherine Sarkisian, and Kevin Tremper. Bias and precision of pulse oximeters and arterial oximeters. *CHEST Journal*, 93(3):515–517, 1988.

[240] Ludwig Nicolai. Über sichtbarmachung, verlauf und chemische kinetik der oxyhämoglobinreduktion im lebenden gewebe, besonders in der menschlichen haut. *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere*, 229(1): 372–384, 1932.

[241] JA Nijboer and JC Dorlas. Comparison of plethysmograms taken from finger and pinna during anaesthesia. *British journal of anaesthesia*, 57(5):531–534, 1985.

[242] Lena Nilsson, Anders Johansson, and Sigga Kalman. Monitoring of respiratory rate in postoperative care using a new photoplethysmographic technique. *Journal of Clinical Monitoring and Computing*, 16(4):309–315, 2000.

[243] Lena Nilsson, Anders Johansson, and Sigga Kalman. Respiratory variations in the reflection mode photoplethysmographic signal. relationships to peripheral venous pressure. *Medical and Biological Engineering and Computing*, 41(3):249–254, 2003.

[244] Lena Nilsson, Anders Johansson, and Sigga Kalman. Respiration can be monitored by photoplethysmography with high sensitivity and specificity regardless of anaesthesia and ventilatory mode. *Acta anaesthesiologica scandinavica*, 49(8):1157–1162, 2005.

[245] M Nitzan, A Babchenko, D Shemesh, and J Alberton. Influence of thoracic sympathectomy on cardiac induced oscillations in tissue blood volume. *Medical and Biological Engineering and Computing*, 39(5):579–583, 2001.

[246] Paul A Obrist, Kathleen C Light, James A McCubbin, J Hutcheson, and J Lee Hoffer. Pulse transit time. *Behavior Research Methods & Instrumentation*, 10(5):623–626, 1978.

[247] Matthew Blake Oefinger. *Monitoring transient repolarization segment morphology deviations in Mouse ECG*. PhD thesis, Citeseer, 2006.

[248] Jörg Ottenbacher and Stephan Heuer. Motion artefacts in capacitively coupled ecg electrodes. In *World Congress on Medical Physics and Biomedical Engineering*, pages 1059–1062. Springer, Sep 2010.

[249] Allan F Pacela. Impedance pneumography - a survey of instrumentation techniques. *Medical and biological engineering*, 4(1):1–15, 1966.

[250] Robert Morris Page. The early history of radar. *Proceedings of the IRE*, 50(5):1232–1236, 1962.

[251] Kent F Palmer and Dudley Williams. Optical properties of water in the near infrared. *Journal of the Optical Society of America*, 64(8):1107–1110, 1974.

[252] Jiapu Pan and Willis J Tompkins. A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, (3):230–236, 1985.

[253] James P Payne and John Wendell Severinghaus. *Pulse oximetry*. Springer, 1986.

[254] Andrew J Peacock. Oxygen at high altitude. *British Medical Journal*, 317(7165):1063, 1998.

[255] Gale Pearson and Heather Duncan. Early warning systems for identifying sick children. *Paediatrics and Child Health*, 21(5):230–233, 2011.

[256] GuoChen Peng and Mark F Bocko. Non-contact ECG sensing employing gradiometer electrodes. *IEEE Transactions On Biomedical Engineering*, 60(1):179–183, 2013.

[257] Barbara Phillips, Kevin Mackway-Jones, and Sue Wieteska. *Advanced paediatric life support: the practical approach*. John Wiley & Sons, Limited, 2007.

[258] Marco AF Pimentel, David A Clifton, Lei Clifton, and Lionel Tarassenko. Probabilistic estimation of respiratory rate using gaussian processes. In *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2902–2905, 2013.

[259] Marcos Pinotti, Nicola Paone, Fernando A Santos, and Enrico P Tomasini. Carotid artery pulse wave measured by a laser vibrometer. In *Third International Conference on Vibration Measurements by Laser Techniques: Advances and Applications*, pages 611–616. International Society for Optics and Photonics, 1998.

[260] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics express*, 18 (10):10762–10774, May 2010.

[261] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions on Biomedical Engineering*, 58:7–11, 2011.

[262] Murray M Pollack, Urs E Ruttimann, and Pamela R Getson. Pediatric risk of mortality (PRISM) score. *Critical care medicine*, 16(11):1110–1116, 1988.

[263] CA Porret, N Stergiopulos, D Hayoz, HR Brunner, and JJ Meister. Simultaneous ipsilateral and contralateral measurements of vasomotion in conduit arteries of human upper limbs. *American Journal of Physiology-Heart and Circulatory Physiology*, 269(6):H1852–H1858, 1995.

[264] RJ Prance, A Debray, TD Clark, H Prance, M Nock, CJ Harland, and AJ Clippingdale. An ultra-low-noise electrical-potential probe for human-body scanning. *Measurement Science and Technology*, 11(3):291, 2000.

[265] John A Prior and Jack S Silberstein. *Physical diagnosis: the history and examination of the patient*. CV Mosby, 1977.

[266] Kari Pulli, Anatoly Baksheev, Kirill Kornyakov, and Victor Eruhimov. Real-time computer vision with OpenCV. *Communications of the ACM*, 55(6):61–69, 2012.

[267] Mohammed Qadeer, John Vargo, John Dumot, Rocio Lopez, Patricia Trolli, Tyler Stevens, Mansour Parsi, Madhusudhan Sanaka, and Gregory Zuccaro. Capnographic monitoring of respiratory activity improves safety of sedation for endoscopic cholangiopancreatography and ultrasonography. *Gastroenterology*, 136(5):1568–1576, 2009.

[268] LR Rabiner and BH Juang. Fundamental of speech recognition liveness verification in audio-video speaker authentication. *University of Canberra*, 1993.

[269] LL Randeberg, JH Bonesrønning, M Dalaker, JS Nelson, and LO Svaasand. Methemoglobin formation during laser induced photothermolysis of vascular skin lesions. *Lasers in Surgery and Medicine*, 34(5):414–419, 2004.

[270] John Rice. *Mathematical Statistics and Data Analysis*. Duxbury Press, second edition, 1995.

[271] Lisa M Richards, SM Shams Kazmi, Janel L Davis, Katherine E Olin, and Andrew K Dunn. Low-cost laser speckle contrast imaging of blood flow using a webcam. *Biomedical optics express*, 4(10):2269–2283, 2013.

[272] Mark Richardson. Principal component analysis, 2009. URL `http://people.maths.ox.ac.uk/richardsonm/SignalProcPCA.pdf`. [Online; accessed 1-Sep-2017].

[273] PC Richardson. The insulated electrode: A pasteless electrocardiographic technique. In *20th Annual conference on engineering in medicine and biology*, volume 9, pages 15–7, 1967.

[274] Andrew L Ries, Lela M Prewitt, and Jeffery J Johnson. Skin color and ear oximetry. *Chest*, 96 (2):287–290, 1989.

[275] Steven Rostedt and Darren V Hart. Internals of the RT patch. In *Proceedings of the Linux symposium*, volume 2007, 2007.

[276] Joseph Ruanaidh and William Fitzgerald. *Numerical Bayesian Methods Applied to Signal Processing (Statistics and Computing)*. Springer, 1996.

[277] Rangsal Ruangsuwana, Gordana Velikic, and Mark Bocko. Methods to extract respiration information from ECG signals. In *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 570–573, 2010.

[278] Mohammed Saeed, Mauricio Villarroel, Andrew T Reisner, Gari Clifford, Li-Wei Lehman, George Moody, Thomas Heldt, Tin H Kyaw, Benjamin Moody, and Roger G Mark. Multiparameter intelligent monitoring in intensive care ii (mimic-ii): a public-access intensive care unit database. *Critical care medicine*, 39(5):952, 2011.

[279] Alan V Sahakian, Willis J Tompkins, and John G Webster. Electrode motion artifacts in electrical impedance pneumography. *IEEE Transactions on Biomedical Engineering*, (6): 448–451, 1985.

[280] Pratik Sahindrakar, G de Haan, and I Kirenko. Improving motion robustness of contact-less monitoring of heart rate using video analysis. Master's thesis, Department of Mathematics and Computer Science, Eindhoven University of Technology, 2011.

[281] Stan Salvador and Philip Chan. FastDTW: Toward accurate dynamic time warping in linear time and space. In *KDD workshop on mining temporal and sequential data*. Citeseer, 2004.

[282] B McA. Saykrs. Analysis of heart rate variability. *Ergonomics*, 16:17–32, 1973.

[283] Lorenzo Scalise, Ilaria Ercoli, Paolo Marchionni, and Enrico Primo Tomasini. Measurement of respiration rate in preterm infants by laser Doppler vibrometry. In *IEEE International Workshop on Medical Measurements and Applications Proceedings (MeMeA)*, pages 657–661, 2011.

[284] Knut Schmidt-Nielsen. *Animal physiology: adaptation and environment*. Cambridge University Press, 1997.

[285] G.L.L.H. Schmitz. Video camera based photoplethysmography using video camera based photoplethysmography using ambient light. Master's thesis, Department of Mathematics and Computer Science, Eindhoven University of Technology, 2011.

[286] GLLH Schmitz. *Video camera based photoplethysmography using ambient light*. PhD thesis, Technische Universiteit Eindhoven, 2011.

[287] Lynn M Schnapp and Neal H Cohen. Pulse oximetry: uses and abuses. *Chest*, 98(5): 1244–1250, 1990.

[288] SB Schumacher. Monitoring vital signs to identify postoperative complications. *Medsurg nursing: official journal of the Academy of Medical-Surgical Nurses*, 4(2):142–145, 1995.

[289] BJ Semmes, MJ Tobin, JV Snyder, and A Grenvik. Subjective and objective measurement of tidal volume in critically ill patients. *CHEST Journal*, 87(5):577–579, 1985.

[290] John W Severinghaus and Joseph F Kelleher. Recent developments in pulse oximetry. *Anesthesiology*, 76(6):1018, 1992.

[291] John W Severinghaus and Karen H Naifeh. Accuracy of response of six pulse oximeters to profound hypoxia. *Anesthesiology*, 67(4):551–558, 1987.

[292] Christopher W Seymour, Jeremy M Kahn, Colin R Cooke, Timothy R Watkins, Susan R Heckbert, and Thomas D Rea. Prediction of critical illness during out-of-hospital emergency care. *Jama*, 304(7):747–754, 2010.

[293] Syed Ahmar Shah. *Vital sign monitoring and data fusion for pediatric triage*. PhD thesis, University of Oxford, 2012.

[294] Frank Shann, Kate Hart, and David Thomas. Acute lower respiratory tract infections in children: possible criteria for selection of patients for antibiotic therapy and hospital admission. *Bulletin of the World Health Organization*, 62(5):749, 1984.

[295] Ershad Sharifahmadian and Alireza Ahmadian. Adaptive signal processing algorithm for remote detection of heart rate (HR) using ultra-wideband waveforms based on principal component analysis. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5717–5720, 2009.

[296] CJ Sherry, DW Blick, TJ Walters, GC Brown, and MR Murphy. Lack of behavioural effects in non-human primates after exposure to ultrawideband electromagnetic radiation in the microwave frequency range. *Radiation research*, 143(1):93–97, 1995.

[297] Jianbo Shi and Carlo Tomasi. Good features to track. In *IEEE Conference on Computer Vision and Pattern Recognition CVPR*, pages 593–600, 1994.

[298] David Shier, Jackie Butler, and Ricki Lewis. *Hole's Human Anatomy*. McGraw-Hill, 2001.

[299] Jonathon Shlens. A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*, 2014.

[300] Bruno Siciliano and Oussama Khatib. *Springer handbook of robotics*. Springer Science & Business Media, 2008.

[301] Ole Siggaard-Andersen, Bent Nørgaard-Pedersen, and Jørgen Rem. Hemoglobin pigments. spectrophotometric determination of oxy-, carboxy-, met-, and sulfhemoglobin in capillary blood. *Clinica Chimica Acta*, 42(1):85–100, 1972.

[302] Neven Simicevic. Exposure of biological material to ultra-wideband electromagnetic pulses: dosimetric implications. *Health physics*, 92(6):574, 2007.

[303] EA Simoes, R Roark, S Berman, LL Esler, and J Murphy. Respiratory rate: measurement of variability over time and accuracy at different counting periods. *Archives of disease in childhood*, 66(10):1199–1203, 1991.

[304] Talia Sirkis, Yevgeny Beiderman, Sergey Agdarov, Yafim Beiderman, and Zeev Zalevsky. Monitoring blood vital bio signs using secondary speckle patterns. *Optics express*, 24(24): 27907–27917, 2016.

[305] Josef Sivic, Mark Everingham, and Andrew Zisserman. Who are you? – Learning Person Specific Classifiers from Video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[306] Ian Smith, John Mackay, Nahla Fahrid, and Don Krucheck. Respiratory rate measurement: a comparison of methods. *British Journal of Healthcare Assistants*, 5(1):18, 2011.

[307] NV Sneed and AD Hollerbach. Accuracy of heart rate assessment in atrial fibrillation. *Heart & lung: the journal of critical care*, 21(5):427–433, 1991.

[308] Leif Sörnmo and Pablo Laguna. *Bioelectrical signal processing in cardiac and neurological applications*. Academic Press, 2005.

[309] Enrique Spinelli and Marcelo Haberman. Insulating electrodes: a review on biopotential front ends for dielectric skin–electrode interfaces. *Physiological measurement*, 31(10):S183, 2010.

[310] Sebastijan Sprager, Denis Donlagic, and Damjan Zazula. Monitoring of basic human vital functions using optical interferometer. In *IEEE 10th International Conference on Signal Processing (ICSP)*, pages 1–4, 2010.

[311] JR Squire. An instrument for measuring quantity of blood and its degree of oxygenation in the web of the hand. *The American Heart Journal*, 4:331–339, 1940.

[312] Enrico M Staderini. UWB radars in medicine. *IEEE Aerospace and Electronic Systems Magazine*, 17(1):13–18, 2002.

[313] Matthias Steffen, Adrian Aleksandrowicz, and Steffen Leonhardt. Mobile noncontact monitoring of heart and lung activity. *IEEE Transactions on Biomedical Circuits and Systems*, 1(4):250–257, 2007.

[314] Karl A Stroetmann, Lutz Kubitschke, Simon Robinson, Veil Stroetmann, Kevin Cullen, David McDaid, World Health Organization, et al. *How can telehealth help in the provision of integrated care?* World Health Organization. Regional Office for Europe, 2010.

[315] Norbert Stubán, Niwayama Masatsugu, et al. Non-invasive calibration method for pulse oximeters. *Periodica Polytechnica Electrical Engineering*, 52(1-2):91–94, 2008.

[316] Thomas J Sullivan, Stephen R Deiss, and Gert Cauwenberghs. A low-noise, non-contact EEG/ECG sensor. In *IEEE Biomedical Circuits and Systems Conference*, pages 154–157, 2007.

[317] Nanfei Sun, Marc Garbey, Arcangelo Merla, and Ioannis Pavlidis. Imaging the cardiovascular pulse. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR*, volume 2, pages 416–421, 2005.

[318] Yu Sun, Sijung Hu, Vicente Azorin-Peris, Stephen Greenwald, Jonathon Chambers, and Yisheng Zhu. Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise. *Journal of biomedical optics*, 16:077010, 2011.

[319] Satoshi Suzuki, Takemi Matsui, Hiroshi Kawahara, Hiroto Ichiki, Jun Shimizu, Yoko Kondo, Shinji Gotoh, Hirofumi Yura, Bonpei Takase, and Masayuki Ishihara. A non-contact vital sign monitoring system for ambulances using dual-frequency microwave radars. *Medical and Biological Engineering and Computing*, 47:101–105, 2009.

[320] David B Swedlow, Vicki Running, and Sandara J Feaster. Ambient light affects pulse oximeters. *The Journal of the American Society of Anesthesiologists*, 67(5):865–865, 1987.

[321] Chihiro Takano and Yuji Ohta. Heart rate measurement based on a time-lapse image. *Medical engineering & physics*, 29(8):853–857, 2007.

[322] He Tan, Dengyu Qiao, and Ye Li. Non-contact heart rate tracking using doppler radar. In *International Conference on Systems and Informatics*, pages 1711 –1714, May 2012.

[323] Lionel Tarassenko, Laura Mason, and Neil Townsend. Multi-sensor fusion for robust computation of breathing rate. *Electronics Letters*, 38(22):1314–1316, 2002.

[324] MB Taylor and JG Whitwam. The accuracy of pulse oximeters. *Anaesthesia*, 43(3):229–232, 1988.

[325] NV Thakor, JG Webster, and WJ Tompkins. Optimal QRS detector. *Medical and Biological Engineering and Computing*, 21(3):343–350, 1983.

[326] William J Tharion, Mark J Buller, Cynthia M Clements, David Dominguez, Christina Sampsonis, Stephen P Mullen, Anthony J Karis, and Adam W Potter. Human Factors Evaluation of the Hidalgo Equivital EQ-02 Physiological Status Monitoring System. Technical report, DTIC Document, 2013.

[327] F Thiel, D Kreiseler, and F Seifert. Non-contact detection of myocardium's mechanical activity by ultrawideband RF-radar and interpretation applying electrocardiography. *Review of Scientific Instruments*, 80:114302, 2009.

[328] Matthew Thompson, Nigel Coad, Anthony Harnden, Richard Mayon-White, Rafael Perera, and David Mant. How well do vital signs identify children with serious infections in paediatric emergency care? *Archives of disease in childhood*, 94(11):888–893, 2009.

[329] WC Tinus and WHC Higgins. Early fire-control radars for naval vessels. *The Bell System Technical Journal*, 25(1):1–47, 1946.

[330] Marko Tkalcic, Jurij F Tasic, et al. Colour spaces: perceptual, historical and applicational background. In *Eurocon*, 2003.

[331] Lloyd N Trefethen and David Bau III. *Numerical linear algebra*, volume 50. Siam, 1997.

[332] Kevin K Tremper. Transcutaneous PO2 measurement. *Canadian Anaesthetists Society Journal*, 31(6):664–677, 1984.

[333] Kevin K Tremper. Pulse oximetry. *CHEST Journal*, 95(4):713–715, 1989.

[334] Gill R Tsouri, Survi Kyal, Sohail Dianat, and Lalit K Mestha. Constrained independent component analysis approach to nonobtrusive pulse rate measurements. *Journal of Biomedical Optics*, 17(7), 2012.

[335] Ethel Tur, Moshe Tur, Howard I Maibach, and Richard H Guy. Basal perfusion of the cutaneous microcirculation: measurements as a function of anatomic position. *Journal of investigative dermatology*, 81(5):442–446, 1983.

[336] Maki Uenoyama, Takemi Matsui, Kouske Yamada, Satoshi Suzuki, Bonpei Takase, Shinya Suzuki, Masayuki Ishihara, and Mitsuyuki Kawakami. Non-contact respiratory monitoring system using a ceiling-attached microwave antenna. *Medical and Biological Engineering and Computing*, 44:835–840, 2006.

[337] Michael S Urschitz, Judith Wolff, Volker von Einem, Pilar M Urschitz-Duprat, Martin Schlaud, and Christian F Poets. Reference values for nocturnal home pulse oximetry during sleep in primary school children. *CHEST Journal*, 123(1):96–101, 2003.

[338] Emilio Vanoli, Philip B Adamson, Gian D Pinna, Ralph Lazzara, William C Orr, et al. Heart rate variability during specific sleep stages a comparison of healthy subjects with patients after myocardial infarction. *Circulation*, 91(7):1918–1922, 1995.

[339] Nielufar Varjavand, Jonathan Kaye, Shaojie Wang, and Frank Primiano. The interactive oxyhemoglobin dissociation curve, 2002.

[340] Jennifer A Veitch and Shelley L McColl. Modulation of fluorescent light: Flicker rate and light source effects on visual performance and visual comfort. *Lighting Research and technology*, 27(4):243–256, 1995.

[341] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson. Remote plethysmographic imaging using ambient light. *Optics express*, 16(26):21434–21445, 2008.

[342] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Conference on Computer Vision and Pattern Recognition CVPR*, volume 1, pages I–511, 2001.

[343] Michael J Vrhel and H Joel Trussell. Color device calibration: A mathematical formulation. *IEEE Transactions on Image Processing*, 8(12):1796–1806, 1999.

[344] Brian K Walsh, David N Crotwell, and Ruben D Restrepo. Capnography/Capnometry during mechanical ventilation: 2011. *Respiratory care*, 56(4):503–509, 2011.

[345] Thomas J Walters, PA Mason, CJ Sherry, C Steffen, and James H Merritt. No detectable bioeffects following acute exposure to high peak power ultra-wide band electromagnetic radiation in rats. *Aviation, space, and environmental medicine*, 66(6):562–567, 1995.

[346] Nick Waltham. CCD and CMOS sensors. In *Observing Photons in Space*, pages 423–442. Springer, 2013.

[347] Chen-Chia Wang, Sudhir B Trivedi, Feng Jin, Serguei Stepanov, Zhongyang Chen, Jacob Khurgin, Ponciano Rodriguez, and Narasimha S Prasad. Human life signs detection using high-sensitivity pulsed laser vibrometer. *IEEE Sensors Journal*, 7(9):1370–1376, 2007.

[348] G. Wang, D. Renshaw, P.B. Denyer, and M. Lu. CMOS video cameras. In *Euro ASIC '91*, pages 100 –103, May 1991.

[349] T Wartzek, S Weyer, and S Leonhardt. A differential capacitive electrical field sensor array for contactless measurement of respiratory rate. *Physiological measurement*, 32(10):1575, 2011.

[350] Tobias Wartzek, Thomas Lammersen, Benjamin Eilebrecht, Marian Walter, and Steffen Leonhardt. Triboelectricity in capacitive biopotential measurements. *IEEE Transactions on Biomedical Engineering*, 58(5):1268–1277, 2011.

[351] David B Wax, Philip Rubin, and Steven Neustein. A comparison of transmittance and reflectance pulse oximetry during vascular surgery. *Anesthesia and analgesia*, 109: 1847–1849, 2009.

[352] RK Webb, AC Ralston, and WB Runciman. Potential errors in pulse oximetry. *Anaesthesia*, 46(3):207–212, 1991.

[353] John G Webster. *Design of pulse oximeters*. CRC Press, 1997.

[354] Greg Welch and Gary Bishop. An introduction to the kalman filter. *Proceedings of the SIGGRAPH Course - Los Angeles*, 2001.

[355] DK Whitaker. Time for capnography–everywhere. *Anaesthesia*, 66(7):544–549, 2011.

[356] Fokko P Wieringa, Frits Mastik, and A FW van der Steen. Contactless multiple wavelength photoplethysmographic imaging: a first step toward "SpO2 camera" technology. *Annals of biomedical engineering*, 33(8):1034–1041, 2005.

[357] FP Wieringa, F Mastik, RH Boks, A Visscher, A Bogers, and AFW Van der Steen. In vitro demonstration of an SpO2-camera. In *Computers in Cardiology*, pages 749–751, 2007.

[358] Earl H Wood and JE Geraci. Photoelectric determination of arterial oxygen saturation in man. *The Journal of laboratory and clinical medicine*, 34(3):387, 1949.

[359] R Andrew Wood. Uncooled thermal imaging with monolithic silicon focal planes. In *SPIE's 1993 International Symposium on Optics, Imaging, and Instrumentation*, pages 322–329. International Society for Optics and Photonics, 1993.

[360] World Health Organization. Electromagnetic fields and public health effects of EMF on the environment, 2005. URL `http://www.who.int/peh-emf/publications/facts/envimpactemf_infosheet.pdf`. [Online; accessed 14-May-2017].

[361] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics*, 31(4):65, 2012.

[362] Michael Wukitsch, Michael Petterson, David Tobler, and Jonas Pologe. Pulse oximetry: analysis of theory, technology, and practice. *Journal of Clinical Monitoring and Computing*, 4(4):290–301, 1987.

[363] Gunter Wyszecki and Walter Stanley Stiles. *Color science: concepts and methods, quantitative data and formulae*. John Wiley & Sons, 1982.

[364] Feng Xiao, Joyce E Farrell, Jeffrey M DiCarlo, and Brian A Wandell. Preferred color spaces for white balancing. In *Electronic Imaging 2003*, pages 342–350. International Society for Optics and Photonics, 2003.

[365] Ming Yang, Qiong Liu, T. Turner, and Ying Wu. Vital sign estimation from passive thermal video. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, Jun 2008.

[366] Fumihiko Yasuma and Jun-ichiro Hayano. Respiratory sinus arrhythmia: why does the heartbeat synchronize with respiratory rhythm? *Chest Journal*, 125(2):683–690, 2004.

[367] I Yoshiya, Y Shimada, and K Tanaka. Spectrophotometric monitoring of arterial oxygen saturation in the fingertip. *Medical and Biological Engineering and Computing*, 18(1):27–32, 1980.

[368] Bill X Yu. An innovative non-contact sensing platform to prevent traffic accident due to driver drowsiness, 2011.

[369] Chenggang Yu, Zhenqiu Liu, Thomas McKenna, Andrew T Reisner, and Jaques Reifman. A method for automatic identification of reliable heart rates calculated from ECG and PPG waveforms. *Journal of the American Medical Informatics Association*, 13(3):309–320, 2006.

[370] Xiaogang Yu, Changzhi Li, and Jenshan Lin. System level integration of handheld wireless non-contact vital sign detectors. In *IEEE Radio and Wireless Symposium*, pages 514 –517, Jan 2009.

[371] Zhen Zhu, Jin Fei, and I Pavlidis. Tracking human breath in infrared imaging. In *Fifth IEEE Symposium on Bioinformatics and Bioengineering*, pages 227 – 231, Oct 2005.

[372] WG Zijlstra, A Buursma, and WP Meeuwsen-Van der Roest. Absorption spectra of human fetal and adult oxyhemoglobin, de-oxyhemoglobin, carboxyhemoglobin, and methemoglobin. *Clinical chemistry*, 37(9):1633–1638, 1991.

[373] Willem Gerrit Zijlstra, Anneke Buursma, and Onno Willem van Assendelft. *Visible and near infrared absorption spectra of human and animal haemoglobin: determination and application*. VSP, 2000.

[374] W Zong, T Heldt, GB Moody, and RG Mark. An open-source algorithm to detect onset of arterial blood pressure pulses. In *Computers in Cardiology*, pages 259–262, 2003.

[375] W Zong, GB Moody, and D Jiang. A robust open-source algorithm to detect onset and duration of QRS complexes. In *Computers in Cardiology*, pages 737–740, 2003.