Ultrasound Segmentation Tools And Their Application To Assess Fetal Nutritional Health

T.M.Rackham

Wadham College

Dept. of Engineering Science

Supervised by
Professor Alison Noble
Submitted: Trinity Term,

October 28, 2015

Thesis

Declaration

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

T. M. Rackham,

Wadham College

Acknowledgements

This thesis would not have been possible without the supervision of Professor Alison Noble, assistance of Sylvia Rueda and data acquired by Caroline Knight and the INTERGROWTH-21st and INTERBIO-21st projects.

The work into fetal face characterisation was achieved with the cooporation of Professor Peter Hammond and Mike Suttie of University College London, and data aquired by Hein Odendaal.

The facial mesh work was conducted in collaboration with Richard Stebbing.

I would also like to thank John and Edna Rackham for their continious and unwavering support of everything I have ever done. Also Phill Brown for playing devil's advocate and providing coffee.

This research has been funded by the RCUK Digital Economy Programme and Wellcome Trust.

Abstract

Maternal diet can have a great impact on the health and development of the fetus. Poor fetal nutrition has been linked to the development of a set of conditions in later life, such as coronary heart disease, type 2 diabetes and hypertension, while restricted growth can result in hypogylcemia, hypocalcemia, hypothermia, polycythemia, hyperbilirubinemia and cerebral palsy. High alcohol consumption during pregnancy can result in Fetal Alcohol Syndrome, a condition that can cause growth retardation, lowered intelligence and craniofacial defects. Current biometric assessment of the fetus involves size-based measures which may not accurately portray the state of fetal development, since they cannot differentiate cases of small-but-healthy or large-but-unhealthy fetuses.

This thesis aims to outline a set of more appropriate measures of accurately capturing the state of fetal development. Specifically, soft tissue area and liver volume measurement are examined, followed by facial shape characterisation.

A number of tools are presented which aim to allow clinicians to achieve accurate segmentations of these landmark regions. These are modifications on the Live Wire algorithm, an interactive segmentation method in which the user places a number of anchor points and a minimum cost path is calculated between the previous anchor point and the cursor. This focuses on giving the clinician intuitive control over the exact position of the segmented contour.

These modifications are FA-S Live Wire, which utilises Feature Asymmetry and a weak shape constraint, ASP Live Wire, which is a 3D expansion of Live Wire, and FA-O Live Wire, which uses Feature Asymmtery and Local Orientation to guide the segmentation process. These have have been designed with each of the specific biometric landmarks in mind.

Finally, a method of characterising fetal face shape is proposed, using a combination of the segmentation methods described here and a simple shape model with a parameterised b-spline meshing approach to facial surface representation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Poor fetal nutrition has been linked to a number of conditions occurring in later life such as coronary heart disease, type 2 diabetes and hypertension [64]. Restricted growth has been shown to correlate with several more immediate morbidities, such as hypogylcemia, hypocalcemia, hypothermia, polycythemia, hyperbilirubinemia and cerebral palsy [83, 110, 124]. Factors that can have a negative influence on fetal development include poor maternal diet or disruptions in placental function; when a mother is in a state of malnutrition, placental functionality is maintained at the cost of fetal development [8]. If the placenta itself is malfunctioning, the passage of nutrients between the mother and fetus is essentially blocked, resulting in growth restriction [178]. It has been suggested that, if detected before the levels of nutrition remain low for an extended period and while still in the fetal stage, the damaging effects of such *in utero* conditions can be reversed and corrected [8]. In animal studies where severe periods of malnutrition were introduced, fetal growth increased with improved nutrition after 9 and 16 days of malnutrition, but remained at a low rate of growth after 21 days (around 15% of the gestational period) [111]. This means that the signs of a fetus experiencing restricted growth due to poor nutrition should be detected as early as possible to allow for rapid intervention before permanent damage occurs.

Similarly, maternal diet in the form of high alcohol consumption during pregnancy can result in the development of long term conditions, specifically Fetal Alcohol Syndrome (FAS), a condition that results in growth retardation, lowered intelligence and craniofacial defects [172]. Burd and Martsolf claim FAS to be one of the leading causes of mental retardation in the United States, and the effects of other conditions symptomatic of FAS form a significant burden on their healthcare system (though it should be noted that this is a rather outdated paper) [24]. However, many of the physical symptoms of FAS have been observed to become less severe with time, meaning that the ability to conduct early diagnosis is essential in correctly identifying the condition [172].

Currently, the purpose of ultrasound (US) scans during pregnancy has been to detect gross fetal anomalies, as well as to date a pregnancy and assess the development of the fetus by taking quantitative size measurements, most often taken of the head, abdomen and femur [154, 158]. While this can inform a clinician about the rate at which the fetus is growing, one dimensional (1D) size measurements do not predict fetal body composition and therefore using such measures would be insufficient to accurately interpret fetal nutrition. Basic size measurements used to estimate weight can lack predictive accuracy, as the ranges of fetal weight that can be considered normal are actually very broad and do not give nutritional information [26]. As a result, classifications lack differentiation between fetuses that are small due to malnutrition and those that are naturally small [99]. Similarly, those diagnosed with FAS tend to be smaller than healthy fetuses, but size measures alone would not be enough to identify this condition, as it would be necessary to differentiate the diagnosis of FAS from Small for Gestational Age (SGA) or Intrauterine Growth Restricted (IUGR) fetuses.

Volume measurements have been viewed as impractical in the past due to the additional effort required to calculate these values, either via computationally-intensive automatic methods or time-intensive manual delineation [100]. In order to give an impression of fetal nutrition and provide more predictive measures of fetal health, determining fetal body composition is necessary. Improvements in the quality of US imaging and advances in image segmentation techniques mean that it is now possible to shift clinical observations towards even more predictive measures of fetal development, descriptive of both size and body composition. Studies such as INTERGROWTH-21st are designed to provide a large data set of standardised, carefully selected, healthy pregnancies, which could help to define what can be considered normal fetal growth [11]. This data set also provides the opportunity to discover new, currently unestablished biometric parameters to use in the assessment of fetal development.

So far, while there have been studies to investigate more effective measures of identifying FAS by performing facial analysis of children from 3D photography, identifying such conditions *in utero* has yet to be explored [103]. It is possible that similar techniques could be used on facial surfaces segmented from 3D fetal ultrasound images thereby providing methods to diagnose FAS far earlier than previously possible. This could be combined with segmentation methods developed for fetal volume measurement in order to simplify the system.

Aims and Thesis Structure

The aims of this thesis are to set the context of alternative methods for the assessment of health related to maternal dietary factors in fetal development studies, to characterise the challenges faced in the segmentation of ultrasound images and ultimately to propose possible solutions for suitable semi-automated segmentation methods. Figure 1.1 outlines the key targets of this thesis, including the different fetal health conditions being observed, imaging modalities, dimensionality of data and key components of the solutions for each of these objectives. This thesis is structured in the following way;

- **Chapter 2** outlines current trends in the assessment of fetal development and nutritional state, as well as studies into the alternative measures which have been shown to contribute to a more explicit view of fetal nutritional development. The need for engineering support in the form of tools for the segmentation of volumes from fetal ultrasound images are justified and existing techniques explored.

- **Chapter 3** reviews the literature of object segmentation in ultrasound images. This section is divided into the main facets that make up semi-automated segmentation methods; optimisation methods and image features.

- **Chapter 4** outlines a method for interactive segmentation, FA-S Live Wire, designed with the purpose of segmenting fetal fat tissue from 2D images in mind. This method is an adaptation of Live Wire, incorporating Feature Asymmetry (FA) and a weak shape constraint in order to better suit the original technique to the setting of ultrasound images. FA-S Live Wire is then validated against clinical segmentations.

- **Chapter 5** expands on the work described in Chapter 3, to allow for the segmentation of 3D liver volumes. An existing 3D Live Wire expansion is discussed and a new method, ASP Live Wire, is developed and compared against the former. This section looks at both segmentation in MRI and US acquisitions of the fetal liver.

- **Chapter 6** investigates the use of the fetal face shape to detect health conditions. ASP and FA Live Wire are applied to the segmentation of facial profiles and surfaces, while an Active Shape Model (ASM) is used to derive meaning from this data. Mesh representations of the facial surface have also been explored. This chapter deals first with the 2D fetal face profile and then the 3D fetal face surface extracted from US images as metrics for assessment of FAS.

- **Chapter 7** discusses the key developments and findings of this thesis and outlines potential avenues for expansion on this work. In particular, the initial stages of potential US imaging protocol which could facilitate the forms of fetal health assessment described here are discussed.

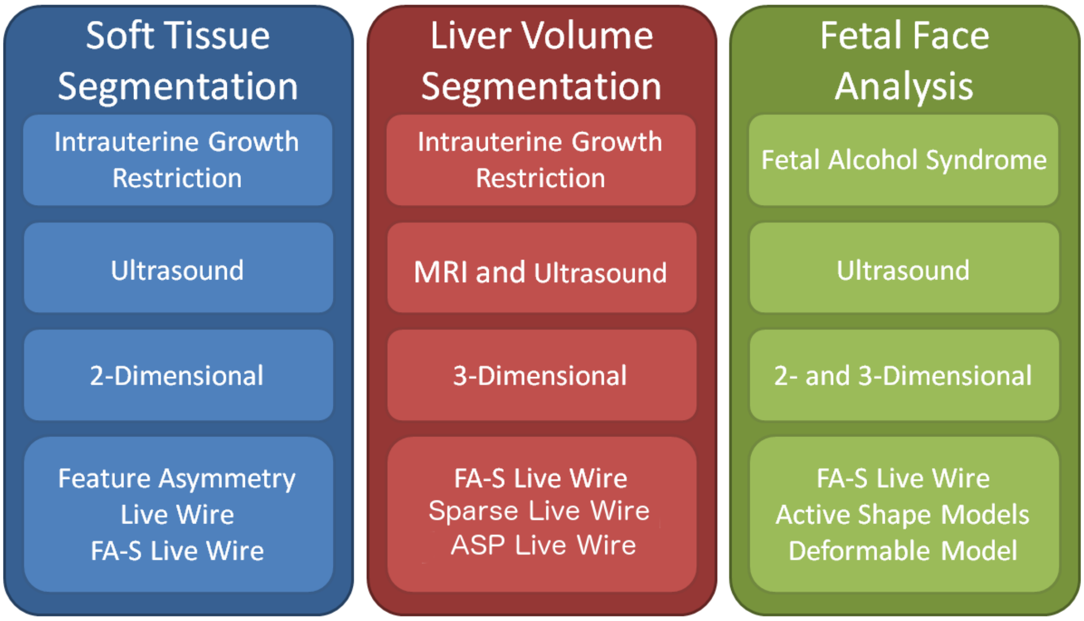


Figure 1.1: *Overview of thesis content, including the fetal health condition being observed, imaging modality used, dimensionality of data used and key components of the solutions*

List of Publications

- T.M. Rackham, S. Rueda, C.L. Knight, J.A. Noble, 'Ultrasound image segmentation using Feature Asymmetry and shape guided Live Wire', Proceedings of SPIE Medical Imaging 2013: Image Processing, 86690P, Orlando, FL, USA, February 2013
- T.M. Rackham, J.A. Noble, M. Suttie, P. Hammond, 'Fetal face segmentation and contour characterisation in 3D ultrasound' (abstract), MECbioeng14, London, 2014
- T.M. Rackham, C. Knight, J.A. Noble 'Volumetric segmentation of the fetal liver in 3D ultrasound' (poster), MECbioeng14, London, 2014

Chapter 2

Clinical Motivation

This literature review chapter discusses the clinical motivation for the objectives set out in this thesis, exploring standard fetal biometry and comparing these to potentially more effective means of measuring fetal health through US imaging.

2.1 Fetal Growth Restriction

2.1.1 Anomalous Fetal Growth

It is important that the distinction is made between fetuses that are SGA and those that are expressing IUGR/Fetal Growth Restriction (FGR). SGA fetuses are defined as those which are in the lower percentiles of expected growth (often thresholded at 5% or 10%) while IUGR fetuses are specifically those which are small due to abnormally restricted intrauterine growth rather than just being naturally small [110]. It is believed that around 70-80% of SGA fetuses are not IUGR, which would represent a highly significant proportion of misdiagnosed cases [124, 167]. Consequently, the confusion between the definitions of SGA and IUGR has been cited as a weakness in many investigations which attempt to quantify the effects of FGR on fetal health [162].

As SGA is described as when a fetal size metric is below a certain percentile compared to growth charts constructed from groups of *normal* pregnancies, by definition that same percentage of normal fetuses will appear within the SGA region of the chart; consequently a fetus being SGA can imply FGR but does not define it [110]. Similarly, a fetus may be growth restricted but remain within the bounds indicative of nominal growth [110]. Mayer and Joseph use the example of a fetus which, in early observations is within the 60th percentile of Estimated Fetal Weight (EFW), but falls to lower percentiles later in the pregnancy while nonetheless remaining above the 10th percentile; this would be an indication of a growth restricted fetus but is never classified as SGA [110].

Catalano *et al.* noted that a wide range of birth weights can be considered Appropriate for Gestational Age (AGA), while Lee *et al.* observed that classifying a fetus as SGA does not necessarily indicate malnourishment [26, 98]. Similarly, Viegas *et al.* stated that good screening practices should identify nutritional risk rather than just aim to predict low birth weights [192]. Measures of functional properties and estimates of fetal body composition rather than just projections of birth weight are required for sufficient sensitivity in identifying fetal nutritional problems and preventing nutritional insufficiency post-birth.

There are a number of influences which can result in a fetus experiencing FGR, which can be categorised broadly under placental, maternal and fetal influences:

- **Fetal:** Non-singleton pregnancies are far more likely to express FGR in their third trimester [124]. However, it is possible that an inaccurate estimate of gestational age (GA) could push the fetus into the lower 10th percentile of expected growth [167].

  Methodological differences can be responsible for such variation; for instance, in a study estimating fetal liver size, CH Chang *et al.* found that their measurements on a Taiwanese population were greatly different to those from a Dutch population, which were generally larger at a given GA [29]. Although there could be factors such as different maternal diet at play, Ioannou *et al.* considered the differences too great, attributing them instead to variation in methodology [80].

  The INTERGROWTH-21st study aimed to an international standard for fetal growth between geographically diverse populations through standardised measurement methodology, enrolment critera and equipment [139]. One outcome of this standardisation was a greatly reduced variation between study centres compared to other investigations [194], supporting the idea that methological differences may have a greater influence over fetal growth projections than geographic location. This also highlights the importance of consistent measures when designing a system for aiding the segmentation of fetal biometric landmarks.

According to some studies, a fetus may be more likely to be identified as SGA due to sex (observing female fetuses to be generally smaller) without expressing FGR [124]. However, taking birth weights of neonates in the INTERGROWTH-21st study against the GA at which they were born, Villar *et al.* showed near identical growth trajectories between the sexes [193], suggesting that even this may not be a factor in fetal development.

If this is indeed the case, the primary cause of deviation from predicted growth charts could be largely health-driven. Altered growth entirely due to fetal factors may be a sign of genetic abnormalities, chromosomal defects or infection [124].

- **Maternal:** Maternal age, both at the extreme older and younger end of reproductive ages, have been shown to express higher instances of FGR [178]. Maternal malnutrition can result in FGR, with fetal growth sacrificed in order to maintain placental functionality [8]. It has been established that poor nutrition during pregnancy can result in lower birth weights, with a study by Viegas *et al.* of 81 pregnancies showing a mean weight of 180g less in neonates that were identified as having a nutritional deficiency compared to those with adequate nutrition [192]. In this example, the distinction was made by conducting skinfold measurements on the triceps of the mother, with a less than 20μm/wk increase indicating poor nutrition.

  However, as the results were derived from a significant change in average weight between two groups, the weight measurement of an individual would not necessarily identify their fetus as being in a nutritionally sufficient or poor environment. The use of tobacco and drugs such as alcohol, heroin and cocaine, as well as medications such as warfarin and folic acid antagonists have also been related to FGR [124,178]. The INTERGROWTH-21st study aimed to model optimal fetal growth and, as a result, included only mothers who fitted a set criteria; healthy, well nourished and with low exposure to pollution, radiation, tobacco smoke and other hazardous substances [193].

It is important that mothers are aware of these risks and take appropriate measures during pregnancy in order to avoid preventable fetal health problems, with additional focus placed on cases where FGR has been identified.

- **Placental:** As the main passage for nutrients between the mother and fetus, severe problems with the placenta can often result in the death of the fetus [118]. For example, if the placenta fails to implant properly, blood flow can be restricted, thereby hindering the transfer of nutrients between mother and fetus [124]. Such conditions affecting the placenta that do not result in fetal death may result in FGR [178]. It is currently not possible for clinical intervention to improve placental insufficiency [124].

2.1.2 Effects of Growth Restriction

FGR is a major cause of perinatal mortality, with an investigation by Frøen *et al.* observing IUGR in 52% of a population of "sudden intrauterine unexplained death" cases compared to only 13% in a healthy population [57]. It should be noted that in this case their criteria is actually identifying SGA neonates (by using birth weight compared to expected values) and therefore only implicitly IUGR, which may alter these proportions. However, in a study of 107 pregnancies identified as IUGR due to placental insufficiency, the mortality of the group was 50%, which is in alignment with the findings of Frøen *et al.* [57,127]. In addition to this, complications during birth were common, with 53% of the mothers suffering from preeclampsia and 13% being affected by HELLP syndrome [127].

FGR is also associated with several newborn morbidities, such as hypogylcemia, hypocalcemia, hypothermia, polycythemia and hyperbilirubinemia [110, 124]. Fetuses identified as IUGR are also more likely to suffer from intrapartum asphxia and premature birth [124]. In a study of 4,503 cases of cerebral palsy in singleton pregnancies, it was found that those with weights below the 10th percentile are far more likely to have the condition [83].

Poor intrauterine growth has also been related to the onset of various conditions in later life; Barker *et al.* described the relationship between low birth weights and the onset of cardiovascular disease [8]. A later, more comprehensive study by Godfrey and Barker showed a correlation between disproportionately small neonates and conditions such as coronary heart disease, hypertension and type 2 diabetes in adulthood [64]. Fetuses identified as being SGA have also been shown to have a higher chance of long-term neurocognitive impairments; neurocognitive disabilities are more common among premature SGA infants, while higher instances of poor academic performance, behavioral issues and depression have been identified in both term and preterm SGA infants [162].

The onset of such conditions have been related to malnutrition rather than to just relative size for GA, for example maternal pelvic size does not correlate with the onset of cardiovascular disease in adulthood [8]. It was proposed that these changes in fact occur in response to poor nutrition during fetal development, favouring short-term survival (reaching reproductive age) at the cost of health in later adult life and that the diet of the mother during pregnancy can have an impact on these mechanisms [64]. Shah refers to a 'dive reflex' wherein the fetus supplies blood to the developing brain, at the detriment of liver development and the growth of subcutaneous tissue [162]. Similarly, in Boito's study into brain/liver volume ratios of SGA fetuses, the 'brain sparing effect' is described as a mechanism which prevents loss of the fetus's natural brain volume but causes retarded growth in the rest of the fetus [15]. Geva *et al.* point out that this preservation of brain growth is often insufficient in preventing neurocognitive disabilities, although the mechanisms behind this disparity are yet to be found [62].

Postnatal growth has been shown to have an important influence on the development of some of the negative outcomes of FGR, with children who had been identified as IUGR but expressing 'catch-up growth' within their first two years, exhibiting reduced instances of neurological impairment compared to those who did not [162]. It has been suggested that this could justify the use of growth hormone in infants who had been identified as IUGR in order to prevent later associated complications, although the long term side-effects of the treatment itself requires further investigation [162]. Detecting FGR can also assist in preventing short-term complications, aiding with delivery planning, which could take the form of induced labour and/or the preparation of facilities for care of a vulnerable neonate [124, 167].

2.2 Effects of Fetal Alcohol Syndrome

Growth trajectory is not the only influence maternal nutrition can have on the fetus; maternal alcohol intake is another factor which can lead to detrimental fetal health outcomes. FAS was initially described by Jones and Smith in 1973, who identified a common set of characteristics in the children of pregnancies associated with maternal alcoholism [84]. Signs of FAS can include, in broad terms, growth retardation (both pre- and post-natally), mental deficiencies and a characteristic set of craniofacial deformities [24, 181]. Failure to identify such conditions will mean a lack of appropriate support, potentially resulting in secondary conditions such as depression [6].

Despite a clearly defined set of possible markers, it is the sheer range of conditions that an individual patient can present with that often makes FAS difficult to diagnose, with clinical disagreement over the exact combination and severity of signs that can merit a diagnosis [109]. Additionally, potential indicators such as growth retardation could be symptomatic of a number of other fetal conditions (such as those outlined in Section 2.1.1). On top of this, studies have shown that the exact expression of FAS can vary with the timing of alcohol exposure; for example, Lipinksi *et al.* found that mice exposed to alcohol at different stages of development exhibited varying facial and brain

Table 2.1: *Results from Suttie* et al.*'s study, showing the prevalence of certain facial features associated with FAS in an FAS-diagnosed group and a non-FAS group [179]*

Sign	Description	FAS	Control
Flat philtrum	Flattening of the region between the base of the nose and top of the mouth	90.9%	21.7%
Thin vermilion border	Thin pigmentation region between the lips and the surrounding skin	90.9%	24.6%
Short palpebral fissures	Shortening of the distance between the extremities of the eyelid	81.8%	11.6%

deformations [103]. Duration and severity of the alcohol exposure can also have an impact on the exact effects of FAS [181]. This means that objective, discriminative methods of diagnosing a condition such as FAS are in high demand.

Due to the fact that craniofacial features have been identified as a qualitative metric for diagnosing FAS, studies have looked into quantifying the differences between FAS and non-FAS facial features. Suttie *et al.* observed the differences in the facial features of young children via 3D photography in a cohort of 192 subjects, with their investigation including individuals who were subject to heavy alcohol exposure during development but did not present with FAS [179]. Their FAS group showed high instances of specific facial dysmorphism compared to the control group, such as flat philtrums, a thin vermilion border around the mouth and short palpebral fissures [179]. Details of these findings are shown in Table 2.1 and a schematic of the different facial features is shown in Figure 2.1. The children observed, who were aged between 6 and 11, were also shorter (127.0 ± 13.1 FAS, 135.6 ± 14.1 control (cm)) and weighed less (25.8 ± 7.2 FAS, 34.2 ± 13.2 control (kg)) than the controls, supporting the notion that growth retardation caused by FAS *in utero* can persist into childhood [179].

Spohr *et al.* studied the continued effect of FAS on a cohort of 60 children diagnosed in infancy over the course of ten years, and showed that while many of the physical abnormalities symptomatic of FAS lessened over time and catch-up growth tended to occur, microcephaly tended to persist (88% of the cohort were microcephalic in the first observation, 65% in the follow-up), as did mental disability [172]. The findings of Suttie

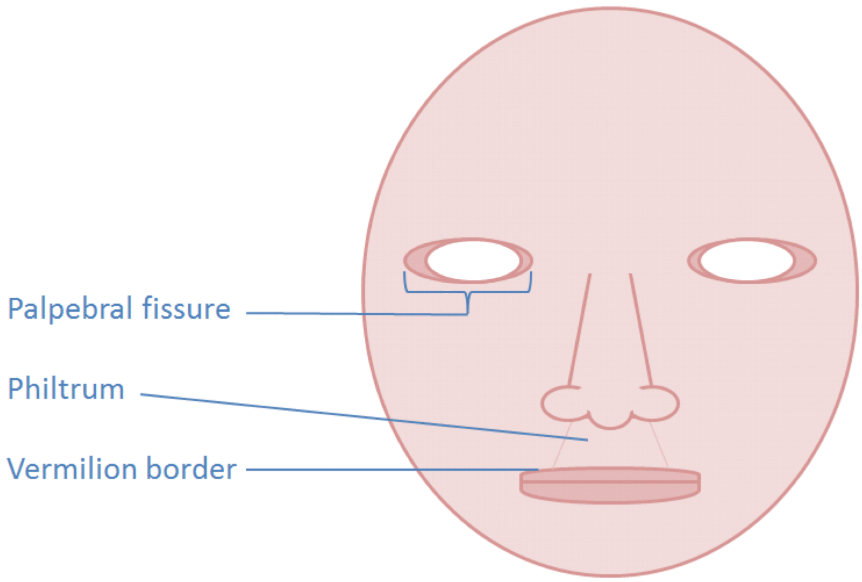


Figure 2.1: *Schematic showing some of the facial regions of interest used in the diagnosis of FAS.*

et al. confirm this, with their classification between FAS and non-FAS subjects showing sensitivity and specificity of 100% and 91% respectively and 5 years of age, but fell to 76% and 83% respectively at 12 years of age [179]. This shows that diagnosis of FAS should be established as early as possible, as later in life many of the physical characteristics of the condition may no longer be apparent. Indeed, Spohr *et al.* note that many of the patients in their study either expressed mild FAS or were difficult to diagnose at all in their follow-up observations [172].

2.3 Observing Fetal Development

Key points at which scans are made during a normal pregnancy are;

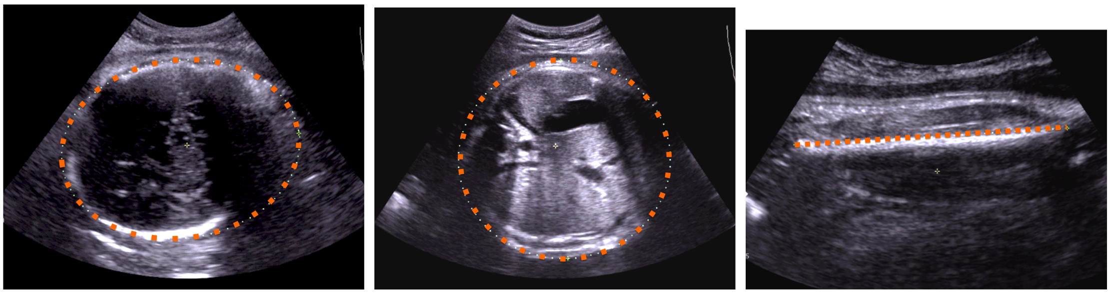
- **First Trimester:** At around 11-14 weeks, the number of fetuses is checked (assessing whether or not it is a singleton pregnancy), an estimation of the fetus's age can be made and conditions such as Down's Syndrome are screened for during this scan by measuring the nuchal translucency (the amount of swelling behind the fetal neck) [124, 169]. Bryan and Hindmarsh place the dating scan at around 10-12 weeks, stating the argument that this early in the pregnancy, inter-fetal variability is small enough to allow for an accurate estimation of GA [22].

Commonly, the crown-rump length is used to estimate the GA of a fetus in the first trimester, which will be around 50-90mm at this stage, in conjunction with the date of the start of the last menstrual period [190]. When calculating GA from the crown-rump length alone, Robinson and Fleming estimated systemic errors to be in the order of 1mm, relating to an error of approximately one day [152]. Further to this, they estimated a 95% confidence of ±4.7 days GA for a single measurement and ±2.7 days GA for three independent measurements [152].

Although it could be argued that an earlier scan means that systemic errors will have a greater effect on the estimation of GA, Bottomley and Bourne point out that several factors make early assessment a better option; specifically that there is low biological variation in fetal size at this age, and the fetus is not yet able to move and extend itself [18].

- **Second Trimester:** At 18-22 weeks, an assessment of fetal growth and of the placenta can be made, and an estimation of amniotic fluid volume is taken [124]. If the GA is estimated at this age, the biparietal diameter (BPD) is often used [186]. At this time a more complete investigation for abnormality is made, which involves both checking fetal structures for deformity and making biometric measurements. There are established guidelines to guide this process from groups such as the International Society of US in Obstetrics and Gynecology (ISUOG) and the Fetal Anomaly Screening Programme (FASP) [88, 157].

- **Third Trimester:** At 28-32 weeks, further assessment of fetal growth can be made (although this is not a part of general practice in the UK), wherein general fetal health is assessed and another estimation of the volume of amniotic fluid is made [124]. Bryan and Hindmash point out that there have been disputes over the validity of scans in the third trimester, and that these procedures are only deemed necessary in circumstances where poor growth is expected or has been observed in



(a) (b) (c)

Figure 2.2: *Examples of ellipse fitting to* ***(a)*** *the head and* ***(b)*** *the abdomen as well as* ***(c)*** *femur length measurement*

the second trimester scan [22]. They argue that in severe cases of late-onset IUGR, either clinical action will have been taken or the fetus will have died, diminishing the utility of such a scan [22]. This means that a high sensitivity of the second trimester scan to deficiencies in fetal growth is essential.

Current clinical fetal growth measurements are 1D, size-based measurements taken from 2D US images, most often taken from the fetal head, abdomen and femur. Their current roles are to estimate fetal weight, to make approximations of GA and to make assessments of fetal size development [154]. These measures are detailed by Salomon *et al.* and have been adopted by ISUOG as guidelines for fetal biometry [154, 158].

The details of these measurements are as follows:

- **Fetal head measurements**: The fetal head circumference (HC) can be estimated by taking measurements of the longest and widest sections of the skull (BPD and Occipito-Frontal Diameter (OFD)) taken on the same plane:

$$HC = \frac{\pi(BPD + OFD)}{2} \tag{2.1}$$

  Alternatively, it can be found by fitting an ellipse to the shape of the skull, as shown in Figure 2.2a.

- **Abdominal Measurements**: Abdominal measures (APAD and TAD) are also taken. The APAD is a measurement across a cross-section of the abdomen, measured from the posterior to anterior outer skin edges, intersecting the spine. The TAD is perpendicular to this, on the widest section of the abdomen. From this, the Abdominal Circumference (AC) is:

$$AC = \frac{\pi(TAD + APAD)}{2} \tag{2.2}$$

Similarly to HC, an ellipse can be fitted as an alternative measure, as shown in Figure 2.2b.

- **Femur Length:** The femur length (FL) is taken along the full length of the femur, observed in the horizontal plane. The landmarks in this case are simply either end of the bone, although in practice these two points can be poorly defined. Figure 2.2c shows an example of a FL measurement being made from an US image.

INTERGROWTH-21st has aimed to create an international standard for fetal growth using such measures [11]. Papageorghiou *et al.* outline the results from recording HC, BPD, OFD, AC and FL in 13,108 healthy singleton pregnancies [139]. These measurements were taken five times between 14 and 42 weeks GA, demonstrating the range of ages at which such measures can be taken accurately [139].

However, these are purely sized based measures and so only implicitly describe fetal nutrition. Another concern for routine clinical use is that inaccuracies could be caused by taking a 1D measurement on a 2D slice of a 3D object.

2.4 Alternative Measures for Fetal Growth

There are a number of different methods for assessing fetal growth, which are not included in standard fetal biometric measurements, but may provide additional insight into the development of the growing fetus. Such measurements include:

- Soft tissue measurement, including limb volume and body fat measurement. The work of Rueda *et al.* has focused on obtaining segmentations of 2D US images of fetal fat [155,156].

- Liver size measurement. The liver is clearly visible in some forms of Magnetic Resonance Imaging (MRI) scan, but can be particularly challenging in US. At the time of writing, no work has been reported on semi-automatic or automatic liver segmentation from 3D US.

- Cranial and brain analysis. This includes head shape, brain volume and the detection and analysis of brain structures. Namburete *et al.* have made efforts to perform fetal cranial segmentation, and also to predict fetal GA by learning the structural changes within the fetal brain through development [119,122].

- Face shape, either in the form of the profile or the full face surface. This is an area which has been studied widely in young children using 3D photography but is yet to expand into obstetric studies [45,70,182].

This section will outline relevant studies which utilise each of these methods, clinical significance and the technical advantages of using these biometric landmarks.

2.4.1 Soft Tissue Estimation

Estimates of the percentage of fetal body fat have the potential to differentiate 'small' and 'undernourished' fetuses, as it has been shown that abnormally thin neonates and altered fat distributions can be indicators of insufficient nutrition [8,100].

Catalano *et al.* were able to show that there is a significant relationship between birth weight and fat mass; in a study of 188 neonates, birth weight, fat mass according to the Dauncey approximation and hence lean mass (subtracting the fat mass from birth weight) were recorded within 24 hours of birth [26]. The Dauncey approximation models the arms, legs and torso as cylinders with a constant fat thickness and the head as a sphere containing no fat [46]. The results showed that while neonatal fat contributed to only 14% of the birth weight, it was responsible for 46% of the variance in birth weight. A poor relationship exists between the size-based ponderal index, defined as:

$$ponderalindex = \frac{weight}{length^3} \times 100 \tag{2.3}$$

and percentage body fat (%BF), demonstrating a clear discrepancy between size and nutritional state [26]. From this, it can be assumed that an EFW based on size measurements alone is unlikely to correlate well with fetal body composition. Conversely, they concluded that fat measurement may be a good predictor for neonatal weight.

Based on postmortem observations, Dauncey *et al.* found that subcutaneous fat accounts for 70-80% of total fat mass in newborn infants, suggesting that fat volume estimates could be suitable for predicting neonatal body composition [46]. Using *in vivo* estimates of fat thickness from US images, Gardeil *et al.* showed that subcutaneous fat around the abdomen could be used as a predictive indicator for fetal weight, with low thicknesses of abdominal fat indicating a significantly higher chance of abnormally low birth weight [59]. Significantly, more cases of low birth weight were found in the group of fetuses with less than 5mm abdominal fat thickness at 38 weeks than those with over 5mm. However, these measures provided only modest sensitivity (76.2%) and specificity (66.6%), showing poorer performance than conventional fetal weight estimations. Although related to fetal size, using a 1D thresholding measure is unlikely to be sufficient in assessing fetal size or body composition. A volumetric measure would allow for altered distributions of abdominal fat. Additionally, taking measurements at 38 weeks prevents

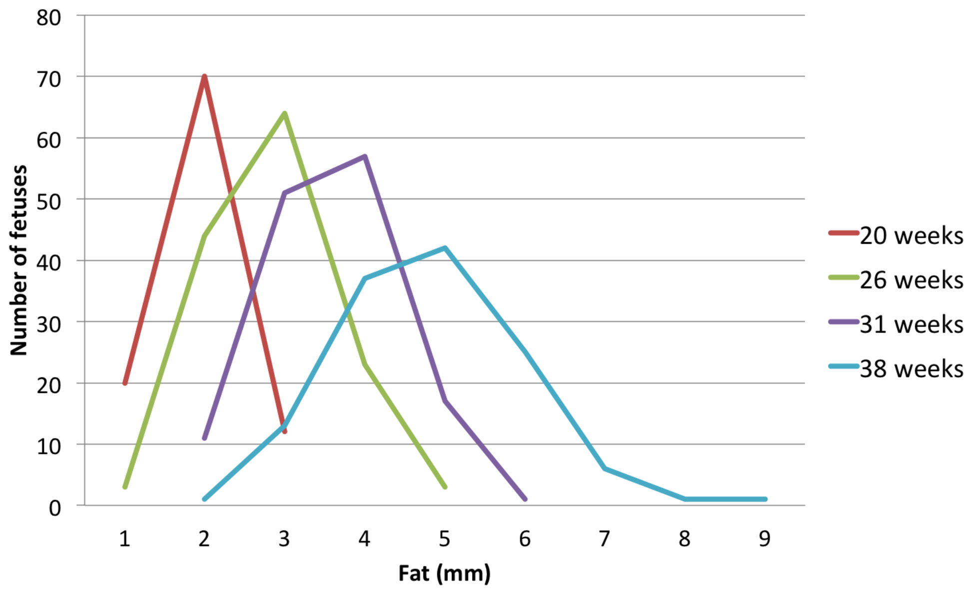


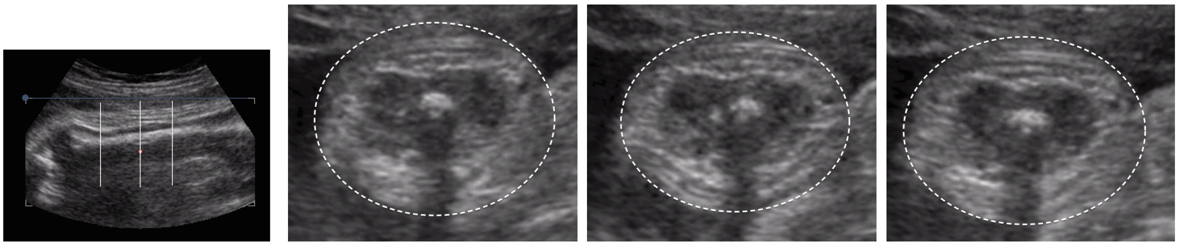
Figure 2.3: *Variation of abdominal subcutaneous fat thickness with GA as reported by Gardiel* et al. *[59]*

any significant ‘predictive’ value to these results; at this stage, the nutritional outcome of the fetus will be set.

Some studies claim that prior to around 30 weeks of gestation, little fat deposition is visible in US images, limiting how early such measurements can be taken and how much differentiation is possible earlier in pregnancy [161]. Nonetheless, Gardiel *et al.* were able to record subcutaneous fat around the abdomen at around 20 weeks GA in 74% of the fetuses in their study [59]. Figure 2.3 shows the spread of fat thicknesses with gestation, and shows how variation increases with GA. This suggests that malnutrition may be easier to detect later, but a balance must be struck between accurate diagnosis and the ability to influence growth trajectory. In a review for clinical interventions for FGR, Gulmezoglu *et al.* comment that nutritional supplements seem to have some impact on fetal weight outcomes, but timings of such interventions and whether long term damage has been avoided is yet to be investigated [67].

Hill *et al.* concluded that it was not possible to form a distinction between normal and abnormal size fetuses using subcutaneous tissue thickness [75]. This may be more indicative of the techniques used in that study rather than the predictive use of tissue measurements. Fat has an uneven distribution within a given body region, making a simple 1D measurement at the abdomen, thigh or arm insufficient to accurately characterise the total amount of subcutaneous fat [161]. Any predictor based on these measures would be unable to cope with physiological differences between fetuses. In that study, measurements were of general tissue thickness rather than fat quantification; a given thickness formed by high muscle and low fat will weigh more than the same thickness of low muscle and high fat. This was cited as a possible reason for the poor performance of the tissue thickness in assessing growth [75]. Finally, the pixel resolution of their US probe was cited as 2-3mm, which was of the order of the fat thickness being measured, contributing to the discrepancies in measurements [75]. The issues faced by that study highlight the importance of good resolution 3D measurements and tissue segmentation in order to make useful predictions of fetal weight outcomes.

Lee *et al.* studied the ability of various fetal parameters to predict neonatal body composition (specifically neonatal %BF), including HC, AC, EFW, mid-thigh volume, Fractional Thigh Volume (TVol) and Fractional Arm Volume (AVol), as well as birth weight [98]. It was found that the image biomarker with the closest correlation to neonatal %BF was TVol, being the most significant factor in variance of %BF (46.1%). This was found to correlate more significantly to body fat measurements than even birth weight (responsible for 44.7% of %BF). However, as with other investigations of this nature (such as that conducted by Gardeil *et al.* [59]), the fetal measurements were taken far too close to birth to be considered predictive, in this case four days, and the quality of measurements could be questioned. Despite this, it does show that fetal growth parameters observed using US can be used to estimate neonatal body composition. With additional work into how fetal body properties change during pregnancy and advances in measurement accuracy, it may be possible to extend these calculations to earlier stages of pregnancy.



(a) **(b)** **(c)** **(d)**

Figure 2.4: *The length of the fetal limb can be seen in **(a)**. A sample of slices are taken along the length of the limb, which can be seen in **(b)-(d)**. Segmentations taken in this way can be used to form an estimate of TVol.*

Fractional Limb Volume

TVol is an approximation based on taking a series of 2D US images. It is calculated by taking an approximated volume across half of the femoral diaphysis length (the shaft of the bone), as viewed on US; five equidistant slices centred around the mid-point of the thigh are taken, their areas measured by manual segmentation and used to approximate the volume [99]. Figure 2.4 shows an example of how such measurements are taken. In addition to identifying TVol as having a close correlation with neonatal body composition, Lee *et al.* conducted research into the technical aspects of TVol, such as acquisition, nominal values and reproducibility of results, as well as the specifics of fetal weight estimation based on TVol [98–100]. Rueda *et al.* have based a number of papers on the segmentation of fetal fat for the purposes of assessing fetal nutrition [155, 156].

Lee *et al.* suggest that the performance of TVol at predicting neonatal body composition could be improved by conducting further analysis on the proportions of fat and lean tissue present within the measured volume [98]. Furthermore, they call for improved segmentation methods between soft tissue in order to conduct this research [98]. This further supports the argument that in order to more accurately assess fetal development, techniques of defining the boundary between fat and lean tissue on US images are necessary.

Mayer and Joseph comment that in the event of TVol providing a more accurate insight into fetal growth, this measure will only be viable if the increased complication and duration of the task is compensated for [110]. As time consuming, manual area traces are required for the calculation of TVol, it is important that supporting software that can delineate between soft tissues accurately and efficiently are made available to clinicians.

2.4.2 Liver Quantification

As an alternative to taking measurements of fat volumes, measures of organ development have been proposed to assess growth. Liver volume (LVol) has been successfully estimated from US images and abnormally low values of LVol have been shown to correlate with FGR [15, 16, 118].

Naeye conducted an extensive study of 53,518 pregnancies, which were found to contain around 120 cases of death due to placental infarction [118]. By observing known cases of placental dysfunction, the investigation allowed for analysis of fetal growth under conditions of poor fetal nutrition. A postmortem study showed that while the body weight of infants who died due to placental infarction were on average around 15% lower than values published for healthy infant weights, livers showed an approximately 25% decrease from nominal values, second only to the thymus in terms of decrease in expected growth of the organs measured [118]. This indicates that estimation of LVol may be useful in assessing fetal development.

Boito *et al.* assessed the use of the ratio between hepatic volume measurements and HC compared to AC by comparing each measurement between sets of healthy and growth restricted fetuses [16]. In a study of 24 growth restricted fetuses and 85 normal pregnancies, the mean values of each measurement on growth restricted fetuses were expressed as a percentage of the median in normal pregnancies, with liver outlines traced manually. This gave a value of 45% for LVol, 90% for HC and 82% for AC, suggesting that LVol is a good indicator of FGR. However, using a z-score, calculated as $zscore = \frac{IUGRMeasure - MeanHealthyMeasure}{StandardDeviation}$, LVol showed an improvement over HC but was deemed to not be a better indicator of FGR than AC [16].

The work of CH Chang *et al.* is very encouraging in the use of liver measurements to assess fetal growth, showing in a study of 42 growth restricted fetuses and 375 normal fetuses that LVol has a sensitivity of 93.6% and specificity of 97.6% in detecting FGR [30]. In contrast to the observations of Boito *et al.* [16] this was shown to be a better estimator than AC (approximate values from the Receiver Operating Characteristic (ROC) curve; sensitivity 88%, specificity 75%) and EFW (approximate values from ROC curve; sensitivity 75%, specificity 87%) in assessing FGR [30].

It was suggested by CH Chang *et al.* that further investigation should be made into the relationship between brain volume and LVol to assess fetal growth, as had previously been assessed in another Boito *et al.* paper [15,30]. This report studied the ratio between brain volume and LVol; in growth restricted fetuses, the brain sparing effect prevents significant loss in brain volume, providing an approximation for the fetuses 'natural size' [15]. This correlates with Naeye's observations, which showed that the brain had one of the lowest changes in size between cases of placental infarcts and normal pregnancies [118]. Comparing brain volume to LVol, which is more significantly affected by FGR, can be assumed to be a useful measure to assess growth. Comparing 47 AGA to 23 SGA fetuses, it was found that the mean brain/liver volume ratio was significantly different; 5.9 ± 1.9 in SGA fetuses compared to nominal values of 3.4 ± 0.7 [15]. LVols were estimated using 10 manually segmented slices on the sagittal plane, while brain volumes were estimated from HC. One consideration to make with this paper is that the definition of a SGA fetus is taken as a weight of below the 5th percentile [15] rather than the more conventional 10th percentile [26, 30, 59, 98, 99]. A high brain/liver volume ratio was also found to be an indicator of fetal outcome for fetuses with a GA of under 29 weeks. This study shows that there is potential for using liver measurements to estimate fetal outcomes and assess growth, although a larger study with a more accurate estimation of LVol (since the original study used an approximation based on only 10 slices) would be required to establish the relationship firmly. Semi- or fully-automated segmentation techniques could be used to improve the efficiency of slice segmentation and allow for the use of 3D US rather than the estimations used in this study. This improved accuracy could compensate for the study's use of the 5th percentile.

The exact method of segmentation of liver slices has varied through different studies. Measurements of LVol taken by Boito *et al.* were based on 10 manually traced sagittal slices, taking the diaphragm as the upper boundary [15]. In a study of LVol in fetuses with chromosome abnormalities, Gielchinsky *et al.* used 12 slices rotated around a fixed axis [63]. In both cases segmentations are manual. CH Chang *et al.* use a 3D US probe, but the exact method of segmentation is not commented on [30]. Automated or semi-automated segmentation techniques could be useful in aiding the segmentation process, especially as the liver is a particularly difficult organ to visualise accurately in US.

Hata and Deter reviewed other organs which can be successfully measured using US, including brain structures such as the cerebrum and cerebellum, the heart, lungs, thymus, spleen, pancreas, stomach, gallbladder, kidney, adrenal gland, intestine and bladder [73]. Some of these may have little use in nutritional assessment, for example the 'brain sparing effect' described by Boito *et al.* makes brain structure measurements alone a poor measure of nutrition [15]. In addition, some of these organs were reported to be difficult to image (the thymus edge was found difficult to determine and bowel loops for intestine measurement were only observable in 30% of fetuses), preventing consistent, reliable measurements [73]. The size and sensitivity to environments of poor nutrition make the liver a good candidate for organ measurements.

2.4.3 Brain and Skull Analysis

Another method of determining the growth trajectory of a fetus is to detect and evaluate the size of the brain and the structures within. It has already been discussed how fetal growth will prioritise the brain, and the resulting 'brain sparing effect' means that brain size can theoretically be used as a benchmark from which to compare the growth of other anatomical features in the fetus [15].

Al-Riyama *et al.* conducted a study into the ability of the ratio between HC and AC to identify growth restricted fetuses [127]. Their investigation observed a cohort of 107 pregnancies identified as growth restricted by placental malfunction, 58% of cases were found to have an abnormally high head-to-abdomen ratio (termed 'asymmetric'), which is a far higher proportion within a group than would be expected [127]. While this shows that the 'brain sparing effect' can potentially be used to confirm diagnosis of FGR, using this metric alone may lack predictive power, with many IUGR cases (according to their study, 42%) *not* exhibiting a high head-to-abdomen ratio [127].

The shape of the fetal head can also be significant in determining instances of fetal abnormalities. Namburete *et al.* segment the fetal skull in US images using superpixels, with the intention of detecting such fetal cranial dysmorphism, as these deformities can be indicative of a number of different fetal health problems [119].

The detection and analysis of structures within the fetal brain can be important tools in assessing developmental health. Namburete *et al.* use machine learning techniques to locate the choroid plexus in US images, which can be used to determine the correct plane from which to calculate standard biometric measurements of the fetal head [120]. Aside from the alignment of the view, detection and visualisation of the choroid plexus is useful as the presence of cysts can indicate instances of trisomy 18 [120].

In an aggregation of results from previous post-mortem studies, Swayze *et al.* identify a number of commonly found fetal abnormalities in cases of FAS, including instances of abnormally small brains and defects in the corpus collosum, specifically cases where it is partially or completely missing [181]. These results were confirmed by their own post mortem and brain MRI study, which were conducted in a cohort of ten patients who had been diagnosed with FAS [181]. It is possible that these features could be identified from US images of the fetus, particularly earlier in gestation, before the skull becomes too much of an acoustic barrier.

However, it is important to note that while there was a set of brain abnormalities apparent in each case of FAS, there did not appear to be a consistent set of brain features which could be used to diagnose the condition without further context, and certainly could not discriminate between conditions. Microcephaly, for example, could indicate a vast range of medical conditions as diverse as genetic mutation, inherited traits, injury and degenerative conditions [116]. Consequently, brain size and features, while a potential method for suspecting a diagnosis of FAS, may not be basis enough for confirmation. Additionally, this means that using the 'brain sparing effect' to compare if the size of other organs are under- or over-sized will not always be a valid assumption in high-risk pregnancies.

2.4.4 Face Analysis

A number of health conditions (both genetic and preventable) have been shown to express themselves through variation in facial shape [70]. Hammond *et al.* were able to show not only that facial analysis can be used to determine between controls and those suffering from a range of syndromes, but that it is also possible to discriminate between these conditions, meaning that face shape has the potential to be a powerful diagnostic tool [70].

Of the conditions which can express themselves in facial structures, FAS is of particular interest within the context of this thesis, as it is both a preventable condition and influenced by maternal diet, relating this problem to that of fetal nutrition. However, detection can only occur after the condition has set in, when corrective measures cannot be taken to reverse the effects of the condition. Despite this, early detection is important as it has been demonstrated that the symptoms of FAS can become less visible over time [172]. Additionally, diagnosis *in utero* allows for treatment planning and can allow for intervention to prevent further damage.

FAS is associated with behavioural difficulties, stunted growth and lowered intelligence and, more significantly to the application of US monitoring, results in physical differences such as flat philtrums, short palpebral fissures and reduced HC, which can theoretically be used to detect it [179]. However, Lipinski *et al.* point out that FAS is often misdiagnosed, as they showed that facial variation in animal studies can vary depending on the timing of exposure during the pregnancy, and this may translate to variation in modes of facial variation in alcohol-exposed human fetuses [103]. It is therefore important that a detailed model to represent the variation of dysmorphism related to FAS can be created.

Studies into facial dysmorphism typically use facial analysis from 3D photographic images to show that it is possible to characterise facial features from extracted surfaces and hence diagnose instances of a number of conditions. Related investigations which have attempted to form objective assessment of the differences between the facial features of those affected by such conditions and control groups include the following:

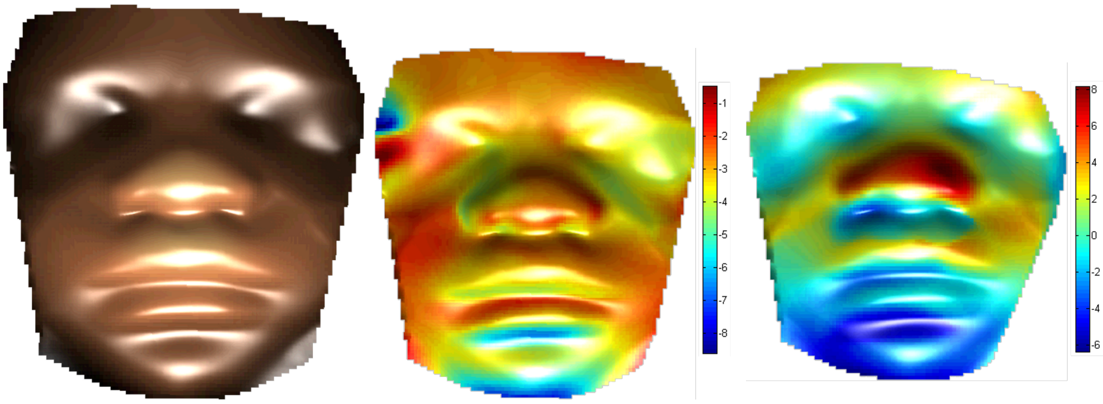
- Hannes *et al.* used a dense shape model created from 200 3D face images of control subjects and 100 images from subjects diagnosed with the genetic condition **Wolf-Hirschhorn syndrome** to inform diagnosis [71].

- **Fragile X syndrome** can also be expressed in facial variations, though these can be subtle and are rarely used to diagnose the condition [74]. However, Heulens *et al.* were able to show the locations and variation of facial differences between affected and unaffected groups by analysing 3D images from 373 control and 75 example cases, even detecting dysmorphic features at earlier ages than previous studies and identifying new differential features [74].

- Suttie *et al.* were able to successfully classify cases of **FAS** from both the full facial surface and facial profile in a study that utilised 3D photography from 192 children of varying levels of maternal alcohol exposure during development [179]. Levels of exposure were established via interviews, with average alcohol consumption across the cohort of 2.8 standard drinks per day and heavier consumption over weekends [179].

- **Bardet-Biedl syndrome** also expressed itself in subtle ways that were difficult to characterise without 3D facial and nasal analysis, which Tobin *et al.* performed on 230 controls and 83 patients, separated by age [184].

- Tassabehji *et al.* conducted a study into the facial dysmorphism of **Williams-**



(a) **(b)** **(c)**

Figure 2.5: *(**a**) Mean fetal face shape, taken from 20 examples (full description in Chapter 5). Figures (**b**) and (**c**) show the corresponding heat map projected onto the mean facial surface. In each case the colour shows the difference at each point between the facial surface height and the mean surface.*

Beuren syndrome, with a cohort of 185 control subjects and 85 afflicted subjects, with ages ranging from 2 weeks to 20 years, finding abnormalities in the skull and jaw [182].

- Bhuiyan *et al.* separated 39 subjects afflicted by **Cornelia De Lange syndrome** into 'classic,' 'mild,' 'possible' and 'definitely not' CdLS phenotypes based on clinical measures which included prenatal growth and skull circumference, facial assessment, limb deformities and behavioral assessment [10]. A subset of 20 subjects had 3D facial images taken, 15 of which were found to be classified as 'classic' or 'mild,' labelled with 21 repeatable landmarks and used to form a plot of face shape against 195 control subjects [10].

- Cox-Brinkman *et al.* constructed their model from 206 3D images of faces from a set of 160 control subjects and 42 with **Fabry disease** [45]. Their studies showed that classification accuracy varied greatly between the sexes, with closest-mean classification. This is a form of classification where a mean face is created from the affected and unaffected groups, and then individuals are classified by finding which of these examples they are most similar to.

The above state the use of Dense Shape/Surface Models (DSM) in order to construct their models, which is described by Heulens *et al.* as a model composed of a set of principal components which are able to describe the variation in shape for 3D surfaces [74]. Another tool frequently adopted by such studies is the 'heat map,' which shows where facial variation between groups is expressed, and the general extent and direction of displacement and is thought to be a useful tool in informing clinical decisions [179]. Figure 2.5 shows an example of such a heat map of two examples of fetal faces (however, it should be noted that it is more common to show the heat map of the mean for different cohorts within a study rather than that of individuals).

Confirmation of the surface models in identifying their respective conditions are varied across these studies. Bhuiyan *et al.* compared five "possible" cases of Cornelia De Lange syndrome to their shape model, finding that two showed similarities to the afflicted group, two were more similar to the controls and one showed unrelated dysmorphism [10]. In their study of facial dysmorphism in Fabry disease, Cox-Brinkman *et al.* performed analysis on results from closest-mean classification of the faces used to construct the model [45].

Suttie *et al.* also confirm their findings by testing the classification accuracy of the facial surface model they created for FAS [179]. Forming 20 classifiers using 90% of the faces as a training set and 10% as a test set, pairs of FAS and control faces (determined by clinical assessment without knowledge of maternal alcoholism) were classified showing a mean agreement with clinical diagnosis of 0.967. However, it should be noted that this is a significantly simpler and far less clinically relevant classification task than classifying individual instances of faces. Since pairs will already consist of one FAS and one non-FAS, the 'more-FAS' face will be classified as FAS, which may actually be closer to the control mean than the FAS mean, and so would not be diagnosed correctly if observed individually. Classification also appeared to be successful using the full face surface (agreement to clinical assessment of 0.967 to 1 depending on the classification method used), facial profile (0.917 to 0.933) and selected regions around the face, specifically around the eyes, nose and mouth, each calculated separately [179]. Although pair-based classification remains unsuitable for clinical use and the results are arguably skewed favourably because of this, the comparability of facial profile and facial surface results at least suggests that 2D facial profiles could be a sufficient metric in diagnosing FAS, which could be translated to 2D US acquisitions.

At the time of writing, no study has used this form of analysis on surfaces segmented from 3D US images in order to delineated shape variation in the fetal face. Kurjak *et al.* provide a detailed explanation of fetal facial development, and state that the face can be reliably rendered between approximately 18 and 36 weeks GA, which suggests that these forms of study would be feasible on fetal faces [94]. Since FAS has been shown to cause the most visible variation in the philtrum (the central region between the upper lip and nose), it may even be possible to classify whether or not a fetus has been affected by FAS using only a 2D contour of the facial mid-line [179]. Therefore, potential avenues for examining whether facial variation can be used to detect fetal health conditions would be to use shape modelling techniques in conjunction with segmented profiles from both segmented 2D contours from US and segmented surfaces from 3D US of the face. This means that methods developed in the segmentation of surfaces in US will translate well as a starting point for the process of determining whether the fetal face can be used to diagnose conditions such as FAS prenatally. From this point, parameterised descriptions of the facial profile and/or surface could be used to differentiate between healthy and unhealthy fetuses.

2.5 Objective Overview

This thesis explores three main objectives, which are as follows;

- **Develop a tool to aid in the segmentation of fetal fat in 2D US images (Chapter 4):** While size at a given GA is often used as a predictor of fetal nutrition, it is not necessarily indicative of fetal health. TVol has been proposed as a measure for assessing fetal development, but it is necessary to delineate the boundary between fat and lean tissue in order to make improved predictions on fetal body composition and therefore nutrition [98]. As time consuming, manual area traces are required for the calculation of TVol, it is important that accurate, semi-automated methods that can delineate between soft tissues are available, and can be used over a range of image qualities.
- **Expand this tool to full 3D liver segmentation in MRI and US (Chapter 5):** In order to accurately determine the level to which liver growth has been impaired, it is important to measure the volume of the organ, as its irregular shape and general lack of internal landmarks makes it poorly suited to estimate based on calculations of cross-sectional area [29]. The liver is particularly difficult to visualise compared to other organs, and so automated methods to find its boundary (in the form of a contour or surface) may be most suitable to aiding clinicians in determining its volume. In some cases, the liver boundary will appear very faintly in an image slice, but the position of the boundary in adjacent slices will help confirm the decision as to where the actual boundary lies.
- **Modify this method to segment both the 2D profile and 3D facial surface in US images, and use this to develop a method of modelling fetal face shape to potentially show a correspondence between FAS and fetal face shape (Chapter 6):** Maternal alcohol consumption has been shown to be related to FAS; a condition which can have long term adverse effects on development [84].

The segmentation and subsequent shape characterisation of the fetal profile and facial surface may provide an insight into diagnosing the condition prenatally.

2.6 Technical Challenges

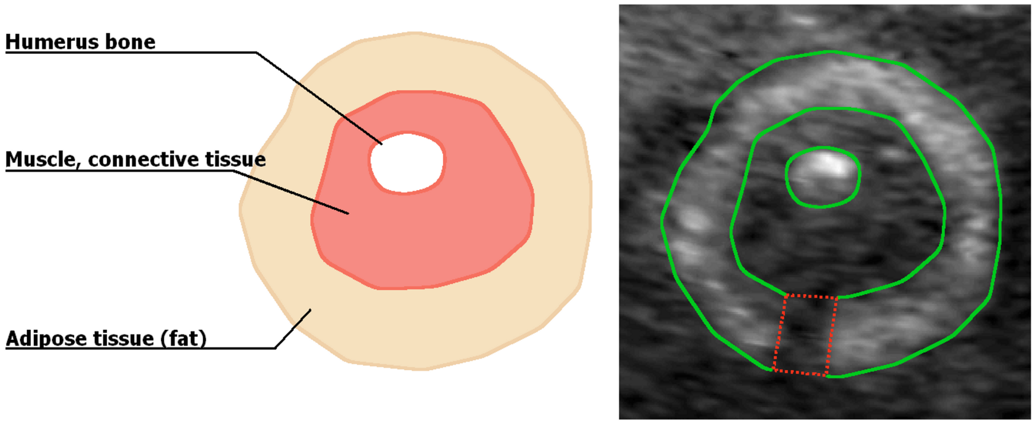
Segmentation of fetal subcutaneous limb fat and fetal liver have been identified as promising metrics for the assessment of fetal nutritional state, and so it is important that tools are developed specifically to target quantifications of these key landmarks. By facilitating the calculation of such volumes, it can be possible to move towards developing techniques to assess fetal growth that rely on explicit rather than implicit markers of fetal nutritional state. US is already commonly used in routine assessments of fetal growth, and other modalities can present an unnecessary risk to development. In the case of x-rays the dosage of radiation is actually well within safe levels but associated maternal anxiety should be avoided, while MRI presents the risk of acoustic damage [35, 150]. It is therefore essential that such volumes can be extracted from US data.

US as an imaging modality can present a number of issues which make conventional imaging techniques fail. Shadow from dense objects can obscure sections of the image and occlude boundaries, speckle can make boundaries between tissue appear unclear, the image attenuates with distance, resulting in poor contrast further away from the transducer and signal dropout can also occur [128]. Variation in the appearance of tissue between fetuses and with GA can make it difficult to train segmentation methods on fetal US images.

US operates by sending a series of acoustic pulses and detecting the time between the sent signal and the received echoes as well as the amplitude of these echoes. Acoustic reflections occur when the pulse bounces off an acoustic boundary, which can either be the boundary between different tissue types or due to the presence of small 'scatterers' in the tissue. This means that the intensity of the signal only shows the difference in acoustic impedance between tissues forming the boundary, and not any properties of either of the tissues individually. Additionally, since multiple echoes will occur as the pulse travels through the scanned object, the power of the remaining signal decreases as the depth of the image increases, and the intensity of boundaries appears reduced (this is known as signal attenuation). In cases where there are objects which present very high acoustic boundaries, such as bone, only a small proportion of the signal is able to pass through, creating a region of very low intensity below the object, resulting in a phenomenon known as shadowing. This means that simple forms of image segmentation, such as intensity thresholding, are largely unsuitable for US, as intensity is representative of a certain kind of tissue.

Similarly, problems may occur when intensity-gradient based techniques are applied to US; these are only useful if speckle and noise are of low amplitude compared to the changes in intensity along boundaries. While this may be appropriate for, as an example, segmenting tissue from surrounding fluid, it can be an issue when segmenting a soft tissue boundary, as the intensities on either side of the boundary will generally be similar. Commonly, imaging techniques employing intensity analysis are less suitable for US images [148].

An additional challenge presented by fetal segmentation is the difference in opinion between operators. Any method that involves user interaction has the potential to result in varying outcomes between different segmentation attempts, potentially impacting on the accuracy of the segmentation. Sarris *et al.* investigated the impact of these differences in a study comparing results calculated by three clinicians on 140 fetuses, finding intraobserver variability of $\pm 3.0\%$ for HC, $\pm 5.3\%$ for AC and $\pm 5.3\%$ for FL, and interobserver variability $\pm 4.9\%$ for HC, $\pm 8.8\%$ for AC and $\pm 11.1\%$ for FL [159]. This scale of variation presents the potential for misdiagnosis of SGA fetuses. The method of segmentation should therefore attempt to limit inter- and intraobserver variability.



(a) (b)

Figure 2.6: *(a) Schematic of fetal arm composition. (b) A typical US image of a fetal arm cross-section. The region outlined by the red dotted line shows an area of shadowing caused by the humerus bone, altering the appearance of the adipose tissue layer in that region.*

2.6.1 Fat Segmentation

Figure 2.6a shows an approximate schematic of the fetal arm composition; the region of interest in our case is the adipose tissue on the outermost layer. This has then been overlaid onto a US image of a fetal arm cross section in Figure 2.6b. The separation of fat from muscle using US images is a rather specific boundary finding problem. The properties of fat and other forms of tissue will be far more similar than between, for example, tissue and fluid, meaning that boundaries will have a relatively low contrast. This could make it difficult to segment using conventional techniques, especially intensity-based methods such as thresholding and intensity-gradients, and statistical measures could also suffer from poor distinction between regions.

Ng *et al.* have attempted to tackle the problem of human fat segmentation and measurement, providing a good summary of the properties of fat when observed by US and key issues that arise from attempting to identify regions of fat compared to other tissues, as well as a comparison of cell structure [126]. In order to identify fat from surrounding tissue, the spectrum along individual radio frequency (RF) scan lines (essentially the raw US signal, unsuitable for visualisation but useful in analysis) is assessed using Fourier analysis. The techniques proposed by Ng *et al.* are used in a context of calculating fat depth from the skin surface, allowing for a number of assumptions (for example, the order of tissue types with image depth) and are therefore unsuitable for our application [126]. The error of the technique against skinfold caliper measurements was found to be in the order of millimeters and it was found to be difficult to detect thin layers of fat, which could have a severely detrimental effect on area calculations [126]. It should be considered that this could be more indicative of the error between US measurements and physical measurements of fat rather than the inaccuracy of the image segmentation techniques used. Their investigation incorporates data from 11 beam angles and 4 probe locations, which is not a common obstetric set-up and may be an issue when imaging a non-compliant subject such as a fetus. However, there is a growing interest in this form of measurement of fetal biometrics. A related but different approach of fusing multiple viewing angles to form an improved image for segmentation has been proposed by Yaqub *et al.* using 6 angles [202] .

Rueda *et al.* have proposed a method that deals directly with the issue of fat segmentation in US [155]. This technique uses the structural information of Local Phase (LP) with Fuzzy Connectedness (FC) to form an automated technique which segments regions of fat in images of fetal thighs and arms. FC is a measure of how strong the 'connection' between any two pixels is by finding the strongest link along all paths between them [155]. Compared against manual segmentations of clinical images, the technique was found to have a precision of 93.5% $\pm$ 1.91% and a recall of 82.8% $\pm$ 5.74%. This shows that there is promise in using LP and FA based methods for fat segmentation in fetal US images. The accuracy of FC segmentation is explored in more detail in Section 4.5.2.

2.6.2 Liver Segmentation

A number of investigations have been made into the calculation of LVol and the prediction of growth trajectories [7, 16, 29, 31, 93, 96]. Table 2.2 gives a summary of these studies, including estimates of LVol at 23 and 31 weeks based on the equations derived from each study. These values of GA have been chosen as they represent the maximum

Table 2.2: *Equations for LVol from GA, as estimated by a number of different population-based studies, along with the scope of each study. Populations have been separated into fetuses identified as 'normal' (AGA), fetuses identified as SGA and fetuses identified specifically as IUGR. Each study's estimate of fetal LVol at 23 and 31 weeks GA have also been calculated.*

Paper	Ref	Year	GA (weeks)	Location	Sample	23wks (ml)	31wks (ml)
Baker	[7]	1995	23+	UK	21 AGA 11 SGA	6.65	75.05
FM Chang	[31]	1997	22-31	Taiwan	55 AGA	22.56	57.63
Laudy	[96]	1998	19-39	Netherlands	25 AGA	17.39	65.95
Kuno	[93]	2002	20+	Japan	14 AGA 10 SGA	19.10	55.75
Boito	[16]	2002	20-36	Netherlands	85 AGA 24 IUGR	19.77	60.05
CH Chang	[29]	2003	20-40	Taiwan	226 AGA	22.16	48.79

and minimum ages that *all* of the studies cover. Boito *et al.* and Laudy *et al.* estimated LVol from ten manual segmentations [16, 96]. Baker *et al.*, FM Chang *et al.*, Kuno *et al.* and CH Chang *et al.* estimated volume from manual slice-by-slice segmentations, with slice widths specified as 10mm [7], 2mm [31], 1mm [93] and 3mm [29] respectively. By manually segmenting dense 2D sections to form an estimate of the 3D volume, the process of finding the LVol can be very time consuming.

CH Chang *et al.* highlight the importance of 3D measures against 2D measures when it comes to assessing the liver [29]. They comment that using a 2D measure as an estimate of a 3D volume relies on the assumption that the target object has a regular, predictable shape, which is notably not the case in the fetal liver [29]. Calculating a 3D measure of volume over 2D is a justifiable objective, but taking this value from a series of 2D slices can be challenging; in the study by Boito *et al.*, which estimated hepatic volume in this way, it was reported that manual volume calculation of the liver took around 15 minutes per case [16]. This can be a massive drain on clinical resources, particularly in studies involving large amounts of data, and could be a significant obstacle to taking LVol as a part of standard fetal monitoring. Additionally, image volumes were not always considered viable for volume calculations; failure to calculate values of LVol occurred in 18% of their sample of normal pregnancies and 23% of their growth restricted cases [16].

This further supports the idea that a move towards automated measures of segmentation is desirable to make LVol measurement a standard metric for assessing fetal health.

Kuno *et al.* observed in their study that while their sample of SGA infants displayed a normal value for liver length in 70% of cases, all were in the lower 10th percentile for LVol, noting that the complex shape of fetal organs means that volume, surface area and weight measurements are far more useful measurements in fetal growth assessment [93]. Similarly, CH Chang *et al.* state that the use of 2D US images in fetal assessment assumes standard geometry, which is particularly unsuitable in the case of the fetal liver [29].

Liver segmentation presents a similarly challenging set of problems for segmentation methods, with few reliable landmarks, irregular geometry and often poor contrast between the liver and surrounding tissue. Very few techniques deal specifically with liver segmentation in the context of fetal LVol measurements, and liver segmentation in adults is also deemed a very challenging problem. However, efforts have been made to characterise liver tissue in order to identify lesions and tumours in adults, which could be applied to the segmentation of the liver in fetuses.

2.6.3 Fetal Face Analysis

At the time of writing, there have been very few studies into the analysis of facial features from fetal US images. Much of the current literature involves manual segmentation or visual assessment. For example;

- Nelson and Pretorius [125] and Lee *et al.* [97] made early attempts at rendering the 3D body and facial surfaces respectively from US to facilitate clinical diagnosis of abnormalities .
- Sonek *et al.* manually measured the nasal bone length of 3,537 fetuses [170].
- Rotten and Levaillant set out to establish protocols for the assessment of the fetal face in 3D US [153], while Kurjak *et al.* outline assessments which can be made with the aid of 4D US [94].

There have also been a number of relevant papers in the two separate tasks of facial morphological analysis and the segmentation of fetal heads and faces. Each of these tasks present a number of challenges which must be overcome.

Feng *et al.* and Chen *et al.* have conducted research into the automatic detection of fetal faces from 3D US volumes, the latter relating the problem to studies of face detection in 2D photography and 3D laser scanners rather than other problems of feature detection in US [34, 55]. The challenges they cite as problematic include many of those already covered in the difficulties presented by segmentation in soft tissue and liver segmentation in US, such as poor contrast, varied appearance, speckle and shadowing. However, additional issues include the potential variation in position and pose of the fetus and the issue of incomplete faces due to the limitations of the sweep of 3D US probes [34, 55].

Feng *et al.*'s solution is to use a sparse face mesh, which is fitted to the fetal face via a combination of 3D feature detection and 2D profile detection, and then objects obstructing the face are removed with a process they refer to as 'carving' [55]. However, these methods were created with the intention of providing obstetricians with the best possible view of the fetal face, rather than a segmentation, and so the facial surface is never actually acquired, but only implicitly found with the 3D mesh. This is reflected in their results, which show an error (measured as the Euler distance between points on the 'ground truth' surface and their segmented surface) in the order of 7mm [55]; potentially acceptable for their application but not sufficient for face analysis.

Chen *et al.* also tackle the problem of fetal head segmentation by starting with a mesh model, but choose to construct a mesh of the entire head rather than just the face [34]. Gabor filters are used to find the eyes as a starting point, then feature based registration aligns the model and finally the boundaries are adjusted using a 3D snake-based deformation [34]. While the final model is sufficient to make standard biometric measurements, further investigation would be required to find out whether this can provide the facial detail necessary to perform accurate facial shape analysis.

Previous investigations into characterising facial structure of children from 3D photography have defined faces using a number of points ranging from 16,000 to 25,000 [10, 74]. The resolution of faces constructed from 3D photography may prove challenging to match in US images, due to both the reduced spatial resolution compared to photographic images and the smaller facial surface of the target object. A further issue is that the field of view of the probe may be insufficient to capture the entire face, particularly later on in gestation. Detection methods such as those described by Feng *et al.* [55] would result in insufficient detail compared to what is required for effective face analysis.

Additionally, there is far less control over the expression of the fetal face; in child and adult studies, it is possible to gain a neutral expression for analysis with relative ease, but in fetal studies this is not the case. Kurjak *et al.* even suggest the use of 3D and 4D US in the observation of fetal facial expressions to monitor behaviour and potentially unlock additional diagnostic capability [94]. For the purposes of structural fetal face analysis however, this can be detrimental.

Change in expression is particularly problematic concerning the mouth, which could take a variety of modes between fully open and fully closed, and actions such as yawning, sticking out the tongue or thumb sucking have the potential to interfere with segmentation. This may not in fact be problematic, since in the photographic study conducted by Hammond *et al.* into the classification power of separate regions of the face of infants (full face, around the eyes, around the nose and around the mouth), it was found that mouth shape was a less effective metric in children than in adults, which may make it a less important region in fetal studies [70]. It is also possible that asymmetrical facial expressions can occur during scanning, which could confuse results, as facial asymmetry can be a sign of a number of fetal health conditions [94].

Segmentation of the fetal face provides a number of challenges, many of which are held in common with the problems raised by liver and soft tissue segmentation. In addition to this, the requirements of the segmentation are high, both in accuracy and resolution, for the facial analysis to be effective. However, one benefit this anatomical feature has is the number of reliable and repeatable landmarks which could be used to guide segmentation more effectively.

Tools for Ultrasound Segmentation

Monitoring Fetal Growth: In order to achieve the goal of accurate, reproducible and fast segmentation of fetal soft tissue, hepatic volumes and the facial surface, it is essential that a set of tools are developed in order to aid clinicians in segmenting these key regions.

Fetal Alcohol Syndrome: The primary target of this area is to segment the curve of the face profile and assess whether or not this can be used to determine whether a fetus is affected by FAS. This can then be extended to the fetal face surface, at which point the most effective facial regions for determining instances of FAS can be found.

Technical Goals: The aim is to create a clinical tool which is highly intuitive to operate and allows for segmentation in the challenging setting of fetal US. The tool should clearly reduce the time and effort required to perform accurate segmentations of the biometric features targeted in this thesis. In the case of FAS, a way of characterising the segmented face shape is required.

The personal contribution in this chapter is a discussion of the clinical issues surrounding the definition and measurement of metrics for fetal development, as well as the justification for the use of segmentation methods in the field.

Chapter 3

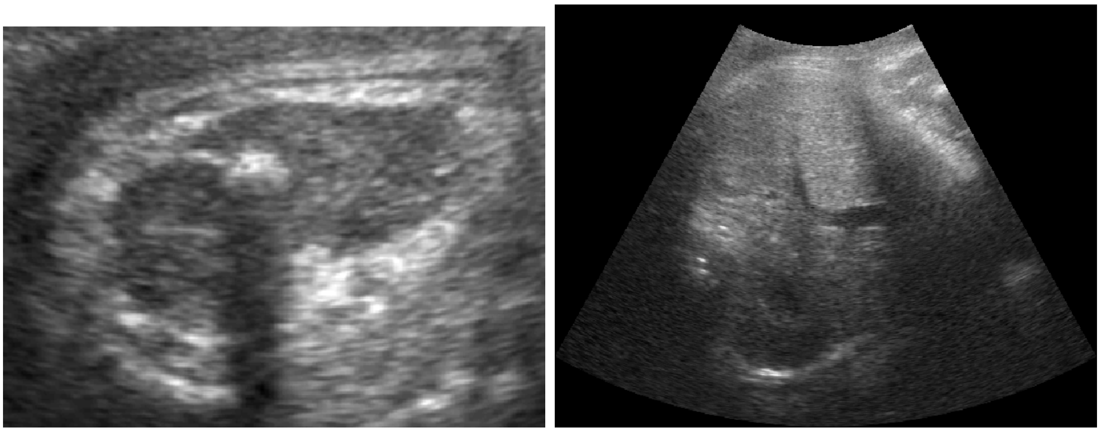
Image Analysis in Ultrasound

This second literature review chapter investigates the state of the art in US image segmentation and evaluates how useful such methods may be in the context of the segmentation of fetal biometric landmarks.

3.1 Segmentation in Ultrasound

In the US image segmentation survey by Noble and Boukerroui [128], a number of different classes of techniques to delineate boundaries between different regions were identified; these include intensity, intensity-gradient, local phase, texture, shape, motion and statistical analysis. US images tend to have poor contrast, low signal to noise ratio and contain speckle, meaning that different techniques to those used in other imaging modalities (such as X-ray, CT and MRI) may be necessary to extract physical boundaries [149].

Intensity thresholding is largely unsuitable for US, as image intensity is based on the difference in acoustic impedance between tissues rather than the properties of the tissue itself. Similarly, problems may occur when intensity-gradient based techniques are applied to US; these are only useful if speckle and noise are of low amplitude compared to the changes in intensity along boundaries. While this may be appropriate for, as an example, segmenting tissue from surrounding fluid, it can be an issue when segmenting a soft tissue boundary, as the intensities on either side of the boundary will have greater similarity. A gradient still exists at large scales, but at small scales, intensity gradients caused by



(a) **(b)**

Figure 3.1: *Examples of challenging imaging conditions often found in fetal US* ***(a)*** *Fetal arm cross section at GA 39 weeks and 1 day* ***(b)*** *Fetal abdomen at GA 32 weeks and 4 days.*

speckle can be on a par with those caused by tissue boundaries. Commonly, imaging techniques employing intensity analysis are less suitable for US images [148]. LP can describe local structures in an image and have been shown to be of use in echocardiography and fetal US [149]. LP can be described as 'intensity invariant', showing structures rather than 'energy,' which is particularly useful in US where image contrast is poor [148]. This method is discussed further in Section 3.3.2.

In a medical image analysis challenge event to automatically find fetal biometric measurements, Rueda *et al.* found that no entrants attempted fetal abdominal measurements, stating that this is the hardest of the biometrics to automatically calculate, owing to variations in internal structures both over GA and between fetuses, as well as poor boundary definitions [156]. Figure 3.1 shows an example of the imaging conditions often found in fetal US. From the image of the fetal arm cross-section, severe speckle, variation in image intensity, disrupted boundaries from shadowing and poor contrast between tissue can all be observed. The image of the fetal abdomen also shows significant shadowing along the right-hand side and signal attenuation with image depth. Any methods used to segment anatomical features of fetal US images will need to take these problems into consideration. Furthermore, it is very difficult to validate the absolute accuracy of US measurements for *in utero* applications. This is an issue that will be faced by all segmentation techniques

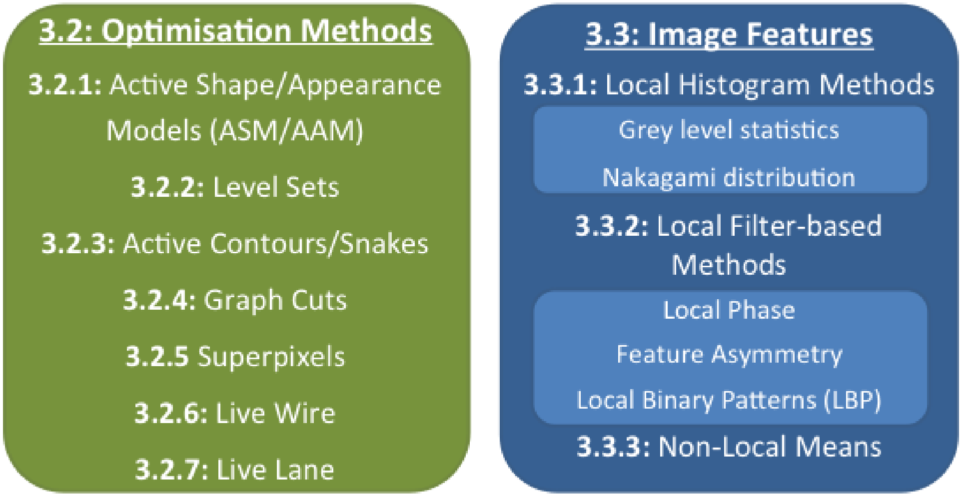


Figure 3.2: *Summary of segmentation methods and tools described in this chapter, and which sections address them.*

described, and should therefore allow for fair comparison.

The methods for image segmentation that are covered in this thesis can be broadly divided into optimisation methods and image features; these two aspects of the segmentation method work hand in hand. Optimisation methods are used to find the most appropriate or optimal boundary in the image, while the image features drive this process and inform how a boundary is defined based on image properties. For optimisation methods, semi-automated methods are primarily investigated, which require some form of user interaction to operate. Figure 3.2 provides a summary of the key areas this chapter covers.

Table 3.1: *Summary of optimisation methods for image segmentation*

Optimisation Method	Application	Strengths & Limitations
Active Shape /Appearance Models	Echocardiography [17, 115, 141] Prostate in US [166] 3D Facial photography [14] Fetal anatomy in US [82]	**Strengths:** Can include the appearance of shadowing as part of the model **Limitations:** Deviations on a mean representation, potentially requires large data set to capture variation across GA in fetal US
Level sets	Echocardiography [9] Left ventricle tracking in MR and US [140]	**Strengths:** Flexible tool for image segmentation, can force rigidity to compensate for shadowing. **Limitations:** Requires the tuning of abstract properties to alter segmentation, may require different parameters across GA, often fail to find a globally optimised solution
Active contours/snakes	Heart structures in MRI [203] Breast cysts in US [203] Bone in CT [203] Echocardiography [114]	**Strengths:** Can compensate for shadowing **Limitations:** Can become trapped in local minima, limited functionality on higher dimensions, may require different parameters across GA, unintuitive user input
Graph cuts	Cervical lymph nodes in US [32] Fetal head in US [40] Breast tumours in US [36] Gallstones in US [4] Liver in CT [50] Prostate in US [206] Photography [20]	**Strengths:** Optimise globally, can be used in nD space **Limitations:** Unintuitive user input, often fail to find a globally optimised solution
Superpixels	Segment the fetal skull in US [119] Tumours in US [72, 145] Photography [3, 58, 151]	**Strengths:** Respects object boundaries in US, good way to simplify segmentation problems **Limitations:** Susceptible to shadowing and attenuation, requires classification of each superpixel, difficult to control
Live Wire	Brain MRI [130] Bone MRI [52] Knee CT [52] Gall bladder in US [101]	**Strengths:** Highly intuitive user input, user can observe potential boundary positions, can adapt as the user adds anchor points, piece-wise inputs are possible **Limitations:** Requires more user interaction than other methods
Live Lane	Bone MRI [52]	**Strengths:** Highly intuitive user input **Limitations:** Difficult to use with a mouse/trackpad, requires more user interaction than other methods

3.2 Optimisation Methods

Semi-automated methods for segmentation offer a number of benefits over purely automated or manual methods, allowing for increased speed and accuracy of operation while allowing for more flexibility when dealing with abnormal image conditions [136]. Investigations into manual segmentation of fetal US images witnessed interobserver deviation in the order of 10%, and could therefore make a significant difference to conclusions made from such measurements [99]. O'Donnell shows a significant reduction (around 30%) in segmentation times of MRI when using an interactive segmentation method [129]. Better standardisation of measurements via systems such as semi-automated segmentation could help to reduce such bias and allow for faster, more accurate classification of fetal development.

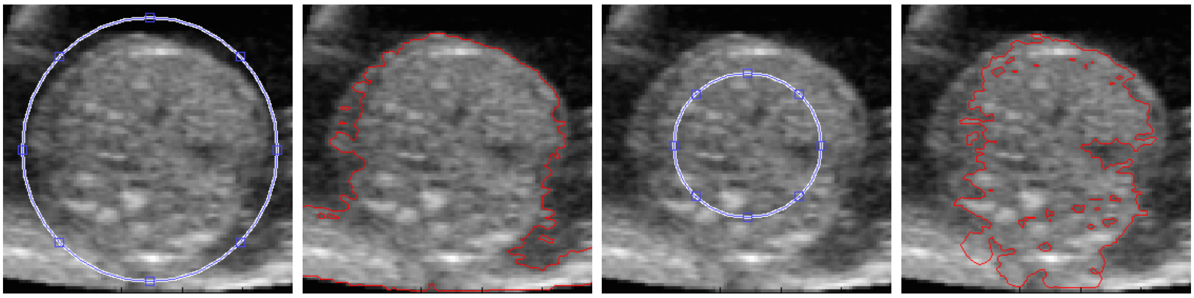
Olabarriaga and Smeulders comment on the importance of semi-automation over automated processes [136]. A significant obstacle to automated imaging is the variability of images. While automated segmentation is suitable for applications where fast, high throughput segmentation is valued over precision, interaction is needed when high accuracy is required and is only possible when a manageable volume of images are to be processed [136]. A system used as an exemplar throughout Olabarriaga and Smeulders' review was Live Wire [52, 136]. Other techniques for user input explored in this review include; free drawn user interaction, the selection of control points, direct curve deformation, the positioning of seed points for an expanding-area and the adjustment of a pre-defined template, as well as more abstract, parameter setting methods [136]. Table 3.1 summarises the optimisation methods explored in this section, detailing their applications as well as their potential strengths and weaknesses in the context of fetal US segmentation.

3.2.1 Active Shape/Appearance Models

ASM build up a statistical shape-based description of the object of interest from a set of prior examples and Active Appearance Models (AAM) expand on this concept by incorporating additional image properties such as intensity distribution to take into account the object's appearance [17]. This model can then be fitted to new instances of the object.

In the field of US, AAM has been applied to segmentation in echocardiography [17, 141] and has been used for instance, in conjunction with Gabor filtering for prostate segmentation from transrectal US images [166]. 3D AAM has also been applied to the segmentation of cardiac MR and US images [115] and the segmentation of fetal US images [82]. The main advantage of ASM/AAM is that the appearance of shadowing can be worked into the design of the model, allowing for accurate segmentation in challenging imaging conditions. Many of the examples in which ASM and AAM have been successful have been in cases of echocardiography and prostate segmentation, where boundaries are generally strong, appearance within the object is consistent and the shape falls within a predictable set of states. Additionally, shape is often a weaker constraint for US analysis in particular, as appearance variation can dominate.

A major disadvantage with ASM and AAM in the application of fetal segmentation problems is that solutions tend to be deviations around a mean representation of the object in question. This may not be appropriate for a method that seeks to segment fat tissue from both AGA and SGA fetuses, before they are identified as such. Additionally, variation in appearance with GA can be vast; shadowing becomes more severe with increased bone density, the shape of objects can change due to compression against the wall of the uterus as the fetus grows and the appearance of tissue changes as it develops. This means that a number of models may need to be built across gestation. This would require more examples than are feasible in the scope of this study to create models with sufficient flexibility. When creating such a model, it would be important that a wide



(a) **(b)** **(c)** **(d)**

Figure 3.3: *Two examples of using level sets on an US image of a fetal abdominal cross section at 16 weeks, 6 days. While **(a)** and **(c)** show initialisations outside and inside of the target boundary respectively, **(b)** and **(d)** show the corresponding segmentation resulting from each initial state.*

variety of fetal growth trajectories are included, to prevent bias towards the 'normal' shape for that age. Taking a standard population may be inappropriate, as SGA fetuses would, by definition, represent only 10% of the examples. This could result in the model not taking SGA variation into account as broadly as AGA examples.

3.2.2 Level Sets

Osher and Sethian presented the concept of building a mathematical model for a closed propagating front which takes into account local curvature, simplifying this process by envisioning the propagation as different layers of a level set [138]. In image segmentation, the zero level set of this higher dimension surface will be the segmented boundary, and the exact form of the level set can be influenced by whatever image parameters the user desires [106]. In the field of US imaging, Belaid *et al.* segmented echocardiography images using level sets in conjunction with LP calculations (described further in Section 3.3.2) [9]. The method has also been used in the segmentation and tracking of the left ventricle in MR and US images by considering a shape-driven approach [140]. Rigidity terms in in the level set could be useful in fetal US, where shadowing can result in incomplete boundaries.

Figure 3.3 demonstrates some of the potential difficulties in using level sets for fetal US, where boundaries are not as strong as some of the situations where this method has been successful. In this example, intensity gradient, line curvature and background/foreground probability (based on the intensity of pixels compared to the mean intensity inside and outside of the boundary) have been used to steer the progression of the level set. However, the level set will assume similar properties along the entire object boundary, which may not be the case in many fetal US images. A potential issue is that initialisation heavily influences the resulting segmentation, with user-defined contours set on the inside and the outside of the boundary of the object of interest producing vastly different results. This means that the user may need to re-segment using different initialisation parameters, which may not behave intuitively.

Figure 3.3b demonstrates a specific issue with using level sets on some fetal US images; in cases where the fetus is in contact with the wall of the uterus, the level set contour can be 'dragged' into adjacent but unrelated regions that have similar appearance (although this could be more of a statement on the use of grey level properties in US applications). In contrast, Figure 3.3d shows an example where regions within the abdomen (such as vasculature) have more in common with the regions outside of the abdomen, hindering the propagation of the boundary. Even with non-intensity based measures, it is likely that such variation within an object of interest will occur. While curvature can be used to limit this, it can result in a loss of detail on object boundaries. The balance between shape preservation and robustness is particularly important in US, where speckle can influence the path greatly.

A related problem is that level sets can become trapped in local minima. For soft tissue this can be a problem because fat/lean tissue boundaries may be adjacent to stronger (fat/fluid or lean/bone) boundaries. Avoiding these could complicate initialisation. This method is particularly unsuitable as a solution to fetal face segmentation, as it requires a closed curve, while the face delineation will be either a line segment in 2D or a plane in 3D.

3.2.3 Active Contours/Snakes

Active contours/snakes are a very well established form of image segmentation in which a line or surface deforms to find the lowest-energy boundary it can occupy; the contour adjusts according to a number of different forces, which can be image-dependent (such as being drawn up or down an intensity gradient) or user-dependent (such as the selection of the elasticity of the contour) [86].

Different methods of defining the 'energy' (or 'cost') that the contour is driven to minimise have been used over the years, depending on the application. In the segmentation of tumours in US images, for example the Gabor response has been used to guide the contour [33]. In other cases, optical flow has been used to achieve segmentation in echocardiography, and tracking is accomplished by using the previous frame in a sequence as the initialisation for the next frame [114]. By combining snakes with features of level set contour evolution (such as adding an inflation term), heart structures have been segmented from MRI, breast cysts from US and bone from CT [203]. These examples do, however, often have stronger boundaries than can be found in the target structures for this thesis. Zouqi and Samarabandu avoid using active contours, reasoning that reducing segmentation to an optimisation problem exposes it to the weaknesses of optimisation methods, such as becoming trapped in local minima (in the case of gradient descent) or having limited functionality on higher dimensions (in the case of dynamic programming) [206]. Yezzi *et al.* also note the potential for snakes to fall into undesired minima, and the unintuitive relationship between user input and the final contour [203].

Initialisation and choice of energy terms are important factors in influencing the final result. The user can influence the final contour through varied initial contour placements and adjusting cost weightings, but the outcome is not necessarily predictable. It may be difficult to force the contour out of local minima, for example where two image boundaries are close together but the weaker of the two is desired. However, the selection of appropriate cost terms and multi-scale analysis should facilitate this process. Further to this, weightings of user-selected parameters such as the contour's elasticity can appear very detached from the final form of the segmentation, and may be inappropriate when a globally determined elasticity is insufficient to define a boundary accurately [203].

3.2.4 Graph cuts

Shi and Malik present the use of 'normalized cuts' (N-cuts) in image segmentation, by treating the pixels of the image as the nodes of a connected graph, with each connection being assigned a different cost [168]. The basic form of graph cut methods involves finding the cut through the image which represents the minimum combined cost through these connections, but Shi and Malik noted that this method favoured the partitioning of smaller regions; the solution, they found, was to express the cut cost as a fraction of the total cost of connections within each graph (hence normalised) [168]. This technique has been used within the field of US image segmentation, and has also been used for the delineation of boundaries for cervical lymph nodes [32], breast tumours [36] and gallstones [4]. Zouqi and Samarabandu favoured graph cuts over active contours, as they can optimise globally and also be used in nD space [206].

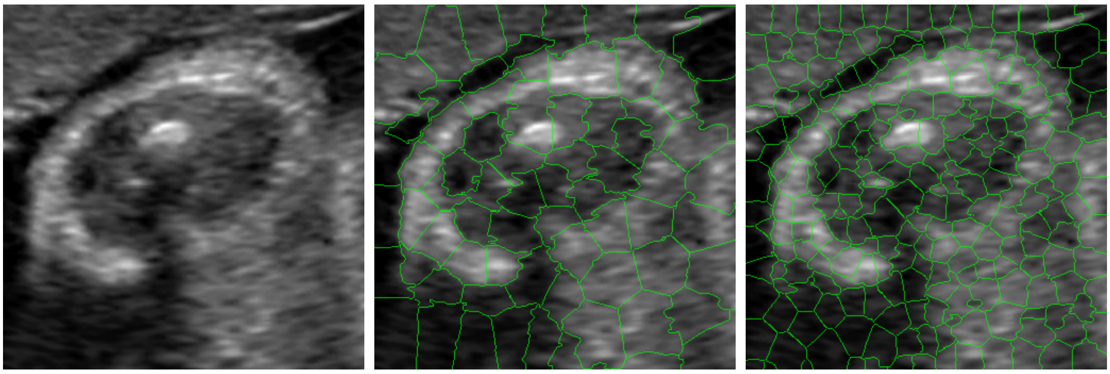
Although the N-cut method of graph cut segmentation is automated, these methods require some additional user input in order to achieve a desired segmentation. For example, in the case of breast tumour segmentation [36], the user is required to crop the image around the tumour boundary to a fairly tight degree. Even though the example images appear to show decent agreement with clinical segmentations, the assessment of their 500 segmentations was vague, claiming that 90% "satisfied clinical demand" [36] while not providing any details of quantification of accuracy. In the case of gallstone segmentation, the resulting boundary contains a fair amount of extraneous information in addition to the gallstone itself, such as the wall of the entire gall bladder [4]. Similarly, the segmentation of cervical lymph nodes included unrelated regions with similar intensity values [32]. This would require some additional manual user-interaction to correct, making the N-cut method, overall, semi-automated.

Boykov *et al.* note that, much like level sets, N-cut methods regularly fail to find a globally optimised solution [20]. They present an alternative graph cut-based system of segmentation in which the user labels different regions as 'foreground' and 'background,' which forces the assignment of the marked pixels to each group, thereby influencing the position of the minimum cut [20]. There have been cases of automated boundary improvement; Zouqi and Samarabandu iterate their results using a Fuzzy Inference System, which adds hard constraints to correct the weakest section of the boundary [206]. Although this interaction can be finer (as the user can exclude specific regions), it still does not allow for an intuitive relationship between the user and the segmentation process.

3.2.5 Superpixels

Ren and Malik present a method of segmenting images which begins with an 'over-segmentation' of the image into small cells with common features, which can then be classified to present a full segmentation [151]. Superpixels were initially calculated using N-cuts graph partitioning, regarding superpixels as the graph nodes rather than pixels [151]. Other methods have used a clustering algorithm by assigning evenly spaced points as initial cluster nodes and using an iterative approach to redefine cluster centres and boundaries until pixel classifications remain stable [3]. Fulkerson *et al.* used 'quick shift' (a non-Euclidean space based method for clustering [189]) in their method over the aforementioned alternatives, as this does not pre-define the arrangement or scale of the superpixels produced (although this may be a disadvantage in some methods) [58].

The concept of superpixels has been successfully utilised in a number of different segmentation problems in US images. Segmentation is often achieved through classification; first the image is split into superpixels, then each superpixel is classified as either object or background [72]. Namburete and Noble used shape information and intensity from superpixels in US images as features for a random forest classification to segment the fetal skull, yielding high segmentation performance on a test set of 10 images from the INTERGROWTH-21st database [11,119]. Hao *et al.* use multiscale superpixels and take



(a) (b) (c)

Figure 3.4: *Superpixels as determined by a MATLAB implementation of the SLIC superpixels algorithm as described by Achanta* et al. *[3] **(a)** An US image of a cross section of fetal arm and the corresponding superpixel boundaries using **(b)** 100 and **(c)** 300 initial seed points*

into account the properties of neighbouring superpixels in order to segment breast tumours in US images [72]. Quan *et al.* have applied superpixels to tumor segmentation in adult liver US by performing N-cuts [145].

Superpixels could present an advantage in the application of texture-based image segmentation, as, when appropriately applied, they are less likely to cross tissue boundaries. This means that texture measures observed within superpixels are more likely to represent the value of a single tissue type, making classification between objects more meaningful than using information from block regions that may encompass more than one tissue type. Figure 3.4 shows the application of a MATLAB implimentation of SLIC (simple linear interative clustering) superpixels, following the description of the algorithm published by Achanta *et al.* [3], when different numbers of initial classification seeds are used. A difficulty with the use of superpixels is that it can be difficult to predict exactly where the final superpixel boundaries will lie; the number of seed points and the number of iterations are the only controlling factors, which may not always provide suitable boundaries. Additionally, the requirement for further classification to form a final segmentation can complicate the process.

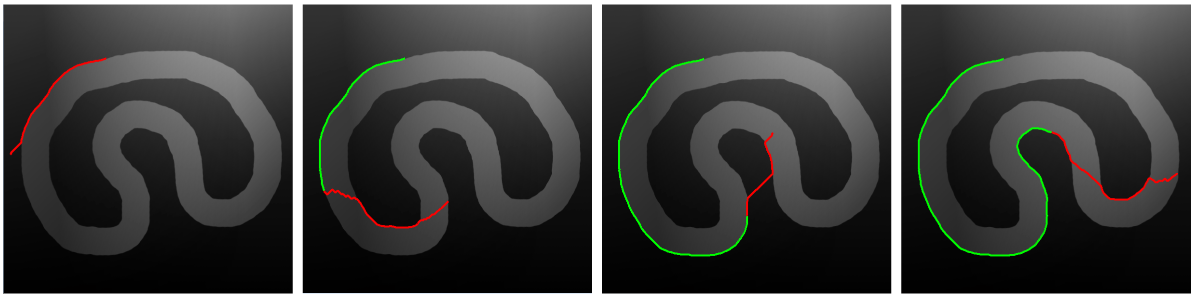


Figure 3.5: *An example of gradient-based Live Wire being applied to a synthetic image. The red line leads from the last placed anchor point to the position of the cursor, finding the shortest path between the two in real-time. The green line shows the previously anchored segment of Live Wire contour.*

3.2.6 Live Wire

Used as an example numerous times in the work by Olabarriaga and Smeulders is a method for interactive segmentation of general images developed by Falcão *et al.*; Live Wire [52, 136]. A benefit of interactive segmentation such as this is that it provides a far more intuitive form of user input when compared to other semi-automated methods. In the Live Wire method, the user selects points known to be on the desired boundary, with the algorithm using cost terms such as local gradient information and the boundary location on previous image slices to trace a path between the two points.

Figure 3.5 shows an example of the process of segmenting a synthetic image using Live Wire, with green lines showing the anchored segments of Live Wire contour and the red showing the shortest cost path between the last placed anchor point and the user's cursor. The constructed contours can be designed to be affected by contours on adjoining slices, meaning that this method can be extended to 3D scans. This allows the user to observe different possibilities for the object boundary's position before committing to any actions.

The structure of the algorithm makes it particularly useful for adaptation; a number of cost terms are taken into account with the plotting of the contour such as image intensity 'inside' and 'outside' of the boundary, the gradient magnitude and gradient direction across the boundary, and the distance from the boundary on a previous slice. Additionally, piecewise measures can be incorporated into the system, as the properties of segments the user has already defined can be used to guide future sections of the segmentation. The modular nature of the algorithm means that it is possible to create additional cost terms which can be tailored to specific applications and improve performance.

Levienaise-Obadia and Gee use a method related to Live Wire for the specific purpose of US segmentation [101]. Additional data gathered for the positioning of the contour between points is the texture on either side of the initial estimate for the line; this takes advantage of the different speckle pattern that will be apparent in different tissue [101]. Unfortunately, there has not been a quantitative assessment of this technique, meaning that the performance of such methods has not been fully explored. However, several important points are raised that should be considered in the creation of a semi-automated US algorithm, such as the use of geometric constraints to compensate for shadowing [101]. This is particularly important for the intended application of segmentation of fat around the thigh and arm, as the bone will block US signals, particularly later in pregnancy when calcification is greater. The main disadvantage of this method is the more advanced level of user interaction that is required when compared to other methods.

3.2.7 Live Lane

Another semi-automated technique developed by Falcao *et al.* and presented alongside Live Wire is Live Lane [52]. This is a method related to Live Wire; in this case the user moves their cursor around the desired contour. The speed and acceleration of the cursor is used to adjust the width of a 'lane' in which the contour is automatically drawn in real time. While this introduces interesting new properties to the segmentation process, such a method is more difficult to use with a computer mouse or trackpad. It is better suited to specialist image segmentation hardware, such as a tablet. Another disadvantage is that it does not allow the user to observe where the segmentation will lie during the interactive process.

Both Live Wire and Live Lane were tested on clinical MRI rather than US images by

Falcao *et al.* [52]. Investigations into the accuracy and efficiency of these methods against standard, manual segmentation, showed that there was a clear increase in repeatability (suggesting increased accuracy) and, in two of the three operators testing the techniques, a significant increase in speed [52]. However, it should be considered that speed will be largely dependent on the user's ability to operate the software; the report highlights that there is potential for improvement as the user becomes more adept at using the software, though this is very difficult to predict accurately. As only three operators and 30 images were used, further investigation is required to alleviate these discrepancies. Another consideration is the expertise of those testing the algorithm; in the case of clinicians, they may not be as familiar with operating the system as engineers involved in its construction and, in the case of engineers, clinical accuracy may be lacking. In order to accurately assess any improvement of speed, tests need to be conducted in a way that ensures clinical relevance, while avoiding bias presented by operators who may be knowledgeable in the inner workings of the system.

Overview: Optimisation Methods

In a review of segmentation methods by Olabarriaga and Smeulders, Live Wire appeared to satisfy the goals of simplicity, intuitiveness, predictability, the computational efficiency and the level of user interaction [136]. In the specific context of US segmentation, Live Wire seems to be a suitable basis for semi-automated segmentation when compared to other methods. By applying parameters that are more relevant to US images than local gradients as cost functions, a useful and versatile tool for US segmentation may be developed.

Table 3.2: *Summary of image features used for image segmentation*

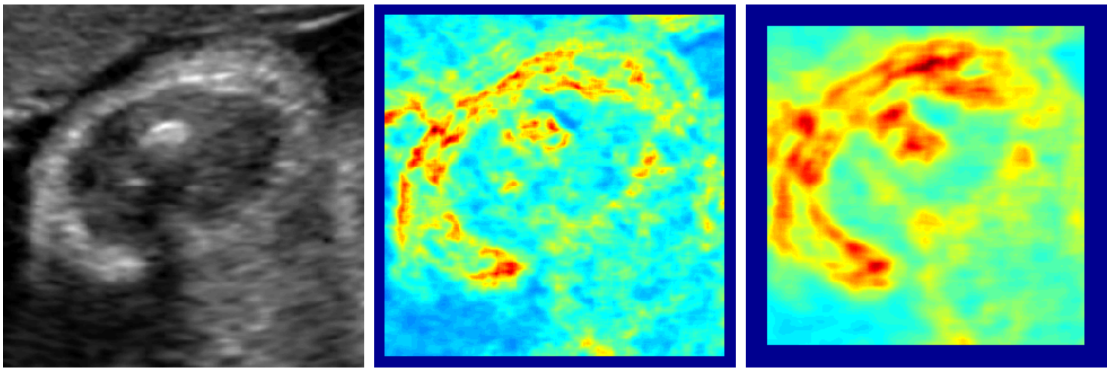
Image Feature	Application	Strengths & Limitations
Grey level statistics	Prostate in US [206] Lymph node in US [32] Breast tumour in US [36] Gallbladder in US [101] Gallstones in US [4] Liver in CT [50]	**Strengths:** Computationally simple **Limitations:** Intensity does not convey tissue properties in US, susceptible to shadowing and attenuation
Nakagami distribution	Masses in US [95] Liver fibrosis in US [76] Breast and liver tumours in US [165]	**Strengths:** Attempts to model US properties **Limitations:** Difficulties at object boundaries
Local Phase	Brain MRI [112, 130] Liver tumour in US [38, 39] Echocardiography [66, 117, 149] MR Cardiovascular [205]	**Strengths:** Intensity invariant **Limitations:** Dependent on good filter selection, can have poor response at object edges in US
Feature Asymmetry/Symmetry	Photography [91, 92] 2D+T and 3D echocardiography [117, 148] Fetal soft tissue segmentation in US [155] Fetal abdomen in US [147] Bone segmentation in US [69]	**Strengths:** Intensity invariant, high response on edges in US **Limitations:** Dependent on good filter selection
Local Binary Patterns	Photographic textures [87, 102, 123, 132–134, 142] Face and palm recognition [5, 68] Thyroid in US [79] Brain MRI [187, 188] Histology [146] Lung tissue analysis in CT [171] Mammography [123] IVUS [144]	**Strengths:** Intensity invariant, computationally simple **Limitations:** Influenced by orientation of texture, can require many different filters to work effectively, requires large training set, more suited to classification than segmentation
Non-local means	Photography [23] Noise reduction in US [42, 43, 204] Noise reduction in MRI [44, 199, 200] Liver in 3D US [43]	**Strengths:** Reduces the appearance of speckle in US, respects image boundaries **Limitations:** Speckle could convey tissue properties

3.3 Image Features

In order to drive the optimisation methods described, appropriate image features are required to define the boundaries of objects of interest. A common opinion as to why segmentation in US can be difficult is the distortion caused by the speckle, which is often referred to as 'noise' and considered an undesirable trait of the imaging modality. Speckle is caused by scatterers suspended in the observed medium, but the appearance of speckle is a function of the transducer settings, meaning that speckle statistics cannot explicitly describe tissue properties [196]. However, it is possible that suitable image features could be used to detect regions of different scatterer density within an image, and provide distinguishing features to separate regions of an image by tissue. The fact that segmentation does not need to characterise tissue appearance, but rather determine differences in texture, makes texture measures a potential option in addressing such problems [128].

Wan and Zhou note that features commonly used for US based texture classification can be summarised as spatial-domain, frequency-domain and model-based, claiming that out of these, frequency based methods tend to perform most successfully [197]. Noble and Boukerroui note that there are a vast number of texture measures used in the literature, observing that texture methods incorporating Haralick's co-occurrence matrices have been successful in a wide range of applications, most notably in the characterisation and classification of liver and breast tissue in US [128]. It is therefore possible that regions of particular tissue can be detected using such classifiers. An alternative approach is to use methods which mitigate the influence of speckle, attenuation and shadowing.

This section will summarise some of the promising image features currently used in segmentation and classification methods. Table 3.2 summarises each image feature explored in this section, detailing their applications as well as their potential strengths and weaknesses in the context of fetal US segmentation.



(a) *Original image* (b) *10 pixel window* (c) *20 pixel window*

Figure 3.6: *The shape parameter for the Nakagami distribution fit to the intensity histogram of different sized sliding windows on an US image of a fetal arm cross section.*

3.3.1 Local Histogram Methods

Grey level statistics

In the simplest case, and fairly commonly in the very earliest of digital image processing, grey level values have been used to segment images by applying a basic intensity level threshold [85]. Quantification of the shape of grey level histograms have been used in the identification of liver cirrhosis against fatty liver in US [81]. In certain applications, texture measures using simple grey level statistics such as magnitude and direction of grey level gradients have been shown to be effective [132].

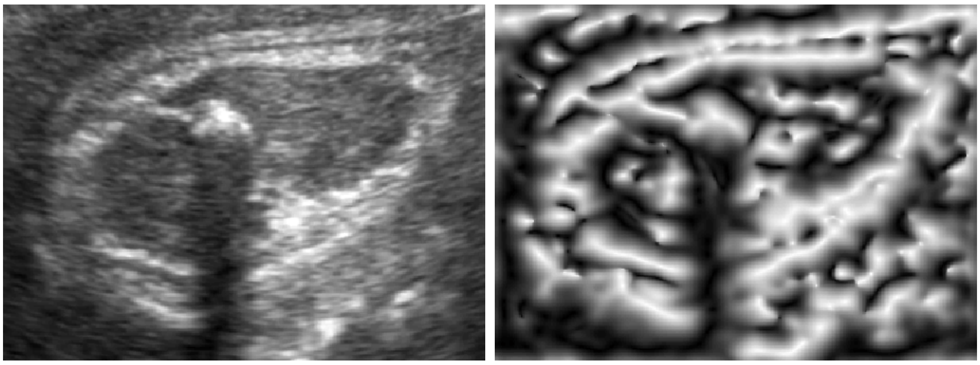
Despite this, low-level grey level statistics are deeply unsuitable to US applications, since they fall into all of the traps presented by the difficulties of the modality; specular appearance can distort histograms, grey level values will be inconsistent and tend to fall with image depth, shadowing will provide anomalous results for any region underneath dense tissue and intensity values for a given anatomical feature are likely to be vastly different for each individual scan taken. However, as speckle is tissue dependent, it is possible that observing the shape of histograms in image patches could provide an insight into where anatomical boundaries lie.

Nakagami distribution

A method of determining the properties of texture in an image is to fit the pixel intensity distribution in a given region with a suitable model, using the parameters of the model to define that region. Previous attempts at achieving this in US have involved fitting to Rayleigh [195] and K-distribution [163] distributions. Shankar, however, proposed that backscattering from tissue could be modelled in a generalised way using a Nakagami distribution, citing analytic simplicity and the ability to depict a wide variety of imaging conditions (varying scatterer sizes and densities) as main advantages of the method [164]. Nakagami images are produced by finding the shape parameter m and scaling parameter Ω of the Nakagami distribution that best fits the intensity histogram of a given pixel region [95]. This method has been applied to the classification of liver fibrosis in rats from the US RF-signal, as well as the classification of breast masses from the B-mode image [76]. The concept of modelling US scattering as a Nakagami distribution has also been used in speckle reduction [165, 204].

Figure 3.6 shows an example of using the Nakagami distribution to characterise fetal US images. In each image, the intensity histogram of a sliding window is fitted to a Nakagami distribution and the shape parameter is assigned to the central pixel of the region. From this, the variation in parameter values obtained depending on the size of window can be seen clearly. Selection of an appropriate window size is an important factor in the success of the segmentation, as too small a window size will not provide enough data points to accurately define the texture of the region, and too large a window may involve boundaries between multiple tissue types.

Larrue and Noble confront the problem of the selection of window size by using Gamma kernel density estimation to provide an estimate of the texture's probability density function from a small window, which ordinarily may not have had enough points to accurately portray this [95]. In applications on US images of breast and liver tumours, it was noted that the usually fuzzy boundaries of objects of interest were more clearly



(a) **(b)**

Figure 3.7: *(a) US image of a cross section of fetal arm (b) LP of the image, using a Gaussian Derivative filter of s = 20*

defined and with a higher contrast in the Nakagami images than the B-mode images [95]. Further to this, application to an US phantom showed a lower standard deviation of Nakagami values within homogeneous regions than without the prior Gamma kernel density estimation [95]. The more clearly defined boundaries and higher spatial resolution provided by this extension means that Nakagami images could potentially be used for segmentation rather than classification [95].

3.3.2 Local Filter-based Methods

Local Phase

There are a number of papers that suggest that the use of LP could be advantageous in the analysis of US images [9,66,117]. Phase-based segmentation has already been applied to echocardiography images in order to enhance segmentation and feature detection [117, 149]. Mellor and Brady also use the technique in multimodal image registration, stating that its use as "conservative information reduction" is useful in an environment of poor intensity relationships but strong structural relationships, such as in US [112].

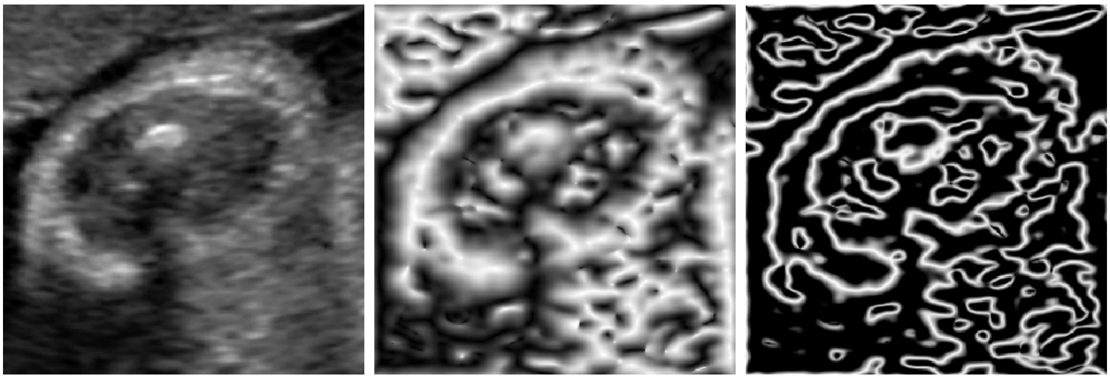
The advantage of using LP is that it is intensity-invariant, proving especially useful when contrast is low [148]. Rather than being affected by the local energy of the signal, LP picks up changes in structure [54]. This is of particular interest in US as it means that signal attenuation can be rectified, making structural changes more prominent. LP has been used in level set segmentation of US by Belaid *et al.*, due to the limited success of conventional gradient based methods used on clinical US images [9]. Their method produced results cited as being robust to speckle and attenuation, which is advantageous to US methods. Figure 3.7 shows the application of LP to an US image.

O'Donnell *et al.* replace the conventional intensity gradient measures of Live Wire with LP calculations to form 'Phasewire' [129,130]. The local energy of the signal is used to quantify the 'certainty' of the LP values and hence modify the Live Wire costs to favour routes of high certainty of edge structures. While O'Donnell *et al.* do not conclusively show the advantage of a phase-based Live Wire technique, they do provide an initial argument for its use [129,130]. Zhang *et al.* have used LP as an image descriptor in the registration of 3D US to 2D MR images [205], while Mellor and Brady have applied LP to registration in both registration of US images to MR-T1, and MR-T1 images to MR-T2 [112]. A measure derived from LP calculation called phase congruency has been used for tumour tracking in 2D US using both Diffeomorphic registration [38] and Log-Demons registration [39].

Maltaverne *et al.* estimate motion in elastography from US images using local *orientation*, an output related to LP, but provides information on the *direction* of edge features; while LP is the angle between the even and odd response, the LO is the angle between the responses of the orthogonal filters [107]. Ojansivu *et al* used LP to classify image textures, using low-frequency features so that texture features are robust to blurring [135]. In the field of segmentation, LP has also been applied to level set segmentation in US images, demonstrated in synthetic phantom images and validated 20 segmentations of the left ventricle in echocardiography, showing high levels of accuracy between the two [9].

Feature Asymmetry

FA is an extension of LP calculations, considering the difference in odd and even response of the LP filters over a number of different scales. FA will have a positive value when the odd response is greater than the even response, therefore indicating image edges. The



(a) (b) (c)

Figure 3.8: *(a) US image of a cross section of fetal arm (b) LP of the image, using a Gaussian Derivative filter of* $s = 20$ *(c) The FA of the image, using a Gaussian Derivative filter and scales of* $s = [18, 20, 22]$

ability to take into account structures at different scales of LP and identify regions of discontinuity can help to differentiate between image features caused by image artefacts and those of clinical significance [117].

FA has been applied to boundary detection in both 2D+T [117] and 3D [148] echocardiography. It has also been utilised in the automatic segmentation of soft tissue in fetal US images [156]. Feature Symmetry (FS) is conceptually the same as FA, but gives a positive value when the *even* filter response is larger than the *odd* filter response [92]. This means that FS will give a response of $FS > 0$ for ridge- and blob-like structures rather than edge-like structures [92]. FS has been applied in fetal US to detect features in abdominal slices, including the umbilical vein and stomach [147]. This concept has been applied to bone segmentation and fracture detection in 3D US [69].

Figure 3.8 shows an example of FA being applied to an US image of a fetal arm cross section. Adding to the difficulty of filter and scale choices in LP, FA presents an additional challenge as the number and value of scales used can be varied, resulting in a practically infinite combination of filters and scales which may be more or less suitable in different applications or image conditions. This means that scale and filter selection can only, in practice, be determined empirically. It is possible that a set of appropriate scales could be estimated on a basis of feature size, speckle size or fetal GA, but an objective 'best' value would be difficult to determine. Nonetheless, FA has become popular in US image processing tasks, allowing for the detection of image features in a way that is robust to the issues of speckle, attenuation and shadowing which prevail in the modality.

Local Binary Patterns

Local Binary Patterns (LBP) were introduced by Ojala *et al.* as a potential simple texture measure [132], as an adaptation of the 'texture unit' presented by Wang and He [198]. In this initial form of LBP, the 3x3 neighbourhood surrounding a pixel is thresholded by the value of the centre pixel, forming an 8-bit binary word which can then be given a numeric value that represents the texture surrounding that pixel [132]. LBP has seen a wide array of applications, including classification of mammographic images [123], thyroid nodule segmentation in US [79] and plaque segmentation in Intervascular US (IVUS) [144].

This has since been modified to be rotationally invariant (where bit-shifted words are seen as equivalent) [133], to discriminate between 'uniform' and 'non-uniform' patterns (defined by the number of bitwise transitions in the binary word) [134], to contain a 'fuzzy' element [79], and has been applied to a varied range of texture classification as well as segmentation problems. Keramidas *et al.* claim that the major advantages of fuzzy LBP over conventional LBP methods are an increased robustness to noise and better differentiation between textures [87]. An advantage of this method includes computational simplicity [134]. LBP is also contrast invariant, though this has been seen as a disadvantage by Ojala and Pietikäinen, overcome simply by comparing the average value of pixels assigned the value 1 compared to the average of those assigned the value 0 [131].

The concept of rotationally invariant LBPs is a popular modification, being used in problems such as photographic textures, face and palm recognition, analysis of brain MRI [187,188], classification of histological images and lung tissue analysis in CT [134,146,171]. Pietikäinen *et al.* point out that rotational invariance is only true in the 'digital sense,' and so textures are only seen as entirely equivalent at 90° [142].

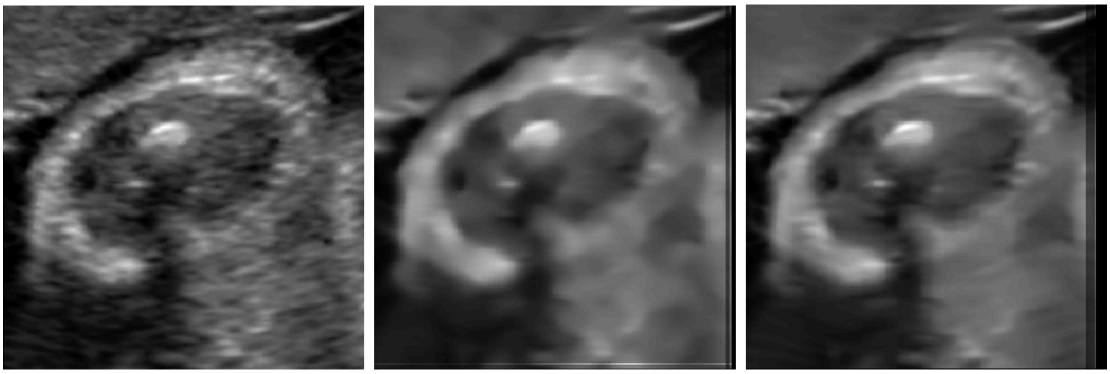
LBP and its many adaptations present an interesting set of computationally simple, grey scale invariant and (arguably) rotationally-invariant texture measures, which could present a possible solution to tissue classification in US images. This, however, would rely on a large amount of training data to properly characterise tissue types. Additionally, this form of image feature is more suited to a classification rather than segmentation problem.

3.3.3 Non-local Means

Presented by Buades *et al.*, Non-local means (NL means) is a method used to reduce image 'noise' but preserve image boundaries, originally demonstrated to reduce noise in photographic images [23]. This is a block-based method which uses the distribution of pixel intensities in a surrounding neighbourhood to inform the calculation of a given image patch, weighting the influence of each surrounding image patch by a weight related to the intensity difference between it and the central patch.

It has since been adopted in medical imaging for its ability to reduce noise without corrupting object boundaries. NL means has been used to improve the appearance of diffusion-weighted and diffusion-tensor MRI by replacing the standard Gaussian-weighted Euclidean distance with a log-Euclidean distance [199]. This work was later expanded on to specifically target the distribution of the noise in MRI, modelled by a Rician distribution and modifying the calculation of the filtered image by introducing the variance of the noise distribution [200]. Further expansions in MRI allow for NL means in 3D, preselecting a subset of voxels to prevent excessive computation of redundant data [44].

More recently, this method has also been adapted to be more suitable to US applications by adopting a Bayesian approach to cost calculation, modelling the noise distribution as multiplicative Gaussian noise ($image_{noisy} = image_{clean} + image_{clean} \times noise$) and using a Pearson distance calculation to weight the image patches [42]. This work was expanded to include qualitative assessment on 3D liver US images in addition to the synthetic images previously used as examples of the method in use [43]. It should be considered in such assessment that examples of synthetic images being used for the



(a) (b) (c)

Figure 3.9: *(**a**) US image of a cross section of a fetal arm and corresponding NL means results using sets of inputs for (**b**) coarse and (**c**) fine filtering*

assessment of a model based approach can be misleading, as they can tend to follow the model more closely than real images.

Figure 3.9 shows two examples of a MATLAB implementation of NL means, following the description of the method by Coupé *et al.* [42]. This has been applied to a fetal arm cross section in US using two different sets of input parameters. It can be seen here how the appearance of speckle has been greatly reduced, but with the preservation of major features and object boundaries. The application of NL means to US images infers that the speckle is merely noise, degrading the signal and therefore of no use to informing segmentation, but could be used to inform non-texture based methods or to extract texture from the background intensity.

Overview: Image Features

A number of different image features have been explored, with the goal of providing meaningful information to drive the Live Wire algorithm in the context of fetal US. While LP has been used as an intensity invariant cost term for Live Wire in MRI [130], FA provides a potential means for detecting object boundaries in the challenging conditions of US, mitigating the influence of speckle and attenuation.

3.4 Overview

3.4.1 Literature overview

This section has outlined a selection of the different optimisation methods and image features that have been used to tackle segmentation problems, focusing in particular on the context of US imaging. Certain methods such as level sets, active contours, snakes and graph cuts have been applied successfully to a number of problems in US imaging, but often require user interaction that is frequently detached from the final segmentation. Live Wire is a fully interactive semi-automated method in which the user directly positions the boundary, which is guided by image properties. Because of this intuitive form of interaction, this thesis will be exploring adaptations that allow for the effective use of Live Wire for soft tissue segmentation in US.

Methods based on the analytic signal may provide a suitable measure in the case of detecting the soft tissue boundary, as it overcomes many of the issues presented by US imaging as a modality, most notably signal attenuation, low contrast boundaries and speckle. Indeed, LP alone has already been applied to Live Wire's application to medical imaging, but FA, a measure derived from the analytic signal, may prove more powerful in the context of US. Texture measures may provide the additional information needed to guide segmentation for liver segmentation, where boundaries in US are often difficult to observe. Tissue-dependent speckle pattern may help to differentiate regions.

Finally, ASMs provide a potential avenue for allowing face shape, once segmented, to be expressed as a form of comparable parameters. This could then be used to classify images depending on whether or not they exhibit features related to FAS. AAMs have already been used in face recognition and modelling, for example, in the construction of 3D models from 2D images [14]. By using a parameterised model for face appearance, segmentations of different faces can be characterised by the model parameters which result in a fit to the segmented surface. This can allow for very simple forms of classification to be applied to determine facial differences between two groups.

3.4.2 Validation of Methods

It can also be concluded from the work reviewed in this chapter that validating the absolute accuracy of US measurements for *in utero* applications is a difficult task, as is quantifying the success of semi-automated methods. Olabarriaga and Smeulders provide a comprehensive overview of how semi-automated segmentation methods should be assessed in terms of accuracy, repeatability, efficiency and interaction [136]. However, in most cases, accuracy can be a difficult measure since there is no such thing as a 'gold-standard,' only collected clinical opinions of where a boundary should lie, which may vary between clinicians and between attempts at segmentation. This means that it is important that several examples of each segmentation are taken (preferably by different individuals) to find such a consensus.

A complication of validating accuracy of semi-automated methods is that accuracy can often be improved with additional user interaction. Given enough time and effort, the system can be made to match the clinically defined boundary exactly. It can be argued that this renders most of standard measures irrelevant when assessing the quality of such a system. The time taken to segment and ease of use are more relevant to the success of the system, but prioritising the former may influence accuracy and the latter can be difficult to assess in a meaningful way. However, the number of actions taken by a user to form a boundary can be recorded and assessed, and be used as an approximation of ease of use, particularly when identical optimisation methods are being compared.

O'Donnell's validation method involved recording the number of user interactions per image set and average per slice, average time per slice and segmented volume, as well as an 'opinion poll' of clinicians using the system [129]. Doctors' opinions of a system designed for use in a clinical environment is an essential aspect, but is rarely considered explicitly in technical papers. This paper does not, however, make any comment regarding the accuracy or repeatability of results [129], which may reflect the fact that interactive methods such as Live Wire intend to facilitate human decision making rather than replacing it. Any attempt to quantify and compare the efficacy of semi-automated segmentation methods should take into account both the accuracy and the level of user interaction required to make that segmentation in order to form a definitive conclusion.

Finally, other aspects of segmentation which should be considered in the evaluation of any method are inter- and intraobserver variability. Sarris *et al.* have shown that differences in clinical opinion can result in differences of up to $\pm 11.1\%$ in certain fetal biometric measures [159]. This has two consequences which are important to the validation process. Firstly, it should be considered that if the manual and assisted segmentations are conducted by different people, there may be a discrepancy which can be attributed to difference of opinion, rather than inaccuracies in the method. Secondly, interobserver differences present a challenge which the segmentation method should aim to limit.

The personal contribution within this chapter is a comprehensive review of the state of the art in image segmentation tools, with respect to their application in addressing the problem of segmentation in a fetal US setting.

Chapter 4

Feature Asymmetry and Shape Guided Live Wire

In this first contributions' chapter a semi-automated segmentation method suitable for use specifically on US images is outlined and tested. This is achieved by modifying an existing Live Wire framework to follow edges defined by a combination of a FA based cost and a weak shape constraint. This ensures that the method will follow paths which favour edges more typical of US images than standard intensity-based methods would allow.

4.1 Live Wire Methodology

Live Wire was introduced in Section 3.2.6. The basis of the Live Wire algorithm is the construction of a cost matrix which defines the 'cost' of the path between each node (defined as a pixel vertex) and all adjacent nodes, and then finding the minimum cost path from one user-defined point to another [52]. In the original implementation, the cost is determined by the pixel values in the 'neighbourhood' around the path between nodes, shown in Figure 4.2a [52].

Any property of the pixel neighbourhood can be used to influence the cost; while Falcão *et al.* use the grey scale pixel intensities [52], Chodorowski *et al.* use colour information, Canny edge values and Laplacian zero crossing edge values [37] while O'Donnell

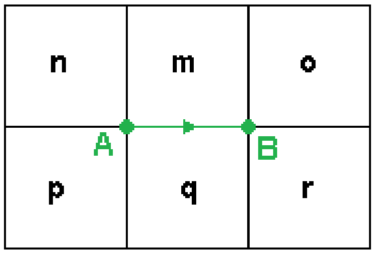
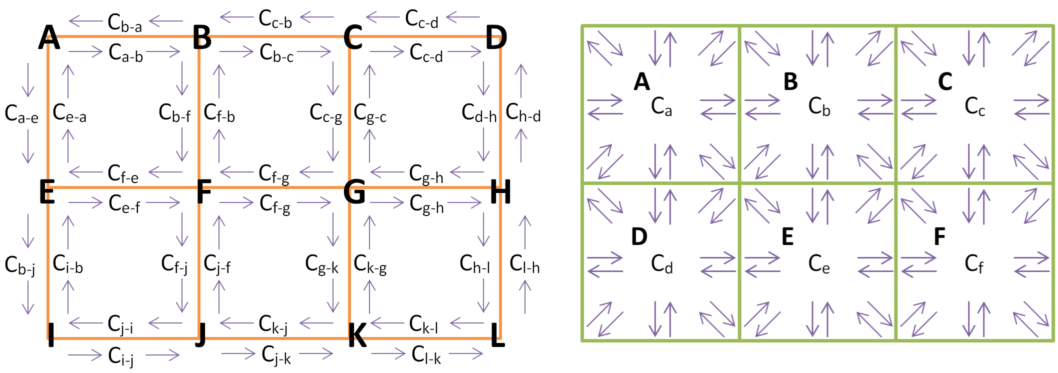


Figure 4.1: *Pixel neighbourhood around the path between nodes A and B.*

uses LP values [130].

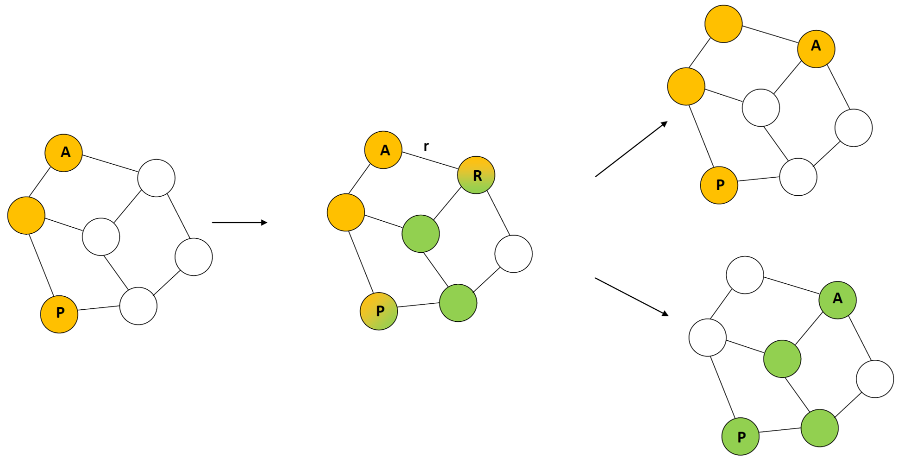
Further to this, the size of the pixel neighbourhood can be adapted. In Chodorowski *et al.*'s implementation, the intensity gradient is used within a cost function (using only pixels m and q from Figure 4.1) [37] while Falcão *et al.* adopt a cost based on the ratio of many surrounding gradient values, including all of those shown in Figure 4.2a [52]. Using a number of different costs allows for the algorithm to be trained by adjusting cost weights to fit against example contours, but cost optimisation decreases the ease of use of such a system.

The method outlined by Falcão *et al.* also introduces directional dependence to paths; costs are different depending on whether a gradient flows from, in the example shown in Figure 4.1, m to n or from n to m [52]. This means that the algorithm will prioritise different closely neighbouring contours depending on the direction in which the path



(a) (b)

Figure 4.2: *The differences between different applications of Live Wire;* ***(a)*** *Using nodes and edge-based costs, with directional dependence* ***(b)*** *Using pixel-based costs.*



(a) **(b)** **(c)**

Figure 4.3: *This diagram shows a simple example of Dijkstra's algorithm in use.* ***(a)*** *A recently found 'shortest path' between* ***P*** *and* ***A*** *is shown in orange.* ***(b)*** *The path* ***r*** *from node* ***A*** *to the adjoining node* ***R*** *is made. In this example, a shortest path between* ***P*** *and* ***R*** *has already been assigned (shown in green).* ***(c)*** *If the new path is shorter than the previous path or a previous path did not exist, this will be the new path to* ***P*** *associated with* ***R*** *(*top*). If the path including* ***r*** *is longer than the original path, it will be rejected as shown (*bottom*).* ***R*** *will then become the new* ***A*** *as the algorithm continues to the next iteration.*

is being drawn. The application of Live Wire from Chorodowski *et al.* deviates from the original methodology by applying costs in terms of pixels instead of vertices [37]. This reduces the search-space and simplifies the path calculation but removes directional dependence.

Figures 4.2a and 4.2b show the differences in cost basis between applications of Live Wire. In Figure 4.2a, the 'wire' travels between pixel vertices (labelled A to L). Travelling from node **A** to node **B** will have the path of cost C_{a-b}. However, travelling from **B** to **A** will be associated with a different cost; C_{b-a}. In Figure 4.2b the path is drawn between pixel centres. The line will accumulate the cost of each pixel it travels through. For example, moving from pixel **A** to pixel **B** will have the cost $C_a + C_b$. As a result, this will be the same cost as moving from pixel **B** to **A**.

4.2 Dijkstra's Algorithm

An important part of the Live Wire algorithm, once the cost matrix has been established, is the shortest path calculation. In their implementation of the Live Wire algorithm, Falcão *et al.* [52] and O'Donnell [129] use Dijkstra's algorithm in order to calculate the shortest path. This algorithm finds the shortest cost path to any given node from a seed point by following these steps:

- Take the initial seed point as **P**, the node thats minimum path was most recently calculated is **A**, the currently investigated node as **R** and **r** is the path between **A** and **R** (as shown in Figure 4.3b).

- The green route shows an existing path between **P** and **R**, while the yellow route shows the current path between **P** and **A**. In this example, the route from **P** to **R** via **A** is being tested.

- If **R** is not already part of a shortest path and the use of **r** results in a path length between **P** and **R** that is shorter than has already been discovered, it replaces the previous path (top option in Figure 4.3c). Otherwise, the path is rejected and the previous one is maintained (bottom option in Figure 4.3c). The new path is also recorded if **R** has not previously been considered in any path at all. In both cases, **R** becomes the new **A**, since it is the most recently calculated path.

- This is repeated until the path from all nodes to **P** has been found.

By following this algorithm, the shortest path to any given point from an initial seed point is calculated [47]. This can be useful in an interactive application such as Live Wire, as once the calculation has been completed the user can observe the shortest path to any point without further computation, allowing for a less interrupted decision making process between point anchoring. O'Donnell cites this as a possible disadvantage when applied to images as it can be "computationally wasteful," including paths which would be unlikely to occur in image segmentation [129]. However, the context of that paper should be considered; this may have been an important factor when the paper was written in 2001, but computational resources are now far more abundant. From experience of implementing this algorithm, this is not a significant issue when Dijkstra's algorithm is run on a 64-bit Windows machine with a 2.27GHz processor and 9GB RAM in a speed-optimised coding language such as C compared to slower environments such as MATLAB. On a standard US image, the path calculation process took approximately one second.

4.3 Feature Asymmetry

LP was introduced and its literature reviewed in 3.3.2. The LP of an image contains structural information and is intensity invariant, making it particularly useful for US, which attenuates with distance from the probe and allows for the detection of weaker, low contrast boundaries [9, 54, 149]. This is derived from the analytic signal, a tool used in 1D signal processing, expanded into a 2D representation known as the monogenic signal. The most important attribute of the analytic signal is the ability to extract phase and amplitude information of the original signal, referred to as the "split of identity" by Felsberg and Sommer [54].

4.3.1 Methodology of Local Phase

Felsberg and Sommer present the monogenic signal as an nD extension of the analytic signal (but in this case, only $n = 2$ is considered), replacing the Hilbert transform with the Reisz transform [54]. The monogenic signal serves a similar function to the analytic signal in 1D, giving information on structures present in a 2D image via the LP, and the energy of the image via the local amplitude [9]. Although FA is the measure used in the segmentation methods outlined in this thesis, LP is a commonly used facet of the monogenic signal in US applications. It is therefore important to establish the context of this value.

Mellor and Brady explain the analytic signal in terms of the following [112]; if a zero-mean 1D signal can be represented in the Fourier domain as

$$f(x) = \int_0^\infty A(\omega)cos(\omega x \ + \phi(w))dw, \tag{4.1}$$

this can also be represented as the real part of an exponential function, as $cos(\omega x \ + \phi(w)) \ = \ real[e^{i\omega x \ +\phi(w)}]$. Including all parts of the exponential function, the analytic signal, $f_A(x)$ is obtained, where

$$f_A(x) = \int_0^\infty A(\omega)e^{i\omega x \ +\phi(w)}dw \tag{4.2}$$

The analytic signal, $f_A(x)$, will have the same amplitude as the original signal, $f(x)$, but there will be a change in phase between the real and imaginary components. The modulus of $f_A(x)$ will represent the amplitude, while the argument will represent the phase [54]. As $e^{i\phi} = cos(\phi) + isin(\phi)$, it follows that

$$f_A(x) = \int_0^\infty A(\omega)cos(\omega x \ + \phi(w))dw + i\int_0^\infty A(\omega)sin(\omega x \ + \phi(w))dw \tag{4.3}$$

which can be written as

$$f_A(x) = f(x) + iH[f(x)], \tag{4.4}$$

where $f(x)$ is the original signal and $H[f(x)]$ is the Hilbert transform of the signal. It is this transform which is replaced by Felsberg and Sommer in order to apply this same concept to 2D signals [54].

For an nD signal, n+1 filters are required [112]. A zero-mean bandpass filter must be applied to the image, $I(x, y)$, in order to limit the range of frequencies being observed [54, 112]. This bandpass version of the image, $I_b(x, y)$ will represent the 'even' part of the signal, or $f(x)$ as described in Equation 4.3.

In order to form the bandpass image $I(x, y)_b$, the original image is convolved with a filter, most commonly Gabor, log-Gabor, Cauchy or Gaussian Derivative, although specialist filters for LP such as the Mellor-Brady filter [112] have also been used in the literature. The bandpass image can now be expressed as

$$I_b(x,y) = I(x,y) \otimes F(s), \tag{4.5}$$

where $I(x,y)$ is the original image, F is a filter of scale s and $\otimes$ is the convolution operation [112].

The 'odd' part of the signal described in Equation 4.3, $iH[f(x)]$, requires a 2D approximation of the Hilbert transform using two anti-symmetric filters [112]. As the Hilbert transform is a primarily 1D transform, a generalised 2D form must be used, known as the Riesz transform [54]. In this case, let the convolution kernels of the Reisz transform be represented by two anti-symmetric spatial filters, o_1, o_2, defined as

$$o_1(x,y) = \frac{x}{2\pi(x^2+y^2)^{3/2}} \qquad o_2(x,y) = \frac{y}{2\pi(x^2+y^2)^{3/2}}. \tag{4.6}$$

Therefore $f_M(x)$, the 2D representation of Equation 4.4, can be represented as

$$f_M(x) = I_b + [(o_1 \otimes I_b), (o_2 \otimes I_b)] \tag{4.7}$$

where I_b is the bandpass filtered version of the original image, I, and $\otimes$ denotes the convolution operation [9, 112]. The LP can finally be expressed as

$$\phi = tan^{-1}(I_b / \sqrt{(o_1 \otimes I_b)^2 + (o_2 \otimes I_b)^2}), \tag{4.8}$$

where values of LP, ϕ, takes a range of $-\pi/2 < \phi < \pi/2$.

4.3.2 Methodology of Feature Asymmetry

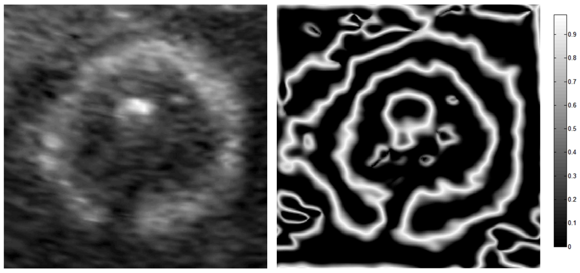
FA is a technique that uses different scales of filter described in Equation 4.5 in order to provide an indicator of the significance of structures in images [91]. By altering the response of the bandpass filter different scales of detail are detected, with significant structural features holding prominence across many scales. FA describes regions where the odd (edge) filter response is greater than the even (ridge) filter response over a number of scales. This has been applied successfully to feature detection in echocardiography and automated segmentation in fetal images [117, 155]. The ability to take into account structures at different scales of LP can help to differentiate between image features caused by image artefacts and those of clinical significance [117]. FA should therefore be useful as a cost term in an US application of Live Wire. The FA of an image is defined by;

$$FA = \frac{1}{N} \sum_{s} \frac{\lfloor |odd_s| - |even_s| - T_s \rfloor}{\sqrt{even_s^2 + odd_s^2 + \varepsilon}}, \tag{4.9}$$

where;

- s is the scale,
- N is the number of scales used to form the FA image,
- ε is a small constant (typically around 0.01) to prevent division by zero,
- $even = I_b$ (from Equation 4.5),
- $odd = (o_1 \otimes I_b, o_2 \otimes I_b)$ (using the spatial filters from Equation 4.6),
- $\lfloor \cdot \rfloor$ represents an operation which sets negative values to zero
- T_s is a noise-controlling threshold value [91, 155]. In this implementation, $T_s = 0.1$.

Figure 4.4.b shows an example of FA being applied to a clinical image, with $N = 3$ and $s = [19, 20, 21]$ using a Gaussian Derivative filter. As can be seen, $FA \approx 1$ for structurally significant boundaries and $FA = 0$ for most other regions of the image. While boundary localisation is good for regions of clear structural importance, there is little to guide the Live Wire algorithm in regions with less clear structural significance. As a result, an additional cost term is required to allow for delineation of regions where $FA = 0$.



(a) **(b)**

Figure 4.4: *(a) A clinical US image of a fetal arm at 27 weeks and 6 days GA. (b) An example of FA image estimated on the same image with* $N = 3$.

4.3.3 Bandpass Filter

When calculating the LP of an image, the selection of both the bandpass filter b and the scale of that filter can have a significant impact on the appearance of the resulting LP and hence FA image. Different sizes of features may appear more prominent depending on the scale, and the response of individual kinds of edges may vary depending on the type of filter. Boukerroui *et al.* summarise the functionality of commonly used filters applied in LP calculation, including Gabor, log-Gabor, Gaussian derivative, Difference of Gaussian (DoG) and Cauchy filters, making an effort to objectively quantify the relative merits of each [19]. Meanwhile, Mellor and Brady developed a filter specifically for LP calculations, aimed at reducing the effect of scale dependence [112].

With the vast number of filter and scale combinations available, each resulting in subtle differences to boundary localisation, methods are often selected empirically. Hacihaliloglu *et al.* [69] and Cifor *et al.* [38, 39], for example, give no explicit justification for their use of the log-Gabor filter in 3D LP calculations. Table 4.1 summarises details of commonly used filters and their use in the field of analysis of medical US images. Figure 4.5 shows the LP of an US image of a fetal arm cross section using the Gaussian Derivative, Cauchy, log-Gabor and Gabor filters at several different scales. Throughout this thesis, the filter scale is referred to as s. This is analogous to ω_o (the centre/peak tuning frequency) for the Gabor and Log-Gabor filters, and σ (standard deviation) for the Cauchy, Gaussian Derivative, DoG and Mellor-Brady filters.

Table 4.1: *Bandpass filters commonly used when estimating the LP of an image, each with example applications within the field of medical US imaging. Equations are as shown by Boukerroui* et al. *in their review of filter choice for the application of LP [19], with the exception of the Mellor-Brady filter [112]. The notation of these equations is shown in Table 4.2.*

Filter	Equation	Applications in US	
Gabor	$B_g(\omega) = n_c e(-\frac{\sigma^2}{2}(\omega - \omega_o))$	*None found at time of writing*	-
Log-Gabor	$B_{lg}(\omega) = n_c e\left(-\frac{ln^2(\omega/\omega_o)}{ln^2(\kappa_\beta)}\right)$	Boundary detection in echocardiography using FA	[117]
		Bone segmentation from 3D US	[69]
		Object detection in fetal US	[147]
		Tumor tracking in 2D US (Diffeomorphic registration)	[38]
		Tumor tracking in 2D US (Log-Demons registration)	[39]
Cauchy	$B_c(\omega) = n_c e(-\sigma\omega)$	Level set segmentation in US images	[9]
Gaussian Derrivative	$B_{gd}(\omega) = n_c e(-\sigma^2\omega^2)$	Boundary detection in 3D echocardiography using FA	[148]
		Automatic segmentation of soft tissue in fetal US images	[156]
Difference of Gaussians	$B_{DoG}(\omega) = n_c(e^{-\frac{\sigma'^2\omega^2}{2}} - e^{-\frac{\sigma^2\omega^2}{2}})$	Motion estimation in elastography from US images	[107]
Mellor-Brady	$b_{mb}(r) = \frac{A}{r^{\alpha+\beta}} - \frac{B}{r^{\alpha-\beta}}$	Multimodal image registration	[112]
		Registration of 3D US to 2D MR	[205]

Table 4.2: *Definition of the notation used in the equations for Table 4.1*

Constant	Definition
ω	Frequency
ω_o	Centre/Peak tuning Frequency
σ	Standard Deviation
n_c	Normalisation constant ($n_c = 1$ in this applications)
κ_β	$\kappa_\beta = \sigma\omega_o$

Constant	Definition
r	$r = \sqrt{x^2 + y^2}$
α and β	Constants selected such that $\beta << \alpha$
A and B	Constants selected such that the function integrates to one, giving zero DC

At the time of writing, no studies concerning US image processing were found to use the Gabor filter for the calculation of LP. This lack of use may be explained by the comments made by Boukerroui *et al.*, finding that the Gabor filter is the least suitable choice of bandpass filter for feature detection, showing poor performance in localising features in comparison to competing filters [19]. Mulet-Parada and Noble favour the log-Gabor filter over the Gabor filter to avoid the latter's non-zero DC component and response at negative frequencies [117].

The log-Gabor filter appears to be the most commonly used filter in the estimation of LP; Rahmatullah *et al.* note its commonality in their selection of this filter [147]. Cifor *et al.*, while using the log-Gabor filter for their application of LP, note that other filters, specifically the Cauchy and DoG filters, "could equally be used," [38]. However, the results of the investigation Boukerroui *et al.* conducted contradict this, with the DoG and Cauchy filters performing better in their localisation for edges, but the log-Gabor filter outperforming the others for detecting lines [19]. Indeed, Belaid *et al.* use Cauchy filters in their phased based level set segmentations based on this analysis [9].

Rajpoot *et al.* compared the log-Gabor and Gaussian Derivative filters empirically, determining that the latter was more appropriate for their application [148]. Rueda *et al.* also uses the Gaussian Derivative filter but without justification [156]. Boukerroui *et al.* note that the Gaussian Derivative filter has invariance to smooth gradients, which may be useful to US applications [19]. Maltaverne *et al.* use a DoG filter, tuned to the scale of the US speckle, but do not offer any reason for this choice of filter in particular [107].

The 'Mellor-Brady' filter, which has the form shown in Table 4.1, has been proposed as an alternative to more generic bandpass filters, and has been designed specifically for the use in LP estimation [112, 113]. This filter was not investigated in the comparative study by Boukerroui *et al.*, having been published a year earlier than the paper detailing the filter [19, 112]. The properties of this filter include zero DC and approximate scale invariance when the value of α is far greater than that of β (true scale invariance when $\beta = 0$, but this reduces the entire function to zero). The filter was deemed more appropriate for

LP analysis over previously used filters as it reduces the problem of interference between close features via its property of near-scale-invariance. Scale-invariance will mean that the response from regions where features are close together will theoretically be equivalent to far features in terms of LP and interference [112]. Zhang *et al.* use this filter due to the scale invariance it provides [205]. However, it should be noted that their application was in image registration, where maintaining the quality of all features regardless of scale is important. In contrast, this scale invariance is not necessarily an advantage in US segmentation, as smaller, compact features may be due to local variation from speckle rather than indicating boundaries of structural significance. In this case, interference between speckle may be advantageous.

Filter Choice

Figure 4.5 shows the scope of the challenge presented in effectively selecting an appropriate filter, in this case showing the impact of different filter and scale combinations for LP calculations. Each filter presents subtle differences in response to image boundaries, and in particular the way nearby boundaries interact with each other.

Even once a particular filter and scale is decided, FA can be calculated using an average response over a number of different scales. This can allow for both large-scale and small-scale filters to be used in tandem. Figure 4.6 shows the further challenge of selecting an appropriate set of scales for the application of FA, in this case using the Guassian Derivative filter. When a larger spread of scales are used, as in Figure 4.6b, smaller image details are preserved alongside stronger object boundaries. This is particularly useful in fetal limb segmentation for cases where the limb is in contact with surrounding tissue, and a strong boundary may not exist. Conversely, using a tighter set of scales as in Figure 4.6c, only boundaries of a similar detail are shown. This is useful in the application of Live Wire, as it will limit the number of potential paths, thereby facilitating the segmentation process. A balance needs to be made between prioritising the visibility of the boundaries on the object of interest while retaining image detail.

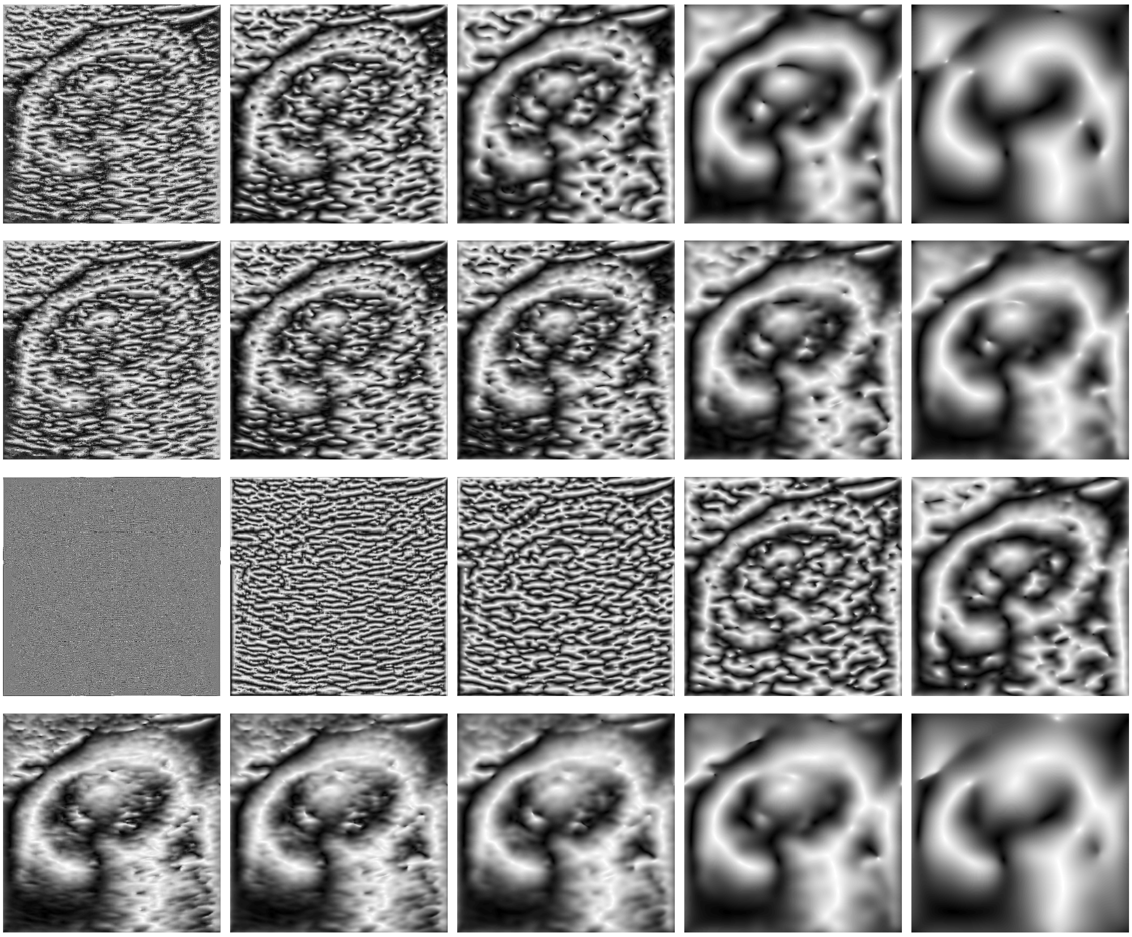
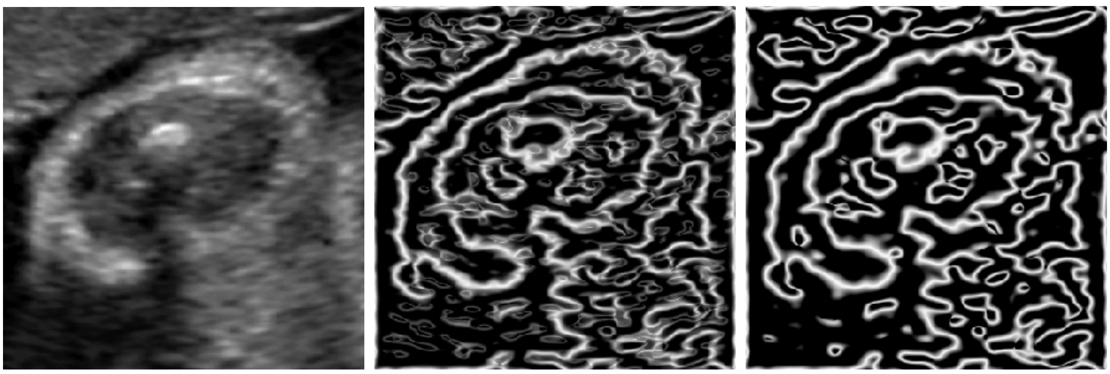


Figure 4.5: *Rows show the LP images of a fetal arm cross section in US using (from top to bottom) the Gaussian Derivative, Cauchy, log-Gabor and Gabor filters respectively. Columns show (from left to right) scales of s = 1, 10, 20, 50 and 100.*



(a) **(b)** **(c)**

Figure 4.6: ***(a)*** *US image of a cross section of a fetal arm and the corresponding FA using a Gaussian Derivative filter and scales of* ***(b)*** $s = [10, 20, 30]$ ***(c)*** $s = [19, 20, 21]$

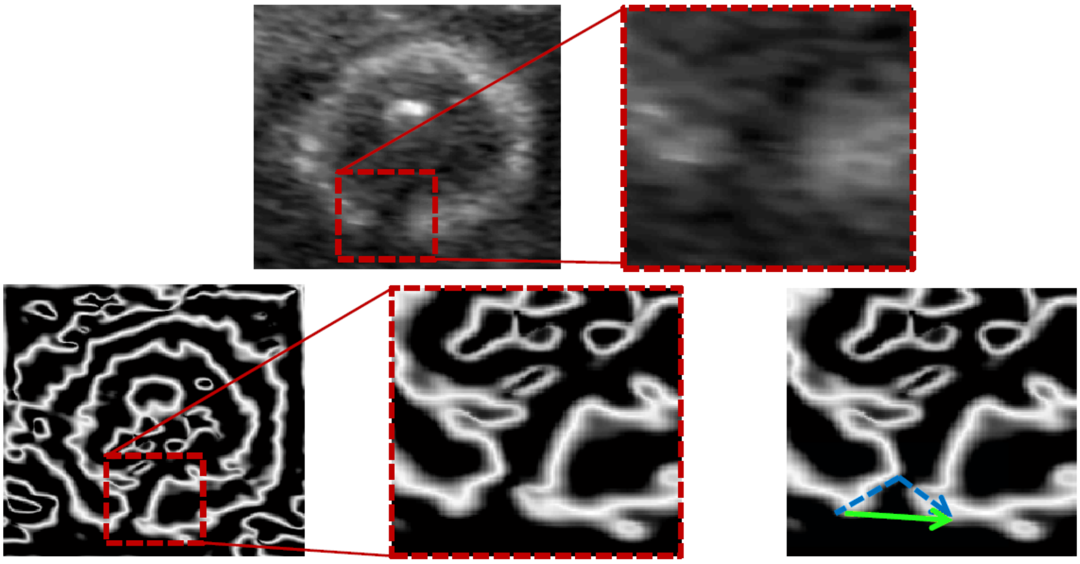


Figure 4.7: *This figure shows an example of a fetal arm cross section in US, and it's appearance using FA. The red dashed boxes highlight the same region of shadowing in both images. Consider the shortest path the Live Wire algorithm using FA alone would take (blue, dashed path), compared to the desired path to delineate the full contour (green, solid path). The shape constraint aims to combat this discrepancy.*

In the specific application of fetal limb soft tissue quantification, it is important to choose a filter that will detect boundaries relevant to soft tissue segmentation, and continue to perform across a range of GAs. Empirically, the Gaussian Derivative filter with scales $s = [25, 30, 35]$ was found to give better FA maps in terms of preserving the boundary of the object of interest while rejecting smaller features caused by speckle and forming a smooth contour, as would be drawn by a clinician.

4.4 Shape Constraint

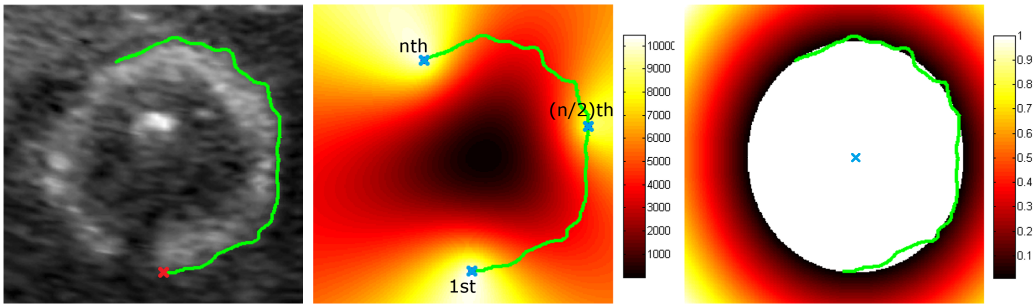
Certain aspects characteristic of US images can cause a semi-automated method such as Live Wire to deviate in undesirable ways, even after FA has removed some of the problems presented by signal attenuation and poor contrast. For example, small structures such as speckle can influence the delineated contour adversely and, in this specific application, shadowing is present in most images, resulting in an incomplete boundary. Figure 4.7 shows an example of how Live Wire would favour a path across a shadowed region, compared to the desired path to accurately estimate the adipose tissue in this section of fetal limb.

In order to make this method more robust, the addition of an additional cost term to the Live Wire method in order to locally guide the shape of the paths calculated at each iteration is needed. This allows Live Wire to estimate the shape of missing pieces of contour, and has the side effect of tending to make other sections of the contour follow a smoother path, penalising sharp deviations in direction.

By using a weak shape constraint (in this case circular), locally smooth contours can be ensured, without the requirement of prior information about the object's overall shape. An advantage of a weak constraint such as this is that the method does not require training sets. This would be particularly difficult to obtain in an application such as fetal soft tissue, where GA and fetal position may alter the shape of the desired object to be segmented.

This weak shape constraint is constructed by fitting a circle to the arc formed by the previous two anchor points;

- When the user places an anchor point on the image to create a new segment of the Live Wire path, the method retrieves the previous n elements of the arc between the previous two anchor points, as shown in Figure 4.8.a.
- Three distance maps are created, considering the first, $(n/2)^{th}$ and the n^{th} pixels forming this segment.
- A variance map is created, with a size equal to the original image, where each pixel displays the variance of distances from each of the three contour points to that pixel. This is shown in Figure 4.8.b.
- The pixel at which the value of this variance map is lowest is taken as the centre of a circle, with the radius of this circle being the mean distance between the three contour points at this pixel.



(a) **(b)** **(c)**

Figure 4.8: ***(a)*** *An US image of a fetal arm with a single Live Wire segment, constructed using a FA cost map.* ***(b)*** *Contour segment over a distance variance map; the first, the $(n/2)^{th}$ and the n^{th} points along the contour are selected. The value of each pixel in the distance variance map corresponds to the variance of distances from each of these three points.* ***(c)*** *The shape-based cost map derived from the location of the minimum distance variance from the three points and the mean distance at that location.*

- The region enclosed by the circle is given a cost of $C_{shape} = 1$. The cost outside this region begins at $C_{shape} = 0$ and increases uniformly with distance from the boundary, set such that the cost outside the circle never exceeds $C_{shape} = 1$. A visualisation of the resulting cost map can be seen in Figure 4.8.c.

- The shape cost is recalculated for each segment that is drawn.

Each time the user places a new anchor point, the size and position of the shape constraint is re-defined, considering only pixels that form the segment between the new anchor point and the previous one, rather than the entire segmented contour to that point. This allows the shape constraint to locally adapt to different sections of the contour. This prevents over-constraining the contour with additional anchor points and hence retains the flexibility and the interactive component of the Live Wire approach.

4.4.1 FA-S Live Wire

The complete FA-S Live Wire method uses FA and a piecewise local shape prior cost map as the costs of the Live Wire algorithm, to form the cost map in the form of the one shown in Figure 4.9. The cost of the wire connecting pixel **A** and pixel **B**, for example,

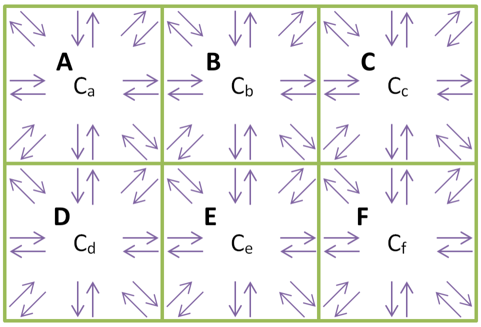


Figure 4.9: *Possible Live Wire paths using pixels rather than vertices as nodes.*

is calculated as $C_{path} = C_a + C_b$. For each pixel centre, the cost is defined as;

$$C = w_{FA}C_{FA} + w_{shape}C_{shape} \tag{4.10}$$

where $C_{FA} = FA$ from Equation 4.9, C_{shape} is the shape constraint cost as described in Section 4.4, and w_{FA} and w_{shape} are their corresponding weights [37]. In this application the costs have been set to $w_{FA} = w_{shape} = 50$ after empirical testing of performance of FA-S Live Wire on example images.

While components of the analytic signal contain directional information, FA does not contain directional information in the same way as a gradient value would, and so there is little value in including a full pixel neighbourhood for FA values. Indeed, the only pixel with relevant information to determine whether a point in an image is on an edge or not is the FA value at that pixel. This means that this application of the Live Wire algorithm does not concern edges and vertices of pixels, but instead is calculated in terms of a path running between pixel centres, as shown in Figure 4.9. As a result, an application such as this in which FA is utilised should follow the Chorodowski *et al.* method of cost-per-pixel [37], rather than the Falcão *et al.* method of cost-per-vertex [52].

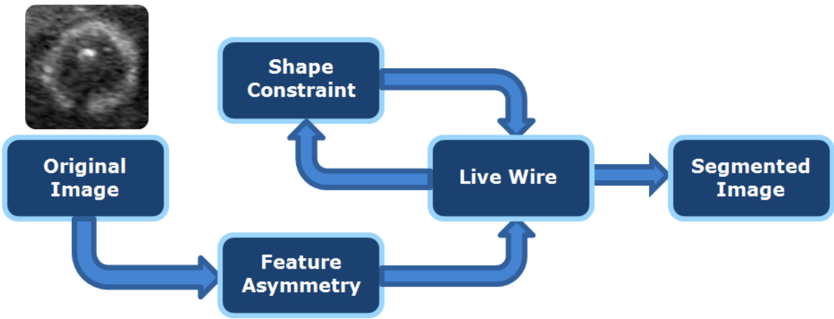


Figure 4.10: *Flow diagram of the FA-S Live Wire method.*

Methodology Overview: FA-S Live Wire

- A cost map based on the FA of the image is calculated.
- A weak shape constraint is applied by approximating a circular cost region from the most recently drawn segment.
- The FA and shape constraint costs are used in order to guide an adaptation of the Live Wire algorithm. The user sets anchor points on the image, and the lowest cost path is shown between the last anchor point and the cursor in real time. These paths are set in place by setting a new anchor point.
- The shape prior helps overcome small missing pieces of boundary in the image and guarantees the local smoothness of contours without making an assumption about the object's overall shape, keeping the application of FA-S Live Wire general.

Figure 4.10 summarises the overall FA-S Live Wire method.

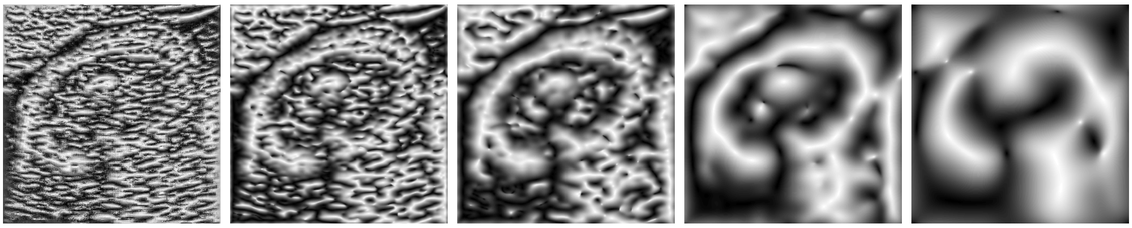


Figure 4.11: *The LP response of an US image using a Gaussian Derivative and (from left to right) scales of* $s = 1, 10, 20, 50$ *and* 100.

4.5 Validation of Method

4.5.1 Experiment Outline

In order to test the ability of FA-S Live Wire to segment the adipose tissue layer in US images of the fetal arm, it is necessary to compare the accuracy and efficacy of this method against existing methods. The proposed method is therefore compared to the intensity-based Live Wire method developed by Chodorowski *et al.* [37] as an extension of the original method by Falcão *et al.* [52], and to a fully automatic segmentation method for the fetal US images, which is also graph-based and incorporates FA and LP and is specifically developed for US image segmentation, which will be referred to as 'FC-Auto' [156]. As an additional experiment, FA-S Live Wire will also be tested with the shape constraint removed, to show the purpose of this extra component. This will be referred to as 'FA Live Wire.'

Phasewire has not been compared, as it was found to be inappropriate for this particular application. The original application of Phasewire was in MRI segmentation [130], where boundaries are strong and phenomenon such as speckle and poor image contrast do not need to be contended with. Applying LP to US images of fetal soft tissue results in poor delineation between adjacent objects and a poorly defined object boundary as shown in Figure 4.11.

The data set consisted of 48 cross-sectional US images of the fetal arm, from a selection of healthy singleton pregnancies studied in the INTERGROWTH-21st project [11]. This

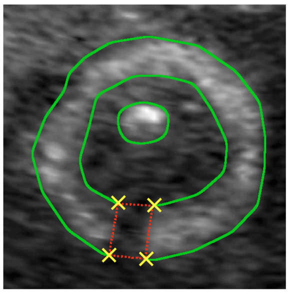


Figure 4.12: *A fetal arm cross-section in US. The red dotted line shows an area of shadowing, while the yellow crosses indicate segmentation start points.*

Table 4.3: *Colour legend*

Method	Colour
Manual	Orange
Live Wire	Blue
FC-Auto	Red
FA Live Wire	Purple
FA-S Live Wire	Green

selection features images from across gestation, ranging between 21 and 38 weeks GA. This presents a significant variability in size, shape and appearance.

In order to assess the accuracy of each method's output, a comparison must be made to a reference segmentation, which can be regarded as an approximation of the 'true' region of adipose tissue. In the assessment outlined here, this reference segmentation was carried out manually by Caroline Knight, a clinician associated with this project. Segmentations were conducted using a bespoke manual tracing tool built in MATLAB. In order to take into account intraobserver variability, this manual segmentation was performed twice for each US image. These were both used as reference segmentations in the evaluation stages, and the results averaged.

The Live Wire, FA Live Wire and FA-S Live Wire method were each applied to all 48 images three times, and performed by me. The entire image set was segmented once before the second segmentation was made, and so on, so that there was a large gap between successive segmentations of the same image. This meant that it was not possible for the segmentation to be learnt by the user over successive attempts, which would improve segmentation speed and accuracy and undermine the validity of the results.

The process has been standardised by starting the segmentation at one side of the shadow cast by the bone, around the region to be segmented and back to the initial point. This ensures that the shape constraint influences the segmentation over the shadowed region. Potential start points for this segmentation are shown in Figure 4.12. FC-Auto was applied to a subset of 20 images; in comparisons involving FC-Auto results, the values from the other methods are taken from the same subset of images. For consistency, all graphs and charts in this section use the same colour legend to display the performance of different methods of segmentation, as displayed in Table 4.3.

Table 4.4: *Evaluation of precision, recall and Dice similarity in terms of area for each of the segmentation methods considered. The highest performing results are in bold.*

Table 1	Method	Precision (%)	Recall (%)	Dice Similarity (%)
20 image set	Live Wire	89.33 ± 3.71	85.08 ± 4.89	86.93 ± 2.78
	FC-Auto	87.64 ± 3.64	80.31 ± 5.72	83.57 ± 2.71
	FA Live Wire	**90.20 ± 4.71**	84.42 ± 5.22	87.02 ± 2.81
	FA-S Live Wire	87.21 ± 6.01	**88.08 ± 3.98**	**87.36 ± 2.89**
48 image set	Live Wire	87.65 ± 6.12	85.82 ± 6.62	86.55 ± 5.62
	FA Live Wire	**89.17 ± 8.57**	84.02 ± 8.16	86.31 ± 7.24
	FA-S Live Wire	86.52 ± 5.24	**88.37 ± 4.07**	**87.20 ± 2.55**

Table 4.5: *Contour agreement between each segmentation method and the reference delineation. The highest performing results are in bold.*

Table 2	Method	≤ 1 pixel (%)	≤ 5 pixels (%)	≤ 10 pixels (%)	Average deviation (pixels)
20 image set	Live Wire	**30.33 ± 7.40**	80.04 ± 8.51	95.50 ± 3.94	**3.41 ± 0.85**
	FC-Auto	25.13 ± 7.93	73.69 ± 15.64	91.60 ± 11.22	4.82 ± 3.88
	FA Live Wire	29.74 ± 7.11	79.94 ± 10.65	95.05 ± 5.35	3.47 ± 1.15
	FA-S Live Wire	27.87 ± 5.71	**80.29 ± 7.54**	**96.02 ± 3.41**	3.43 ± 0.72
48 image set	Live Wire	**29.51 ± 6.86**	**76.44 ± 10.28**	92.79 ± 7.73	4.07 ± 2.57
	FA Live Wire	28.12 ± 7.00	76.18 ± 12.00	92.12 ± 8.75	4.12 ± 2.87
	FA-S Live Wire	26.05 ± 6.13	75.80 ± 9.75	**93.40 ± 5.35**	**3.94 ± 1.04**

Table 4.6: *Time required and number of anchor points needed to complete the segmentation, averaged across all images, as well as the repeatability of the semi-automated and manual segmentations. The highest performing results are in bold.*

Table 3	Time (s)	Anchor pts.	Precision (%)	Recall (%)	Dice (%)
Manual	28.0 ± 5.1	-	87.55 ± 5.21	94.28 ± 4.28	90.61 ± 2.06
Live Wire	24.9 ± 4.8	15.7 ± 2.4	93.51 ± 3.08	94.49 ± 2.19	93.94 ± 1.97
FA Live Wire	18.5 ± 3.9	8.5 ± 1.6	96.17 ± 6.26	97.28 ± 6.00	96.67 ± 5.81
FA-S Live Wire	**13.9 ± 2.6**	**7.8 ± 1.5**	**96.62 ± 2.82**	**97.43 ± 2.19**	**96.98 ± 1.89**

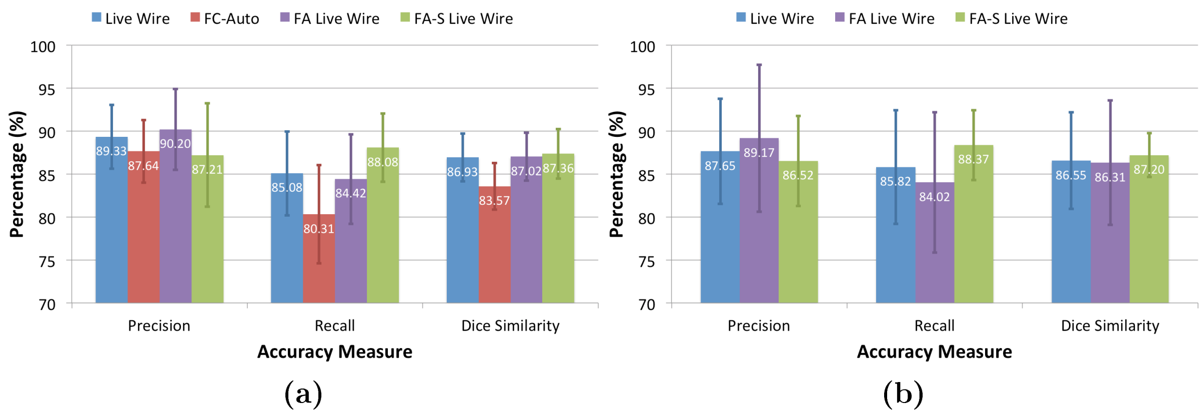


Figure 4.13: *The precision, recall and Dice similarity for;* (a) *The 20 image set for all segmentation methods,* (b) *The 48 image set for Live Wire and FA-S Live Wire*

4.5.2 Results: Accuracy

To assess the agreement between the areas segmented by each method with the reference segmentation of adipose tissue, a comparison was made by calculating the precision, recall and Dice similarity for each segmentation method, defined as;

$$\text{Precision:} \quad P = \frac{|M \cap S|}{|S|} \times 100 \tag{4.11}$$

$$\text{Recall:} \quad R = \frac{|M \cap S|}{|M|} \times 100 \tag{4.12}$$

$$\text{Dice similarity:} \quad D = \frac{2 \times |M \cap S|}{|S| + |M|} \times 100 \tag{4.13}$$

where M and S refer to the manual segmentation (which is used as a reference) and the semi-automatic segmentation method, respectively [156]. These were calculated for each segmentation method against the two corresponding manual segmentations, with an average value being recorded for each image. The mean and standard deviation of these values across all images can be found in Table 4.4, and a graphical representation of these results can be seen in Figure 4.13.

FA-S Live Wire showed a significant improvement in accuracy over the automated method in terms of recall and Dice similarity ($p < 0.01$ in both cases), while remaining comparable in terms of precision ($p = 0.39$). Live Wire, FA Live Wire and FA-S Live

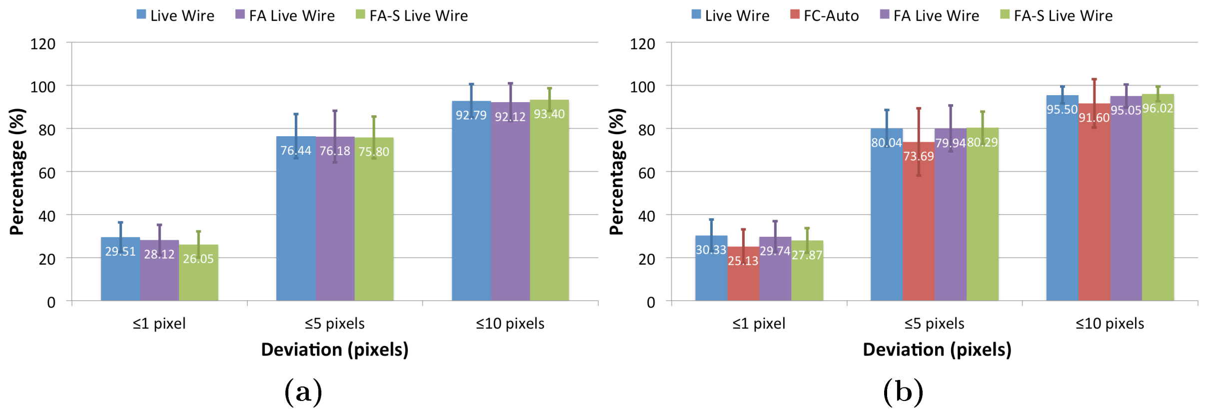


Figure 4.14: *Average deviations of segmented contours from the manually segmented contours*

Wire have shown near-identical results in terms of accuracy across the board, with FA-S Live Wire showing slightly better performance in terms of recall and Dice similarity ($p = 0.02$ for recall in the 20-image set, but $p > 0.05$ for all other cases), suggesting that FA-S contours are generally over-estimating the segmented area. This could be due to the shape constraint causing the curve as chosen by the user to 'inflate' slightly, or overshoot sharp corners to fit a more circular path. The lower values for recall and higher values for precision in FA Live Wire compared to FA-S Live Wire in particular could indicate this. However, it is important to note that a similar level of accuracy would be expected between multiple interactive segmentation methods such as Live Wire and its derivatives, as the output is influenced heavily by the user's judgement. For these experiments, the manual segmentation was conducted by a clinician while I performed all of the semi-automated segmentations. Therefore, it is expected that an investigation involving the same operator for each method, but a different operator than the source of the manual segmentations, would have a similar level of error.

Another evaluation metric consists of the relative position of the semi-automated contour against the clinical standard. A suitable metric to quantify the similarity in contours is to calculate the distance of each pixel on the contours segmented by each method from the manually segmented contours. This is calculated as the percentage of pixels that fall within certain distances of the pixels on the reference contour. In this case, shown in Table 4.5, the percentage of pixels which fall within 1, 5 and 10 pixels of

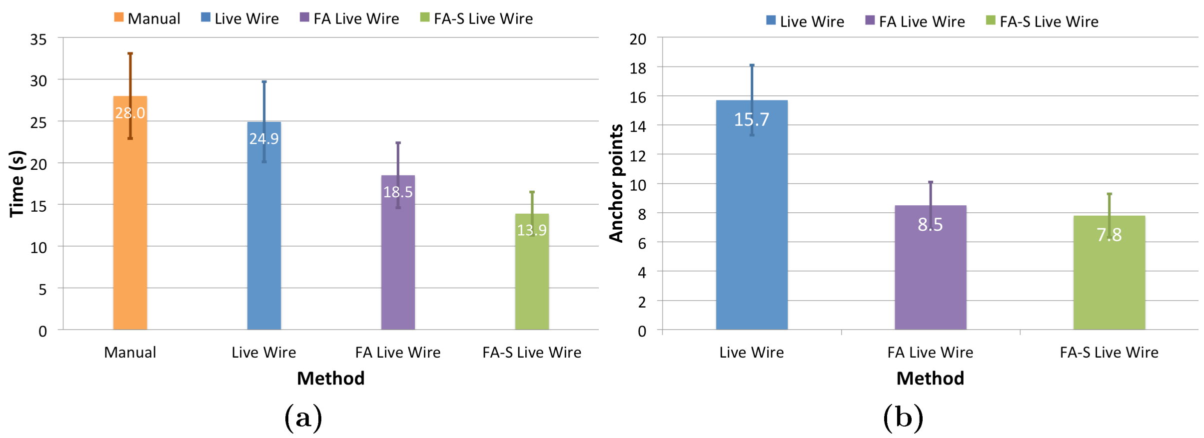


Figure 4.15: (a) *A comparison of the time taken to perform a segmentation manually, and using the Live Wire based methods* (b) *A comparison of the number of anchor points required to complete a segmentation using Live Wire based methods*

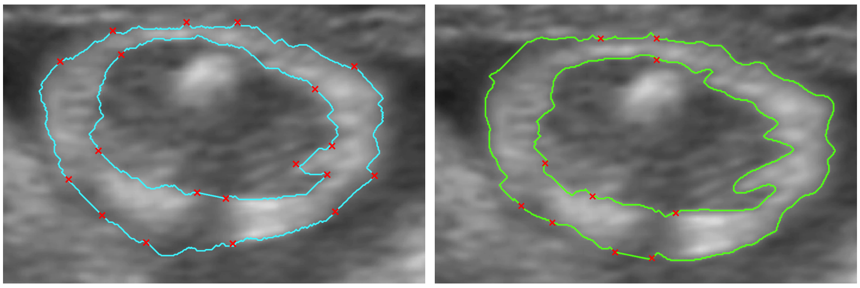
the manually delineated contour are determined. Figure 4.14 illustrates these results.

As with the accuracy in terms of area overlap, contour deviation showed comparable results to traditional Live Wire and more significant improvement over FC-Auto. It is important to note that these results also show that FA-S Live Wire generally has a lower standard deviation compared to the other methods. The lower standard deviation exhibited by the FA-S Live Wire results may suggest more consistent segmentations.

4.5.3 Results: Efficiency

Investigations into the accuracy of FA-S Live Wire compared to alternative methods have shown that this method demonstrates similar, though in some cases marginally better, performance. This is not, however, sufficient to justify the use of FA-S Live Wire over competing methods. As semi-automated methods are intended to facilitate segmentations, allowing for a higher throughput of data in as quick and as easy a manner as possible, it is important to assess each method on these terms. Consequently, this section will detail the number of anchor points required and the time taken to complete the segmentations carried out in Section 4.5.2.

Table 4.6 displays the average time taken and the number of anchor points required for the manual segmentation, the intensity-based Live Wire and the proposed technique. Figure 4.15 presents a graphical representation of these results. The figure for the manual



(a) **(b)**

Figure 4.16: *Delineations and anchor points needed for segmenting an US image of the fetal arm at 27 weeks GA using* ***(a)*** *Live Wire (19 anchor points) and* ***(b)*** *FA-S Live Wire (10 anchor points). The anchor points are displayed as red crosses.*

segmentation average time for segmentation is taken from a subset of 40 segmentations. As can be expected, the time taken and anchor points required for FA-S and FA Live Wire are much lower than Live Wire, since FA provides a very strong path around the object's boundary. FA-S Live Wire's ability to refine the number of paths available to the user appears to further reduce the time taken and anchor points required.

Figure 4.16 visualises the difference in anchor points required to complete the same segmentation using Live Wire and FA-S Live Wire. Note that the FA-S Live Wire required around half the time and half the number of anchor points than Live Wire to segment the adipose tissue layer on the same US image, while Live Wire did not present a significant reduction in time compared to manual segmentation.

Also in Table 4.6, the precision, recall and Dice similarity were calculated comparing subsequent segmentations using the same technique and on the same image, in order to provide an indication of the repeatability of each method. This has been achieved by using equations 4.11, 4.12, and 4.13 but replacing M with the initial segmentation (S_1 or S_2), and S with repeated segmentations (S_2 or S_3), such that all segmentations are compared against each other. In the case of the manual segmentations, which were conducted twice, only S_1 and S_2 could be used. Figure 4.17 gives a visualisation of these results.

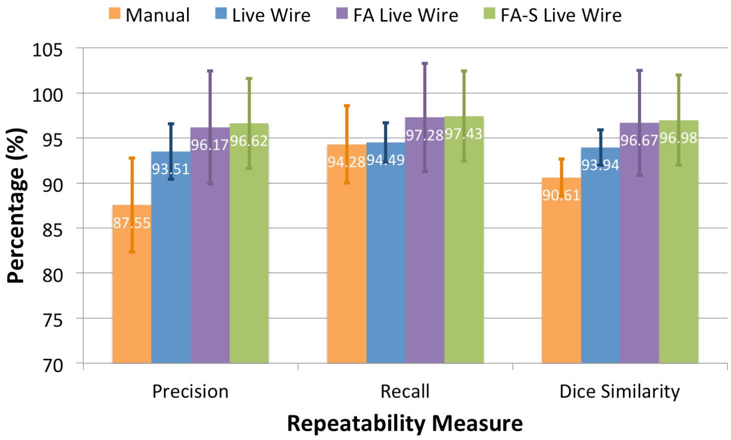


Figure 4.17: *The repeatability of the manual, Live Wire based methods*

4.5.4 Results: Significance testing

In order to assess the relevance of the results shown in this chapter, it is important to determine the mathematical significance of the improvements described. This is to prove that the positive results obtained in the assessment of these tools are more likely to have occurred due to inherently better properties of the FA-S Live Wire segmentation method, rather than due to random chance; falling within the expected deviation of the values obtained. In order to achieve this, the p-value for each result must be calculated. For two groups of mean μ_1 and μ_2, variance ω_1 and ω_2, and number of samples N_1 and N_2, the unequal-variance Students' t-test can be calculated using;

$$t' = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\omega_1^2}{N_1} + \frac{\omega_2^2}{N_2}}} \tag{4.14}$$

From this the associated p-value can be found. Since segmentations on the same sets of images are being compared, $N_1 = N_2 = N$, except in the case of the time measure for manual segmentations, where 40 of the 48 images had corresponding time data.

Table 4.7 outlines the significance values of the results for precision, recall and Dice similarity, as well as for contour agreement. Since the FA-S Live Wire, FA Live Wire and standard Live Wire results were conducted on both the full and partial image set, significance for both of these sets have been considered. Table 4.8 shows the same for the

Table 4.7: *Significance values of the results for precision, recall, Dice similarity and contour agreement. The highest performing results are in bold.*

Measure	FA-S Live Wire	Live Wire	Samples	t-value	p-value
Precision (%)	87.21 ± 6.02	**89.33 ± 3.71**	20	-1.347	*0.09752*
Recall (%)	**88.08 ± 3.98**	85.08 ± 4.89	20	2.128	**0.02299**
Dice Similarity (%)	**87.36 ± 2.89**	86.93 ± 2.80	20	0.478	0.31895
Precision (%)	86.52 ± 5.24	**87.65 ± 6.12**	48	-0.972	0.16803
Recall (%)	**88.37 ± 4.07**	85.82 ± 6.62	48	2.273	**0.01376**
Dice Similarity (%)	**87.20 ± 2.55**	86.55 ± 5.62	48	0.730	0.23456
≤1 Pixel (%)	27.87 ± 5.71	**30.33 ± 7.40**	20	-1.177	0.12650
≤5 Pixels (%)	**80.29 ± 7.54**	80.04 ± 8.51	20	0.098	0.46132
≤10 Pixels (%)	**96.02 ± 3.41**	95.50 ± 3.94	20	0.446	0.33009
Average dev. (pixels)	3.43 ± 0.72	**3.41 ± 0.85**	20	0.080	0.46840
≤1 Pixel (%)	26.05 ± 6.13	**29.51±6.86**	48	-2.606	**0.00609**
≤5 Pixels (%)	75.08 ± 9.75	**76.44 ± 10.28**	48	-0.665	0.25461
≤10 Pixels (%)	**93.40 ± 5.35**	92.79 ± 7.73	48	0.450	0.32753
Average dev. (pixels)	**3.94 ± 1.04**	4.07 ± 2.57	48	-0.325	0.37335
Measure	FA-S Live Wire	FA Live Wire	Samples	t-value	p-value
Precision (%)	87.21 ± 6.01	**90.20 ± 4.71**	20	-1.751	*0.09523*
Recall (%)	**88.08 ± 3.98**	84.42 ± 5.22	20	2.494	**0.02153**
Dice Similarity (%)	**87.36 ± 2.89**	87.02 ± 2.81	20	0.377	0.70999
Precision (%)	86.52 ± 5.24	**89.17 ± 8.57**	48	-1.828	*0.07381*
Recall (%)	**88.37 ± 4.07**	84.02 ± 8.16	48	3.305	**0.00180**
Dice Similarity (%)	**87.20 ± 2.55**	86.31 ± 7.24	48	0.803	0.42576
≤1 Pixel (%)	27.87 ± 5.71	**29.74 ± 7.11**	20	-0.917	0.18501
≤5 Pixels (%)	**80.29 ± 7.54**	79.94 ± 10.65	20	0.120	0.45286
≤10 Pixels (%)	**96.02 ± 3.41**	95.05 ± 5.35	20	0.684	0.25099
Average dev. (pixels)	**3.43 ± 0.72**	3.47 ± 0.72	20	-0.176	0.43116
≤1 Pixel (%)	26.05 ± 6.13	**28.12 ± 7.00**	48	-1.541	*0.06490*
≤5 Pixels (%)	75.08 ± 9.75	**76.18 ± 12.00**	48	-0.493	0.31217
≤10 Pixels (%)	**93.40 ± 5.35**	92.12 ± 8.75	48	0.865	0.19576
Average dev. (pixels)	**3.94 ± 1.04**	4.12 ± 2.87	48	-0.409	0.34235
Measure	FA-S Live Wire	FC-Auto	Samples	t-value	p-value
Precision (%)	87.21 ± 6.01	**87.64 ± 3.64**	20	-0.274	0.39357
Recall (%)	**88.08 ± 3.98**	80.31 ± 5.72	20	4.987	**0.00004**
Dice similarity (%)	**87.36 ± 2.89**	83.57 ± 2.71	20	4.278	**0.00018**
≤1 Pixel (%)	**27.87 ± 5.71**	25.13 ± 7.93	20	1.254	0.11216
≤5 Pixels (%)	**80.29 ± 7.54**	73.69 ± 15.64	20	1.700	*0.05232*
≤10 Pixels (%)	**96.02 ± 3.41**	91.60 ± 11.22	20	1.686	*0.05370*
Average dev. (pixels)	**3.43 ± 0.72**	4.82 ± 3.88	20	-1.575	*0.06544*

Table 4.8: *Significance values for the time required and number of anchor points needed to complete the segmentation. The highest performing results are in bold.*

Measure	FA-S Live Wire	Manual	Samples	t-value	p-value
Time (s)	**13.9 ± 2.6**	28.0 ± 5.1	40	-15.853	**0.00000**
Measure	FA-S Live Wire	Live Wire	Samples	t-value	p-value
Time (s)	**13.9 ± 2.6**	24.9 ± 4.8	48	-13.961	**0.00000**
Anchor points	**7.8 ± 1.5**	15.7 ± 2.4	48	-19.339	**0.00000**
Measure	FA-S Live Wire	FA Live Wire	Samples	t-value	p-value
Time (s)	**13.9 ± 2.6**	18.5 ± 3.9	48	-6.799	**0.00000**
Anchor points	**7.8 ± 1.5**	8.5 ± 1.6	48	-2.211	**0.01591**

time required and the number of anchor points needed to complete the segmentation. In both cases, significance values are shown in bold for p-values of $p \leq 0.05$ (indicating a strong significance) and italics for p-values of $0.05 < p \leq 0.1$ (indicating a weak, but still mathematically relevant significance).

All instances of FA-S Live Wire performing better in terms of recall showed a strong significance ($p \leq 0.05$), while instances of other methods outperforming FA-S Live Wire in terms of precision showed weak significance at best ($0.05 < p < 0.08$). All results for the time taken and the number of anchor points, which were all favourable for FA-S Live Wire, showed very strong significance with $p \approx 0$. In contrast, none of the p-values related to the contour agreement were in the range of $p \leq 0.05$, which were among the weaker results for demonstrating the benefits of FA-S Live Wire. Most improvements over FC-auto, however, have been shown to be within $0.05 < p \leq 0.1$.

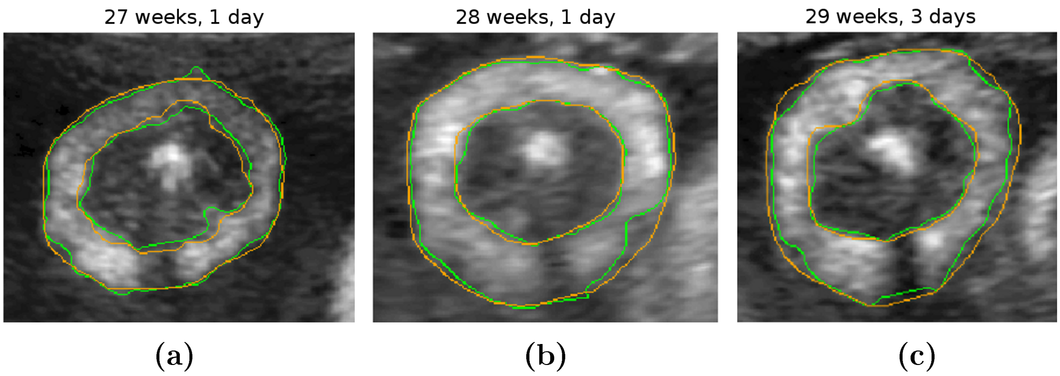


Figure 4.18: *Three examples of FA-S Live Wire on three fetal arm cross-sections for different GAs. The manual segmentation is shown in orange while the segmentation obtained by FA-S Live Wire is displayed in green.*

Performance Overview: FA-S Live Wire

- Overall improvement in area recall and Dice similarity over other methods, comparable precision and contour deviation to conventional Live Wire.
- Improved repeatability over manual segmentation and conventional Live Wire.
- Around twice as fast as manual segmentation.
- An average reduction of over 40% in the time and number of anchor points required to complete the contour compared to conventional Live Wire.
- The shape constraint improves the performance of the method compared to using FA alone

4.6 Discussion

4.6.1 Assessment of FA-S Live Wire

FA-S Live Wire is an adaptation of the Live Wire algorithm which is more suited to US images than standard intensity-based methods by incorporating intensity invariant structural information (FA) and a piecewise weak local shape prior. This method is particularly suited to use with US images, where attenuation can reduce the efficacy of intensity-based measures, speckle can influence the appearance of the object boundary and shadowing can remove sections of the contour entirely. While the bandpass filter of FA can reduce the influence of speckle and provide an intensity invariant indication of boundaries, the weak shape constraint can ensure that boundaries are locally smooth and fill in missing sections of contour based on the trajectory of the previously delineated segment of the object boundary. The method presented is general and not related to the shape of the object to be segmented, preserving the flexibility of the Live Wire method. Figure 4.18 shows examples of the FA-S segmentations against manual clinical segmentations.

When applied to examples of fetal adipose tissue segmentation, it had been demonstrated that the accuracy in terms of agreement with both the contour and area of the segmented area is comparable to and, in some cases, exceeds that of alternative methods. FA-S Live Wire gave a precision of 87%, recall of 88% and Dice similarity of 87% compared to manual segmentations; this discrepancy may be an issue for the clinical application of this method. Assuming the difference between AGA and SGA measures of fetal fat are similar to those of other biometrics, an area calculation error of 10-15% could easily result in a heavily SGA fetus being identified on the low end of AGA or vice versa.

However, this may have been influenced by interobserver variability, which can be supported by the fact that Live Wire and FA-S Live Wire results showed very similar levels of error. Meanwhile, repeatability measures suggest that successive measurements of the same fetus are more highly comparable with FA-S Live Wire than manual measures. It is possible that if all fat measures were taken using the same method, and a model for fetal growth were built using these results, accuracy of SGA diagnosis would actually be increased since a more consistent definition of the fat boundary is being used.

The main improvement on conventional Live Wire, against which accuracy is generally the same, is the great reduction in time and anchor points required to complete the segmentation. FA-S Live Wire has been shown to have an average reduction of around 40% in both the time and number of anchor points required compared to Live Wire, with a reduction in time exceeding 50% compared to manual methods. Additionally, the advantages of this method over alternatives include the fact that it allows the user to have a direct influence over the contour while greatly reducing the amount of time and number of anchor points required to complete the contour over other interactive methods. Additionally, the operation of this method allows for the segmentation of shapes without the use of specialist equipment, which is often required for manual segmentation. The weak shape constraint keeps the application general while allowing for the segmentation of objects prone to shadowing, a common feature of US images.

In the testing of FA-S Live Wire against other methods, assessment via ROC curve has not been attempted. This is because varying parameters such as the cost weightings for FA-S Live Wire and conventional Live Wire would make the segmentation more difficult, but not necessarily have a significant or meaningful impact on the accuracy of the resulting segmentation. As the user has direct control over the ultimate appearance of the segmentation, varying the parameters of Live Wire to become less indicative of the features of interest will only require the user to apply more anchor points, tending towards a manual segmentation as the ability of Live Wire to describe edges decreases.

The evaluation of FA-S Live Wire also showed improvement over FA Live Wire, which itself outperformed standard Live Wire. This demonstrates both the utility of using FA in US images over traditional gradient based methods and the benefit provided by the additional shape constraint. The reduction in the time taken and the anchor points used is likely tied to the fact that the shape constraint effectively limits the number of potential paths around the object to those which would be consistent with the current trajectory of the Live Wire path. This means that the segmentation is less likely to deviate to a non-boundary region.

The improved repeatability of results is particularly important when considering the potential application of such a segmentation method; to build up a library of measurements of fetal adipose tissue in healthy and malnourished pregnancies in order to establish a metric to distinguish between the two. As a potential extension, it could be possible to apply automatic scale selection used for bandpass filter b by taking into consideration the relative size of important structures and speckle in the US images.

4.6.2 Position in Clinical Workflow

One important consideration for the utility of this method is its placement in clinical workflow, particularly in an environment where a number of alternative segmentation methods exist, many of which may involve less intensive user interaction. Rueda *et al.*, for example, have developed segmentation methods for the specific task of segmenting fetal fat in a form that requires minimal user interaction, requiring only the positioning of a seed point within the region of interest to initialise segmentation [155, 156].

FA-S Live Wire provides the user with the opportunity to directly control the position of the contour, which is of particular use in challenging imaging conditions where automated methods may fail and are difficult to alter to find the desired boundary. This also makes the method useful for diagnostically critical applications when clinicians would be more comfortable interacting directly with their data, rather than passively acquiring results.

An early version of the work described in this chapter has been published in the proceedings of SPIE Medical Imaging 2013:

T.M. Rackham, S. Rueda, C.L. Knight, J.A. Noble, 'Ultrasound image segmentation using Feature Asymmetry and shape guided live wire,' Proceedings of SPIE Medical Imaging 2013: Image Processing, 86690P, Orlando, FL, USA, February 2013

The personal contribution in this chapter is a previously-unexplored combination of FA with Live Wire, as well as the creation of a shape constraint which works specifically with Live Wire to improve performance on the segmentation of generic closed-contour shapes.

Chapter 5

FA-S Live Wire for Liver Volume

In this second contributions' chapter the utility of Live Wire is explored in a 3D setting, both in MRI and US. Previous methods have involved using sparse segmentations along two orthogonal axis being used to guide segmentation along the third. This chapter will explore the limitations to and potential improvements on this method in the context of fetal liver segmentation in 3D imaging modalities.

5.1 Fetal Liver Volume Estimation Studies

Fetal hepatic volume is another potential biometric for assessing the growth of a fetus. A number of investigations have been made into the calculation of LVol and the prediction of growth trajectories, the results of which can be found in Table 5.1 [7, 16, 29, 31, 93, 96]. Figure 5.1 and Figure 5.2 show the corresponding graphical representations of the relationship between LVol in millilitres and GA in weeks, as determined by these studies.

These measurements have been obtained primarily via the manual segmentation of 2D areas through each volume. The deviation in trajectory for fetal liver growth between these studies are, in part, a result of attempting to form a relationship between GA and liver size through differing methods of approximation. Additionally, different modalities were used to capture the data.

As an example of the scope of the contrast, Baker *et al.* calculated LVol from MRI

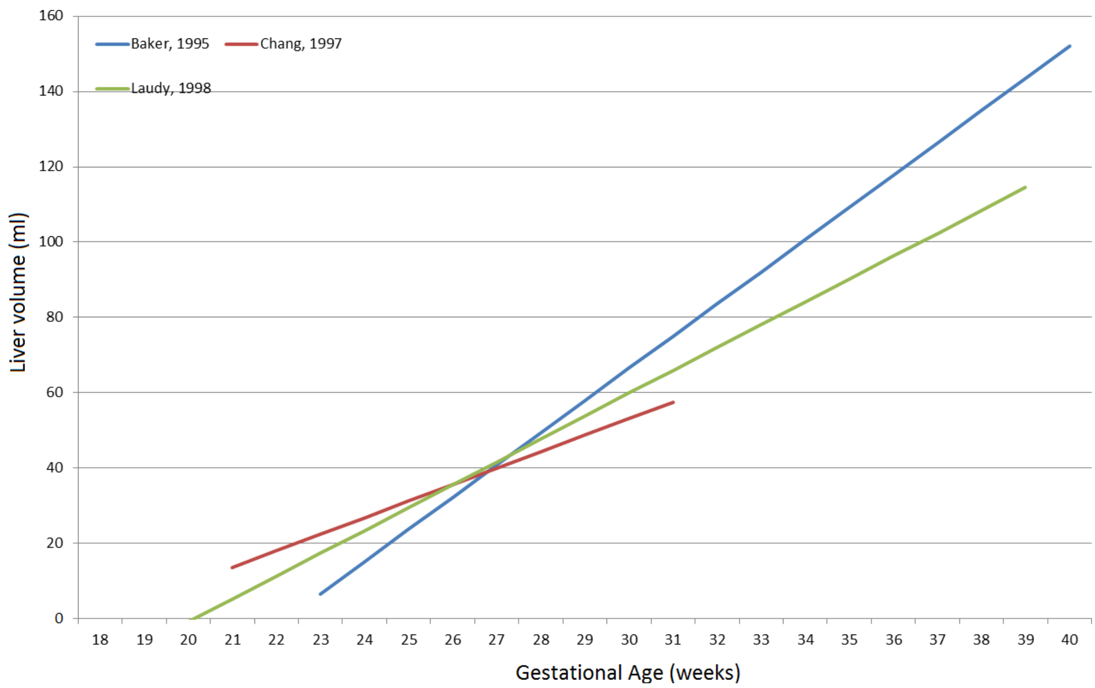


Figure 5.1: *Plot of the relationship between LVol (ml) and GA (weeks) for studies that found a linear relationship. Each line is drawn across the range of GAs used in each study. [7, 31, 96]*

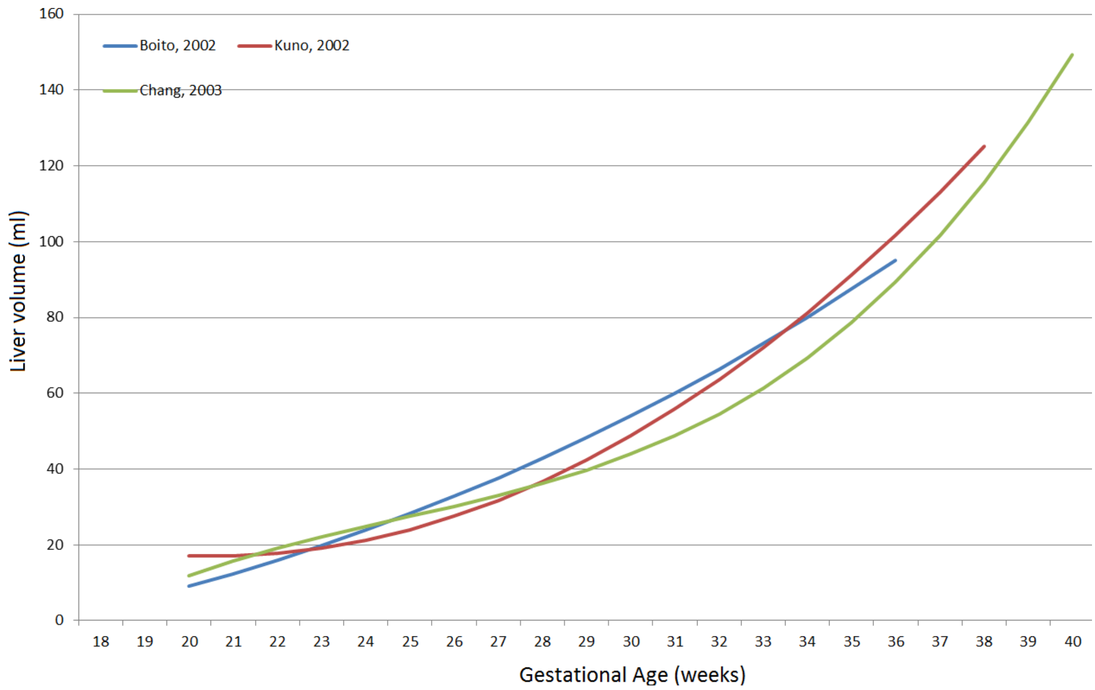


Figure 5.2: *Plot of the relationship between LVol (ml) and GA (weeks) for studies that found a quadratic or cubic relationship. Each line is drawn across the range of GAs used in each study. [16, 29, 93]*

with a 10mm slice separation [7], while Kuno *et al.* segmented 3D US images at 1mm intervals [93]. A further consideration is that many of the studies had relatively small data sets (around 25 fetuses total), which may be due to the difficulties involved in calculating such volumes.

In order to collect more data to establish effective growth charts from which to assess fetal growth, and to allow liver volumetry to be a viable option in clinical practice, it is essential that methods are developed for the segmentation of the fetal liver. This can be achieved through either MRI volumes or US volumes. MRI has the advantage that, under certain settings, the boundary between the liver and surrounding tissue can be clear (as shown in Figures 5.3a and 5.3b). However, it is not a routine part of clinical assessment, with scans usually being taken either specifically for studies or in order to confirm the diagnosis of severe fetal conditions such as spina bifida. US, in contrast, is a standard part of clinical procedure for pregnancy but presents a far more challenging modality to identify liver boundaries (as shown in Figures 5.3c, 5.3d and 5.3e). This chapter will address the issues of segmenting the liver in both of these modalities.

5.2 Data

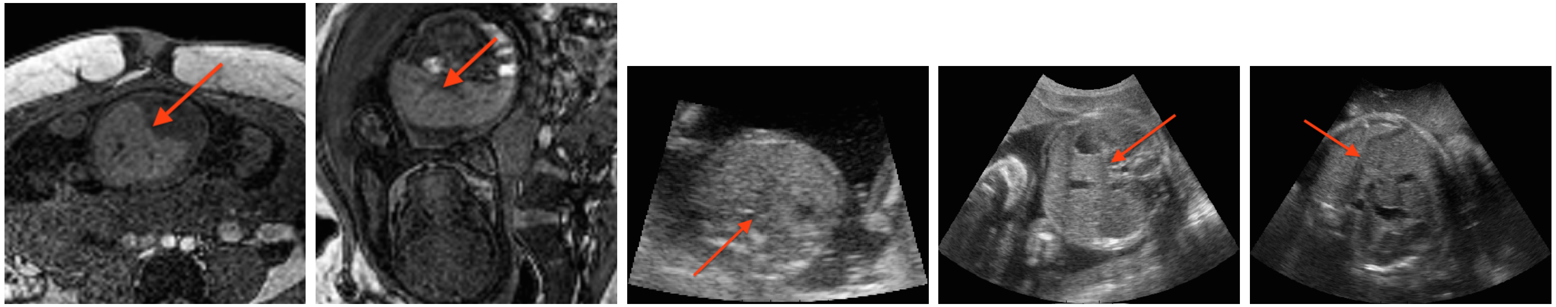
5.2.1 INTERBIO-21ST

MR images of the fetal liver for volume measurement have been obtained from the INTERBIO-21st study [2]. The intake requirements for INTERBIO-21st are not as specific as those used for INTERGROWTH-21st (which sought to eliminate as many risk factors for malnutrition as possible), but at the time of data collection the population still predominantly consisted of healthy, singleton pregnancies.

The MRI volumes have a spatial resolution of $1.95mm \times 1.95mm$ with a slice separation of $4mm$, acquired using the e-THRIVE protocol. Spatial resolutions of the US images were not available from their DICOM files, and so this chapter will be considered a proof of concept for the segmentation process rather than a study into volume measurement.

Table 5.1: *Equations for LVol from GA, as estimated by a number of different population-based studies, along with the scope of each study.*

Paper	Year	GA (weeks)	Sample size	Equation
Baker *et al.* [7]	1995	23+	21*,11†	$LVol = -190 + 8.55GA$
FM Chang *et al.* [31]	1997	22-31	55*	$LVol = -78.285 + 4.3844GA$
Laudy *et al.* [96]	1998	19-39	25*	$LVol = -122.22 + 6.07GA$
Kuno *et al.* [93]	2002	20+	14*, 10†	$LVol = 167 - 14.6GA + 0.3552GA^2$
Boito *et al.* [16]	2002	20-36	85*,24§	$LVol = -18.268 + 0.0443GA^2 + 0.0012GA^3$
CH Chang *et al.* [29]	2003	20-40	226*	$LVol = -398.26 + 46.199GA - 1.7567GA^2 + 0.0236GA^3$



(a) *Patient9 Transverse* **(b)** *Patient18 Coronal* **(c)** *16 weeks, 6 days* **(d)** *23 weeks, 1 day* **(e)** *35 weeks, 2 days*

Figure 5.3: *Examples of fetal abdominal cross sections across GA in both MRI ((**a**) and (**b**)) and US ((**c**), (**d**) and (**e**))*

*Fetuses identified as AGA
†Fetuses identified as SGA
‡Fetuses identified specifically as IUGR

5.2.2 US Challenges

As with all forms of US imaging, typical challenges in image segmentation include speckle, attenuation, low contrast boundaries and shadowing [128]. Liver boundaries are particularly indistinct compared to other structures and can often be difficult to judge by eye. Furthermore, the problem of shadowing is more apparent as GA increases, as structures such as the ribs become greater acoustic boundaries and cause greater attenuation to the signal.

The liver presents an additional problem in that it is unlikely to be the most predominant feature of a given US image. Clearer boundaries can be seen between the liver tissue and its internal vasculature, which may distort many conventional segmentation methods by presenting stronger boundaries than those of the object of interest. Similarly, there is a thin region of tissue between the liver and surrounding amniotic fluid, which presents a strong but incorrect boundary in close proximity to the target boundary.

Figures 5.3c, 5.3d and 5.3e show the appearance of the fetal liver in a slice of 3D US at 16, 23 and 35 weeks GA respectively. The appearance of the liver can also change drastically through gestation, meaning that many shape- and appearance-based methods would require very large training sets throughout gestation in order to build an accurate model. Even in modalities that are generally easier to interpret than US, such as CT, the liver is identified as difficult due to shape variation and inconsistent boundary conditions, due to the the organ forming boundaries with many different tissue types [77].

Additionally, US presents the issue that appearance can change greatly between examples due to factors such as shadowing. The setting of obstetric US increases variation in liver appearance, since GA, growth trajectory, the maternal fat layer and US settings can also influence how it appears. As with the issue of fetal fat segmentation, any assumptions that could influence the overall volume of the segmented shape would be ill-suited to an application in which volume is critical to diagnosis.

5.2.3 MRI Challenges

MRI is another imaging modality which is often used during pregnancy, although not with the same ubiquity of US. Fetal MRI is most frequently used in circumstances where US is insufficient to accurately diagnose conditions, and so is often applied to observe the fetal brain which can become occluded in US during later gestation due to increased bone density [60]. Additionally, it is not performed before approximately 20-24 weeks GA.

Figures 5.3a and 5.3b give two examples of the fetal liver as observed in MRI. It can be seen that boundaries here are far clearer than in the US examples, with a more distinct difference in intensity for the liver tissue compared to surrounding areas. Shadowing and attenuation are not issues in MRI. Potential disadvantages of the modality are that images take time to acquire, meaning that maternal and fetal movement can potentially impact imaging quality [60]. In the examples used for this thesis, e-THRIVE acquisition times were in the order of 10-15 seconds but other protocol can take even longer. As mentioned earlier, boundary conditions can be inconsistent for the fetal liver, regardless of the modality used [77].

In INTERBIO-21st, MR images were taken in a small subset of volunteers for the purposes of research once during pregnancy, after 21 weeks GA. These were taken using the e-THRIVE image setting, as this was found by clinicians to give the best appearance of the fetal liver. e-THRIVE is an 'enhanced' version of THRIVE, which is a Phillips specific imaging protocol, equivalent to FAME and VIBE on GE and Siemens machines respectively [25]. These images have a pixel spacing of 1.953mm in the x and y direction, with a slice thickness of 4mm and slice separation of 2mm (this overlap essentially makes each slice represent 2mm in real terms). A number of the images from the set were rejected on the basis that the sweep, which typically amounted to 40-90 slices, did not completely capture the full volume of the liver.

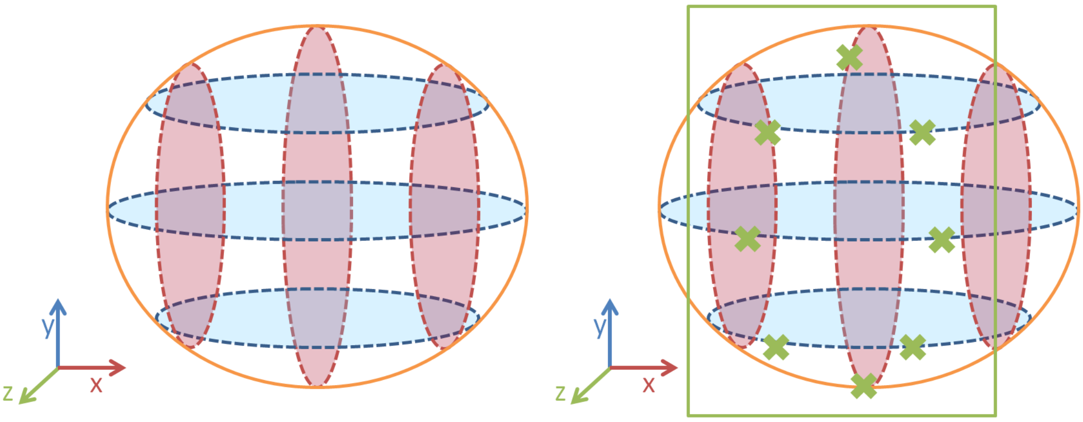
FA-S Live Wire is suited to this application, as fetal liver can exhibit some of the problems faced with segmenting features seen in US in MRI as well, such as low contrast and varying boundary contrast, although in this case the latter depends on what surrounding tissue the liver is in contact with rather than signal attenuation. As the volumes are a lower resolution than many of the US images, the scales used for the FA calculation must be modified accordingly.

5.3 Objectives

The main objective was to develop a method for the segmentation of the fetal liver, a potentially complex 3D structure, in fetal US and MRI. As with Chapter 4, the use of user-guided techniques have been explored, to allow for a greater, intuitive control over the final segmentation. Additionally, an important design constraint for this method is that it must be able to cope with the challenges brought by the imaging modalities used, particularly US (both as the primary method for fetal observation and the more challenging modality). For similar reasons to those described in Chapter 4, it has been decided that an interactive segmentation scheme would be well suited. The solution to this task focused primarily on expansions of the Live Wire algorithm, as a logical expansion of previously discussed work. Facets of the monogenic signal are also utilised, as they have been shown to be useful in solving US-related segmentation problems [128].

The secondary objective was to reduce the amount of user interaction required to perform the segmentation in 3D. This is arguably more essential than in the 2D case, as it is likely that multiple segmentations will be required to form the entire volume. This would increase the time taken to perform the segmentation and could increase the chance of user fatigue during the process [136]. As a result, methods that allow for 3D surfaces to be accurately represented, using a small number of 2D segmentations to initialise the segmentation through the volume, have been investigated.

At the time of writing, there appears to be very little research into automating the segmentation of fetal liver in US and MRI, presumably because of the difficulty involved in observing object boundaries. Segmentation problems often involve segmenting internal



(a) **(b)**

Figure 5.4: *(a) A sparse set of segmentations are taken in the x and y directions (b) Intersections between each layer of the z plane and the existing segmentations are used as the anchor points for Live Wire*

liver structures rather than the organ itself, for example mapping liver vasculature in CT [77]. However, attempts have also been made in the segmentation of the full liver in 3D, although this has been limited to adult livers, again in CT [191].

5.4 Existing 3D Live Wire Methods

In forming a volumetric segmentation of the fetal liver, typically around 100 slices are required to form a full segmentation of the liver in US and around 40 slices in MRI. Even with the increased speed of 2D segmentation time given by using a semi-automated method, volume segmentation remains a lengthy process. This can be addressed by adapting Live Wire to cater specifically for 3D applications rather than just using the 2D method on a slice-by-slice basis.

5.4.1 Sparse 3D Live Wire

A number of attempts have already been made at extending the functionality of Live Wire into 3D imaging problems, primarily by using a sparse set of 2D segmentations (often in two orthogonal directions) to automate the creation of a dense set of 2D segmentations which are combined to form a complete volume [185].

Falcão and Udupa [51], Lu and Higgins [105], and Poon *et al.* [143] all conduct 3D segmentations in this manner, although none of them address the problem of US applications in particular. For simplicity, this method is referred to as Sparse 3D Live Wire.

Figure 5.4 shows a simplified example of how this concept operates in practice; in Figure 5.4a, a sparse set of segmentations are taken in the x and y directions. In Figure 5.4b, every image slice in the z direction is considered, with the intersections between the current z plane and the existing sparse segmentations forming the anchor points for the Live Wire segmentation for that slice.

5.4.2 Point Ordering

An important part of forming an algorithm to construct a Sparse 3D Live Wire volume is automating how the points selected on the plane are ordered as inputs for the Live Wire algorithm.

Poon *et al.* address this issue by using a so-called Turtle algorithm [143]. Since two orthogonal planes are used to select the points on the third orthogonal plane, seed points will always occur in groups (usually pairs) in line with each other on the image axis, as shown in Figure 5.5. This means a connected map of seed points can be constructed with relative ease. Each point on this map can be assigned different values, flagging them as either path, intersection or seed point, an object (referred to here as a 'turtle') can traverse the map, assigning an order to the seed points.

As long as the 'turtle' is on a pixel marked as a path, it will continue to travel in the same direction. When a pixel labelled as an intersection is reached, the 'turtle' will take the leftmost available path with respect to its current direction (prioritising directions in the order; left, forwards, right then back). When a pixel labelled as a potential seed point is reached, the 'turtle' operates as follows;

Set $n = 1$

if The turtle is on a path

\- Move forwards

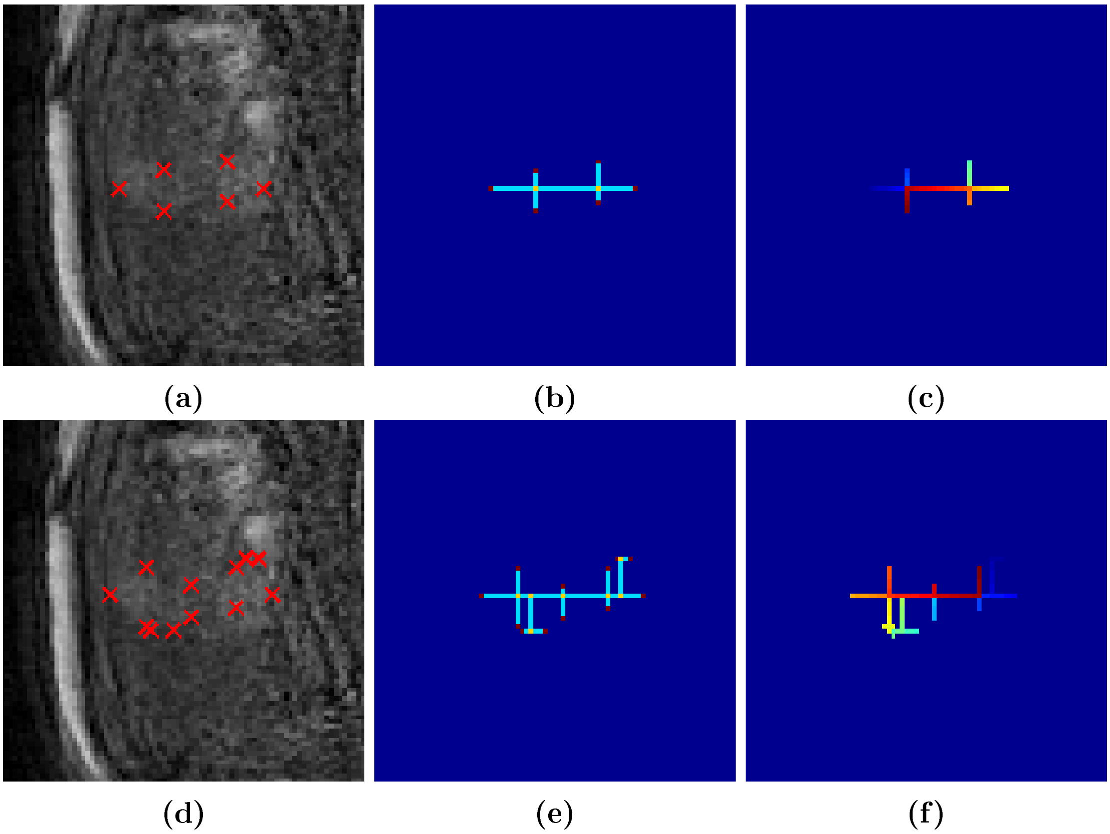


Figure 5.5: *MR image of the fetal liver with Sparse Live Wire points, the 'turtle' map used to order the points and path taken by the 'turtle' to define the point order. This is shown for three ((**a**)-(**c**)) and six ((**d**)-(**f**)) slices in an MRI segmentation. In figures (**c**) and (**f**), the colour of the pixels indicates the path of the turtle, with blue as the first-visited point and red as the last-visited point.*

else

if The turtle is on an intersection

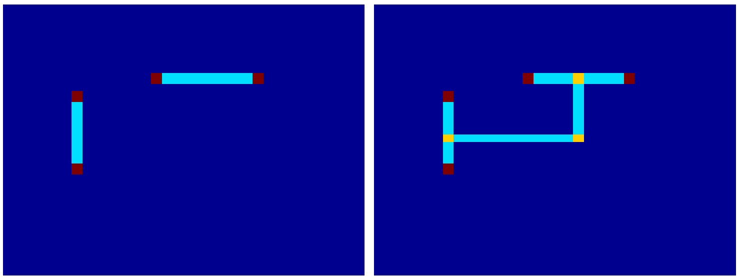
- Take the leftmost unexplored path

else

if The turtle is on a seed point

- Mark the current location as the n^{th} seed point
- Set $n = n + 1$
- Reverse direction of travel

This is repeated until a closed loop is formed. If there are any remaining anchor points, a new one is chosen and the process is repeated until all possible closed loops



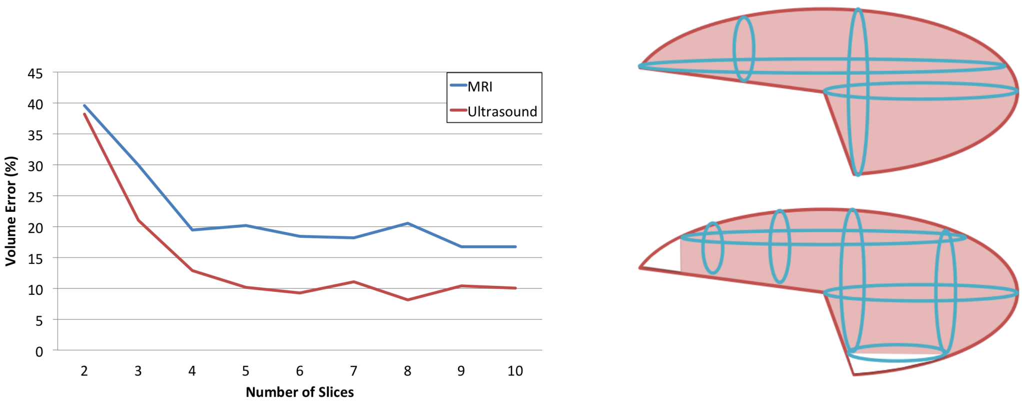
(a) (b)

Figure 5.6: *An example of where the two sample slices are not connected, and must be bridged for the point ordering algorithm to operate properly.*

are found. Figure 5.5 shows an example of this algorithm being applied to two sets of segmentations of the same liver MRI volume, but with different slice separation. Figures 5.5b and 5.5e show the label image for the anchor points; $label = 1$ (cyan) for the pixels joining anchor points in the x and y directions, $label = 2$ (yellow) for intersections and $label = 3$ (red) for anchor points. Figures 5.5c and 5.5f show the path taken by the turtle in this example, and from this the order of the points as they will be applied to the Live Wire algorithm can be seen (the path of the turtle starts with the blue pixels and ends with the red pixels). Here, it can also be seen how this process assures appropriate point selection, even in special cases such as a T-junction. In some cases, sections of the path are not initially connected. Figure 5.6b shows how the mid-points of such sections can be joined, completing the map.

5.4.3 Utility of Sparse 3D Live Wire

One of the major drawbacks of using sparse 2D segmentations to inform a 3D segmentation is that it may not be clear as to how many segmentations are required to complete the full surface. Difficult imaging conditions or adjoining structures may mean that more segmentations are required; but this will not be known until the surface is generated and the user observes that the segmentation process has behaved as expected. In the limit of this case, the user may be required to conduct a number of 2D segmentations close to the order of that required for a slice-by-slice segmentation, but this would be unknown at the outset. Figure 5.7a shows how segmentation accuracy (in terms of volume error)



(a) (b)

Figure 5.7: *(a) Variation in segmentation accuracy (in terms of volume error) of a fetal liver changes with the number of slices in both MRI and US. **(b)** This schematic demonstrates how more slices does not necessarily improve segmentation accuracy. Blue elipses represent a 2D segmentation, while regions in which a 3D segmentation can be performed are shaded in red.*

of a fetal liver changes with the number of slices in both MRI and US. This experiment is described in full in Section 5.6.2, which explores the limitations of using sparse 2D segmentations to inform 3D segmentation.

Even when ideal imaging conditions are available, a reasonably large number of slices would be required in order to give enough information in the third plane to make a segmentation throughout the entire object. This is especially true in the case where the object has unusual topology, for example the liver, since it is essential that multiple points persist through the entirety of the third plane. Consider the object in Figure 5.7b, with a set of segmentations made in two orthogonal planes. Slices in the third plane which intersect with at least two points are able to be segmented. Regions in which this is the case are shaded in red. In the top diagram, four segmentations are used, but the entire object is covered by segmentations in the third plane. In the lower example, seven segmentations are made, but there are regions of the object which have no information from which to inform the 3D segmentation. This may explain the fluctuation in results shown in Figure 5.7a, and how fewer segmentations can occasionally provide better accuracy.

A better way of forming a 3D segmentation from 2D information would be to create a method that ensures that the entire object can be delineated, regardless of the number of segmentations made or where they are placed. The objective is to use 2D segmentations to inform 3D segmentations with entirely arbitrary topology and aiming at using a minimal number of 2D segmentations to achieve this.

5.5 ASP Live Wire

There are alternatives to extending Live Wire to a 3D context which could reduce the number of initialisation slices required. One potential way of bringing automation to a process such as Live Wire is to use the information obtained in an initial segmentation in order to inform subsequent segmentations. Schenk *et al.*, for example, have developed a method for using Live Wire to segment liver CT images by using *shape-interpolation*, in which sparse segmentations in two dimensions are used to inform the location of the position of the segmentations between these user-defined contours [160]. In order to improve the results of the interpolated shapes, user-defined seed points are selected from adjacent slices to calculate what the corresponding Live Wire result would have been on that slice, and the user is able to insert additional seed points to improve this segmentation result [160].

Taking this idea in a different direction, it could be possible to use image data to determine the position of seed points in adjacent slices rather than shape information, iterating through the image volume. If points are selected such that they lie on the same boundary in consecutive images, far fewer 2D segmentations would be required to complete the object boundary in 3D. This method will be referred to as Automated Seed Point (ASP) Live Wire.

FA has been successfully utilised to localise edges in US images, from which it can be inferred that it may be suitable to identify point locations in consecutive slices. However, using FA alone may result in such an algorithm trending towards the closest, strongest boundary, regardless of whether or not it is the same boundary as the one followed in

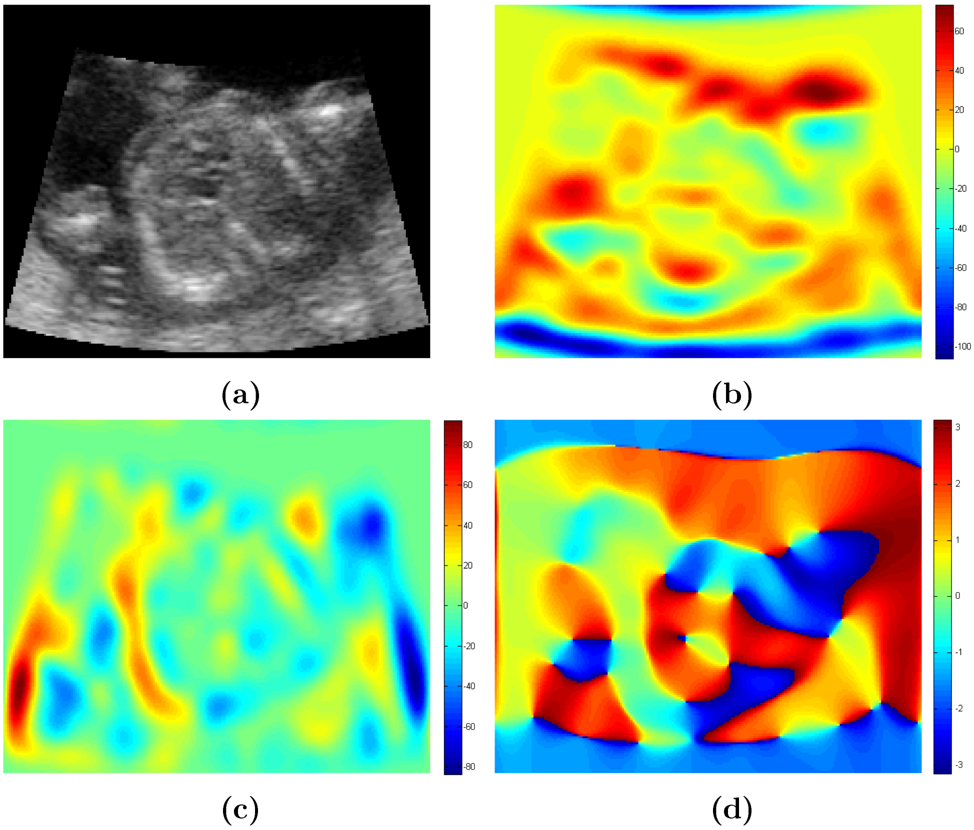


Figure 5.8: *(**a**) Original US image (**b**) First orthogonal filter response (**c**) Second orthogonal filter response (**d**) LO*

the previous slice. In order to guard against this, additional image information must be acquired; the LO of the image may be a useful tool in determining which boundary is the same through different image slices.

An additional advantage of conducting the segmentation in this way is that the user need only define the search window size and number of points around the contour required, altering these parameters to compensate for difficult imaging conditions or complex object topology rather than by conducting further 2D segmentations.

5.5.1 Local Orientation

Chapter 4 introduced the concept of FA-S Live Wire, utilising a circular shape constraint and FA as cost functions to guide the Live Wire segmentation. In this example, the shape constraint introduced by FA-S Live Wire is not appropriate, since the boundary lies on a generally linear path rather than a circular one. The FA aspect of FA-S

Live Wire remains relevant, but any other constraint must consider the properties of this particular segmentation problem in order to be successful.

The FA of an image can provide information on the location of boundaries within it, but in order to find a place that is structurally similar between images, this may not be sufficient. The LO may be able to provide additional context by which to identify similar points between images. Recall from Chapter 4 that the equation for the monogenic signal of a 2D image can be expressed as,

$$f_M(x) = I_b + [(o_1 \otimes I_b), (o_2 \otimes I_b)] \tag{5.1}$$

where I_b is the bandpass filtered version of the original image, I, o_1 and o_2 are odd orthogonal filters and $\otimes$ denotes the convolution operation [9,112]. In contrast to the LP of the signal, which is a measure of the angle between the odd and even filter responses, the LO of the signal is the angle between the responses of the odd filters on the image, and can be defined as

$$\theta = tan^{-1}\left(\frac{o_1 \otimes I_b}{o_2 \otimes I_b}\right), \tag{5.2}$$

giving the angle between the orthogonal filter responses [112]. Figures 5.8b and 5.8c show the appearance of $(o_1 \otimes I_b)$ and $(o_2 \otimes I_b)$ respectively for the US image displayed in Figure 5.8a, while Figure 5.8d illustrates the resulting value of θ. This information alone appears unintuitive, but could prove useful when used in conjunction with other image information.

5.5.2 Automated Seed Point Process

The goal of this process is to select a point which has been established to occur on the boundary of the object being segmented (in this case the fetal liver) and to find where the same point occurs on adjacent slices. These points can then be used to automatically guide a FA-based Live Wire contour. The method operates as follows;

- Initially, a set of contours are drawn throughout the object of interest, utilising

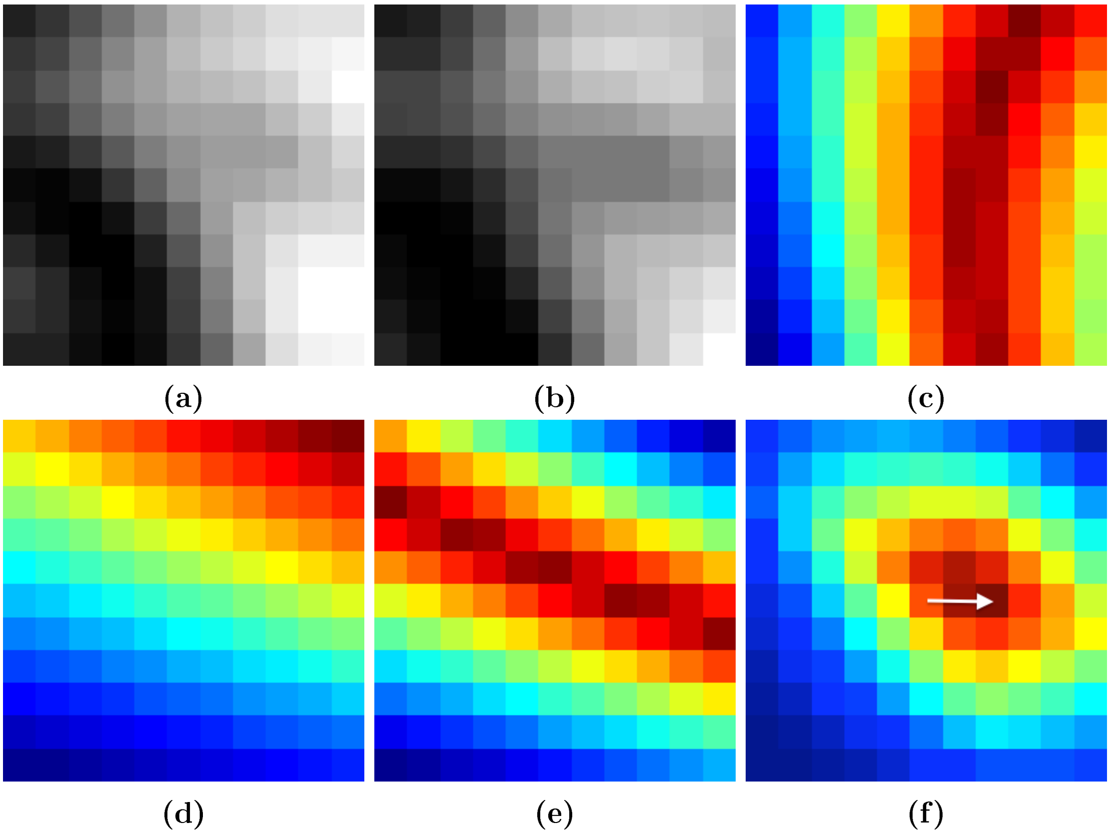


Figure 5.9: *Neighbourhood for **(a)** an image patch on slice $n-1$ **(b)** an image patch on the slice n, where the localisation is occurring **(c)** FA **(d)** LO **(e)** LO of slice n with respect to the LO of the centre point of the neighbourhood in slice $n-1$ (LO_{mod}) **(f)** $FA \times LO_{mod}$*

a tool such as FA-S Live Wire (or, in the case of an entirely arbitrary shape, FA alone may be more suitable). It is important that the extremities of the object are included, as these will be the stopping point of the process.

- One of the slices near the centre of the object is selected as the start point of the automated 3D segmentation process.

- This contour is sub-sampled to somewhere in the region of 10-20 points, evenly spaced around the contour. These are used to determine a number of pixel locations which can be used as initialisation points for the anchor points in the next slice. Schenk *et al.* suggest the use of points corresponding to the user's initial control inputs, since they may be indicative of regions which require greater guidance [160]. However, evenly spacing a fairly large number of control points around the segmented contour will increase robustness to any change in shape and boundary conditions through the image volume.

- These locations are translated onto the next adjacent slice, and a neighbourhood is determined around each pixel, modified to ensure they do not extend beyond the image boundaries. Figure 5.9a shows the intensity for the neighbourhood around the Live Wire contour point on the previous image slice, which can be referred to as slice $n-1$. This is the reference neighbourhood which will help inform the position of the new anchor point in slice n. In this example, an 11 by 11 pixel neighbourhood is used. Figure 5.9b, meanwhile, shows the intensity for the neighbourhood at the same location but on slice n.

- FA is calculated for the neighbourhood in the current slice, n, as shown in Figure 5.9c. This determines where the image boundaries are located for the new image slice.

- The LO for the patch in the slice n is also calculated, showing the orientation of the odd filter responses. This is shown in Figure 5.9d. Note that this forms a line perpendicular to the image boundary identified by FA. The goal is to find the intersection between these two regions, therefore finding a structurally similar location between the previous image slice and the current one.

- Next, the difference in LO between each pixel in the image shown in Figure 5.9d and the LO value at the central pixel of the slice $n-1$ neighbourhood shown in Figure 5.9a is calculated. Figure 5.9e shows $\pi - |\delta\theta_{mod}|$, where $|\delta\theta_{mod}|$ is the magnitude in the difference of LO. This gives a high response where the LO is similar to the value of the LO at the previous Live Wire contour location.

- Finally, the FA value for each point in the image patch is multiplied by the corresponding LO difference. The location of the maximum value of this region is taken as the new anchor point for the Live Wire algorithm, as shown in Figure 5.9f.

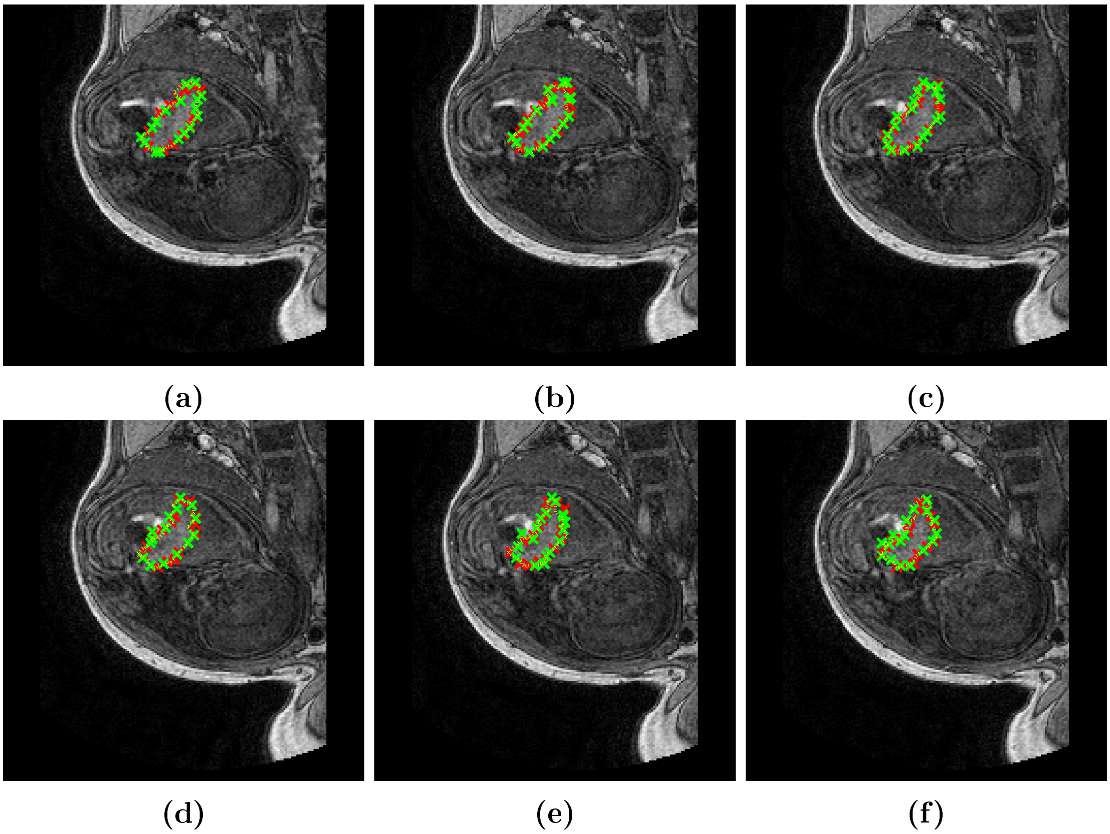


Figure 5.10: *Example of how the ASP Live Wire curve deforms as it propagates through a liver MRI volume as seed points are re-defined*

- These new locations are used as the next inputs for the Live Wire algorithm, allowing for the segmentation to be conducted automatically from this point. The process is repeated for all other image slices. In the case where multiple slices are used throughout the volume, the contours are propagated up to the mid-point between the user-defined segmentations.

In the case of ASP Live Wire, it is not necessary to perform any kind of point-ordering, since they will remain in the order that they appear on the original contour. Figure 5.10 shows an example of the algorithm being applied to several slices within a 3D US image of a fetal liver. Figure 5.10c is the initial contour from which the others are derived. One point to make when calculating the difference between values of LO is that, as is the nature of angles, they follow a cyclic pattern, meaning that values of $\theta = -\pi$ and $\theta = \pi$ actually represent the same value, while $\theta = -\frac{\pi}{2}$ and $\theta = \pi$ have a separation of $\frac{\pi}{2}$ rather than $\frac{3\pi}{2}$. Therefore, the modified difference in LO, $\delta\theta_{mod}$, is the original LO, θ, altered by the previous orientation, φ, using the equation,

$$\delta\theta_{mod} = \begin{cases} \theta - \varphi - 2\pi & if \quad \theta + \varphi > \pi \\ \theta - \varphi & if \quad \pi > \theta + \varphi > -\pi \\ \theta - \varphi + 2\pi & if \quad \theta + \varphi < -\pi \end{cases} \tag{5.3}$$

This ensures that zero represents a point where $\theta=\varphi$ and the greater the value of $|\delta\theta_{mod}|$, the greater the angular deviation.

Methodology Overview: ASP Live Wire

- The user segments a selection of evenly spaced slices throughout the volume, ensuring that they segment at the extremities of the object. This will be used as the stopping conditions.

- A subsample of points are taken on the central user-defined 2D segmentation and projected onto the next slice in the 3D sequence (propagating in either direction).

- FA and LO images of the new slice, and LO of the original slice are then calculated.

- A window around each point is taken, giving a map of the FA and LO in the point's neighbourhood. The FA map is multiplied by the LO difference map, which will take a high value for the smallest difference between the LO map values and the original LO value of the point on the previous slice (giving LO_{diff}).

- The points are then moved to occupy the highest value of the $FA \times LO_{diff}$ region and a new FA Live Wire contour is calculated using these adjusted point positions.

- This process is repeated until the final user-segmented slice is reached and the full object volume has been covered.

5.6 Validation of Method

5.6.1 Experiment Outline

In order to explore the limits of this method, it is important to observe how the segmentation results were influenced by the number of slices used to define the volume. To do this, the fetal livers were segmented on a slice-by-slice basis using FA-S Live Wire. This segmentation was used as the standard by which the others are judged, as it defines the 'best possible' segmentation possible with interactive segmentation. Using these boundaries as a base, a sparse set of segmentations were selected and spread evenly through the segmented volumes. These were used as inputs for the Sparse 3D Live Wire and ASP Live Wire algorithms.

For the MRI group, ten images were used, with the segmentations conducted on every slice in which the liver was visible. Each of these images had a corresponding US image, taken of the same fetus on the same day. However, due to difficulties in determining the exact location in the liver in most of the examples, only four of these were used for the US group. An additional US image with no MRI analogue was also used. Although it would not be possible to compare the results from this example between the two modalities, it can still provide an additional example of the automated tool's use in different US conditions.

Precision, recall and Dice similarity were used to assess the quality of segmentation achieved by each of these methods, comparing automatically segmented results to a volume in which every slice has been manually segmented once using FA-based Live Wire. These values are calculated on a slice-by-slice basis for all of the example images. Additionally, the volume error has been calculated; this is a particularly important value in the context of liver segmentation, as the aim is to measure LVol and hence infer the trajectory of a fetus's development.

For both the MRI and US volumes, 1-10 user-defined contours were selected for the initialisation of ASP Live Wire and 2-10 for Sparse Live Wire, in order to determine the maximum separation these methods can suitably take while retaining segmentation accuracy. In these experiments the number of user-defined slices is denoted by d. MRI volumes were made up of a total of 40-90 slices total, while US volumes were made up of around 50 slices.

In both cases, the 2D segmentations were performed using FA-driven Live Wire and taken in the sagittal and coronal planes, with the automation taking place in the transverse plane. These have been chosen since they present the easiest segmentation problem for human interaction. There are clearly visible landmarks in both the coronal and sagittal planes, since the stomach and lungs can be used as the upper and lower bounds for nearly any given slice. There is slightly more judgement required to see exactly where the liver stops and starts in the transverse plane, whereas in the sagittal and coronal planes the extremities are approximately at the edges of the abdomen. Additionally, the coronal plane has the more complex geometry, making either transverse or sagittal good candidates for the automated segmentation.

For the ASP Live Wire case, the user-defined contours were distributed equally through the volume. It is important to note that while Sparse 3D Live Wire segmentations require the use of user-defined segmentations in two orthogonal directions, ASP Live Wire only distributes them along one axis. Therefore, in the application of Sparse 3D, the number of user-defined slices is split between the x and y axis of the object.

Other variables to investigate to assess the functionality of ASP Live Wire are the filter scales and window sizes used. Filter scale is also relevant to the application of Sparse Live Wire but window size is not a feature of its operation. While the utility of the monogenic signal filter scale is dependent on how large the image features are, the selection of window size is related to the imaging modality's slice separation. An image with poor spatial resolution can mean that features move further between adjacent slices compared to an image with better resolution, requiring a larger window size to be able to more accurately follow the contours. However, too large a window size can result in the automated seed points being moved out to structurally similar but incorrect locations. In order to test this, a variety of window sizes are used (5×5, 7×7, 9×9 and 11×11 pixels for MRI, 11×11 pixels only for US).

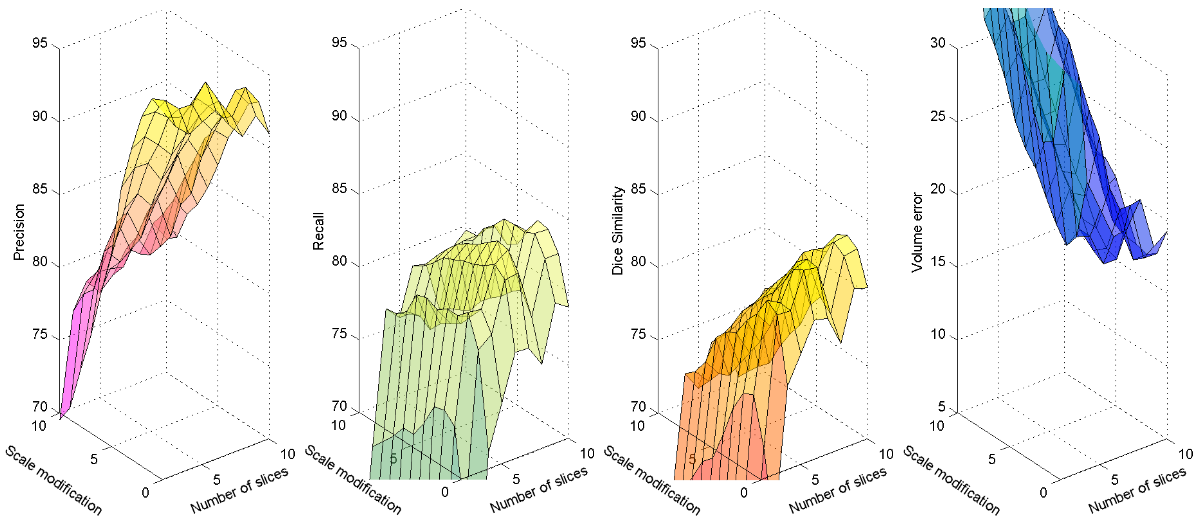
In the calculation of the MRI contours, the scale for FA and LO calculation is defined as $s = [1, 3, 5] + k$ where k is varied between 0 and 10. For US, the scale for FA and LO calculation is defined as $s = [5, 10, 15] + k$ where k is varied between 0 and 10. It is likely that a better result would be obtained by selecting k to specifically suit each example, but this would complicate the assessment process.

Table 5.2: *MRI Sparse and ASP 3D Live Wire*

Method	d (slices)	Precision (%)	Recall (%)	Dice Sim. (%)	Volume Err. (%)
Sparse	1	-	-	-	-
Live Wire	2	94.02 ± 10.53	59.12 ± 36.15	64.54 ± 35.85	39.55 ± 36.23
	3	93.98 ± 10.09	72.32 ± 35.83	74.55 ± 34.95	26.96 ± 35.51
MRI	4	93.23 ± 12.80	79.82 ± 32.50	80.70 ± 32.04	19.43 ± 32.04
	5	92.93 ± 12.56	79.68 ± 33.92	80.16 ± 32.80	20.19 ± 33.02
$s = [4, 6, 8]$	6	93.92 ± 10.95	82.43 ± 33.25	82.11 ± 32.69	18.41 ± 32.92
	7	92.96 ± 11.60	83.07 ± 32.31	82.47 ± 31.58	18.20 ± 31.96
	8	91.86 ± 16.51	79.96 ± 35.84	79.51 ± 35.14	20.53 ± 35.08
	9	93.34 ± 12.97	83.48 ± 33.20	82.92 ± 32.41	16.70 ± 32.19
	10	92.08 ± 18.88	83.43 ± 33.54	83.05 ± 32.93	16.74 ± 32.91
ASP	1	80.75 ± 19.81	88.12 ± 15.54	81.75 ± 15.82	30.59 ± 83.67
Live Wire	2	81.73 ± 17.48	92.78 ± 8.73	85.28 ± 11.81	23.38 ± 45.25
	3	87.88 ± 12.79	91.83 ± 9.52	88.84 ± 9.07	13.32 ± 27.17
MRI	4	87.51 ± 13.61	92.87 ± 7.88	89.22 ± 9.17	12.85 ± 25.14
	5	88.73 ± 13.26	93.58 ± 7.46	90.23 ± 8.95	12.59 ± 27.19
$s = [4, 6, 8]$	6	89.58 ± 12.33	93.27 ± 7.24	90.70 ± 8.65	10.61 ± 25.13
$w = 7 \times 7$	7	89.54 ± 12.42	93.73 ± 7.27	90.95 ± 8.87	10.21 ± 24.43
	8	89.91 ± 12.24	93.59 ± 7.60	91.06 ± 8.78	10.15 ± 24.43
	9	90.82 ± 11.52	94.45 ± 7.03	92.03 ± 8.31	8.88 ± 21.31
	10	90.92 ± 11.44	94.57 ± 6.81	92.15 ± 8.13	8.77 ± 20.85

Table 5.3: *US Sparse and ASP 3D Live Wire*

Method	Slices	Precision (%)	Recall (%)	Dice Sim. (%)	Volume Err. (%)
Sparse	1	-	-	-	-
Live Wire	2	88.98 $\pm$ 21.95	56.28 $\pm$ 31.77	63.39 $\pm$ 32.62	38.20 $\pm$ 30.25
	3	95.19 $\pm$ 7.43	76.81 $\pm$ 27.30	80.85 $\pm$ 26.82	21.02 $\pm$ 27.53
US	4	94.18 $\pm$ 12.29	87.07 $\pm$ 22.77	88.43 $\pm$ 22.05	12.86 $\pm$ 22.39
	5	96.28 $\pm$ 6.04	88.89 $\pm$ 21.59	90.12 $\pm$ 20.86	10.15 $\pm$ 21.46
$s = [8, 13, 18]$	6	96.18 $\pm$ 5.71	90.47 $\pm$ 20.69	90.98 $\pm$ 20.47	9.25 $\pm$ 20.80
	7	94.75 $\pm$ 15.67	88.55 $\pm$ 25.01	89.47 $\pm$ 24.24	11.03 $\pm$ 24.54
	8	95.73 $\pm$ 9.73	92.64 $\pm$ 18.44	92.72 $\pm$ 17.67	8.13 $\pm$ 18.81
	9	94.73 $\pm$ 14.18	90.22 $\pm$ 24.34	90.24 $\pm$ 23.66	10.41 $\pm$ 24.32
	10	95.83 $\pm$ 11.53	90.07 $\pm$ 24.81	90.34 $\pm$ 23.70	10.05 $\pm$ 23.94
ASP	1	74.59 $\pm$ 27.33	83.94 $\pm$ 13.48	74.24 $\pm$ 20.25	80.82 $\pm$ 162.09
Live Wire	2	83.22 $\pm$ 19.63	86.35 $\pm$ 16.76	81.69 $\pm$ 16.38	40.78 $\pm$ 100.22
	3	87.10 $\pm$ 16.25	90.50 $\pm$ 9.71	87.42 $\pm$ 11.99	21.81 $\pm$ 47.89
US	4	86.41 $\pm$ 15.67	91.22 $\pm$ 13.57	87.02 $\pm$ 13.27	24.28 $\pm$ 45.64
	5	91.16 $\pm$ 12.11	90.23 $\pm$ 12.06	89.73 $\pm$ 10.61	14.42 $\pm$ 31.94
$s = [5, 10, 15]$	6	90.51 $\pm$ 15.01	90.22 $\pm$ 9.48	89.29 $\pm$ 11.13	17.82 $\pm$ 38.47
$w = 11 \times 11$	7	90.95 $\pm$ 14.22	92.09 $\pm$ 9.53	90.49 $\pm$ 10.69	16.56 $\pm$ 36.11
	8	89.83 $\pm$ 14.78	93.61 $\pm$ 8.26	90.72 $\pm$ 10.85	16.53 $\pm$ 36.36
	9	91.23 $\pm$ 12.28	94.36 $\pm$ 8.37	92.11 $\pm$ 9.35	11.91 $\pm$ 23.03
	10	92.91 $\pm$ 11.06	94.17 $\pm$ 8.73	92.94 $\pm$ 8.81	10.75 $\pm$ 21.71



(a) **(b)** **(c)** **(d)**

Figure 5.11: *Sparse Live Wire results on MRI of the fetal liver* ***(a)*** *precision* ***(b)*** *recall* ***(c)*** *Dice similarity and* ***(d)*** *volume error for scale modifications of* $k = [0 : 10]$ *and sample slices of* $d = [2 : 10]$

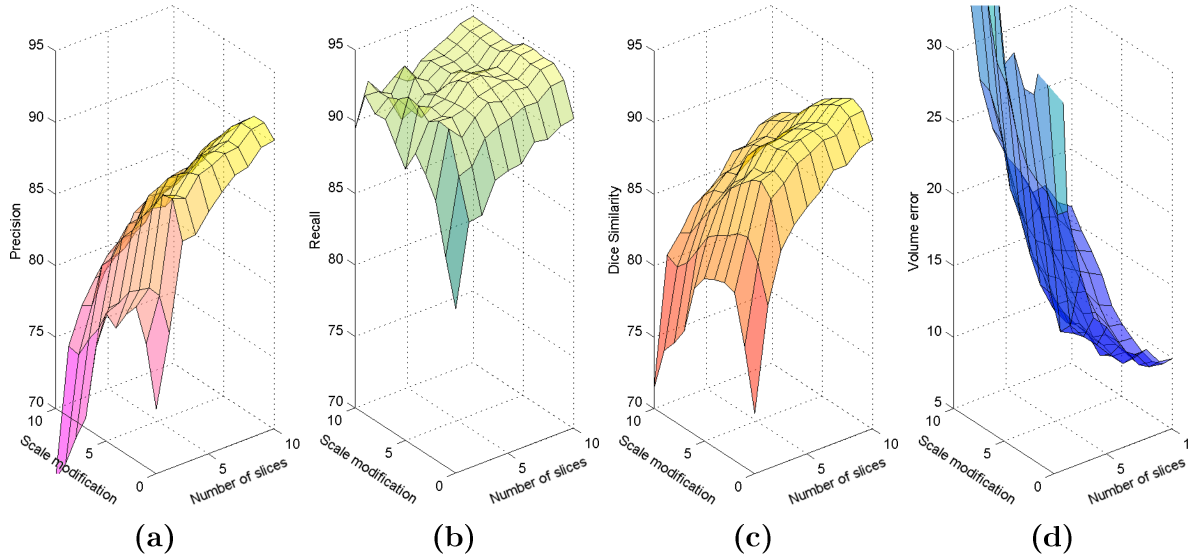


Figure 5.12: *ASP Live Wire results on MRI of the fetal liver precision, recall, Dice similarity and volume error for scale modifications of $k = [0 : 10]$ and sample slices of $d = [1 : 10]$ and $w = 5 \times 5$.*

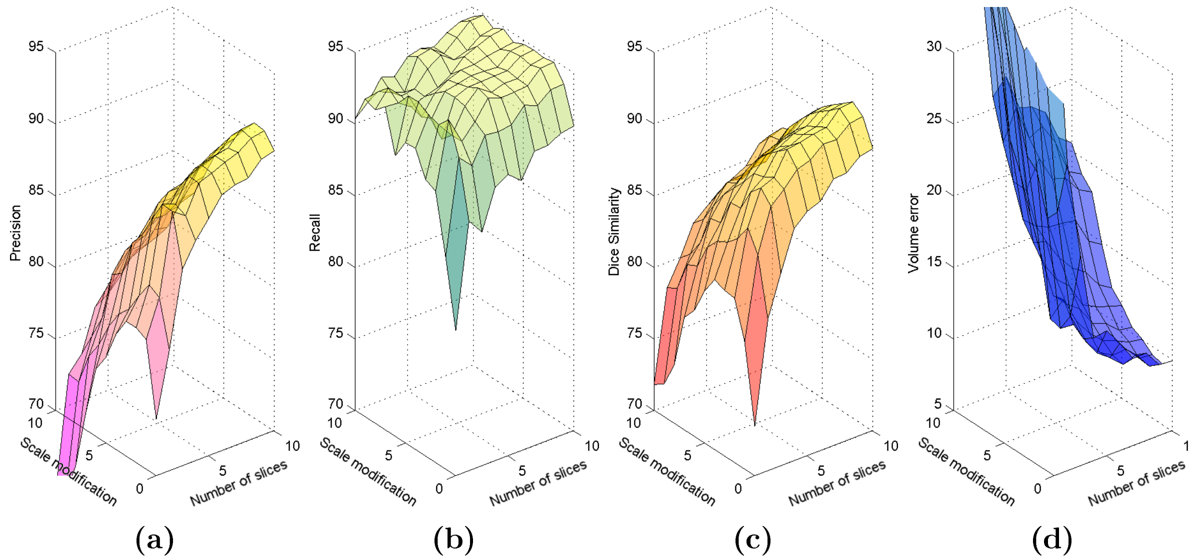


Figure 5.13: *ASP Live Wire results on MRI of the fetal liver precision, recall, Dice similarity and volume error for scale modifications of $k = [0 : 10]$ and sample slices of $d = [1 : 10]$ and $w = 7 \times 7$.*

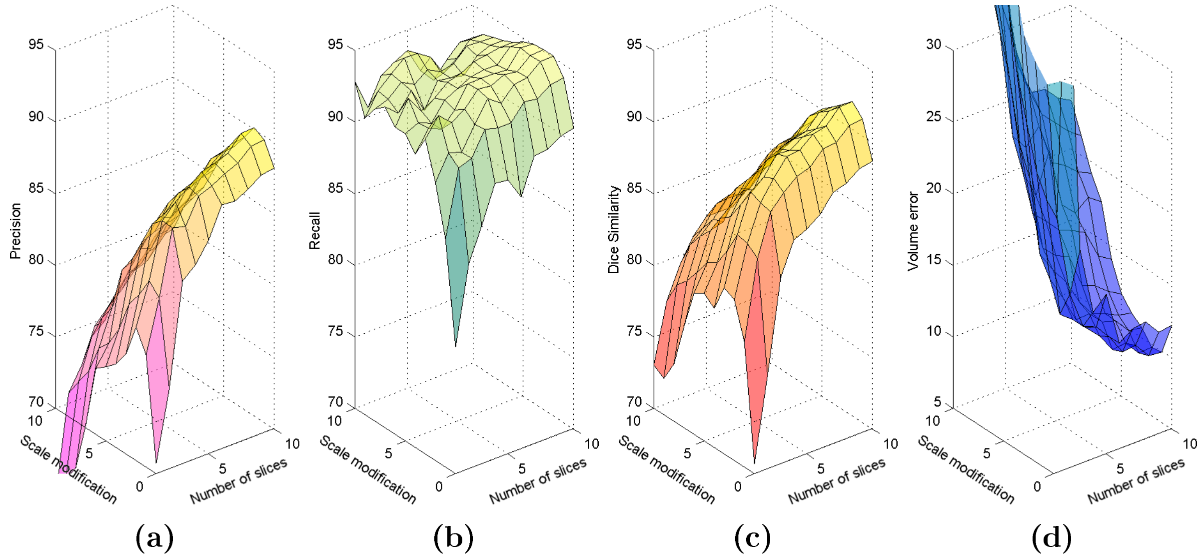


Figure 5.14: *ASP Live Wire results on MRI of the fetal liver precision, recall, Dice similarity and volume error for scale modifications of* $k = [0 : 10]$ *and sample slices of* $d = [1 : 10]$ *and* $w = 9 \times 9$.

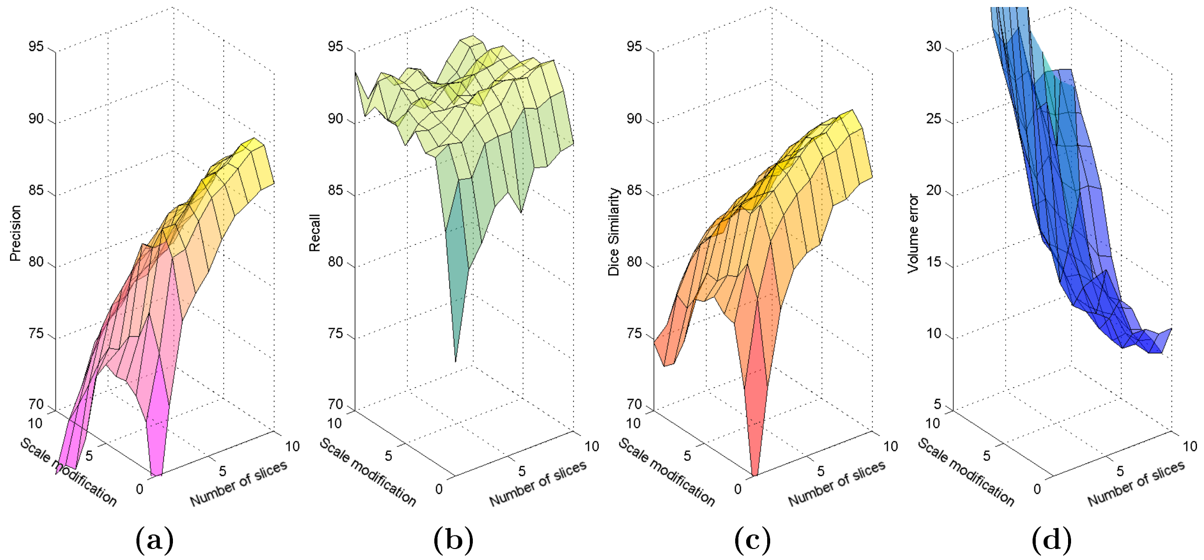
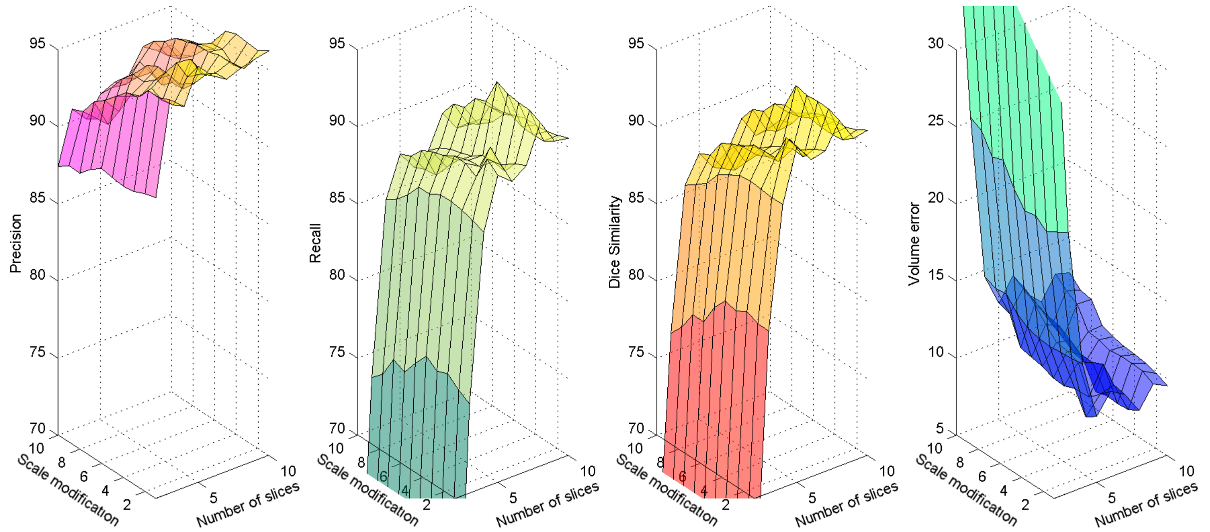
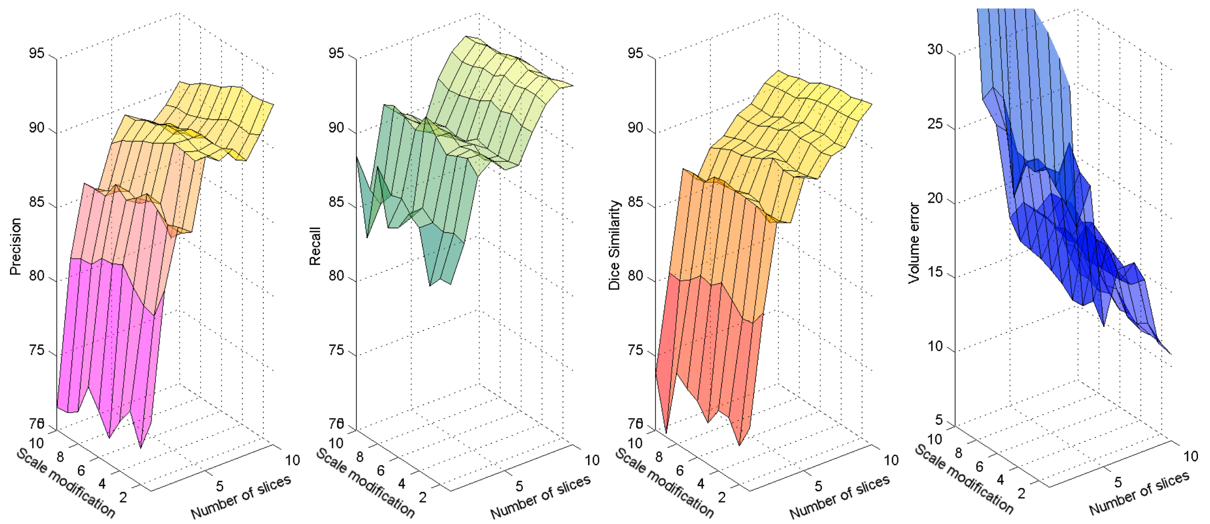


Figure 5.15: *ASP Live Wire results on MRI of the fetal liver precision, recall, Dice similarity and volume error for scale modifications of* $k = [0 : 10]$ *and sample slices of* $d = [1 : 10]$ *and* $w = 11 \times 11$.



(a) **(b)** **(c)** **(d)**

Figure 5.16: *Sparse Live Wire results on US of the fetal liver **(a)** precision **(b)** recall **(c)** Dice similarity and **(d)** volume error for scale modifications of* $k = [0 : 10]$ *and sample slices of* $d = [2 : 10]$



(a) **(b)** **(c)** **(d)**

Figure 5.17: *ASP Live Wire results on US of the fetal liver **(a)** precision **(b)** recall **(c)** Dice similarity and **(d)** volume error for scale modifications of* $k = [0 : 10]$*, window size* $w = 11 \times 11$ *and sample slices of* $d = [1 : 10]$

5.6.2 Results: Sparse vs. ASP 3D Live Wire

First the robustness of Sparse 3D Live Wire to a reduction in input segmentations was assessed. The ability for Sparse Live Wire to segment the full volume, varying slice separation and filter scale, can be seen in Figure 5.11. The first half of Table 5.2 shows the corresponding numerical results from performing Sparse 3D Live Wire on the 10 MR images for a scale of $s = [4, 6, 8]$. Using a single slice with Sparse 3D Live Wire produced zero-area segmentations. This should not be surprising since, in our implementation of Live Wire, the lowest-cost path from point A to point B will be the same as the lowest-cost path from point B to point A. As a single slice in Sparse Live Wire will produce a maximum of two points for each image slice, this will result in a single line, encompassing zero pixels.

Results from applying the ASP Live Wire method to 3D MR images with different window sizes can be seen in Figures 5.12, 5.13, 5.14 and 5.15. The second half of Table 5.2 shows the corresponding numerical results from performing ASP 3D Live Wire on the 10 MR images for a scale of $s = [4, 6, 8]$ and window size $w = 7 \times 7$ pixels.

In MRI, ASP Live Wire outperforms Sparse Live Wire at all levels of assessment for recall, Dice similarity and volume error. In contrast, Sparse Live Wire produces high levels of precision in almost all cases. This is most likely a sign that Sparse Live Wire is consistently under-estimating the boundary of the liver, creating a volume which exists entirely within the manually delineated example. This is potentially because Live Wire will attempt to find the lowest-cost path between points which, in a situation of multiple low-cost paths, will be the shortest. In ASP Live Wire, a large number of points are propagated between slices, regardless of the size of the area to be segmented. In contrast, as the object tapers off at extremities, Sparse Live Wire will have a reduced number of points from which to create a Live Wire contour. This is because fewer of the initialisation slices will be intersecting at this point. This may result in under-constrained contours, which will tend to be smaller than the desired areas due to the cost-minimising nature of Live Wire.

Table 5.3 shows the performance of Sparse and ASP Live Wire on the five example US images, while Figures 5.16 and 5.17 show their respective results over a range of scales and the number of slices used. Results between the two are far closer on these examples, with Sparse Live Wire giving better results in a number of cases. Once again, Sparse Live Wire has better results for precision, but in the case of US it also performs well for volume error. ASP Live Wire may be less effective in these examples due to the difficulty in segmenting the fetal liver manually in US. Since the object boundaries are far more open to interpretation, it is possible that between slices, the contour of one slice is not as close to the next as would be expected. This could cause the core assumption of ASP Live Wire (that deviation of the contour between slices will be small) to break down.

Further development is required to allow ASP Live Wire to function with larger motion between slices. This could be done in a number of ways. First of all, a linear path between known segmentations could be assumed, giving an expected search region for the location of the segmentation boundary, which will be different than simply a window around the last known boundary pixel. Alternatively, or in conjunction with this, a multi-scale approach could be taken. By finding the FA and LO values at a pixel for a number of different scales, a more accurate definition can be formed when trying to locate the same position on the following slice. This can allow for a much larger window or even for point searching at a global level.

Table 5.4: *Significance values for the 3D Live Wire methods in MRI*

S	Measure	ASP Live Wire Result	Sparse LW Result	t-value	p-value
1	Precision (%)	**80.75 ± 19.81**	-	-	-
	Recall (%)	**88.23 ± 15.54**	-	-	-
	Dice Similarity (%)	**81.75 ± 15.82**	-	-	-
	Volume Error (%)	**30.59 ± 83.67**	-	-	-
2	Precision (%)	81.73 ± 17.48	**94.02 ± 10.53**	-7.34	**0.0000**
	Recall (%)	**92.78 ± 8.73**	59.12 ± 36.15	15.89	**0.0000**
	Dice Similarity (%)	**85.28 ± 11.81**	64.54 ± 35.85	9.50	**0.0000**
	Volume Error (%)	**23.28 ± 45.25**	39.55 ± 36.23	-5.70	**0.0001**
3	Precision (%)	87.88 ± 12.79	**93.98 ± 10.09**	-4.03	**0.0012**
	Recall (%)	**91.83 ± 9.53**	72.32 ± 35.83	9.16	**0.0000**
	Dice Similarity (%)	**88.84 ± 9.07**	74.55 ± 34.95	6.81	**0.0000**
	Volume Error (%)	**13.32 ± 27.17**	29.96 ± 35.51	-6.65	**0.0000**
4	Precision (%)	87.51 ± 13.61	**93.23 ± 12.80**	-3.52	**0.0028**
	Recall (%)	**92.87 ± 7.88**	79.82 ± 32.50	6.49	**0.0000**
	Dice Similarity (%)	**89.22 ± 9.17**	80.70 ± 32.04	4.20	**0.0009**
	Volume Error (%)	**12.85 ± 25.14**	19.43 ± 32.04	-2.75	**0.0102**
5	Precision (%)	88.73 ± 13.26	**92.93 ± 12.56**	-2.61	**0.0129**
	Recall (%)	**93.27 ± 7.46**	79.68 ± 33.92	6.68	**0.0000**
	Dice Similarity (%)	**90.23 ± 8.95**	80.16 ± 32.80	4.93	**0.0003**
	Volume Error (%)	**12.59 ± 27.19**	20.19 ± 33.02	-3.10	**0.0057**
6	Precision (%)	89.58 ± 12.33	**93.92 ± 10.95**	-2.84	**0.0087**
	Recall (%)	**93.27 ± 7.24**	82.43 ± 33.25	5.39	**0.0002**
	Dice Similarity (%)	**90.70 ± 8.65**	82.11 ± 32.69	4.22	**0.0009**
	Volume Error (%)	**10.61 ± 25.13**	18.41 ± 32.92	-3.24	**0.0045**
7	Precision (%)	89.54 ± 12.42	**92.96 ± 11.60**	-2.21	**0.0259**
	Recall (%)	**93.73 ± 7.27**	83.07 ± 32.31	5.36	**0.0002**
	Dice Similarity (%)	**90.95 ± 8.87**	82.47 ± 31.58	4.22	**0.0009**
	Volume Error (%)	**10.21 ± 24.43**	18.20 ± 31.96	-3.36	**0.0036**
8	Precision (%)	89.91 ± 12.24	**91.86 ± 16.51**	-1.15	0.1384
	Recall (%)	**93.59 ± 7.60**	79.96 ± 35.84	6.54	**0.0000**
	Dice Similarity (%)	**91.06 ± 8.78**	79.51 ± 35.14	5.51	**0.0001**
	Volume Error (%)	**10.15 ± 24.43**	20.53 ± 35.08	-4.26	**0.0008**
9	Precision (%)	90.82 ± 11.52	**93.34 ± 12.97**	-1.61	*0.0692*
	Recall (%)	**94.45 ± 7.03**	83.48 ± 33.20	5.47	**0.0001**
	Dice Similarity (%)	**92.03 ± 8.31**	82.92 ± 32.41	4.51	**0.0006**
	Volume Error (%)	**8.88 ± 21.31**	16.07 ± 32.19	-3.11	**0.0055**
10	Precision (%)	90.92 ± 11.44	**92.08 ± 18.88**	-0.67	0.2602
	Recall (%)	**94.57 ± 6.81**	83.43 ± 33.54	5.55	**0.0001**
	Dice Similarity (%)	**92.15 ± 8.13**	83.05 ± 32.93	4.49	**0.0006**
	Volume Error (%)	**8.77 ± 20.85**	16.74 ± 32.91	-3.44	**0.0032**

Table 5.5: *Significance values for the 3D Live Wire methods in US*

S	Measure	ASP Live Wire Result	Sparse LW Result	t-value	p-value
1	Precision (%)	**74.59 ± 27.33**	-		
	Recall (%)	**83.94 ± 13.48**	-		
	Dice Similarity (%)	**74.24 ± 20.25**	-		
	Volume Error (%)	**80.82 ± 162.09**	-		
2	Precision (%)	83.22 ± 19.63	**88.98 ± 21.95**	-2.82	**0.0090**
	Recall (%)	**86.35 ± 16.76**	56.28 ± 31.77	13.65	**0.0000**
	Dice Similarity (%)	**81.69 ± 16.38**	63.39 ± 32.62	8.27	**0.0000**
	Volume Error (%)	40.78 ± 100.22	**38.20 ± 30.25**	0.71	0.2457
3	Precision (%)	87.10 ± 16.25	**95.19 ± 7.43**	-5.26	**0.0002**
	Recall (%)	**90.50 ± 9.71**	76.81 ± 27.30	7.12	**0.0000**
	Dice Similarity (%)	**87.42 ± 11.99**	80.85 ± 26.82	3.33	**0.0038**
	Volume Error (%)	21.81 ± 47.89	**21.02 ± 27.53**	0.29	0.3897
4	Precision (%)	86.41 ± 15.67	**94.18 ± 12.29**	-4.65	**0.0005**
	Recall (%)	**91.22 ± 13.57**	87.07 ± 22.77	2.18	0.0273
	Dice Similarity (%)	87.02 ± 13.27	**88.43 ± 22.05**	-0.75	0.2352
	Volume Error (%)	24.28 ± 45.64	**12.86 ± 22.39**	4.38	**0.0007**
5	Precision (%)	91.16 ± 12.11	**96.28 ± 6.04**	-3.80	**0.0017**
	Recall (%)	**90.23 ± 12.06**	88.89 ± 21.59	0.73	0.2409
	Dice Similarity (%)	89.73 ± 10.61	**90.12 ± 20.86**	-0.22	0.4152
	Volume Error (%)	14.42 ± 31.94	**10.15 ± 21.46**	1.85	**0.0472**
6	Precision (%)	90.51 ± 15.01	**96.18 ± 5.71**	-3.94	**0.0014**
	Recall (%)	90.22 ± 9.48	**90.47 ± 20.69**	-0.14	0.4442
	Dice Similarity (%)	89.29 ± 11.13	**90.98 ± 20.47**	-0.95	0.1821
	Volume Error (%)	17.82 ± 38.47	**9.25 ± 20.80**	3.52	**0.0028**
7	Precision (%)	90.95 ± 14.22	**94.75 ± 15.67**	-2.20	**0.0263**
	Recall (%)	**92.09 ± 9.53**	88.55 ± 25.01	1.90	**0.0430**
	Dice Similarity (%)	**90.49 ± 10.69**	89.47 ± 24.24	0.55	0.2986
	Volume Error (%)	16.56 ± 36.11	**11.03 ± 24.54**	2.25	**0.0243**
8	Precision (%)	89.83 ± 14.78	**95.73 ± 9.73**	-3.77	**0.0018**
	Recall (%)	**93.61 ± 8.26**	92.64 ± 18.44	0.59	0.2830
	Dice Similarity (%)	90.72 ± 10.85	**92.72 ± 17.67**	-1.18	0.1318
	Volume Error (%)	16.53 ± 36.36	**8.13 ± 18.81**	3.58	**0.0025**
9	Precision (%)	91.23 ± 12.28	**94.73 ± 14.18**	-2.15	**0.0285**
	Recall (%)	**94.36 ± 8.37**	90.22 ± 24.34	2.29	**0.0225**
	Dice Similarity (%)	**92.11 ± 9.35**	90.24 ± 23.66	1.03	0.1638
	Volume Error (%)	11.91 ± 23.03	**10.41 ± 24.32**	0.69	0.2531
10	Precision (%)	92.91 ± 11.06	**95.83 ± 11.53**	-0.89	0.1964
	Recall (%)	**94.17 ± 8.73**	90.07 ± 24.81	2.24	**0.0246**
	Dice Similarity (%)	**92.94 ± 8.81**	90.34 ± 23.70	1.44	*0.0899*
	Volume Error (%)	10.75 ± 21.71	**10.05 ± 23.94**	0.33	0.3750

5.6.3 Results: Significance Testing

In order to establish whether the differences between Sparse 3D Live Wire and ASP Live Wire are of importance, their mathematical significance must be determined. This has been conducted in the same way as it was in Section 4.5.4. Tables 5.4 and 5.5 give significance values comparing the results from Sparse Live Wire and ASP Live Wire methods, for MRI and US respectively, in order to assess the quality of the improvements described. In both of these, significance values are shown in bold for p-values of indicating a strong statistical significance ($p \leq 0.05$) and italics for p-values of a weak, but still notable significance ($0.05 < p \leq 0.1$).

Although precision values of Sparse Live Wire in MRI were consistently greater than those of ASP Live Wire (Table 5.4), this advantage is lost at higher numbers of slices, when the significance value of the difference rises to above 0.05. In all other MRI cases, ASP Live Wire's improvement over Sparse Live Wire remains firm. In contrast, ASP Live Wire was found to be lagging behind Sparse Live Wire in a number of cases for the US volumes. This may be due to the weakness of boundaries found in many of the US examples, causing the ASP seed points to become 'lost' though successive image slices. Sparse Live Wire, meanwhile, would maintain user-defined inputs throughout the image volume, maintaining a baseline accuracy even when image boundaries are weak. However, Table 5.5 shows that in many cases, the differences between the two methods could not be considered significant, with almost half of the results giving $p > 0.05$. Despite Sparse Live Wire appearing to perform better overall, this shows that the advantage in US is far from conclusive.

Performance Overview: ASP Live Wire

- In the specific application of segmenting e-THRIVE MRI images of the fetal liver, ASP Live Wire appears to segment the liver contour more accurately and with fewer user-defined initialisation contours than Sparse Live Wire.
- In the US examples used, results produced by Sparse and ASP Live Wire were found to be very similar, though Sparse Live Wire arguably performed better overall. This could be due to the large deviation of the segmented object boundary between slices, which is not accounted for by the ASP Live Wire algorithm, and could be considered a potential limitation.
- US has proven to be a difficult modality in which to segment the fetal liver manually, hampering attempts to compare the results between MRI and US volume measurement of the same fetus.

5.7 Discussion

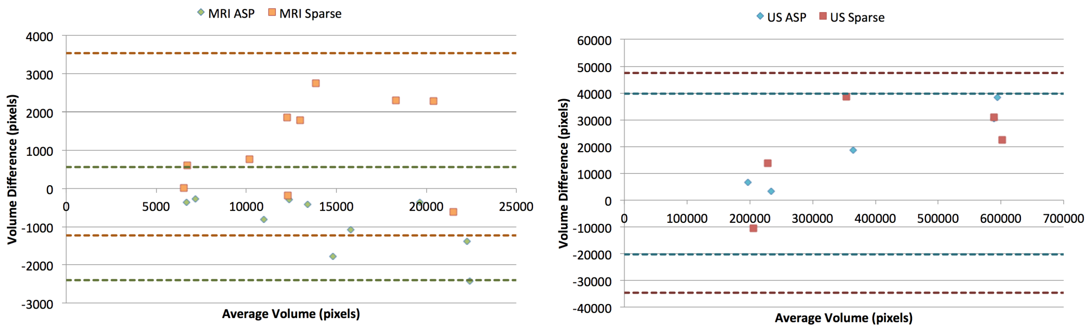
5.7.1 Performance of ASP Live Wire

In this chapter, ASP Live Wire has been introduced as an alternative method for the interactive segmentation of 3D objects. This has then been compared to Sparse Live Wire in both US and MRI scans of the fetal liver. As would be expected, a greater number of user-defined slices results in a greater similarity to manually determined surface for both methods of segmentation. After approximately four segmentations distributed throughout the MRI volume, ASP Live Wire appeared to show diminishing returns in terms of accuracy for the amount of additional user interaction. This could therefore be considered the minimum number of user-delineated boundaries required to produce a fairly accurate 3D surface for this specific acquisition protocol. This suggests that, at this density of slices, the same optimum high-FA path was being followed.

The fluctuation of results as the number of slices being semi-automatically segmented increases could be the result of varying utility of the specific slices used. Since the slices are spread equally between the volume, there may be cases where a particularly difficult to navigate section of boundary is delineated by the user, whereas in other cases it may need to be done automatically, and could fail. This may result in lower accuracy when a larger number of slices are used. The method appears to be robust to changes in window size, with results from the 5×5, 7×7, 9×9 and 11×11 pixel window sizes (9.75-21.45mm) on MRI volumes all producing relatively similar values. As with the results from the density of slices through the volume, this suggests that the same features are being selected as the contour iterates through the surface, regardless of the size of region used to search for features.

ASP Live Wire showed greater agreement with manually defined boundaries than Sparse Live Wire in MRI, with the latter only performing better in terms of precision, though this could have been due to the method consistently under-estimating the volume. Meanwhile, in US, there was not any particular benefit in using ASP Live Wire over Sparse Live Wire. This similarity in results may be due to the large variation in the contour position from one slice to the next, as a direct result of larger slice spacing and liver boundaries in US being far more difficult to localise manually. In this case, having the kind of boundary definition throughout the entire volume that Sparse Live Wire provides may be more beneficial. Further to this, it is possible that the method struggles with some of the challenges presented by US imaging. The presence of speckle could influence the FA and LO results, forcing the propagation of the boundary to move in unexpected directions, though this could be mitigated by appropriate filter scales. Shadowing may cause the method to lose track of the boundary altogether. This latter problem is further reason to look to a global optimisation solution over point-wise matching.

For additional comparison between the two methods, Figure 5.18 shows a set of Bland-Altman plots, showing the difference in volume (in terms of pixels) between each of the



(a) (b)

Figure 5.18: *Bland-Altman plots for the calculation of LVol using both Sparse and ASP Live Wire, comparing each to manual segmentations for* ***(a)*** *MRI and* ***(b)*** *US when ten slices are used.* ± 2 *standard deviations from the mean difference are also shown.*

segmentation methods and the manually segmented volumes [13]. This appears to show greater agreement with the overall liver segmentation for ASP than Sparse Live Wire in both modalities. In the case of ASP Live Wire, the user must define the location of the extremities of the image. This is not a major disruption to the segmentation process, especially as in most instances the extremities of the object were close to the extremities of the image, but this must be viewed as additional user interaction to consider the comparison fair.

5.7.2 Potential Modifications to ASP Live Wire

Tailoring the value of the scale modifier, k, could improve the performance of the segmentation, but it would require an additional, unintuitive input parameter from the user. A potential way of avoiding this could be to base the value of the scale modifier on the size of the segmented area, since this will give an impression of the size of features relative to noise, and thus the filter scales which should be used. It may appear that, given enough sample points taken around the contour, it would be possible to conduct the segmentation process without the aid of any form of Live Wire at all; the points could merely be connected directly or used with such density that a contour forms naturally.

However, there are several problems with this; firstly, using a large number of points can impact on the assumption that the points will remain in the same order after they have been perturbed, thereby adding the potential for error in plotting the contour. Secondly, in the case of using all points in the contour as sample points, a local minima close to the contour could attract a number of different sample points, with the same minima appearing in all of their neighbourhoods. This would result in gaps along the contour, and therefore requires points to be more spaced out. Connecting them directly without the use of Live Wire would be undesirable, as it would likely avoid key features. Although the blob-like structure of the liver may produce similar enough volumes without the use of Live Wire, in the examples that are explored in Chapter 6, it is essential that the shape of the object resembles its appearance in the image as closely as possible.

An alternative direction for this form of segmentation could be to consider the deformation of the boundary as it propagates through the volume as a registration problem. LP measures for similarity have already seen use in US [201] and multimodal [112] registration applications. It is possible that redefining the point re-location as a registration problem could be advantageous, and allow for a global optimisation of the curve rather than an unconnected, point-wise optimisation of a selection of pixels around the contour.

5.7.3 Application of ASP Live Wire

Unfortunately, volume error remained at around 10% throughout the application of ASP Live Wire, which may make the approximation unsuitable for use in clinical practice in its current form. This magnitude of error could easily cause a volume measure of an undersized fetal liver to appear normal-sized or vice-versa. However, ASP Live Wire has shown itself to be a 3D Live Wire extension that is comparable and competitive with an existing method. Further development would be required to bring this method up to the standard required for clinical applications in US. This is the case for both methods.

In its current form, ASP Live Wire could be used as a first-pass for fetal liver size assessment. If a segmented liver is found to be within a certain tolerance of the volume considered under-sized when ASP Live Wire is performed, it could indicate to the clinician that further segmentations are required to hone the result. This would prevent the need

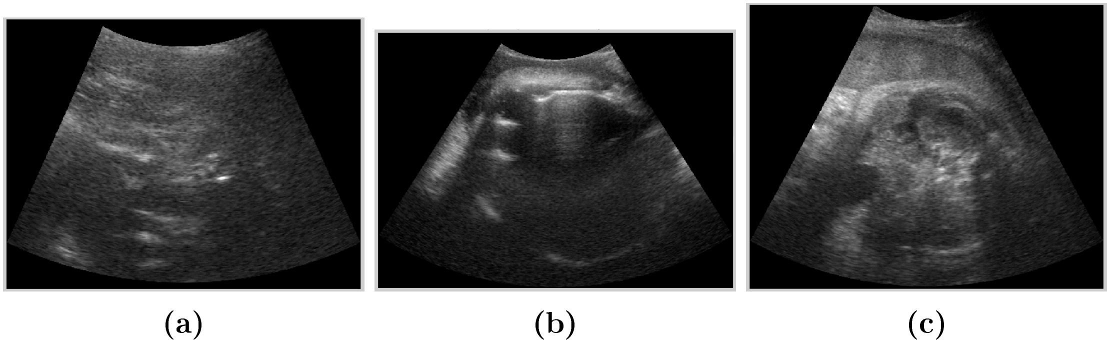


Figure 5.19: *A number of particularly challenging images of the fetal liver in US, as taken from our data set. These prove very difficult to segment even for manual delineation.*

for excessive liver segmentations in cases where the volume is well within a 'safe' value. The tracking of the same image feature between multiple image slices is essentially an image registration problem, where the same image features are tracked between different views or modalities. It is possible that ASP Live Wire could be improved on by drawing upon methods developed for this field.

5.7.4 Liver Volume in US

In order to establish whether the calculation of LVol via US can be obtained automatically or semi-automatically, consideration has to be given as to whether the liver can be reliably segmented from US in the first place. In many of the US examples corresponding to the MRI volumes, it was found that the full volume of the liver was not visible, with attenuation and shadowing often obliterating the appearance of any internal organs.

One of the main problems of viewing the liver in US is that it is an outstandingly difficult region to image effectively, with the abdominal area containing various tissue of very similar appearance. Furthermore, the appearance of the lung relative to the liver changes through gestation, with contrast between the two shifting. Shadowing from the ribs can easily obliterate portions of the image and contrast with surrounding tissue is exceedingly poor. In some cases, judging where to place a boundary manually was found to be guesswork at best, and thus excluded from the analysis. Some examples of particularly poor images of the fetal liver are displayed in Figure 5.19.

Even in situations where the abdominal slice appears clearly, the boundaries can be subjective at best, relying on experience over visual assessment to determine their location. In some cases, the fact that corresponding MRI were available helped in the decision making process which, of course, would not be possible in a clinical environment. It may be that certain age criteria or a set of protocol for acquiring accurate liver segmentations could be developed. In the event of US being used as a modality for visualising the liver on a regular basis, protocols could be developed to ensure that the full fetal liver can be accurately delineated. Alternatively, US technology may require further advancement to allow for accurate fetal liver visualisation.

Beyond being able to prove the efficacy of a segmentation method in a different modality, a significant reason to strive for more accurate images of the fetal liver is to compare the results obtained from MRI data with those from US data, and to determine whether they are congruent. If this is the case, US measurements can be considered to be a viable alternative to MRI for liver volumetry. MRI currently remains an expensive modality, and compared to US, can be a difficult experience for the mother during pregnancy. Currently the modality is reserved for situations where severe complications are suspected in the pregnancy and confirmation is needed. In these situations, calculating liver size for the purposes of monitoring nutritional development would not serve any purpose.

5.7.5 Position in Clinical Workflow

A final consideration is how this method could potentially fit into existing clinical workflow. One of the main objectives of this chapter has been to establish whether measuring the fetal liver in US is viable, as this would allow such measurements to be included in standard prenatal scans. MRI remains an expensive resource, and so should not be considered as standard procedure for monitoring pregnancy, even if it is found to give more accurate volumes.

Unfortunately, at this stage, acquiring suitable US images to be assessed visually is demanding, showing immense variation in the appearance of the abdomen, both between examples and slices in the same volume. A possibility is that, given enough data, patterns in textural information may provide an insight into the location of liver tissue regardless of shadowing, but experience of these images has shown that in many cases shadowing may be too severe and appearance too varied to be able to achieve such a classifier. Additionally, this would require manually delineated segmentations to form the training set, which in some cases is guesswork at best. It may therefore be necessary to establish a protocol specifically to allow for liver images suitable for 3D segmentation for manual and/or semi-automated segmentation. It may be necessary for advances to occur in US technology before accurate liver segmentation can be pursued fully.

The process of ASP Live Wire has been summarised and exhibited as a poster in the following conference: T.M. Rackham, C. Knight, J.A. Noble 'Volumetric segmentation of the fetal liver in 3D US', MECbioeng14, London, 2014

The personal contribution is the methodology for expanding Live Wire to 3D via ASP Live Wire, involving a combination of FA and LO to track object edges between slices. This had not been used before in the context of Live Wire segmentation.

Chapter 6

Fetal Face Analysis in Ultrasound

In this final contributions' chapter a modification of Live Wire is explored, with an aim to form a method suitable for the extraction and quantitative characterisation of fetal facial profiles and surfaces. This is broadly achieved by introducing LO as a cost function, iterating between slices using FA information and using sparse segmentation along one axis to inform segmentation in an orthogonal axis. Finally, methods for modelling the segmented contours and surfaces are explored, and meshing fitting is used to form a parameterised model of the facial surface. The overall goal of this chapter is to allow for the characterisation of FAS via properties of fetal profiles and facial surfaces.

Work in this chapter has been conducted as a part of a collaborative study between *University of Oxford* and *University College London*, in an attempt to extend the latter's existing study into childhood facial dysmorphism to include fetal facial information.

Part of this work involves describing a smooth facial surface as a coarse mesh for use in shape characterisation. This has been conducted in collaboration with Richard Stebbing of *University of Oxford*, who was responsible for the parameterised subdivision surface representation of the face and fitting of this mesh to the segmentation results.

6.1 Face Analysis in Developmental Studies

Section 2.4.4 outlined some of the conditions which have been studied for their tendency to cause facial dysmorphism. Previously, efforts were made to characterise whether faces exhibited signs of FAS from 2D photography using, for example, palpebral fissure length and luminosity to approximate the shape of the philtrum [6]. More recent methods often involve the taking of 3D photographic images of the faces of children and performing analysis via DSM (See Section 2.4.4). This is possible due to the high resolution and model accuracy possible with 3D photography. However, this is clearly only possible for children and adults.

Within the aims of this thesis it is desirable to obtain such information *in utero*, allowing for the diagnosis of these conditions before birth. This is particularly important as some studies have suggested that the physical symptoms of conditions such as FAS can become less apparent over time [179], and so a more accurate diagnosis can be made if it is performed as early as possible. While the damage of FAS may not be reversible, early detection with a more accurate diagnosis can allow for postnatal care planning; for instance, psychiatric treatment to prevent or alleviate the high potential for mental illness presented by those affected by FAS [53].

This means that methods must be designed to extract either the fetal facial profile from 2D US or fetal face surface from 3D US and analysis performed on this data set, in order to draw conclusions about fetal health. This presents the joint problems of obtaining accurate facial surfaces from 3D US images and forming a corresponding model using the decreased resolution, field of view and accuracy that will be inherent in using such data.

At the time of writing, very little research has been conducted into the automatic assessment of facial dysmorphism in fetuses from US images. Visual assessment of fetal US images has been used to identify deformities such as cleft lip and skin tags, but this has generally not progressed past qualitative assessment [94]. In order to deal with the subtleties of facial dysmorphism, quantitative measures for facial shape are required. Specifically dealing with FAS, Matthews *et al.* made a number of quantitative measures on a cohort of 60 control and 60 high-risk fetuses of between 24 and 28 weeks GA, finding that the distance between the eyes and frontal lobe measurements showed the most significant differences between the two groups [108]. However, the ability of these measures to act as a discriminator between the two groups is not expanded on. A full-face approach to the analysis of fetal face dysmorphism may be more appropriate.

6.1.1 Objectives

The main objectives in identifying the use of fetal US in face analysis to identify instances of FAS are:

- To segment the fetal face profile in 2D US images. This is covered in Section 6.2 and is later expanded from a semi-automated to an automated process.

- To expand the segmentation process to 3D face surfaces in US images. This includes finding the density of surface points and the coverage required for specific anatomical features which have particular significance for FAS. This process is developed in Section 6.3.

- To develop a method for to characterising the fetal face shape. The aim is to create a shape model, the parameters of which can potentially be used to determine identify if a fetus is affected by FAS. The method for making constructing a shape model in both the 2D and 3D cases are detailed in Section 6.4. The results of this process are discussed in Section 6.6.2 for 2D and Section 6.6.3 for 3D.

- To find which regions are sufficient to discriminate between fetuses with FAS and those without. One question is whether it is possible to characterise conditions using only a limited portion of the fetal face (such as the philtrum), simplifying analysis and lowering image acquisition requirements. This final question will be left open since, at the time of writing, the FAS risk-level of fetuses used in this study are unknown.

In the solution described in this thesis, the segmentation is largely achieved by modification of the FA-S Live Wire algorithm developed in Chapters 4 and 5, and shape characterisation via the construction of a shape model. Figure 6.1 shows a planned workflow for achieving these aims.

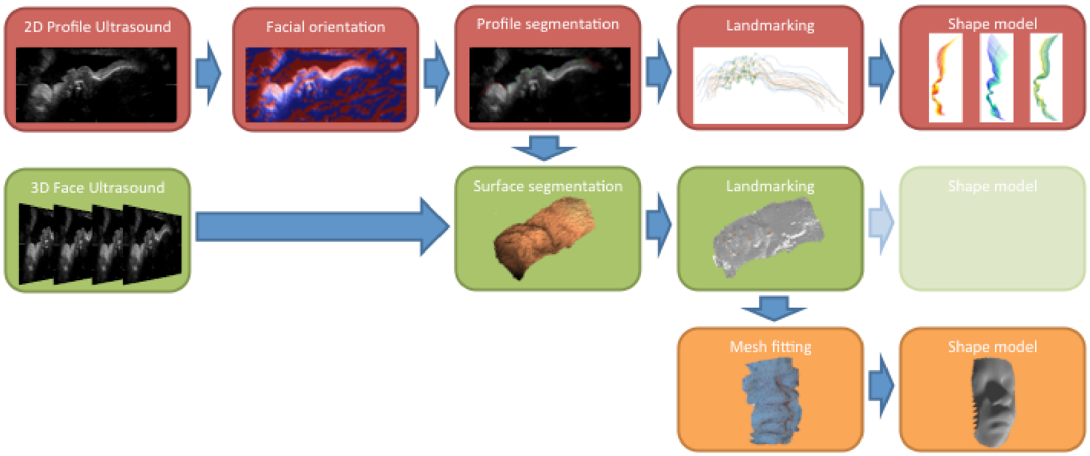


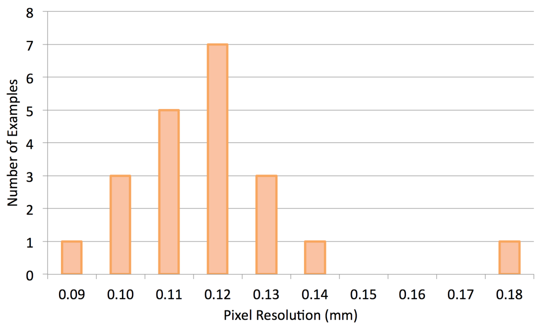
Figure 6.1: *Workflow diagram for fetal face characterisation*

6.1.2 Data

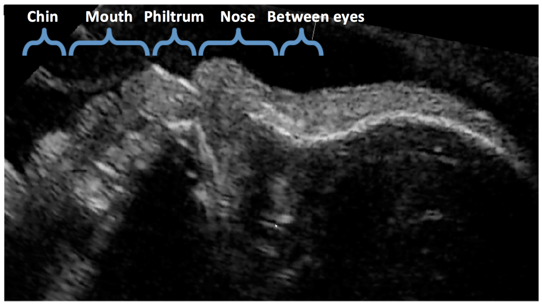
3D US of the fetal face has been obtained through the 'Prenatal Alcohol in SIDS and Stillbirth' (PASS) Network [1]. This is a study following over 10,000 pregnancies in South Africa and America, observing development through pregnancy and for the child's first four years after birth, with FAS as one of their specific targets. While this provides the opportunity to compare the predictions made during pregnancy with predictions made from 3D photographic images after birth (which are easier to acquire), PASS will not release information pertaining to alcohol exposure during pregnancy until after the study has been completed.

The data set nonetheless allows for an opportunity to provide a proof of concept for a method of automating face characterisation. Additionally, since the data set contains both control and FAS affected fetuses, it can be expected that there will be a full spectrum of FAS-related features such as varying philtrum shapes and palpberal fissure lengths. In the absence of a diagnosis, the effect of variation in these features on model parameters can still be observed, which may not be apparent in an all-control or all-FAS data set.

The data acquired for this particular investigation consists of 3D US scans of 25 individuals, all around 27 weeks GA. Kurjak *et al.* suggest that it is broadly possible to construct an accurate model of the fetal face from US images from around 18 to 36 weeks GA, while the study conducted by Matthews *et al.* into FAS included subjects between



(a)



(b)

Figure 6.2: *(**a**) Spatial resolution of the fetal facial profiles used in this thesis (**b**) The key anatomical features of the fetal profile*

24 and 28 weeks GA [94, 108]. This puts our data in an excellent position with regards to facilitating the segmentation of the fetal face.

As the data is of the *.vol* format, conversion to a MATLAB compatible format involved a workaround; conversion from the displayed *.vol* image to an *.avi* video file, which contains one 0.5mm slice per frame. This *.avi* file was then read into MATLAB and converted to a *.mat* volume, representing the US data. However, as the resolution of the resulting image was not the native US resolution, a post-processing method was used to scale the axis correctly once the surfaces had been extracted. Quantitative analysis will therefore be performed on a pixel-wise basis, rather than using absolute measures. The subsequent pixel resolution of the facial profiles is shown in Figure 6.2a.

The quality of the images was generally high, with little shadowing occluding the facial profile. Figure 6.2b shows the key anatomical features of the fetal profile. One problem to note is that, due to the GA and limitations imposed by the sweep of the 3D US scan, while the 2D facial profiles appear clearly and well within the limits of the image boundaries, the extremities of the face for the 3D facial surface were restricted. Many investigations into facial characterisation use the outer extremities of the eyes as landmarks and take into account the full surface of the face. However, in many of the examples used here, the outer edges of the eye were not necessarily visible, with severe image degradation or even total lack of coverage in these regions.

Additionally, there were cases where limbs were pressed against the fetus's face. While this provides an additional challenge to the segmentation process (which is possible to overcome with the methods described here) a greater obstacle was formed by the possibility of the facial features being squashed or distorted. Exactly how this impacts on facial characterisation is explored later in the chapter. It should be noted that these US scans were not optimised for this study. The implications of this are examined in the discussion section.

6.2 Fetal Profile Segmentation

The first stage of characterising the fetal facial profile was to acquire a set of contours describing the profile for a number of different fetuses. This can be obtained manually but, as has been established throughout this thesis, manual segmentation presents a time consuming and user-dependent problem. Semi-automated methods such as FA-S Live Wire can provide a useful tool in these circumstances in order to reduce the amount of user interaction required to perform an appropriate segmentation. However, since the fetal profile across a narrow age band allows for a number of assumptions to be made, which are simply not possible for liver and subcutaneous fat measurements, it could be possible to entirely automate this process with some modifications.

6.2.1 FA-O Live Wire

Chapter 4 introduced the concept of FA-S Live Wire, utilising a circular shape constraint and FA as cost functions to guide the Live Wire segmentation. In this example, the shape constraint introduced by FA-S Live Wire is not appropriate, since the boundary lies on a generally linear path rather than a circular one. The FA aspect of FA-S Live Wire remains relevant, but other constraints which consider the properties of this particular segmentation problem can be used in order to be successful.

In the case of fetal face segmentation, the surface is, broadly speaking, on a single, straight plane without significant folds or overlaps. This means that it is possible to make assumptions of where the facial surface lies based on the orientation of the boundaries present. Recapping the methodology explained in Chapter 4, if monogenic signal of a 2D image can be expressed as,

$$f_M(x) = I_b + [(o_1 \otimes I_b), (o_2 \otimes I_b)] \tag{6.1}$$

where I_b is the bandpass filtered version of the original image, I, o_1 and o_2 are odd orthogonal filters and $\otimes$ denotes the convolution operation, the LO is given as,

$$\theta = tan^{-1}\left(\frac{o_1 \otimes I_b}{o_2 \otimes I_b}\right), \tag{6.2}$$

giving a value between $-\pi$ and π radians [9, 112]. The cost function for Live Wire can then be redefined with the LO in mind, turning this equation into,

$$C = w_{FA}C_{FA} + w_\theta C_\theta \tag{6.3}$$

where C_{FA} is the FA-based cost function, C_θ is the LO-based cost function, and w_{FA} and w_θ are the weightings of the FA and LO costs respectively. In order to favour 'upwards facing' boundaries, the LO cost can be defined as,

$$C_\theta = \begin{cases} 1 & if \quad -\pi < \theta_{mod} < 0 \\ 0 & if \quad 0 \leq \theta_{mod} \leq \pi \end{cases} \tag{6.4}$$

where θ_{mod} is the LO of the image, rotated such that the profile lies along $\theta_{mod} = 0$. By assigning a high cost to regions where the boundary is downwards-facing, irrelevant boundaries are generally excluded. The greatest difficulty of approaching the segmentation of fetal profiles in this way occurs when the fetus is positioned such that it is in contact with the wall of the uterus or has its limbs drawn towards the face. At this point, the LO becomes a less reliable metric (the user may be required to change C_θ for $\theta \geq 0$ to a value slightly above zero), but FA allows these contours to be found nonetheless.

With the modification of this form of Live Wire involving the use of FA and LO as

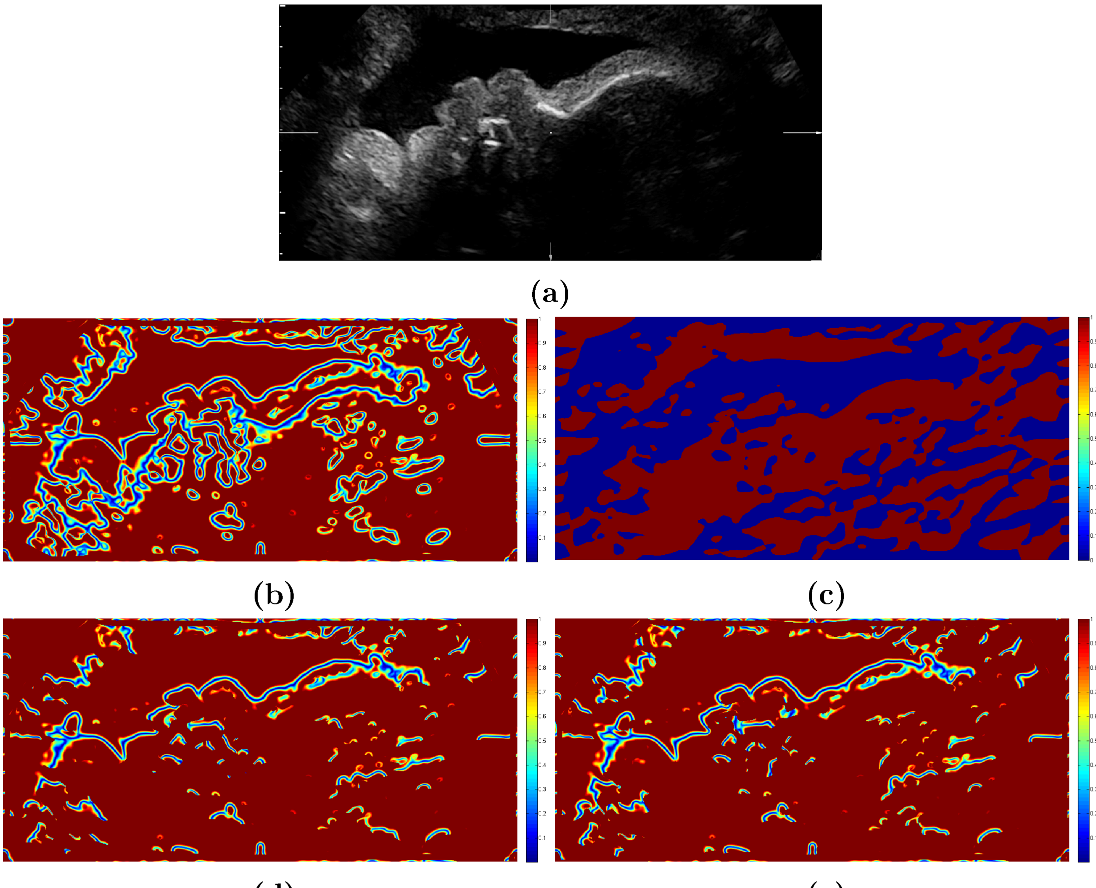


Figure 6.3: *(**a**) Fetal profile in US (**b**) FA of the image (**c**) Thresholded LO of the image (**d**) Cost map C, created by multiplying the thresholded image by the FA image (**e**) Cost map C when θ_{mod} is used to define the threshold*

the costs, this method will be referred to as FA-O Live Wire. Figure 6.3 shows the actual process involved in finding the path with FA-O Live Wire; in this case the user inputs are two initial seed points and the angle between those points. As in FA-S Live Wire, the FA of the image is found (Figure 6.3b), but in this case the LO is also calculated. The LO is then thresholded at $\theta = 0$ as shown in Figure 6.3c. The cost C derived from a combination of FA and the LO image is shown in Figure 6.3d. Figure 6.3e shows the cost when the θ is altered to θ_{mod} by taking into account the orientation of the fetal face. How θ_{mod} is obtained will be described in the following sections.

In order to use LO effectively as a cost for segmenting the facial profile, there are two possible approaches:

- Use the LO of existing segments of the Live Wire segmentation in order to inform the possible values of LO for the next segment.

- Perform pre-processing on the image in order to rotate the profile to a common plane, against which the expected orientation of the facial profile can be assumed.

The first option would operate much in the same way as the shape constraint for FA-S Live Wire, altering the expected value for LO as the user selects new segments of the boundary. However, in many cases it was found that it was possible to segment the boundary using only two anchor points, which results in only one segment, making the first option redundant. Once the LO restriction is added, many more of the examples are capable of finding the correct boundary using only an anchor point at either end of the facial profile. This means that it would be more useful to use a pre-processing method in order to orientate the face.

It is possible to use FA-O Live Wire as it is, as a part of a semi-automated segmentation method for the extraction of the fetal profile that requires minimal user interaction. This method involves the simple pre-processing step of the user drawing a line from one extreme of the face to the other (from just below the chin to the top of the head). This step has two influences on the segmentation process; firstly, it allows the user to explicitly define the angle by which the image should be rotated and secondly, it gives the FA-O Live Wire algorithm a start and end point by which to operate. After this line has been defined, the image is rotated and the extremities of the line are used as the two seed points for the FA-O Live Wire algorithm, giving a full fetal face surface. However, it is possible to expand this to a fully automated method.

6.2.2 Automatic Profile Segmentation

It can be beneficial to form an automated approach to fetal profile segmentation to allow for a seamless inclusion into standard clinical practice. Automation could allow for fetal profile characterisation to be part of the scanning software rather than requiring additional analysis by clinicians. Additionally, this could be useful in the collection and analysis of larger data sets, since a wide selection of faces will be required for model construction to be successful.

A number of assumptions can be made about fetal face images which can be used to aid automation. US images of fetal faces will generally already be facing towards the probe, since the skull would attenuate the facial signal too heavily if the fetus is facing away. When the correct plane is chosen, the skull will often be the brightest part of the image and the facial contour will be visible for the entire length.

However, there are cases which do not obey the assumptions made in the automated method, either because of unusual fetal position, a poor facial boundary or that a fetal limb appears in the view. The latter provides a potential acoustic boundary that could interfere with the facial surface and could also present an additional high-intensity ridge or occluding the boundary of the fetal face. In these cases, the semi-automated version can be used.

Bone Extraction

In order to find a suitable angle through which to rotate the face automatically, it is important to identify relevant, repeatable landmarks and use them to align the profiles. One such feature which occurs consistently on the fetal face profile is the highly reflective section of bone on the brow ridge. Since bone forms such a strong acoustic boundary, it can be expected to have the highest intensity value of the region being observed. This form of feature can be detected using the Local Energy (LE), which is another image property that can be estimated from the monogenic signal and can be defined as;

$$LE = \sqrt{{I_b}^2 + (o_1 \otimes I_b)^2 + (o_2 \otimes I_b)^2} \tag{6.5}$$

where definitions are the same as those used in Equation 6.1 [112]. The LE gives a high response where the magnitude of the response for all filters (the even filter and both orthogonal odd filters) is high. It is important to note that the orthogonal filters have the potential to cancel out if the responses in either direction are not both positive or both negative.

Detection using this method does, however, rely on certain aspects of the image. The skull appears with a high intensity due to;

- bone presenting a barrier of high acoustic impedance compared to surrounding tissue.
- the bone often being perpendicular to the incoming signal, resulting in a reflection directly back to the probe.
- the skull being in such a position that there is often little attenuation before the US signal reaches it.

Therefore, the detection could fail if the skull is in a significantly different orientation or if a highly attenuating object (such as a fetal arm bone) is positioned between the skull and the probe. In these cases, manual positioning of the facial contour may be required.

However, in order to ensure the correct section is selected, LE alone will not be enough; Figure 6.4a shows an example where there are other regions, such as the philtrum, which are also highly reflective. In order to select regions which are more likely to be associated with the bone ridge, FS can be used, as described by Kovesi *et al.* [91]. Since it will give a high response for strong ridges, and bone will tend to have a very strong but thin boundary in US, this will be better represented as a ridge than an edge. Recall from Chapter 4 that FA can be expressed as,

$$FA = \frac{1}{N}\sum_{s}\frac{\lfloor |odd_s| - |even_s| - T_s \rfloor}{\sqrt{even_s^2 + odd_s^2 + \varepsilon}}, \tag{6.6}$$

where s is the filter scale, N is the number of scales used to form the FA image, ε is a small constant (typically around 0.01) to prevent division by zero, *even* is the bandpass filtered image, *odd* is the image convolved by two orthogonal anti-symmetrical filters, $\lfloor \cdot \rfloor$ represents an operation which sets negative values to zero, and T_s is a noise-controlling

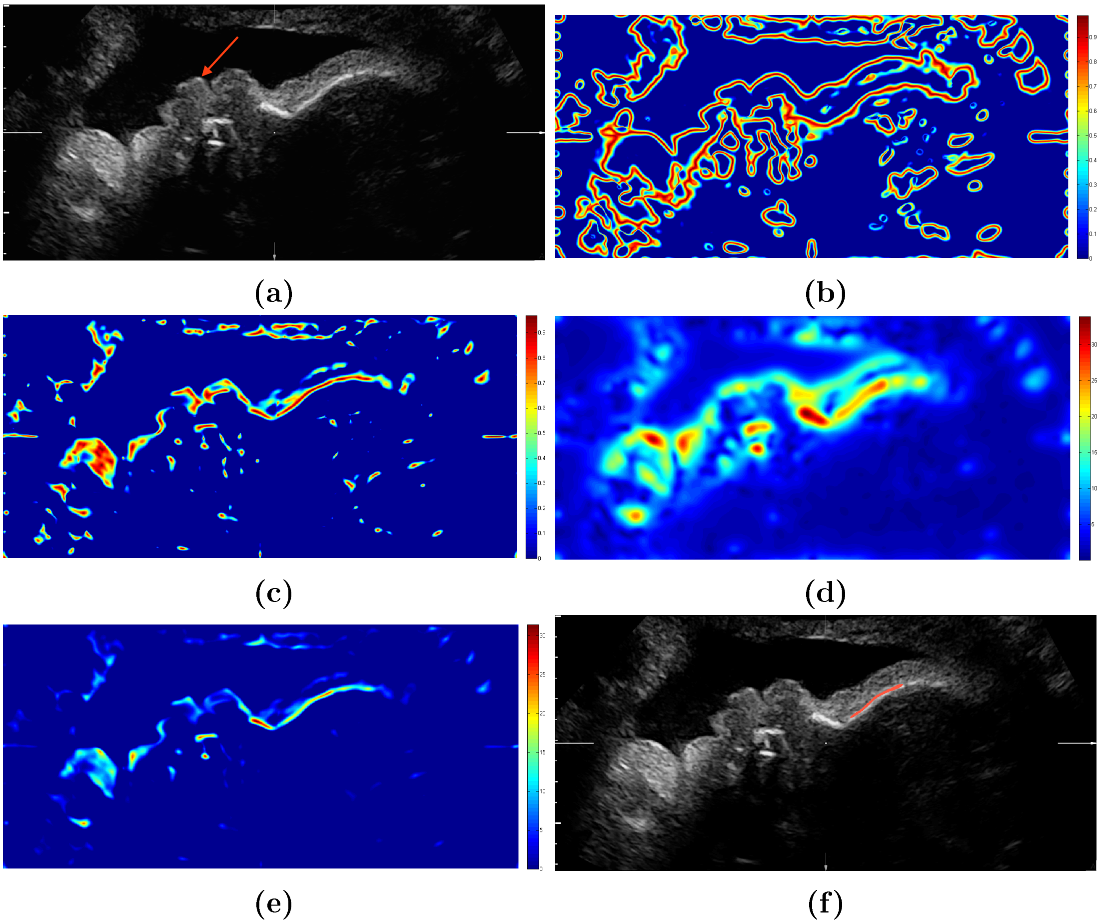


Figure 6.4: *(**a**) An US image of a fetal face profile. The highly reflective philtrum edge is highlighted (**b**) FA response, highlighting the edge-like structures of the image (s = [30,35,40]) (**c**) FS response, highlighting the ridge-like structures of the image (s = [30,35,40]) (**d**) LE of the image (**e**) LE multiplied by FS, identifying high intensity ridges (**f**) The largest connected section*

threshold value [91, 155]. FS is, in contrast, described as;

$$FS = \frac{1}{N}\sum_{s}\frac{\lfloor |even_s| - |odd_s| - T_s \rfloor}{\sqrt{even_s^2 + odd_s^2} + \varepsilon}, \tag{6.7}$$

so that the value of FS is high when the even filter response is high and the odd filter response is low [91]. This results in a response which is high for ridges but not for edges. Figures 6.4b and 6.4c illustrate the FA and FS responses of Figure 6.4a respectively. Figure 6.4e shows the result of multiplying the LE by the FS. This can further be refined by thresholding and then considering the largest connected structure in the image, as shown in Figure 6.4f.

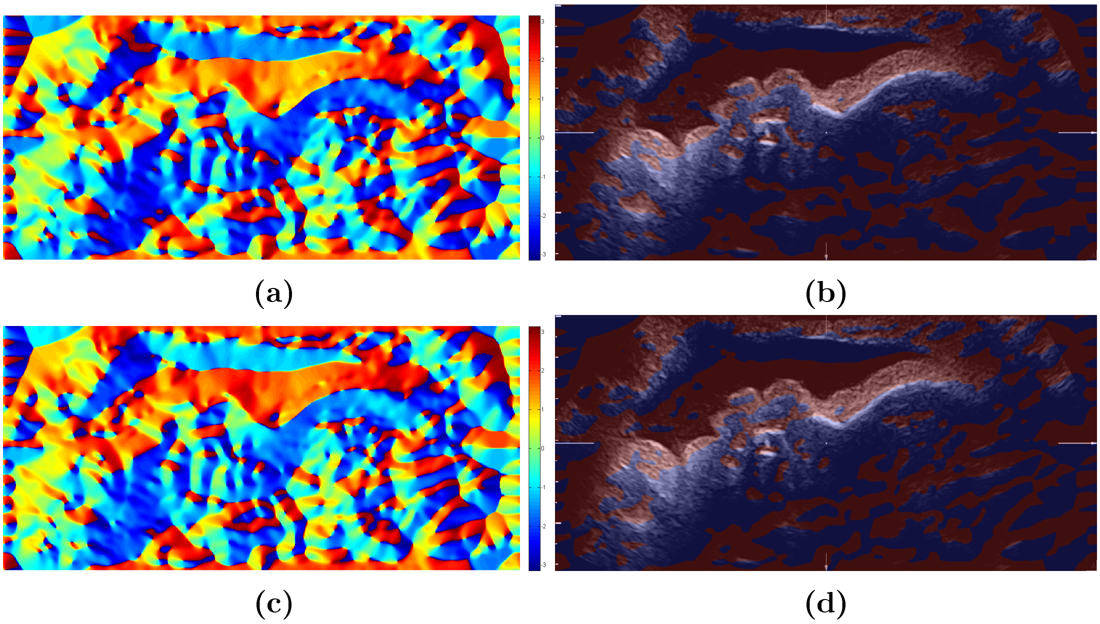


Figure 6.5: *(a) The LO of the fetal profile as it appears normally (b) $\theta > 0$ for the standard LO overlaid on the fetal profile (c) LO modified by the fetal profile orientation (d) $\theta_{mod} > 0$ for the modified LO overlaid on the fetal profile*

Bone Orientation

Once the bone ridge has been segmented, the orientation of this section can be calculated and from this, the LO-based costs derived. Since an edge that runs perfectly along the x axis will have a high odd filter response in the y direction, this hypothetical edge will have a LO of $\theta = \pi/2$. Any edge where $\pi > \theta > 0$ can be considered upwards-facing. Recall Equation 5.3, which found θ_{mod} for a previous LO value φ. If φ is taken to be the angle between the facial profile's primary axis and the x axis, positive values of θ_{mod} will represent upwards facing sections of fetal profile. As the fetal face will most likely be bounding amniotic fluid, which will appear as a dark region on the US image, and is itself reflective and therefore light in appearance, then the assumption can be made that the fetal profile will tend to display an upwards facing section. Any downwards facing edge-like feature can be disregarded as part of the profile. Figure 6.5 shows the effect of thresholding $\theta > 0$ compared to $\theta_{mod} > 0$ on an example of a facial profile.

There are, however, a number of different ways to determine the value φ associated with the bone segment. Initially, the principal orientation of the binary mask for the bone was used to determine the orientation of the fetal facial profile. While this generally produces an appropriate rotation for the image, one of the disadvantages of this method is that the length of visible bone can vary from example to example, with the result that some segmented sections of bone do not reliably represent the angle by which the profile should be rotated.

Another potential method for determining an appropriate rotation angle which is less susceptible to varying amounts of visible bone ridge is to use the Hough Transform to find the predominant straight edge. Since the variable sections of bone ridge tend to either curve up to the top of the head or curve towards the nose, the common feature of the ridge is most frequently the flat section between these points of curvature. By using a Hough Transform to identify the position and angle of this flat section of bone, an accurate rotation can be achieved more reliably.

The Hough Transform is a method used to find structures such as lines in an image by means of defining straight lines by an angle (θ) and a radius (ρ) from the origin, then identifying the number of pixels which owe membership to each line, represented in the feature space $[\theta, \rho]$ [49]. When the values assigned to $[\theta, \rho]$ are over a threshold, they can be said to represent lines in the image. In this case, the point at which $[\theta, \rho]$ is the highest is taken and θ can be used directly to alter the orientation of the image. In the calculation of this angle, the MATLAB implementation of the Hough Transform is utilised.

A final alternative for considering the orientation of the fetal face is to find the LO associated with the segmented bone structure. As the mean value would be an unrepresentative metric for angular values (since $\theta = \pi$ and $\theta = -\pi$ represent the same angle, and this would not be accurately represented by a mean), the median value in this region is considered. The results section of this chapter will consider the results obtained by these three different automated methods of calculating facial orientation.

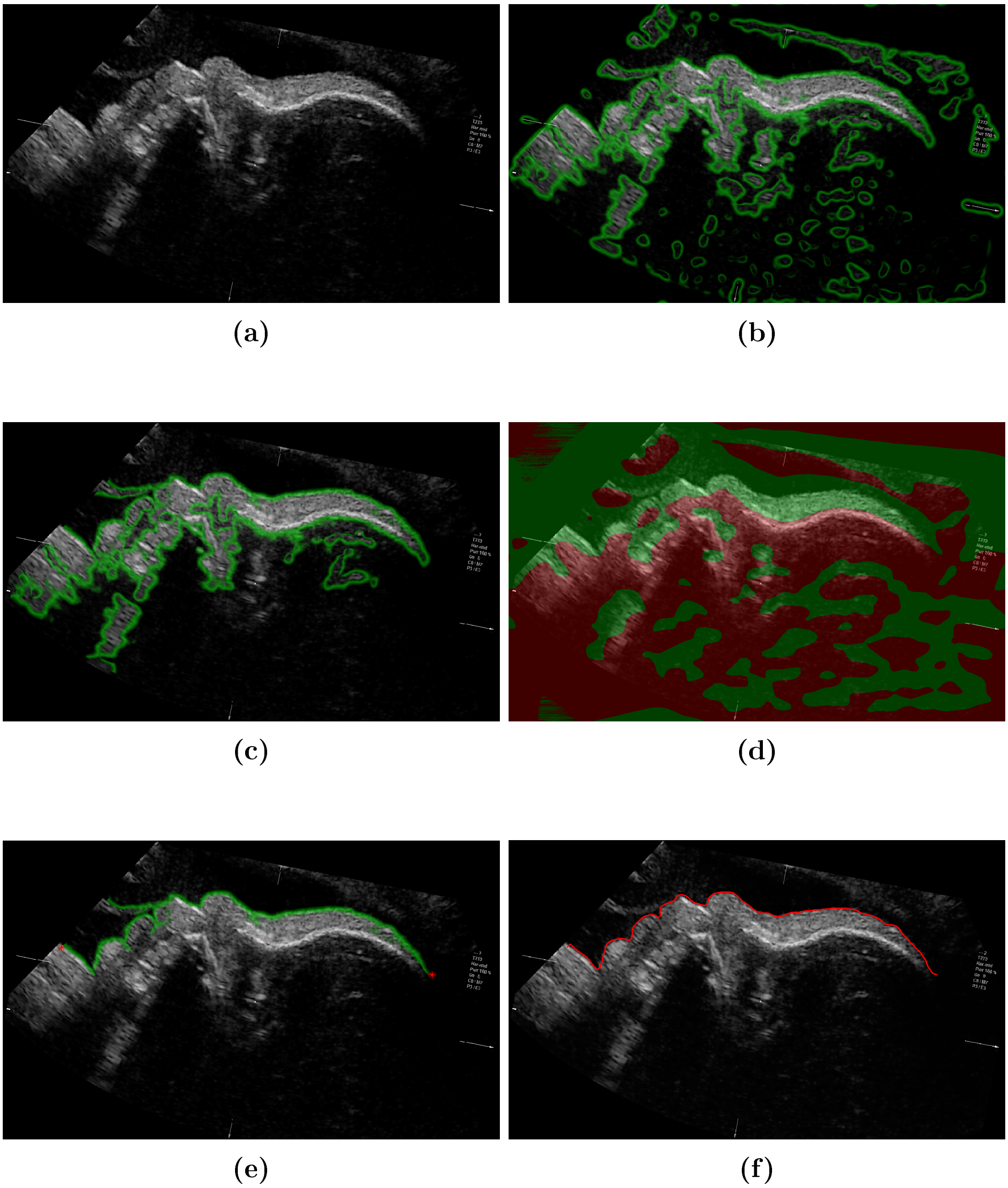


Figure 6.6: *The process by which the fetal profile can be acquired automatically. For ease of visualisation, the image has been manually rotated so that the facial profile lies horizontally.* ***(a)*** *Original US image of a fetal face profile* ***(b)*** *Fetal facial profile with FA at s = [40,45,50] overlaid* ***(c)*** *The largest object of connected FA>0 regions is isolated* ***(d)*** *LO: $\theta_{mod} > 0$ is in green while $\theta_{mod} < 0$ is in red* ***(e)*** *Upwards facing section of the FA object are retained* ***(f)*** *FA-S Live Wire is then applied to the image*

Automated Fetal Profile Segmentation

Now that LO has been set up as a cost function for Live Wire and the image has been automatically rotated such that the face is in a consistent orientation, it is possible to automate the segmentation of the fetal profile.

Figure 6.6 shows how the properties of the facial profile can be further exploited in order to facilitate its segmentation. Due to the stage in gestation at which the data acquired for this study has been taken, certain assumptions can be made about the appearance of the fetal face; primarily that it will be the largest feature of this form of US image. While Figure 6.6b shows the FA response of the image, Figure 6.6c shows only the largest connected component, thereby reducing the problem.

In addition, the known orientation of the face can be used to further reduce the area of interest; Figure 6.6d positive (green) and negative (red) values of LO. In this set of images, angle correction is achieved by physically rotating the image, while the actual implementation corrects the angle by subtracting the facial angle from the LO image. Figure 6.6e shows the FA, thresholded by the LO. The extremities of this object will be the start and end points of the FA-O Live Wire algorithm. Figure 6.6f shows the resulting profile segmentation, while Figure 6.7 shows some of the segmentations obtained by this method on other examples of fetal face profiles.

In a preliminary test of the automated method, 17 out of a set of 20 images used could be segmented using the Hough-based automated tool. The main obstacle to automation is that the presence of fetal limbs in the central imaging plane can also interfere with the assumptions made regarding the extremities of the face. Figure 6.8 shows two examples where the adjoining tissue causes the automated method to fail by incorrectly assigning end points. These examples also used the non-Hough based alignment; note that they have not been rotated fully.

This performance could potentially be improved by adjustment of the different filtering parameters for the different steps of the automated procedure, but this would arguably

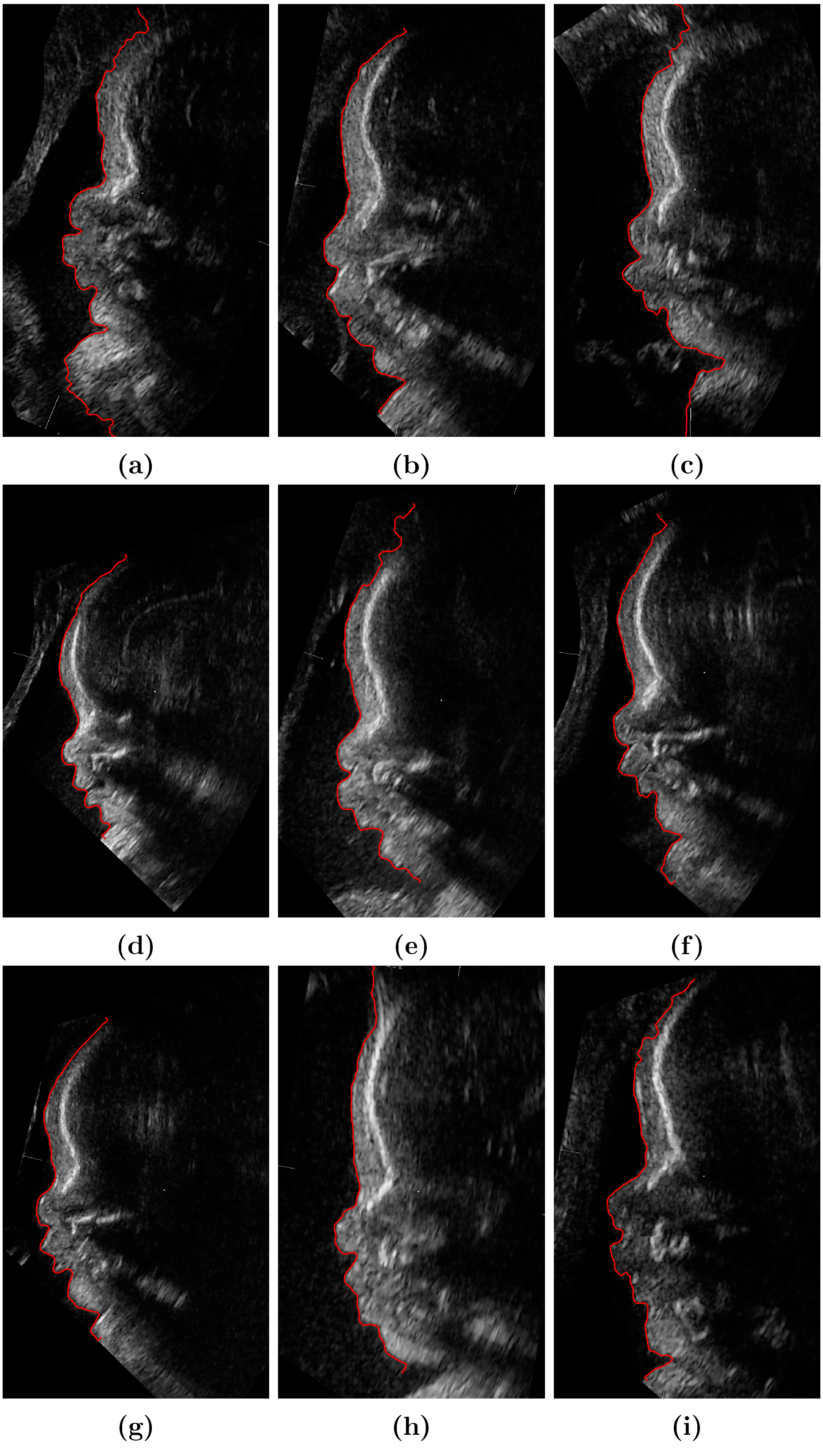
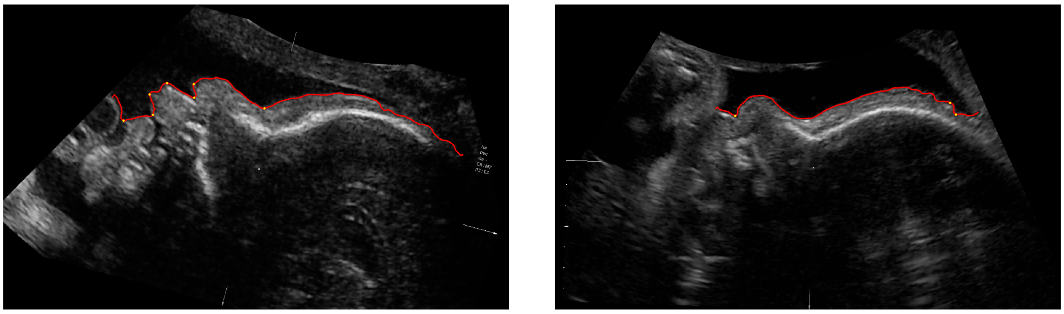


Figure 6.7: *Examples of successful, fully automated fetal profile segmentation. These have been deemed successful as the segmentations include the full facial profile and coincide with visual assessment of the boundary.*



(a) (b)

Figure 6.8: *Examples where the automated segmentation method has not functioned correctly*

require more effort from the user than the interactive method. All of these images were successfully segmented with the user-defined rotation and end points.

6.3 Surface Acquisition

In order to allow for a relevant comparison between current work in facial dysmorphism to identify FAS, it is necessary to expand the work described here from a 2D metric based on the fetal face profile to a 3D metric based on the entire facial surface. This is also significant since it can be difficult at times to accurately obtain the fetal profile; Kurjak *et al.* describe some of the difficulties encountered when attempting to find the 2D mid-line of the fetal face, since small deviations can still appear to represent the profile but may contain different information (particularly in diagnosing conditions such as cleft lip) [94]. Forming a model for the full facial surface of the fetus could avoid this form of problem.

Table 6.1: *Number of surface points and landmarks used by prior investigations into face analysis*

Condition	Landmarks	Surface Points	Reference
Wolf-Hirschhorn syndrome	22	-	[71]
Fragile X syndrome	24	28,000	[74]
FAS	24	25,000+	[179]
Cornelia De Lange syndrome	21	16,000	[10]
Fabry disease	25	25,000	[45]

There are a number of challenges presented by using a surface segmented from 3D US, rather than the standard method of using 3D photography. Table 6.1 shows the number of landmarks and surface points used for previous studies into modelling face shapes. Studies involved in characterising faces of children and adults tend to use around 25,000 points. Aside from restrictions of resolution caused by the imaging modality, the ages of subjects in many of these investigations range up to around 20 years, meaning that faces will be larger and with more spaced out features than that of a fetal face.

Additionally, it is important to observe the position and the number of landmarks used to align these surfaces accurately, which are independent of the aforementioned discrepancies between 3D photography and 3D US. The largest concern in this respect is the field of view available in 3D US, which may not be wide enough to view certain facial features. In the data acquired for this study, the 3D volumes obtained are generally centered on the facial mid-line, but it is not always possible to see the outer extremities of the eyes (a standard landmark used in these investigations).

The greatest challenge in the characterisation of fetal facial profiles is the segmentation of the facial surface. In the expansion of these methods to the segmentation of 3D facial surfaces, it is important to automate as much of the surface acquisition phase as possible. Even using tools like FA-S or FA-O Live Wire to segment each slice can be incredibly time consuming, as each face has in the order of 100 slices. It is, however, possible to expand on the methods described here to guide the segmentation to cover the full facial surface.

6.3.1 3D Live Wire Segmentation

As outlined in Chapter 5, a method of using Live Wire in the context of 3D segmentations has been discussed by Poon *et al.*, in which sparse Live Wire segmentations along two orthogonal axis are used to guide a dense Live Wire segmentation in the third axis [143]. The points at which the prior segmentations intersect the planes along the third axis are used as automatically selected anchor points to guide the final Live Wire segmentation. Arguably, the most complex aspect of this work is how to order these points most effectively to complete the segmentation.

As previously explored, it is possible to use this approach to create a full dense segmentation of the fetal face using a sparse set of segmentations. Since the objective is to segment a generally planar surface rather than an arbitrary volume, segmentations along just one axis may be sufficient to segment the full facial surface. Using such a method on a plane rather than a volume has the further advantage of possessing a very simple solution to the point-ordering problem, as segmentations will already be ordered along one axis. In this case, FA-O Live Wire can be used to perform the segmentation.

6.3.2 ASP Live Wire

A simple method of forming a 3D surface such as the fetal face (which is approximately planar) is to use the FA of adjacent slices to guess the location of the next contour. Schenk *et al.* use the shape of the previous slice to inform the next in their 3D application of Live Wire [160]. This method largely operates from the assumption of the shape of the fetal face surface. Since the face should be facing in an expected direction after the alignment conducted in the preprocessing step, variation in the contour on a slice-by-slice basis should be small and roughly perpendicular to the contour.

As with all uses of FA, scale selection is an important factor in achieving the appropriate results. Figure 6.9 shows the examples of the different Live Wire paths taken using only the extremities of the face as seed points, showing the different paths taken depending on the scale chosen. Similarly, the peak FA value in the neighbourhood will be altered by this value.

An initial implementation of contour propagation involved taking all contour points to the adjacent slice and perturbing them in a neighbourhood restricted to the y axis, taking the highest FA point as the new contour pixel location. The main problem with using FA alone to form a segmentation in this way is that it can be prone to error, particularly if two contours come close together in the image; if a set of pixels are following the wrong surface at one plane, they will continue to follow that surface in subsequent planes, resulting in a poor segmentation. A corrective step can be to filter the contour, removing

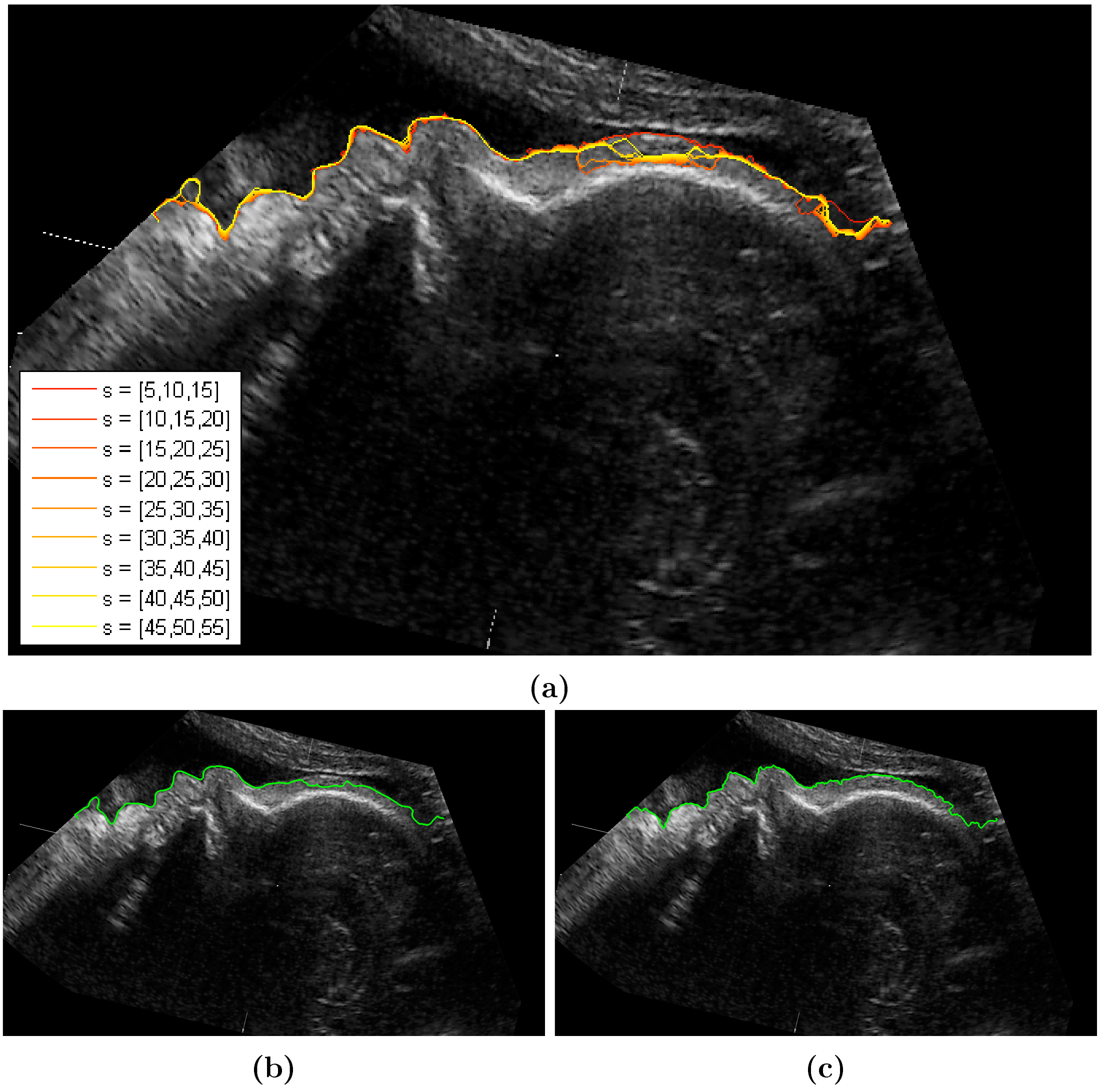


Figure 6.9: *(a) The variation of fetal profile segmentation result with varying filter scale (b) Higher extreme of filter scale, showing a smoother contour but deviation to the stronger boundary formed by the skull (c) Lower extreme of the filter scale, showing a rougher contour but following the correct boundary*

any obviously erroneous pixel relocations, but this can distort the facial shape which is highly undesirable in our particular application.

Chapter 5 introduced ASP Live Wire; a method of segmenting 3D objects using Live Wire in which seed points for the Live Wire algorithm are automatically placed in adjacent slices using a set of points from the initial contour, offset to a location which best represents structurally equivalent points. This uses a combination of FA and LO to localise a subset of points, which are then used to draw an FA Live Wire contour. This has a number of advantages over the previous method, most predominantly that it is robust against being distracted by nearby but incorrect contours and it preserves the smoothness of the contour without artificially filtering the segmentation.

6.3.3 Mesh Shape Model

Approaching the modelling of the fetal face in 3D from a dense point cloud may not be the best method for accurate comparisons between faces since there is no explicit anatomical equivalence in the position of points. A more useful representation of the fetal face for the purposes of forming a shape model could be as a deformation of a common mesh, thereby allowing variation in shape to be expressed more directly.

Initially used in computer graphics work, the field of subdivision surfaces is now being applied to segmentation problems as a form of surface representation [27, 175]. The subdivision surface method, in basic terms, involves using a coarse mesh to define a smooth underlying surface, which is obtained by refining the initial surface in an iterative manner [174]. Stebbing has worked on the use of subdivision surfaces to aid the problem of segmentation in echocardiography [175].

A surface can therefore be represented as an underlying control mesh for an essentially continuous surface, much in the same way a b-spline surface is represented by a set of points in 2D. In 3D, a coarse control mesh can be refined to define a smooth surface by means of Doo-Sabin [48], Catmull-Clark [27] or Loop [104] subdivision.

Catmull-Clark subdivision has been utilised by Chandrashekara *et al.* for the 3D segmentation of the left ventricle in MR images, while Doo-Sabin subdivision surfaces have been applied by Orderud and Rabben in the same segmentation problem but in echocardiography [28, 137]. Outside of medical imaging, Loop subdivision surfaces have been applied to creating 3D models of hands from 2D images [183]. The 'subdivision limit position' (i.e. the surface to which the mesh converges) has also been used as a means of representing a set of data points by Suzuki *et al.* [180]. The Loop-based subdivision surface fitting method developed by Stebbing has seen application in the field of US segmentation of the fetal cranium [121, 175].

The benefit of representing the facial surfaces in this way is that it is now possible to form an explicit definition of surface features. This means that all points are now inherently aligned to the same anatomical feature across examples. Also, since the problem has been reduced to around 300 points (and is therefore of a similar dimensionality to the profile contours), the method for model construction used for the 2D profile will be easily translatable to the meshed faces.

It should be noted that it is possible to conduct the entire segmentation process using subdivision surfaces, as has been conducted in previous studies. However, this requires the appropriate initialisation of the control mesh to the US volume, and so the coarse mesh needs to be aligned properly to the facial surface by selecting a set of landmarks corresponding to mesh positions. For simple shapes such as the left ventricle and fetal head, this form of initialisation is possible. A more complex object such as the fetal face may present a greater challenge since, as noted previously, it can be difficult to localise the exact position of facial features from 2D slices. Aligning the mesh to a pre-segmented surface allows for landmarks to be selected with far greater ease, as the user is able to observe the full facial surface when selecting these points. Additionally, this maintains the use of highly intuitive, interactive segmentation at the core of fetal health assessment.

The creation of the surface mesh representation of the fetal face has been conducted as a collaboration with Richard Stebbing, also from *University of Oxford*.

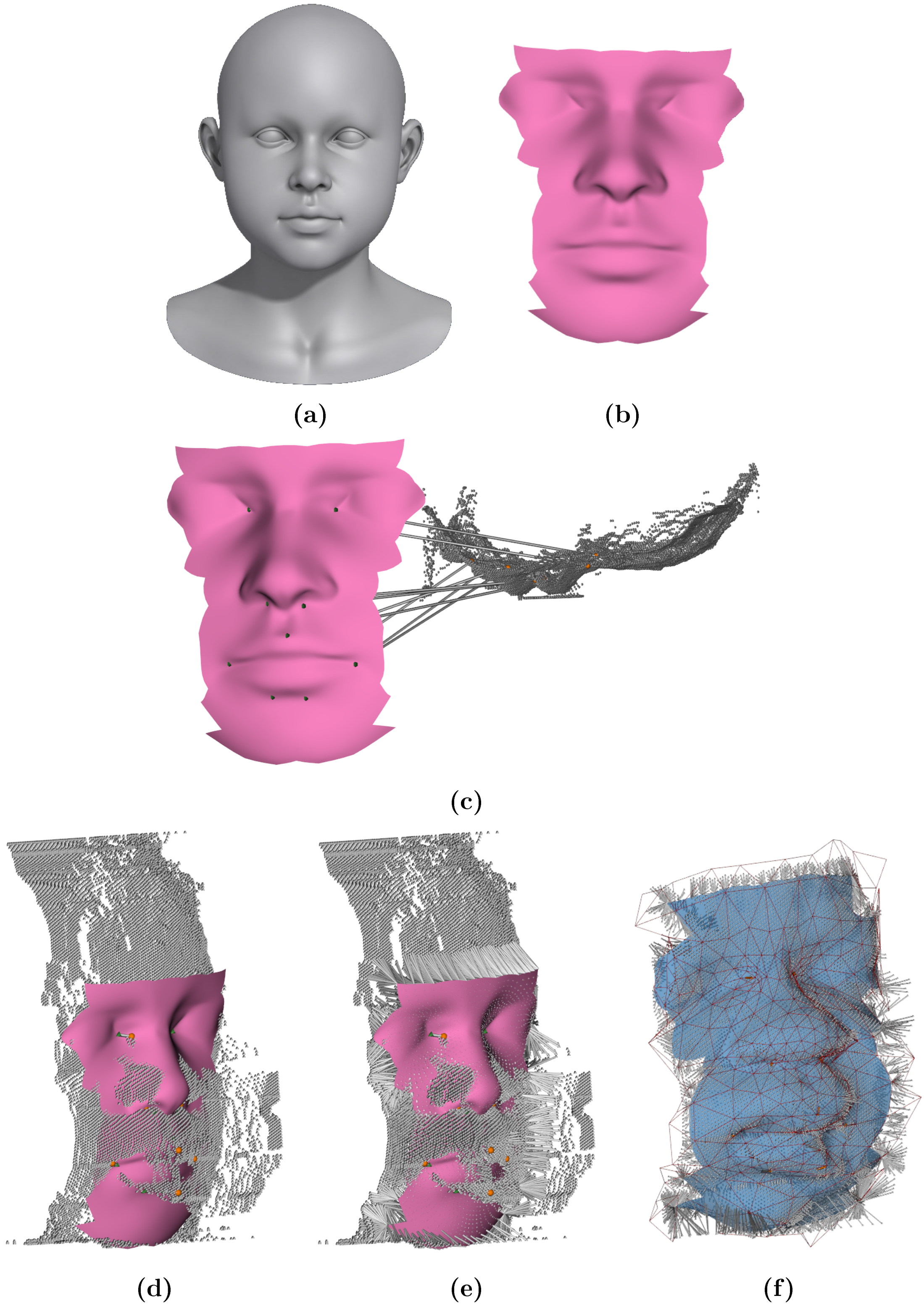


Figure 6.10: ***(a)*** *Original stock surface model used to create the initial face surface* ***(b)*** *Cropped and simplified surface to be used as a base for the face surface* ***(c)*** *Correspondence between landmark locations on the initial surface and an example segmentation of a fetal face* ***(d)*** *Alignment of the initial surface to the point cloud* ***(e)*** *Influence on the point cloud data on the mesh nodes defining the initial surface* ***(f)*** *Final deformed surface, shown with underlying mesh*

Subdivision surfaces are described in more detail in Appendix Chapter B.

The Stebbing implementation of Loop subdivision surfaces is used to fit a control mesh to the point clouds generated by the initial 3D fetal face surface segmentation, in a similar manner to his previous work in echocardiography [175]. Figure 6.10 shows the various steps required to form a mesh representation of the segmented fetal face surface. These are as follows:

1. **Figure 6.10a**: In order to create such a mesh, an initial, generic surface must be created and fitted to each example. The initial face mesh is created by modifying a stock media face model from the Turbo Squid 3D model archives, as shown [12].

2. **Figure 6.10b**: The surface has been cropped to the region of interest, containing all of the important features for accurate facial characterisation. This surface is defined by an underlying mesh of around 300 control points. It has also been modified to contain only triangular faces, as is required by the Loop subdivision process. Less important regions such as the cheeks have been omitted to constrain the number of parameters. This surface will be used as the initialisation for all facial surfaces.

3. **Figure 6.10c**: For each point cloud defining a segmentation of a fetal face, nine landmarks are defined and matched to their corresponding positions on the initialisation surface.

4. **Figure 6.10d**: The initialisation surface is aligned to the segmentation surface as closely as possible via the landmarks.

5. **Figure 6.10e**: The shape fitting process is described by Stebbing *et al.* [175] and can be summarised as follows; a set of candidate points are selected uniformly along the segmented surface and each node of the mesh is matched to the closest candidate point perpendicular to the surface. The Levenberg-Marquardt algorithm, a combination of gradient descent and Gauss-Newton minimisation [61], is used to minimise

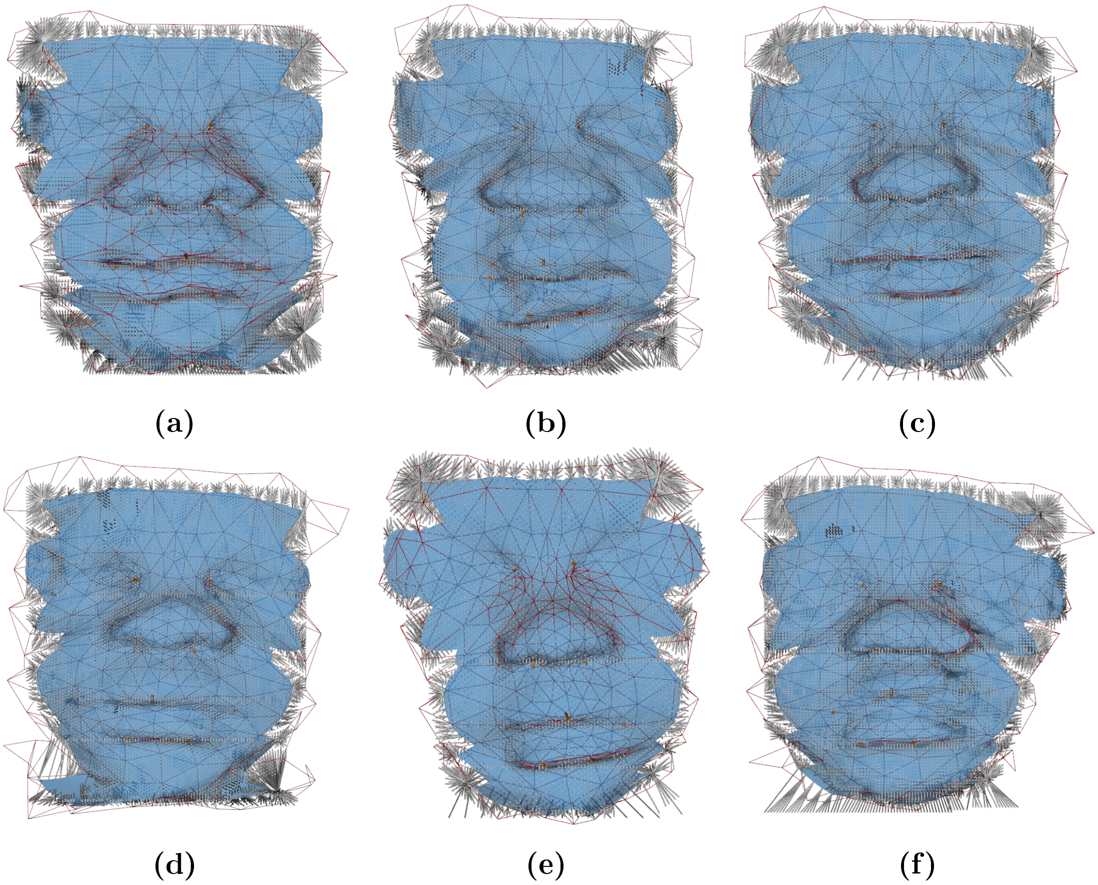


Figure 6.11: *Examples of coarse meshes and the surfaces they represent, fitted to data obtained from fetal face segmentations*

an energy function. This function includes the sum of all orientation, translation and scale errors between the candidate and mesh surface points, a term to limit folding and intersection of the surface with itself and thin-plate regularisation to ensure that the final surface is smooth [175]. Correspondence between candidate and mesh surface points are not fixed between iterations, and a Huber loss function is used to reduce the influence of poorly selected boundary candidates [175].

6. **Figure 6.10f**: When the solution has converged, the final face surface appears as shown. The facial surface is shown in blue while the underlying mesh defining this surface is shown in red.

Steps 4-6 are repeated for all other surfaces until each has a mesh representation. Once the mesh representation for the faces have been obtained, it is possible to obtain a dense surface of arbitrary resolution. Figure 6.11 shows a number of examples resulting from this process.

6.4 Fetal Face Shape Characterisation

6.4.1 Shape Model Construction

In order to characterise the segmented curves and surfaces, the next step is to form some kind of parameterised model which can be used to describe each example and, more importantly, form a metric for comparing them against each other. One potential way of doing this is to characterise the segmentations using an ASM.

It is important to note that the aim here is not to use ASM for the segmentation process itself. This is because the potential modes that the shape constraint will take are defined by the input shapes, which means that instances of abnormal faces may not be segmented as such, as they fall outside of the expected variance of face shapes as defined by the model. This is of particular concern in this study as, at the time of writing, it is not known whether or not individual profiles are from fetuses with FAS; if a model were constructed using part of the data set to train and part to test, it is entirely possible that the FAS cases will either be skewed into one group, or that they will be too small a presence within the training group to make a difference in testing. In order to avoid this, the processes of shape characterisation and segmentation are separated entirely, allowing for the segmentation of completely arbitrary facial shapes.

Cootes *et al.* outline how an ASM model can be constructed, and most of the mathematics described in the model construction section of this chapter will follow their described methods [41].

Annotation

The first stage in constructing a shape model is to align all of the data sets to each other. This is done by taking a set of common landmarks which appear in each data set and aligning these landmarks across all examples, transforming the data solely by translation, rotation and scale. This ensures that the deviation between each example is made up of purely structural features.

For the 2D curves, landmarks were chosen to ensure that they are easily identifiable across different face shapes. Due to variation in mouth position and shape, some structures are not easily identified between examples. It is also important to ensure that the philtrum is aligned in all images. This is one of the most important features of the profile in identifying FAS [179]. Therefore, it is important that this section in particular is in alignment across examples of the fetal face, meaning that landmarks should be positioned to ensure this. As a result, the landmarks used for the fetal face profiles were:

1. Just under the nose at the top of the philtrum.

2. At the lower edge of the philtrum, on the boundary with the upper lip.

3. At the point where the two lips join (although in cases where the mouth is open slightly, this is the middle of the gap between the lips).

4. At the lower edge of the lower lip.

The selection of these landmarks can be applied manually, but could be expanded into an automated form, selecting landmark points based on common, objective properties. In this work, manual landmark selection was used, but the possibility of automation can be explored.

One of the defining features of these landmarks is that they all occur at high curvature locations of the contour. This means that it could be possible to select n landmarks by selecting the locations that centre on the n^{th} highest curvature points. One way of approximating curvature at a point is to consider the angle created by a triangle formed at a point on the curve and two points k values either side of the point in question.

Consider a contour is defined as $\mathbf{x} = [\mathbf{i}_1, \mathbf{i}_2, ...\mathbf{i}_m]^T$ where $\mathbf{i} = [x_i, y_i]$ and m is the number of points in the curve. The cosine law can be used to find the length of a side on an arbitrary triangle, and is of the form;

$$a^2 = b^2 + c^2 - 2bc\cos\theta, \tag{6.8}$$

where a, b and c are side lengths of the triangle and θ is the angle between sides b and c [78]. This can be rearranged to find the angle θ, forming the equation;

$$\theta = \cos^{-1}\left(\frac{b^2 + c^2 - a^2}{2bc}\right). \tag{6.9}$$

The curvature at the N^{th} point in the curve can then be approximated using 6.9 with the values;

$$a = \sqrt{(x_{N+k} - x_{N-k})^2 + (y_{N+k} - y_{N-k})^2} \tag{6.10}$$

$$b = \sqrt{(x_N - x_{N-k})^2 + (y_N - y_{N-k})^2} \tag{6.11}$$

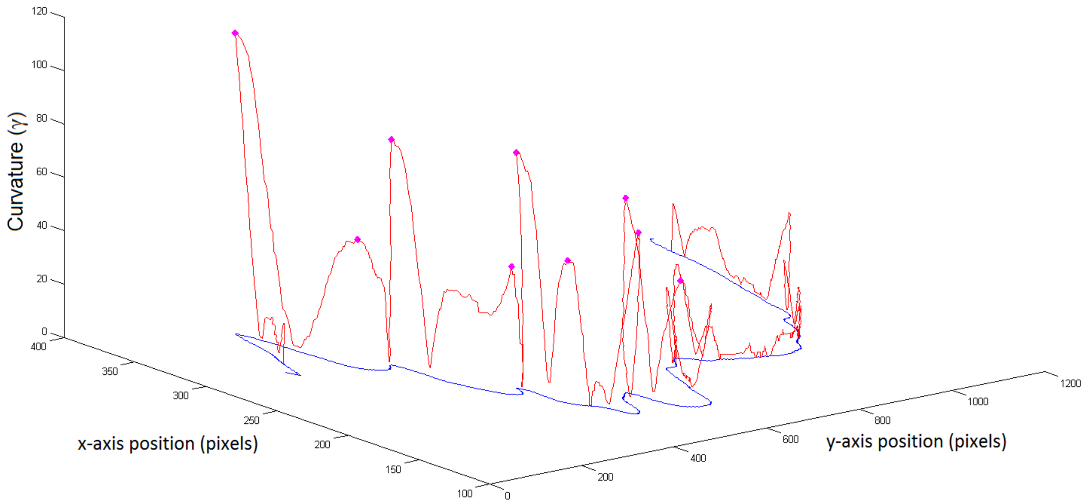
$$c = \sqrt{(x_{N+k} - x_N)^2 + (y_{N+k} - y_N)^2}, \tag{6.12}$$

where k defines the window size used to determine curvature (the full window size will be $2k + 1$). The selection of k is important in the final result as too small a value will take into account curvature likely to be associated with noise, whereas too large a value may miss vital landmarks.

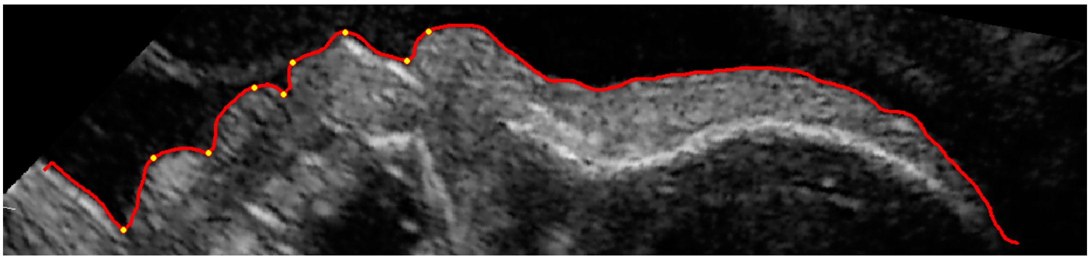
Figure 6.12 shows how the peak values of the curvature, γ, defined as;

$$\gamma = 180 - \theta, \tag{6.13}$$

vary along the facial profile. The highest curvature points of the contour $\mathbf{x}$ can then be identified by finding the peak values of γ. More importantly, it can be seen that all of the concave-curvature peaks (from the perspective of the face as 'inside' and the surrounding area as 'outside') coincide with the landmarks selected earlier in this chapter. The only discrepancy in this respect between the landmarks shown in Figure 6.12 is the point shown on the eye. This was disregarded as a landmark due to the possibility of a wide



(a)



(b)

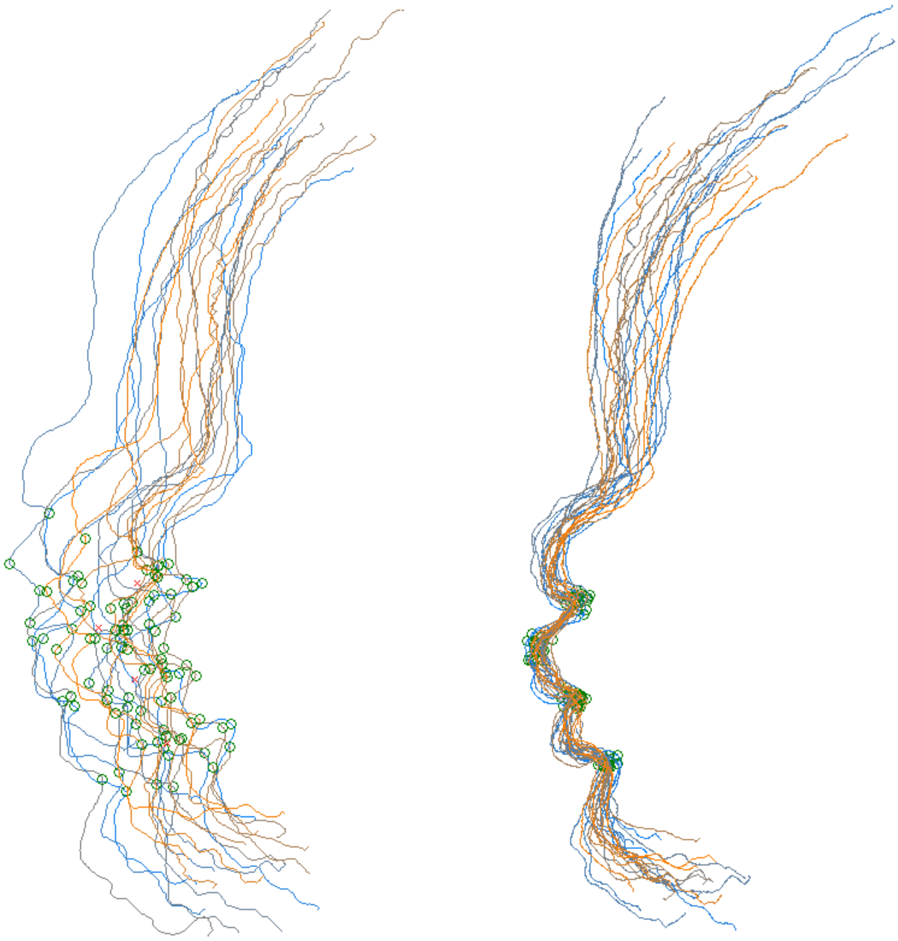
Figure 6.12: *(a) The curvature of a fetal face profile, defined as $\gamma = 180 - \theta$. In this example $k = 5$. (b) The peak curvature points, as defined in (a), marked on the original contour as it appears on the fetal facial profile.*

variation in judgment concerning its location.

In the case of creating a model for the 3D meshes of the fetal face, specific nodes on the original model have been deformed to fit anatomically-based landmarks. This means that each of the nodes on the mesh are already on approximately anatomically equivalent points of the face and additional landmarking to align between meshes is not required.

Shape Alignment

The task of aligning each of the face contours and surfaces is achieved by means of Procrustes analysis [65]; this is conducted by first aligning each set of landmark points to the average landmark locations (averaged across all examples) and applying the resulting transformation to each respective facial contour/surface. A new average landmark

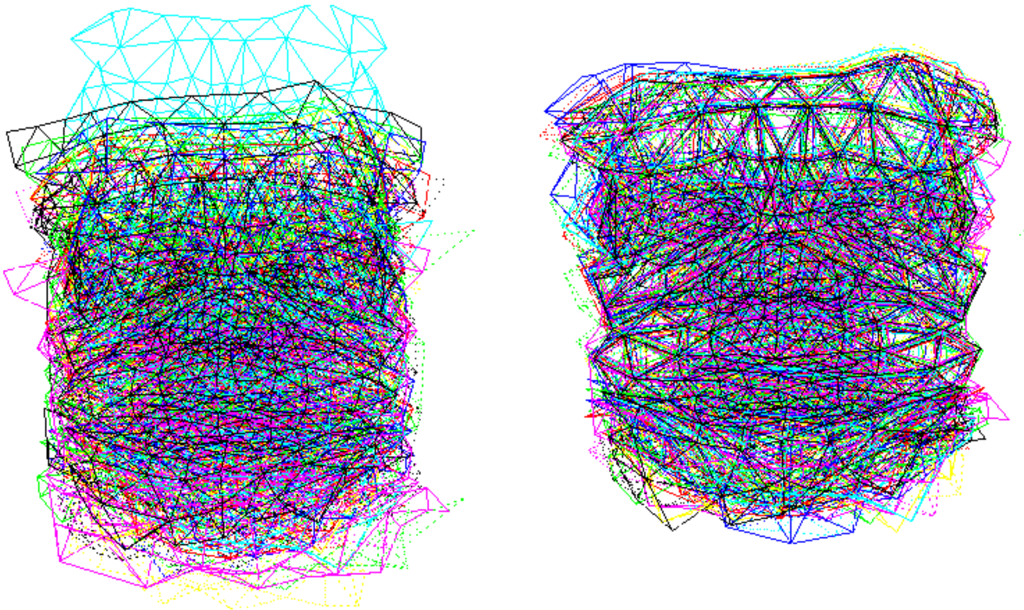


(a) **(b)**

Figure 6.13: *Facial profiles* ***(a)*** *before and* ***(b)*** *after alignment*

location is found and this process repeated iteratively until the differences in positions between the mean and examples converge. In the case of the facial surfaces, all nodes on the facial mesh are used as landmarks. The MATLAB function for Procrustes analysis was used in all cases described here.

Figure 6.13 shows the set of 2D contours extracted from fetal face profiles before and after alignment. From this, it can be seen that the most vital regions for identifying FAS have been successfully aligned; primarily the philtrum and lips. Regions that deviate greatly are the chin and forehead. However, this could in itself be a useful metric for identifying dysmorphism, since the curvature of the forehead will be greater for smaller heads where the face is comparatively larger and flatter for larger heads. The results of



(a) **(b)**

Figure 6.14: *Facial face meshes **(a)** before and **(b)** after alignment*

this process for the 3D surface meshes can be seen in in Figure 6.14.

Model Construction

Once the examples have been aligned, the model can be constructed. Cootes *et al.* describe the formation of an ASM in the following terms [41]. In 2D, a shape can be defined as $\mathbf{x}$, where $\mathbf{x} = (x_1, x_2, ..., x_n, y_1, y_2, ..., y_n)^T$ and $[x_n, y_n]$ is the coordinate location of the nth point in the shape. The mean shape is therefore defined as;

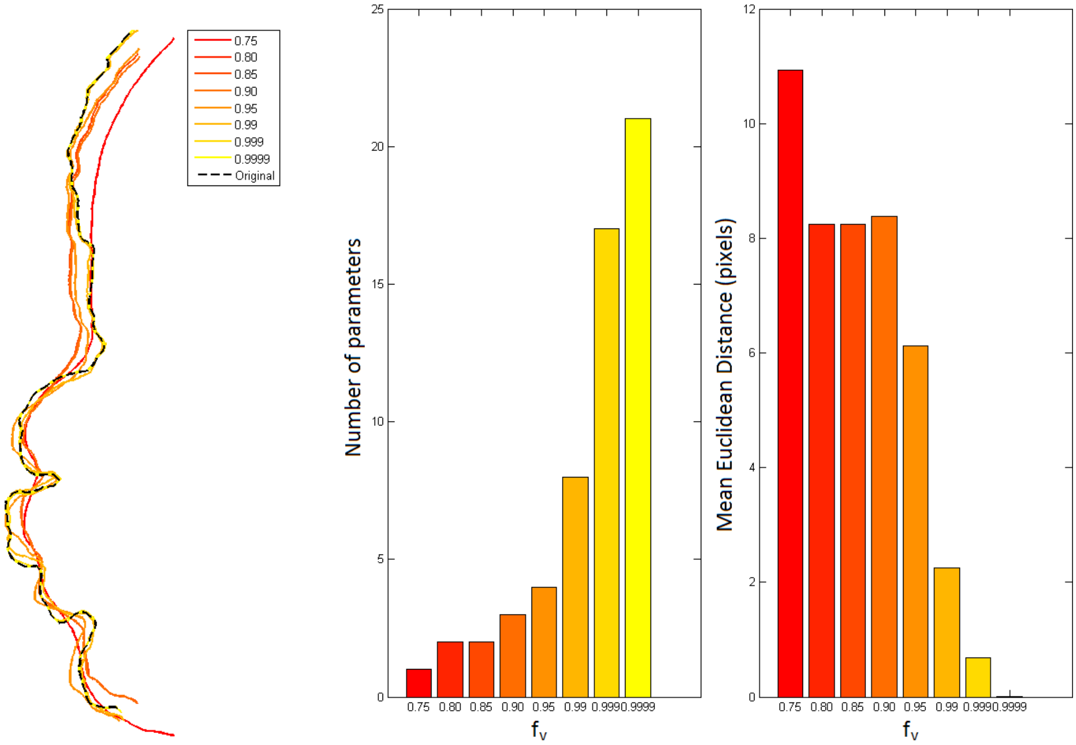
$$\bar{\mathbf{x}} = \frac{1}{s} \sum_{s}^{i=1} \mathbf{x_i} \tag{6.14}$$

Any shape can then be defined as a deviation from the mean, resulting in;

$$\mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{Pb} \tag{6.15}$$

where;

$$\mathbf{b} = \mathbf{P}^T(\mathbf{x} - \bar{\mathbf{x}}) \tag{6.16}$$



(a) (b) (c)

Figure 6.15: *(a) Resulting ASMs (b) Number of parameters required (c) Mean Euclidean distance*

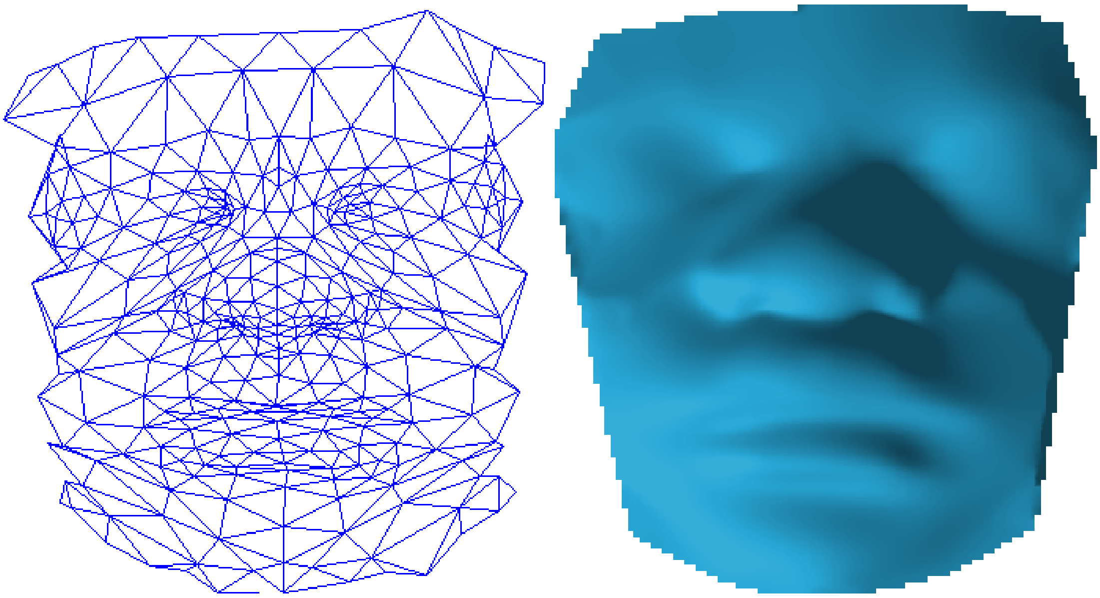
$\mathbf{P}$ is defined as $\mathbf{P} = (\mathbf{p}_1|\mathbf{p}_2|...|\mathbf{p}_t)$ where the values p_n are the eigenvectors of the covariance, $\mathbf{S}$, of the set of curves. This, in turn, can be defined as;

$$\mathbf{S} = \frac{1}{s-1}\sum_{s}^{i=1}(\mathbf{x_i} - \bar{\mathbf{x}})(\mathbf{x_i} - \bar{\mathbf{x}})^T \tag{6.17}$$

Eigenvectors to be used for the model are chosen by selecting the t largest corresponding eigenvalues (defined here as λ_i) such that;

$$\sum_{i=1}^{t}\lambda_i \geq f_v V_T \tag{6.18}$$

where $V_T = \sum_t \lambda i$ and $0 \leq f_v \leq 1$, describing the proportion of the variance that the model is intending to describe. Figure 6.15 shows the effect of varying f_v on both the appearance and accuracy of the approximated 2D curve x as defined by Equation 6.15, as well as the corresponding length of the vector $\mathbf{P_t}$. As would be expected, using a larger



(a) **(b)**

Figure 6.16: *This figure shows **(a)** the average mesh according to point location and **(b)** the corresponding surface defined by this mesh*

value of f_v results in a more accurate curve, but it is important not to use too high a value, or the model can become over-fitted and respond poorly to unseen examples.

At this point, $\bar{x}$ and $\mathbf{P}$ are known values and $\mathbf{b}$, which will be t dimensional, can be determined via Equation 6.16. For any given set of curves $\mathbf{x}$, which must be sampled to be the same length as $\bar{x}$, $\mathbf{b}$ can be found and compared for each of these values. This can provide a quantitative assessment of their shape, since this value gives the degree to which the curve is represented by each of the principal eigenvectors defined by the model. It is these values which can be used to identify principal differences between each group.

If the goal were to create a model for the full 3D facial surface before meshing, the 2D shape model would need to be expanded. This is conceptually simple, but has an inherent flaw when it comes to the actual application of this method. The size of the covariance matrix used in ASM will be $(nD)^2$, where n is the dimensionality of the data and D is the number of data points. For the 2D example, where only points on the profile are used, this figure remains manageable, with $n = 2$, $D \approx 400$ and therefore $(nD)^2 \approx 640,000$.

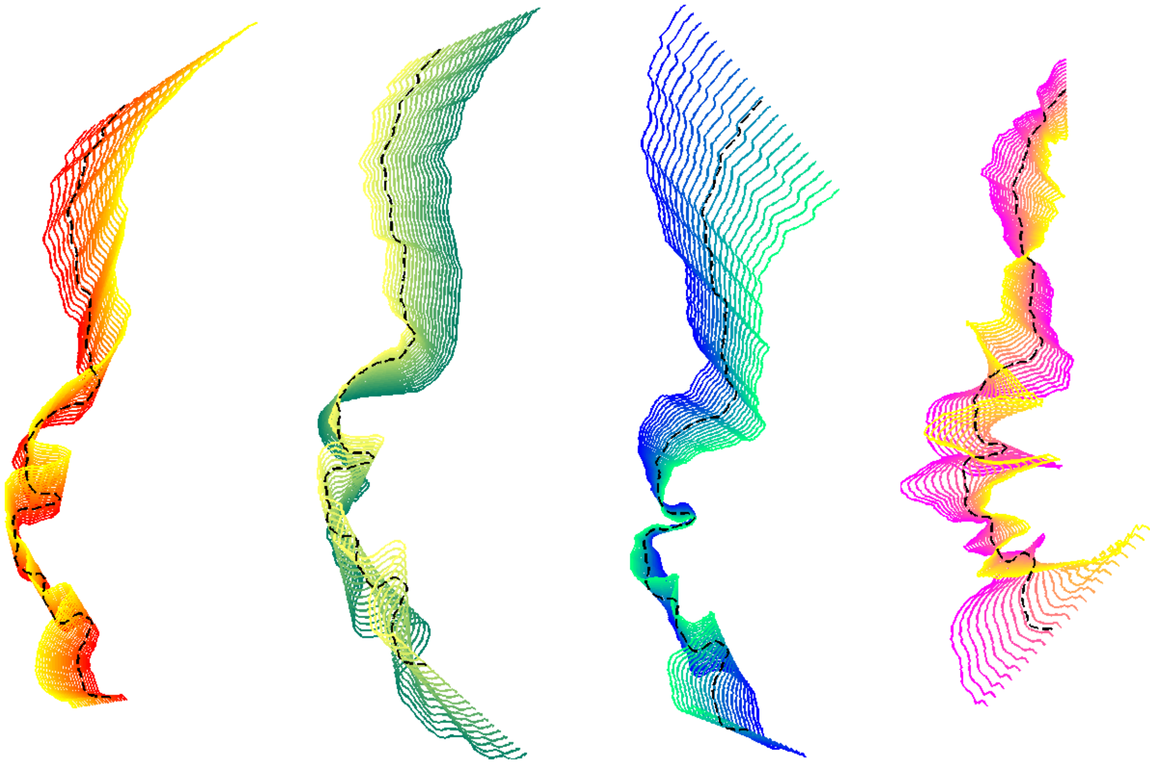
However, this is not the case with 3D data, with typical investigations into face analysis for FAS using around 25,000 points using 3D photography [179], although it is possible to gain such images of a higher resolution. Similarly, 3D US has a spatial resolution of around 0.1mm $\times$ 0.1mm per slice with a 0.5mm slice separation. A window of approximately 50mm $\times$ 50mm will cover the main features of the fetal face, giving a total of around 50,000 points when the probe is perpendicular to the facial surface.

Following this guideline, the resulting covariance matrix can easily contain in the order of 5,625,000,000+ points. Computational limitations in passing around a several-gigabyte-matrix mean that the shape model must be constructed in a different way. Simple down-sampling risks losing fidelity of the image and missing key facial structures. However, since the use of subdivision surfaces has reduced the definition of the facial surface to around 300 points, the method for model construction used for the 2D profile can be used for the meshed faces.

Figure 6.16b shows the continuous surface function represented by the average facial mesh, $\bar{\mathbf{x}}$. The most poorly defined regions of the face are at the sides of the head and at the outer extremities of the eyes, although this can be expected considering the limitations of the field of view of fetal US imaging at this stage in gestation.

6.4.2 Shape Characterisation

As has been described in the previous section, an arbitrary facial profile can now be expressed in terms of a single t dimensional matrix, $\mathbf{b}$. Figure 6.17 shows the effect of altering values in this vector on the resulting fetal profile shape. In this example, the value of $\mathbf{b}$ was found for the curve $\mathbf{x}$ as per Equation 6.16. The four examples shown here each represent an alteration in one of the four values associated with the four eigenvectors with the highest eigenvalues. From Figure 6.17b it can be seen, for example, that changes in the second parameter appear to be the most significant in determining the angle of the upper lip and appearance of the philtrum. Figures 6.17a and 6.17c seem to show changes in nasal shape and head size with changes in influence of the first and third eigenvectors respectively. This means that it is possible that the value associated with the second eigenvector will help determine whether a facial profile is indicative of FAS.



(a) **(b)** **(c)** **(d)**

Figure 6.17: *The first, second, third and fourth parameters are altered in **(a)**, **(b)**, **(c)** and **(d)** respectively*

Figure 6.18 shows variation in 3D facial surfaces from the mean when altering different parameters within the matrix $\mathbf{b}$, when $\mathbf{P}$ for the first two eigenvectors ($n = 1, 2$). In these examples, $\mathbf{b}$ is set to zero and $\mathbf{b}(n) = k$ ranging from $\mathbf{k} = -200$ to $\mathbf{k} = 200$, and then $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{bP}$ is used to find the final mesh.

6.4.3 Training Sets

The eventual goal of this contribution is to create a form of classification metric for diagnosing FAS. There are a number of different ways that a model can be formed, dependent on the type of data used. The options are to form a model exclusively on normal data, a single model based on a mix of FAS and non-FAS data and two separate models for both FAS and non-FAS data.

In the case of a single 'normal' model being used, FAS would be identified in fetuses that exhibit model parameters outside of the expected range, which could be indicative

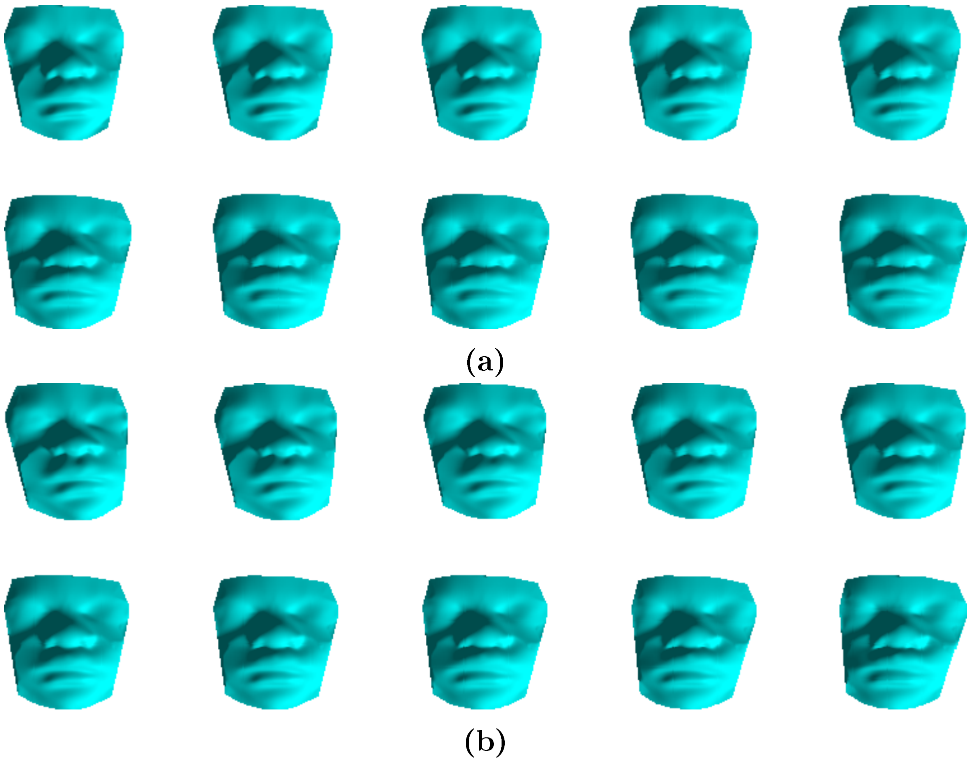


Figure 6.18: *Variation of the* ***(a)*** *first and* ***(b)*** *second eigenvectors of the facial mesh*

of any form of facial variation which is not included in the original data set, not just FAS. This may be useful if the data set for normal faces is sufficiently large enough so that any deviation from the model can be attributed to clinically important abnormality but for specific FAS diagnosis this may not be appropriate. In the case of the two-model system, there may be cases where a face falls outside of the expected range of both models, exhibiting facial variation unknown to both. Similarly, this may be appropriate if both data sets are sufficiently large, allowing the user to distinguish between FAS, healthy non-FAS and abnormal non-FAS.

However, assuming that all known variation in fetal faces cannot at this time be captured, a more appropriate form of model would be to include both FAS and non-FAS fetuses. From these, a classifier can be designed on the values of $\mathbf{b}$, which can then be used to identify new examples as more indicative of FAS or non-FAS face shape. This classifier can be created with a learning algorithm such as *AdaBoost*, which forms a strong binary classifier from a set of weak classifiers [56], or Random Forests, which use a random set of successive classifiers [21]. The latter has been deemed to be comparable to the former in terms of error but more robust to the effects of noise [21]. This may be useful in cases where the segmentation or model fitting are imperfect.

In order to construct and use the model to identify FAS, the following process is proposed;

1. Collect a set of contour/surface segmentations for the faces of a large number of FAS and non-FAS fetuses. In the case of the facial surfaces, a surface mesh can be formed.

2. Find the value of $\mathbf{P}$ for the data set.

3. For each example, find the value of $\mathbf{b}$ via Equation 6.16.

4. Define a classifier to most appropriately separate the two classes according to their corresponding values of $\mathbf{b}$. This can be achieved using a classification algorithm such as Random Forests (in this case, defined by each value of the vector $\mathbf{b}$).

5. For an unseen case, find the value of $\mathbf{b}$ for the segmentation via Equation 6.15 and use the model-generated value of $\mathbf{P}$.

6. The curve/surface can then be classified as either FAS or non-FAS based on the output of the classifier when the value $\mathbf{b}$ is applied.

Figure 6.19 shows the resulting values of $\mathbf{x}$ that are obtained by the model, with different colour contours representing varying values of f_v in Equation 6.18. Due to the small number of test images with which to construct a model, the ASM created within this thesis will not be as adept at modelling the fetal profile as one made as a result of a more comprehensive study. Additionally, assessment of the ability to classify images will be hampered by the lack of clinical diagnosis associated with the images. Since whether each case exhibits FAS or not is unknown, unsupervised classification must be considered. This form of classification intends to cluster together groups of data based on common

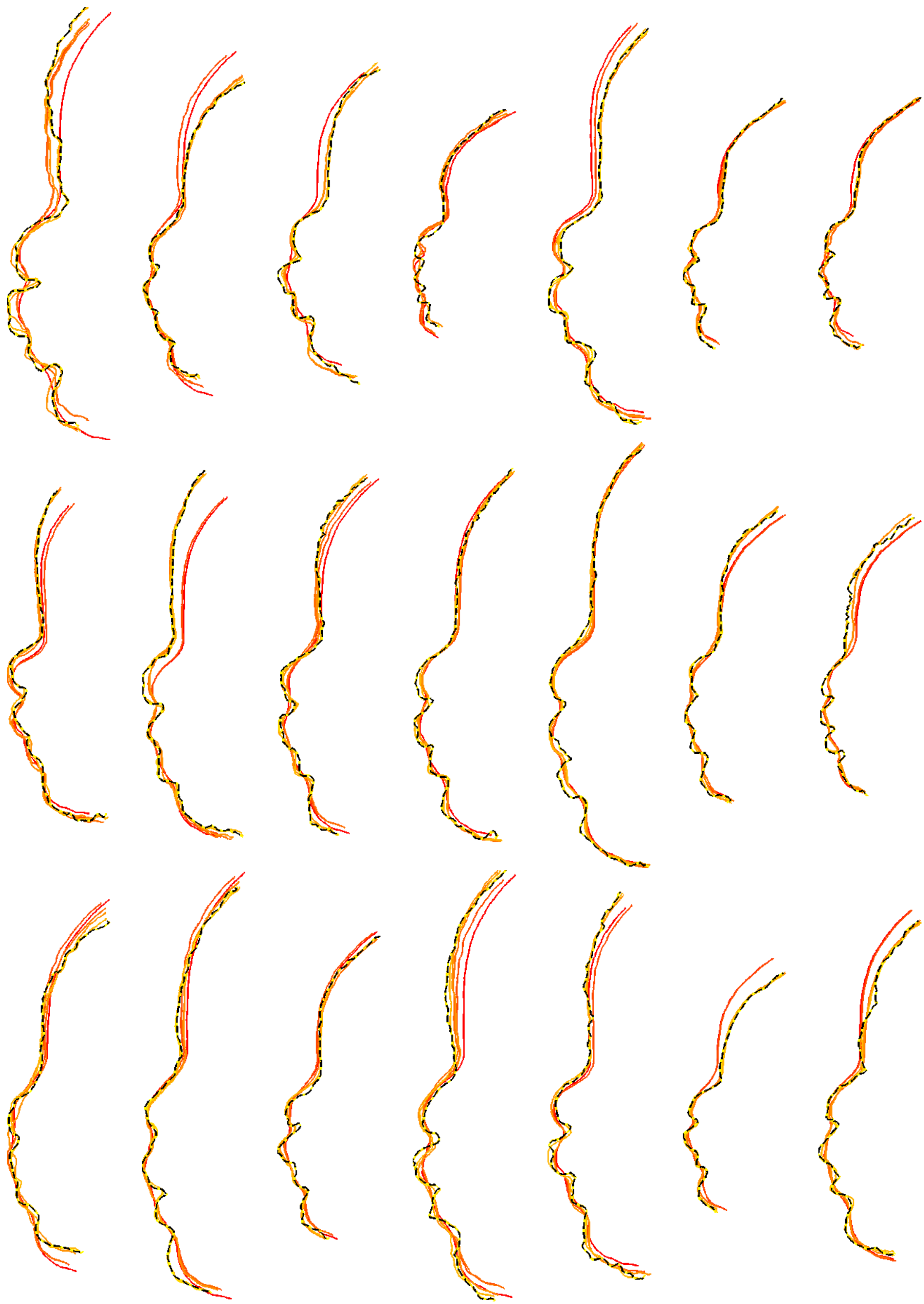


Figure 6.19: *This shows the fitting of the 2D profile model to the input examples, using a varying number of eigenvectors*

properties. Since the data set contains both FAS and non-FAS examples, it is possible that the primary modes of variation will separate these groups naturally. This may, however, be more apparent in a larger data set than the one available. Once this initial classification has been made, future examples of fetal face shape could be identified.

Methodology Overview: Fetal Face Shape Characterisation

- Starting from a central slice in the facial surface, the fetal profile orientation can be determined by localising the visible bone ridge (via a combination of LE and FS) or from user input.
- From this initialisation process, a Live Wire can be applied using FA and LO modified by the profile orientation as cost terms.
- In the case of the automated process failing, FA-O Live Wire can be used to segment the profile using only two seed points at the extremities of the face.
- Once the profile has been obtained, ASP segmentation is applied to extract the full facial surface. Only the central slice is used as an input.
- In the case of the surface extraction failing, FA-O Live Wire can be used to re-segment slices with undesirable results.
- A b-spline mesh, with around 300 vertices, is fitted to the facial surface after being positioned by a set of landmark points selected by the user.
- For both the profile and facial surface mesh, each face can be expressed as $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}\mathbf{b}$. By finding the mean facial profile/surface $\bar{\mathbf{x}}$ and the eigenvector describing the principal components of variation, $\mathbf{P}$, each face shape can be described quantitatively with the matrix $\mathbf{b}$.

6.5 Validation of Segmentation

6.5.1 Experiment Outline

There are a number of different facets of the face acquisition and shape characterisation paths that require evaluation. These include:

- Segmentation accuracy of the 2D facial profile acquisition method.
- Segmentation accuracy of the 3D surface segmentation methods.
- 3D surface accuracy maintained after the mesh process has been conducted.
- Assessment of the ability of shape models derived from these segmentations to group faces based on whether or not they exhibit signs of FAS.

In each experiment, methods are conducted on 21 3D US volumes of the fetal face. In the case of the 2D segmentations, the profile slice has been manually selected. The facial profiles have then each been segmented entirely manually three times by the same user. For the 2D experiments, the automatic contours are compared to each of the three corresponding manual contours and averages are taken of all 63 assessments.

Each 3D facial surface consists of around 100 slices, making manual segmentation an incredibly time consuming task. For practicality, facial surfaces were acquired on a slice-by-slice basis using FA Live Wire, for a range of 40 slices on either side of the central profile, making a total of 1,701 segmentations. This was used as a 'gold-standard' segmentation against which the other segmentation surfaces were compared. Surface comparisons are conducted on a slice-by-slice basis, following the same procedure as for the 2D contours. Since, in all cases, two surfaces or two contours which do not represent a volume/area are compared, it is not possible to use measures such as precision, recall and Dice similarity to compare examples. Distance based methods are more appropriate for comparison, such as calculating the average distance between surfaces and the percentage of points which lie within different range bands compared to the reference segmentation. It is necessary to calculate these values in terms of both the manual contour compared to the segmented contour and the segmented contour compared to the manual contour, since if these contours vary in length (which is very likely), the two measures will provide different values. Using both measures will also help to give an impression as to whether the automatically segmented contours over-shoot or under-estimate the extremities of the region of interest.

6.5.2 Results: 2D Profile Segmentation

Fetal facial segmentations have been obtained manually, three times for each profile.

For the evaluation of the 2D profile segmentation, a number of different methods of obtaining the facial angle have been compared, which also define how the end points of the contour are chosen. The following methods are used:

- **User Guided:** Semi-automated. The user draws a line from below the chin to the top of the brow. The angle formed by this line is taken as the facial orientation, the end points are taken as the extremities of the face.

- **Hough Transform:** Fully automated. The angle is obtained by performing a Hough Transform.

- **Bone Orientation:** Fully automated. The angle of the segmented bone structure is taken as the facial orientation.

- **LO:** Fully automated. In this case, the median LO of the segmented bone structure is used to provide the segmentation orientation.

Tables 6.2-6.5 show the results of the profile segmentation process. These have been conducted for both uncropped, including the entire segmented contour in all cases, and cropped, taking into consideration only the segment of face which would be used for the shape model. Tables 6.2 and 6.3 give the uncropped results while Tables 6.4 and 6.5 give the results for the cropped contours.

Table 6.2: *Profile segmentation accuracy determined from the distance of the uncropped automated contours from the manually segmented contours*

Distance (pixels)	Guided	Hough Transform	Orientation	LO
0 pixels (%)	16.50 ± 4.54	12.93 ± 3.92	12.89 ± 4.02	13.14 ± 4.16
$\leq$ 1 pixel (%)	47.12 ± 8.92	36.85 ± 8.19	36.83 ± 8.50	37.48 ± 8.88
$\leq$ 5 pixels (%)	92.85 ± 5.80	72.52 ± 8.02	72.40 ± 8.80	74.24 ± 10.20
$\leq$ 10 pixels (%)	98.21 ± 3.17	77.37 ± 8.15	77.18 ± 8.69	79.28 ± 10.44
$\leq$ 15 pixels (%)	99.27 ± 2.13	79.06 ± 8.27	78.79 ± 8.51	80.85 ± 10.31
Average (pixels)	2.37 ± 0.71	18.52 ± 9.48	19.15 ± 10.38	17.82 ± 11.88

Table 6.3: *Profile segmentation accuracy determined from the distance of the manual contours from the uncropped automated contours*

Distance (pixels)	Guided	Hough Transform	Orientation	LO
0 pixels	12.27 ± 3.41	11.87 ± 3.64	11.85 ± 3.79	11.44 ± 4.12
$\leq$ 1 pixel (%)	44.45 ± 9.01	42.83 ± 9.75	42.78 ± 10.37	41.26 ± 11.93
$\leq$ 5 pixels (%)	90.98 ± 5.27	86.92 ± 8.75	86.73 ± 10.30	83.32 ± 14.58
$\leq$ 10 pixels (%)	96.65 ± 3.51	92.09 ± 9.30	91.88 ± 10.52	88.48 ± 15.07
$\leq$ 15 pixels (%)	97.78 ± 2.85	93.11 ± 9.04	92.97 ± 10.07	89.61 ± 14.91
Average (pixel)	2.96 ± 1.05	7.32 ± 11.10	7.44 ± 11.44	13.64 ± 25.12

Table 6.4: *Profile segmentation accuracy determined from the distance of the cropped automated contours from the manual contours*

Distance (pixels)	Guided	Hough Transform	Orientation	LO
0 pixel	16.65 ± 4.53	16.70 ± 4.71	16.76 ± 4.80	16.57 ± 4.87
$\leq$ 1 pixel (%)	47.59 ± 8.96	47.50 ± 9.28	47.79 ± 9.44	47.15 ± 9.77
$\leq$ 5 pixels (%)	93.55 ± 5.52	93.06 ± 5.79	93.62 ± 6.34	92.80 ± 5.84
$\leq$ 10 pixels (%)	98.67 ± 2.95	98.03 ± 3.72	98.57 ± 3.71	97.90 ± 3.74
$\leq$ 15 pixels (%)	99.47 ± 1.98	98.81 ± 3.07	99.25 ± 2.50	98.60 ± 3.07
Average (pixels)	2.29 ± 0.67	2.57 ± 1.31	2.32 ± 0.81	2.62 ± 1.32

Table 6.5: *Profile segmentation accuracy determined from the distance of the manual contours from the cropped automated contours*

Distance (pixels)	Guided	Hough Transform	Orientation	LO
0 pixel	12.24 ± 3.41	11.81 ± 3.61	11.79 ± 3.76	11.40 ± 4.09
$\leq$ 1 pixel (%)	44.37 ± 9.02	42.63 ± 9.76	42.59 ± 10.36	41.12 ± 11.88
$\leq$ 5 pixels (%)	90.72 ± 5.49	86.43 ± 8.85	86.25 ± 10.37	82.92 ± 14.51
$\leq$ 10 pixels (%)	96.38 ± 3.86	91.67 ± 9.37	91.47 ± 10.58	88.12 ± 15.00
$\leq$ 15 pixels (%)	97.57 ± 3.20	92.79 ± 9.09	92.66 ± 10.10	89.34 ± 14.85
Average (pixels)	3.06 ± 1.26	7.56 ± 11.09	7.67 ± 11.43	13.84 ± 25.06

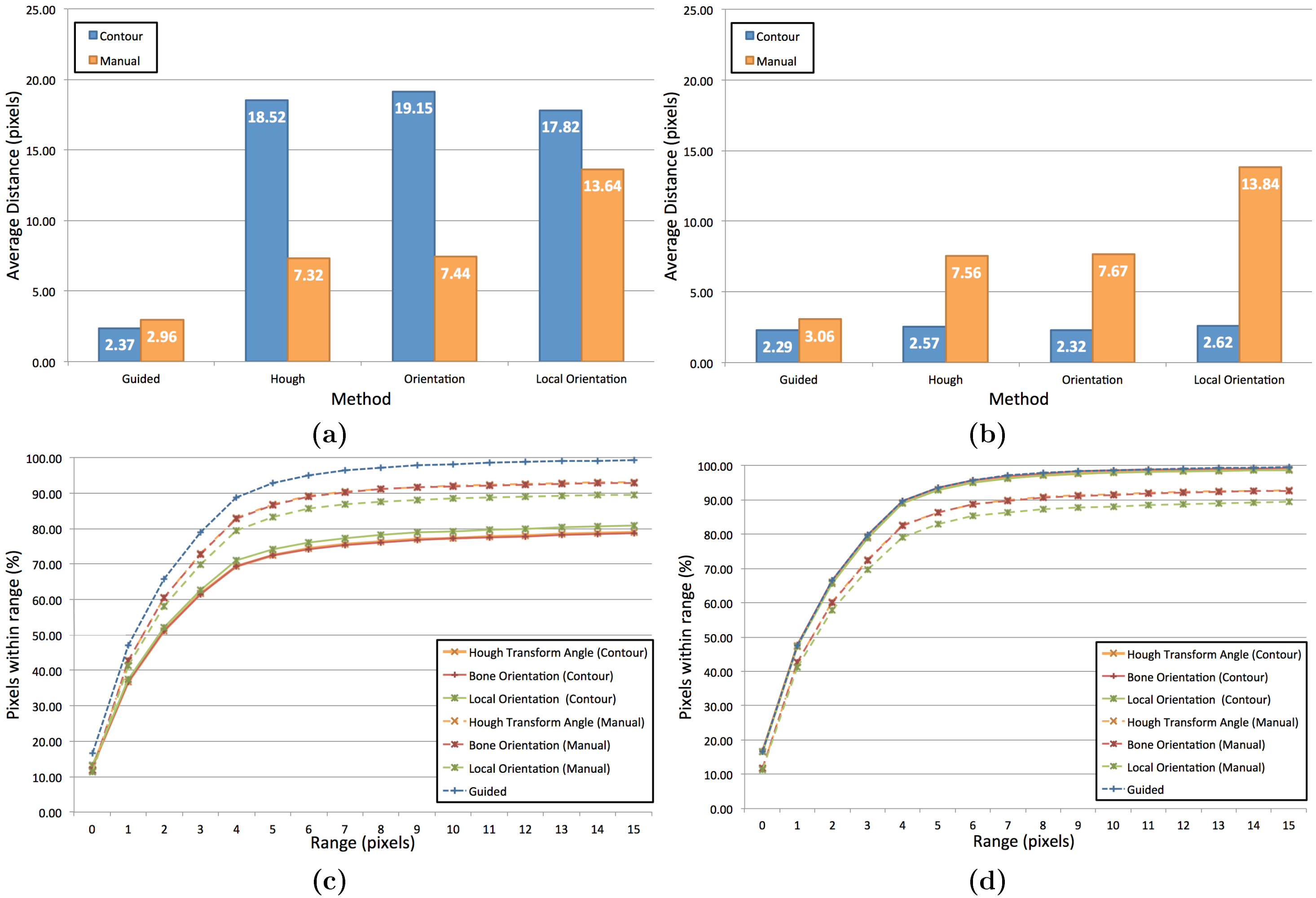


Figure 6.20: *Results of the 2D fetal face profile segmentation; the average distance for the* ***(a)*** *uncropped and* ***(b)*** *cropped contours, and the range of pixels from each segmentation for the* ***(c)*** *uncropped and* ***(d)*** *cropped contours. 'Contour' refers to the results by calculating the distances of pixels on the segmented contour from the manual contour, while 'manual' refers to the results from calculating the distances of pixels on the manual contour from the segmented contour.*

The distance metrics have also been measured in terms of the distance of the pixels on the manually segmented contour from the automatically segmented contours and vice-versa, in order to give a fuller picture of the similarity between them. Tables 6.2 and 6.4 give the results calculated as the automated contours' distances from the corresponding manual contours while Tables 6.3 and 6.5 give the results calculated as the manual contours' distances from the automated contours. These can be considered analogous to precision and specificity for area-based measures. Figure 6.20 summarises these results.

Results for the automated contours relative to the manual contours on the uncropped examples showed the worst results overall. This reflects the fact that, in many of the examples used, the automated contour greatly overshot the boundaries of the face, segmenting the facial region but extending far beyond it (recall examples from Figure 6.7). There is therefore a great improvement in accuracy when the boundaries are well defined. Similarly, the error in the manual contours with respect to the automated contours reflects situations where the boundary is incomplete, as shown in Figure 6.8.

The LO of the bone segmentation was found to be least effective at determining the facial orientation for profile segmentation. This should not be surprising, as the bone is found using the FS of the image (which relies on a high even filter response and a low odd filter response) while the LO uses the angle between the odd filter responses. Therefore by definition, the LO at the bone structure will be weakly defined. Since the time taken to perform the segmentation in the user guided case is minimal (only around a second or two) and requires less user interaction than FA guided Live Wire alone, it is recommended that this form of segmentation should be used to obtain the fetal profile.

All of these calculations have been conducted in a pixel-wise manner but it should be considered that across these examples, one pixel represents in the order of 0.09-0.18mm, (mean $\approx$ 0.12mm). This means that errors obtained here (particularly that of the Guided segmentation scheme) represent significantly small physical dimensions.

Table 6.6: *Surface segmentation accuracy of the automatic segmentation for a number of different filter scales against a surface segmented semi-automatically on a slice-by-slice basis. Highest performing results are shown in bold.*

Reference	Scales	Average (pixels)	$\leq$ 0 pixels (%)	$\leq$ 1 pixels (%)	$\leq$ 5 pixels (%)	$\leq$ 10 pixels (%)	$\leq$ 15 pixels (%)
Segmented	$[10, 15, 20]$	**3.97 ± 2.96**	19.07 ± 4.92	62.30 ± 12.38	**89.73 ± 7.82**	**92.49 ± 6.51**	**93.79 ± 5.92**
	$[20, 25, 30]$	6.83 ± 9.02	48.92 ± 10.51	**84.39 ± 11.46**	87.69 ± 9.94	89.06 ± 9.46	90.29 ± 9.12
$w = [7 \times 7]$	$[30, 35, 40]$	7.37 ± 8.94	**49.84 ± 11.98**	83.35 ± 11.91	86.51 ± 10.37	88.06 ± 9.95	89.31 ± 9.62
	$[40, 45, 50]$	8.88 ± 10.06	22.23 ± 9.74	64.20 ± 16.37	83.87 ± 14.95	86.03 ± 14.54	87.28 ± 14.19
$p = 5$	$[50, 55, 60]$	10.80 ± 11.86	12.61 ± 7.30	43.85 ± 16.04	78.40 ± 16.42	83.63 ± 15.88	85.59 ± 15.64
	$[60, 65, 70]$	15.41 ± 13.81	6.61 ± 4.47	27.55 ± 14.11	70.30 ± 18.95	76.73 ± 18.98	79.23 ± 18.54
Manual	$[10, 15, 20]$	3.10 ± 3.06	20.36 ± 4.97	67.75 ± 12.62	93.32 ± 8.22	95.38 ± 6.28	96.32 ± 5.36
	$[20, 25, 30]$	2.36 ± 3.40	53.10 ± 9.97	**91.96 ± 10.40**	**94.90 ± 7.42**	**95.86 ± 6.44**	**96.55 ± 5.70**
$w = [7 \times 7]$	$[30, 35, 40]$	**2.23 ± 2.98**	**53.69 ± 10.46**	91.22 ± 10.52	94.69 ± 7.06	95.74 ± 6.17	96.41 ± 5.53
	$[40, 45, 50]$	3.45 ± 4.48	23.27 ± 8.94	69.02 ± 15.07	92.08 ± 12.31	94.06 ± 10.88	95.01 ± 9.84
$p = 5$	$[50, 55, 60]$	4.40 ± 4.92	13.25 ± 7.13	46.55 ± 14.39	86.92 ± 15.08	92.95 ± 12.50	94.46 ± 11.26
	$[60, 65, 70]$	6.71 ± 6.68	7.19 ± 4.61	29.85 ± 13.70	79.37 ± 16.89	87.77 ± 16.57	90.49 ± 15.04

Table 6.7: *Surface segmentation accuracy, comparing the dense surface mesh of the semi-automated slice-by-slice surface segmentation to automatically segmented surfaces of varying filter scale. Highest performing results are shown in bold.*

Reference	Scales	Average (mm)	$\leq$ 0mm (%)	$\leq$ 0.1mm (%)	$\leq$ 0.5mm (%)	$\leq$ 1.0mm (%)	$\leq$ 1.5mm (%)
Manual	Manual Mesh	4.03 ± 1.04	0.07 ± 0.04	1.60 ± 0.91	16.23 ± 6.68	32.46 ± 8.72	45.96 ± 7.81
Mesh	Manual Mesh	1.07 ± 0.29	0.18 ± 0.12	3.04 ± 1.91	26.76 ± 15.06	55.21 ± 19.58	77.41 ± 13.95
Mesh	$[10, 15, 20]$	0.47 ± 0.13	0.34 ± 0.32	7.45 ± 4.70	72.24 ± 16.05	94.14 ± 6.47	98.37 ± 1.73
	$[20, 25, 30]$	0.40 ± 0.09	0.68 ± 0.73	12.29 ± 8.24	81.13 ± 10.71	96.76 ± 3.31	98.89 ± 1.14
	$[30, 35, 40]$	**0.34 ± 0.10**	1.17 ± 2.01	19.04 ± 14.26	**86.99 ± 9.92**	**97.12 ± 2.44**	**98.70 ± 1.18**
	$[40, 45, 50]$	0.35 ± 0.15	3.60 ± 10.73	24.01 ± 25.65	84.31 ± 13.08	96.63 ± 2.82	98.41 ± 1.35
	$[50, 55, 60]$	0.35 ± 0.16	**5.23 ± 13.16**	**26.58 ± 26.37**	83.42 ± 14.83	96.58 ± 2.98	98.22 ± 1.70
	$[60, 65, 70]$	0.46 ± 0.30	1.36 ± 2.75	15.94 ± 15.10	75.09 ± 21.48	94.34 ± 10.78	96.71 ± 8.04

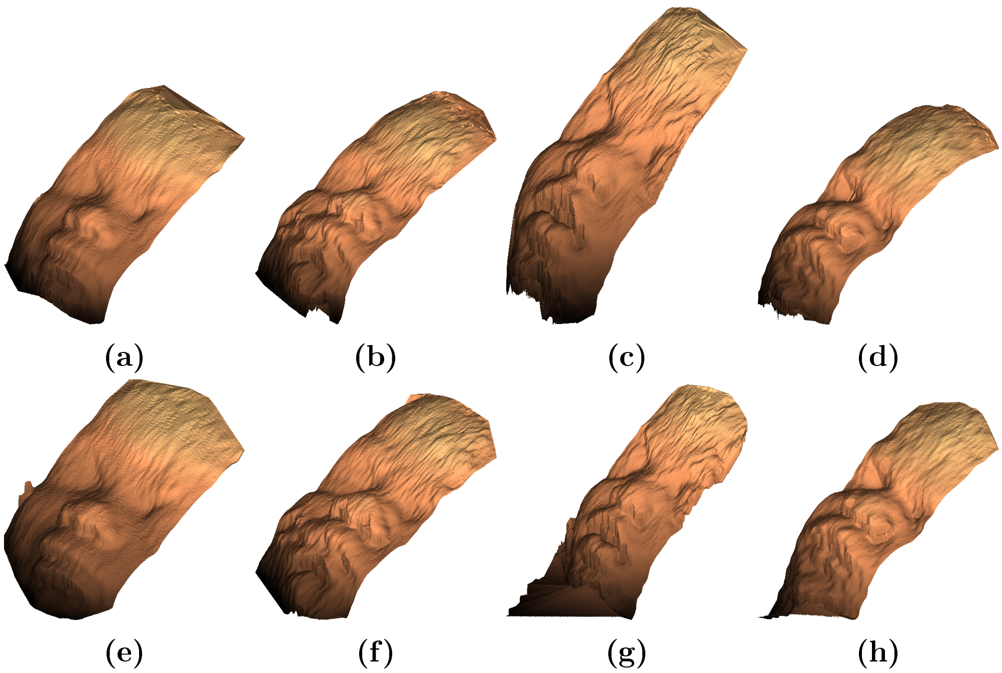


Figure 6.21: *A number of examples of the fetal face surface, segmented on a slice-by slice basis using FA Live Wire (***(a)***-***(d)***) and automating the process by using ASP Live Wire on the same examples, using the central profile segmentation as the sole input (***(e)***-***(h)***)*

6.5.3 Results: 3D Surface Segmentation

The results for the accuracy of the 3D profile segmentations have been taken against respective reference surfaces. The automated surface segmentation has been conducted with varying values of filter scale. Results for this can be seen in Table 6.6, and examples of both manual and automated facial surfaces can be viewed in Figure 6.21. As with the fetal face profile, results have been calculated both with respect to the reference and segmentation contours. The average distance and percentage distances of pixels that lie within 0, 1, 5, 10 and 15 pixels have been calculated.

Many of the faces appear to show a ripple-effect along the faces. This is likely due to the speckle which appears in the facial tissue. A higher-scale filter could reduce the influence of speckle on the path of the wire. However, this could also have a negative influence on detecting more detailed features of the face and prevent effective segmentation in cases where the face is close to another surface such as the wall of the uterus or a fetal limb. Instead, the mesh fitting process intends to correct this phenomenon, by attempting to fit a minimum-cost smooth surface through the point-cloud formed from these facial surfaces.

Although a scale of $s = [10, 15, 20]$ gave the smallest average deviation from the semi-automated surface, $s = [20, 25, 30]$ arguably gives a more accurate segmentation, with a far higher percentage of pixels lying within $0 - 1$ pixels of the original contour. The reason for this is possibly due to the nature of the filters used in FA. The highest FA path will only be in the exact same place in the image if the filters used are the same type and scale, since filter responses will change and shift spatially. The curve for $s = [20, 25, 30]$ is more accurately following the lay of the fetal face (since it is less likely to be deviated by speckle) but the localisation of the edge is slightly off compared to the semi-automated method. This example demonstrates the importance of providing a variety of measures to assess how close contours lie to each other but also the weakness in comparing results from Live Wire methods with different costs.

One consideration when comparing the semi-automated and automated surface contours is that, in the semi-automated case, only contours which are clearly visible are segmented. Regions of extreme attenuation are not included, as contours would be drawn purely on personal judgment rather than from any image data. In the automated case, the contour is determined for the full length of the face, regardless of whether or not there is any clearly visible contour. This, firstly, results in a contour which overshoots the semi-automated version, hence the better result when calculating the average distance of the manual contour to the automated contour rather than the reverse. The second effect is that there may be some extraneous regions of segmentation. While this does have the potential to influence the meshing process, it is an issue which is more likely to influence the extremities of the face, away from the main features.

6.5.4 Results: 3D Surface Meshing

Once the meshes have been fitted to the facial segmentation surfaces, a dense surface can be created of arbitrary point density. While the underlying mesh contains 319 vertices, the dense surface can be represented by a point cloud in the order of 20,000 points. Unlike the original segmentations, which were calculated in a pixel-wise basis, the meshes have been calculated using physical distances. The difference between these fitted dense surfaces can provide insight into the variation after the meshing process has been completed. In all cases the same initialisation landmarks have been used.

Table 6.7 shows the results from comparing the mesh created from the semi-automated segmentation to the mesh created from each of the automated segmentations. For completeness, also shown is the difference between the slice-by-slice semi-automated contour (labelled as 'manual' since it is the analogous 'gold-standard' in this case) and its resulting mesh, to show the deviation in surface shape caused by converting the segmented surface into a mesh representation.

This appears to show low accuracy, but two aspects are at play here. Firstly, the density between the two surfaces are different, resulting in a larger distance between points than may be represented by a continuous surface. Secondly, the mesh does not cover the entire segmented surface, disregarding regions at the extremities (the forehead is a particular example of a large area of segmented surface which does not appear on the mesh surface).

Since the meshes result in surfaces of known resolution, with direct correspondence to each other, there is no need to calculate the differences in surfaces with respect to the manual and automated segmentations separately. However, it may be useful to observe the results obtained from finding the distances between the control points and the dense surface separately, since it will give an impression of how deviation in initial segmentation will impact the final mesh.

Interestingly, the differences in mesh surface produce a fairly different picture of accuracy compared to the differences between the semi-automated and automated segmentation surfaces, with higher filter scales appearing to give closer-fitting results. This could be due to, as seen in the segmentation results, many of the larger filter scales having small regular deviations whereas smaller filter scales have larger but less frequent deviations.

Table 6.8: *Results for the mesh accuracy when white Gaussian noise is applied to the original, manually segmented facial surfaces*

SNR	Control Point (mm)	Mesh Surface (mm)	Model Error (mm)
50	0.01 ± 0.00	0.02 ± 0.00	0.01 ± 0.01
40	0.04 ± 0.00	0.06 ± 0.00	0.02 ± 0.02
30	0.13 ± 0.01	0.12 ± 0.02	0.06 ± 0.05
20	0.98 ± 0.41	1.20 ± 0.69	0.35 ± 0.26
15	2.01 ± 1.12	2.39 ± 1.85	0.71 ± 0.52

These larger variations are more likely to distort the facial mesh than subtle changes, producing a very different end result. The mesh process also appears to be fairly robust to deviation, with errors appearing smaller than their original segmentations (one pixel represents around $0.16mm$). This robustness to segmentation error can be investigated further.

6.5.5 Robustness to deviation: 3D Contour

One potential use for meshing is compensating for an incomplete scan or particularly noisy image, as it should be robust to random variation which does not correlate with surrounding structure. It is therefore necessary to establish how errors in the profile segmentation can influence the final **b** value obtained. For the first instance, the values along the full point-cloud were held in the x- and y-axis, but with white Gaussian noise of varying signal to noise ratio (SNR) applied to the z-axis values (perpendicular to the facial surface). Landmark locations were kept the same in all three dimensions. Signal power was calculated by MATLAB's *awgn* function. The results from this experiment are shown in Table 6.8. In the second case, the noise was applied to the locations of the landmarks (in all dimensions), while the surface was unaltered, with results shown in Table 6.9.

In these experiments, the distance between the manually segmented and noise-adjusted surfaces are calculated for the initial point-cloud segmentation, to give an idea of the distance the noise has perturbed the surface in physical terms. In the case of landmark perturbation, the distance between the original and noise-adjusted landmarks is deter-

Table 6.9: *Results for the mesh accuracy when white Gaussian noise is applied to the original landmark positions, using the unmodified manual facial surface segmentation*

SNR	Control Point (mm)	Mesh Surface (mm)	Model Error (mm)
55	0.10 ± 0.00	0.12 ± 0.01	0.05 ± 0.03
50	0.19 ± 0.01	0.23 ± 0.03	0.09 ± 0.07
45	0.30 ± 0.03	0.37 ± 0.07	0.16 ± 0.12
40	0.51 ± 0.08	0.60 ± 0.16	0.28 ± 0.20
35	1.00 ± 0.32	1.19 ± 0.61	0.55 ± 0.38

mined. The distances between the control points between meshes, the dense mesh surfaces defined by the control points and the resulting model fitting are then calculated, showing the effects of segmentation error on subsequent parts of the system.

The results from fitting the face meshes and subsequent shape models to the noisy data can be seen in Table 6.8, for noise applied to the segmented surface, and Table 6.9 for noise applied to the landmarks used to initialise the facial surface. From these results, it can be seen that even with modest noise applied to the facial surface, there is little change in the facial mesh. This is to be expected since the lowest cost path through a set of points under Gaussian noise is likely to be close to the original surface. For larger levels of noise, the mesh is deformed an appreciable amount. This can subsequently result in a potentially large difference in shape model parameter values.

6.6 Validation of Classification

6.6.1 Experiment Outline

The overall aim of segmenting features from the fetal face is to obtain a means of identifying fetal conditions. The shape modelling process provides a means of giving quantifiable values to the shapes of segmented structures. Recall from Figure 6.19 how segmented contours can be approximated in the form $x \approx \bar{x} + \mathbf{Pb}$. It is these values of $\mathbf{b}$ which can be used to separate fetal profile and facial surface shapes, in our case by whether or not they exhibit signs of FAS.

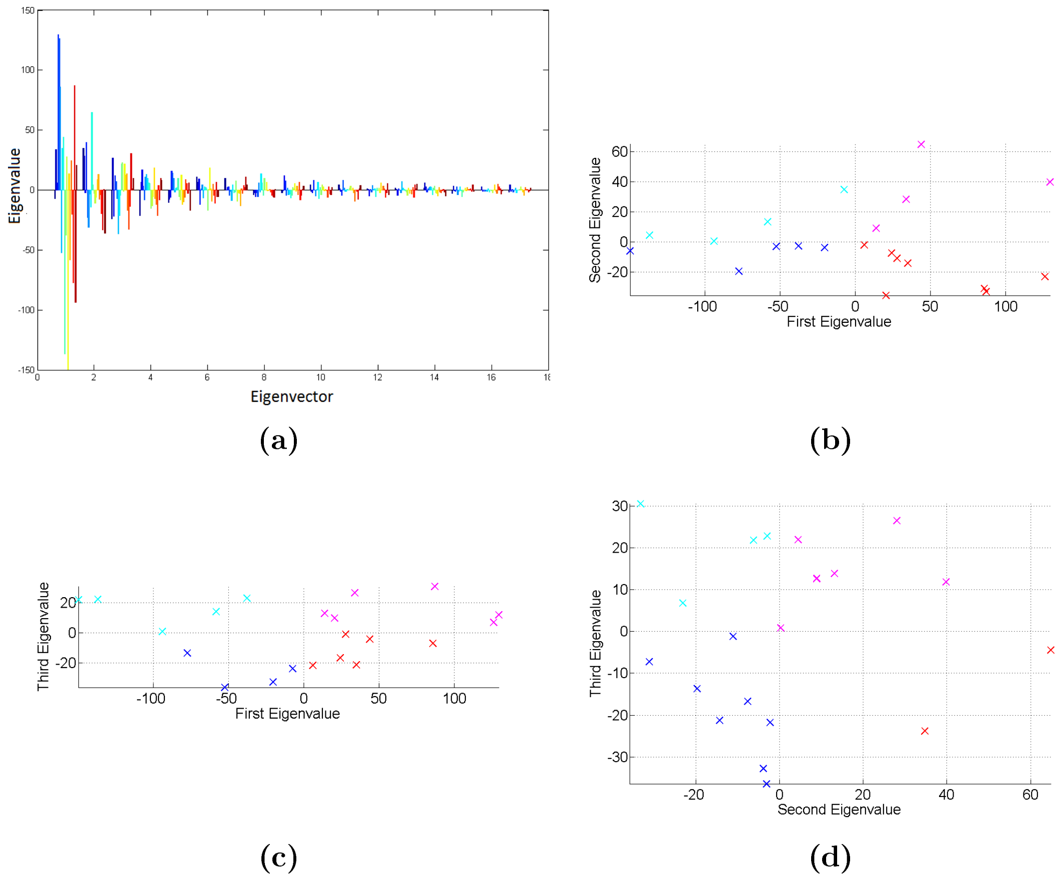


Figure 6.22: *(**a**) The eigenvalue of each eigenvector for all facial profiles. Comparing the eigenvalues for **b**) the first and second (**c**) first and third and (**d**) second and third eigenvectors.*

Unfortunately, at the time of writing, the maternal alcohol intake for each example is not known, meaning that it is not possible to explicitly observe the differences between high and low FAS risk groups. However, as the faces used have a number of potential variables in common (for example, approximate GA), FAS may be one of the more defining features by which the examples can be separated.

6.6.2 Classification of FAS: 2D Profile

Here, the shape parameters derived from each of the facial profile contours are observed. In the first instance, the values of $\mathbf{b}$ are found for each of the fetal profiles. Figure 6.22a shows the value of each element in $\mathbf{b}$ for each facial surface. It can be seen that the values corresponding to the first four eigenvectors have a far larger influence on the shape than all other eigenvectors. This suggests that there are fairly few modes of variation in

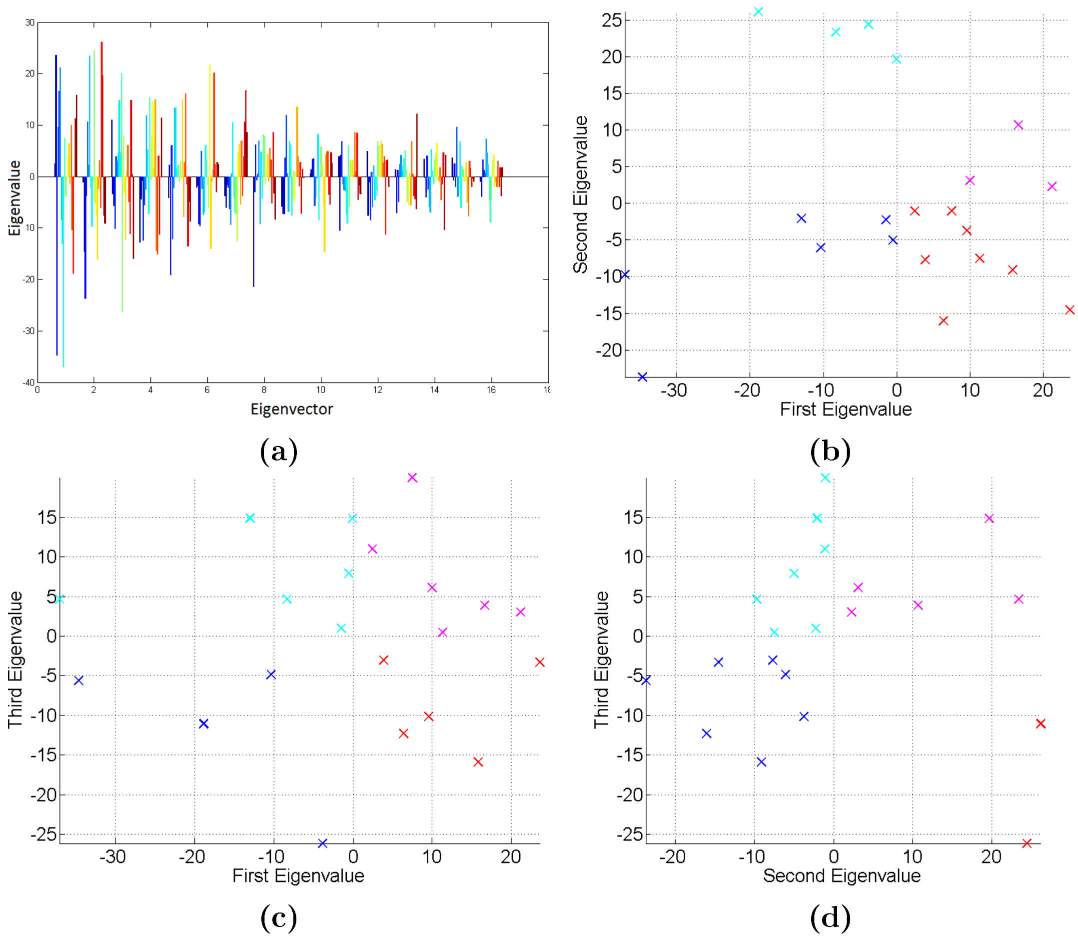
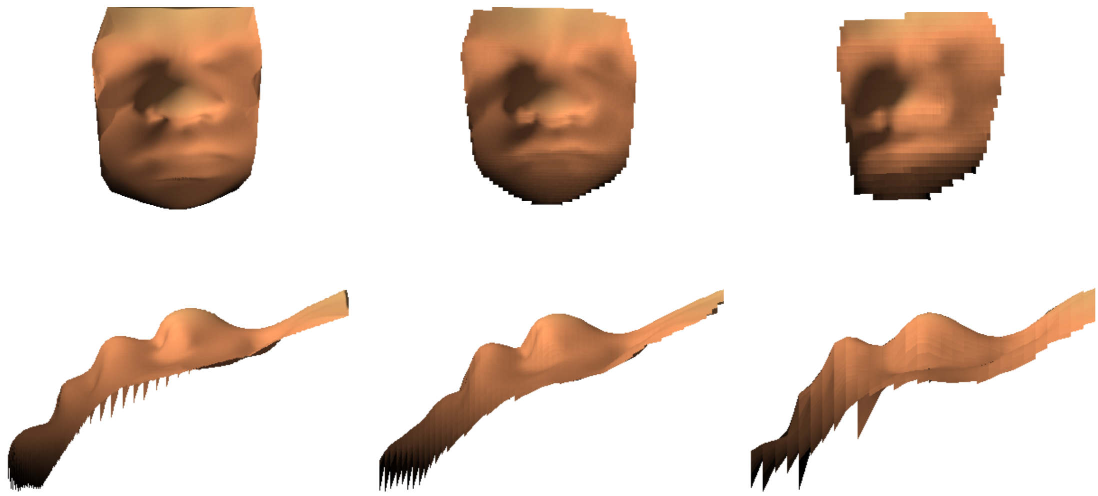


Figure 6.23: *(a) The eigenvalue of each eigenvector for all facial surface meshes. Comparing the eigenvalues for* ***b)*** *the first and second* ***(c)*** *first and third and* ***(d)*** *second and third eigenvectors*

shape that the profiles can generally take.

Figure 6.22b shows values of **b** associated with the first eigenvalue against those associated with the second eigenvalue, while Figure 6.22c shows the first plotted against the third and Figure 6.22d shows the second against the third. Point colours correspond to the quadrants they occupy.

There is not necessarily a distinct separation between groups in this set of fetal profiles since, as established by Lipinksi *et al.*, facial dysmorphism can appear with different severity depending on the extent and timing of alcohol exposure during pregnancy [103]. By taking a section of high-risk individuals, it is likely a spectrum of the condition exists within the set. This means that classification may not be as simple as a threshold on



(a) (b) (c)

Figure 6.24: *Faces resulting from grouping outliers in the eigenvector domain*

one of the shape parameters. Additionally, interaction between modes of variation may result in a more complex model being formed to accurately characterise FAS-related dysmorphism.

6.6.3 Classification of FAS: 3D Surface

As models generated directly from the point clouds were found to be ineffective, computationally intensive and sensitive to error, only 3D models generated from the segmentation meshes will be considered. As these are limited to around 300 vertices, they are of similar complexity to the facial profiles, but can be used to define the entire facial surface. The values of **b** across all meshes can be seen in Figure 6.23a. From this it can be seen that shape parameter values show huge variation across a broad range of eigenvalues. This is possibly due to the greater potential variation in a 3D system of points. A breakdown of the values associated with the first three eigenvalues plotted against each other are shown in Figures 6.23b-6.23d.

There do appear to be fairly distinct groups emerging from these shape parameters, but, as with the profiles, it is unlikely that separation on a single parameter can accurately define FAS. Figure 6.24 shows the faces resulting from averaging the position of facial mesh points for the average of the main group (Figure 6.24a), the average from a group

(a) (b) (c)

Figure 6.25: *Faces from the original data set with a range of values for the second eigenvalue. These examples have values of **(a)** 26.15, **(b)** −1.08 and **(c)** −23.69*

of four faces that showed high second eigenvalue (Figure 6.24b) and the average from a group of two faces that showed very low first eigenvalue (Figure 6.24c).

The variation between the first two averages appears to be influencing the shape of the lower lip and chin. It is possible that a smaller feature, such as the philtrum, may be dominated by a combination of shape parameters that control the position of other, more significant facial structures. Figure 6.25 shows three faces with a high, mid and low second eigenvalue. From this, it appears that this value indicates depth of the face as a whole, and particularly the prominence of the lips.

The last group's average face appears to show examples of the face that have been particularly affected by distortion on the right side of the face. Figure 6.26 shows mesh, fitted surface, manual segmentation and a transverse US slice taken across the eyes of the two faces which exhibited the lowest value for the first eigenvector. It can be seen across all of these stages that there is a deformation on the right side of the face, which the US images reveal to be caused by severe shadowing in that region.

Although the resources available do not allow for the detection of FAS, this shows that a common anatomical feature can be indicated by similar eigenvalues. However, the small size of the data set, and subsequently the coarse nature of the captured variation in face shape, limits the extent to which this can be observed. Since there is variation in

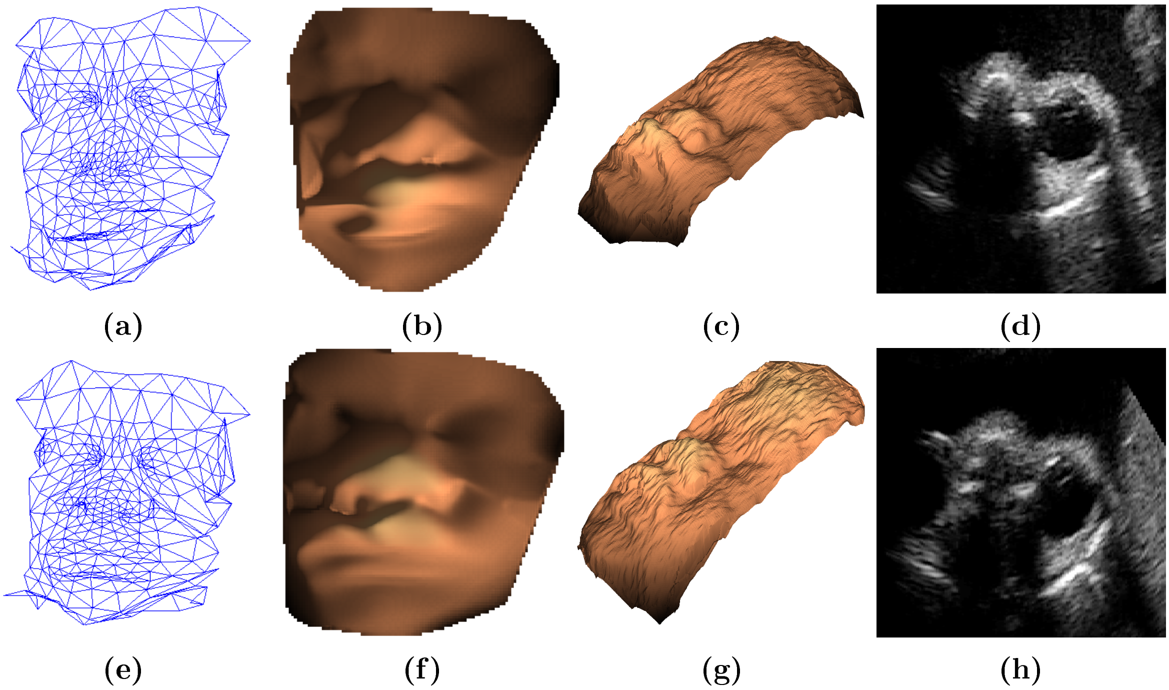


Figure 6.26: *For the two faces with a low value for the first eigenvector, a similar mode of distortion can be seen. This is visible across all stages of the process, including;* ***(a),(e)*** *the face mesh* ***(b),(f)*** *the surface fitted to the segmentation* ***(c),(g)*** *the manual surface segmentation, and* ***(d),(h)*** *a transverse US slice taken across the eyes, showing severe shadowing on the side of the face.*

multiple components of the vector $\mathbf{b}$ and interactions between these components will be complex, it is difficult to identify which values are responsible for subtle changes such as philtrum shape and palpberal width. Section 6.6.4 investigates this further.

6.6.4 Model Properties

Each facial profile and surface can be expressed in terms of $\mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{Pb}$; the mean facial shape, $\bar{\mathbf{x}}$, and the eigenvector matrix, $\mathbf{P}$, are both used to define the shape model, while $\mathbf{b}$ defines the shape of each instance of a fetal face. It is the variable $\mathbf{b}$ that will be used to classify whether faces are indicative of FAS or not. Since the shape of the philtrum is a potential indicator for FAS, and one of the primary modes of variation within this population should be FAS/non-FAS features, it follows that some of the variables that define the shape of the face should indicate the shape of the philtrum, irrespective of structural changes or noise elsewhere. However, the influence of each point on the value of $\mathbf{b}$ depends on the value of $\mathbf{P}$ (generated in the model construction phase) and

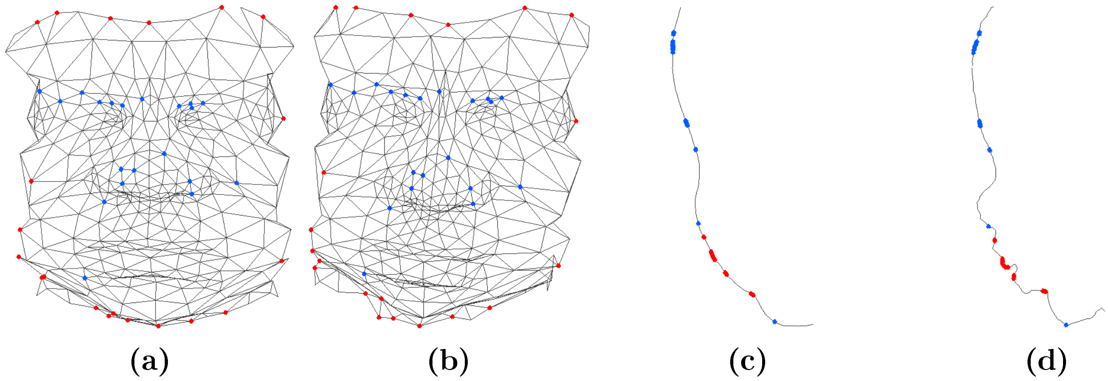


Figure 6.27: *The 20 points which cause the most (red) and least (blue) variation in the shape model when they are perturbed perpendicular to the facial surface for **(a)** the mean shape mesh **(b)** an example face mesh **(c)** the mean facial profile and **(d)** an example of a facial profile.*

the coordinates of each point in $\mathbf{x}$ and it is unlikely that an element of $\mathbf{b}$ will be solely influenced by the position of the philtrum.

The influence of each point of the profile and surface models can be assessed by shifting each of the points individually across all face examples and observing the average difference in the shape descriptor $\mathbf{b}$. Figure 6.27 shows the 20 points with the most (red) and least (blue) variation in $\mathbf{b}$ when they are shifted, displayed on the average mesh and profile, as well as an example face and profile. This highlights regions which require the greatest and least deformation from their mean representations in order to be fitted to real-life examples, since $\mathbf{x} - \bar{\mathbf{x}} \approx \mathbf{Pb}$.

In the case of the facial profile, this appears to be potentially beneficial to the characterisation of FAS; the forehead and chin have less of an impact on the value of $\mathbf{b}$, while regions around the mouth have a greater influence. This suggests that noise or poor segmentation around the forehead may not interfere with the classification of faces using the value of $\mathbf{b}$. In contrast, the value of $\mathbf{b}$ in the 3D facial mesh appears to be controlled by the points on the edge of the face, with more robust features around the brow and nose. This could be partly due to the choice of landmarks for aligning the facial surfaces, and the vast variation at the extremities of the facial models, where it is least constrained to the underlying segmented surface.

This can be examined in greater detail, since it is possible to observe the average variation in each individual element of $\mathbf{b}$ across facial examples. From this an impression can be formed of which elements are more likely to contain information pertaining to FAS-related structures and which may be less useful to analysis. Figure 6.28 shows the average percentage variation of each value in $\mathbf{b}$ when each point on the facial surface is perturbed perpendicular to the facial profile. Figure 6.29 shows this same concept applied to the facial mesh nodes. From these it is possible to observe the varying influence of each parameter on different parts of the fetal face.

In the facial profile, good candidates for the characterisation of FAS using this model ($f_v = 0.999$) are $\mathbf{b}(8)$ and $\mathbf{b}(10)$ (Figures 6.28h and 6.28j respectively), which show a much higher variation in value with movement around the philtrum than anywhere else on the facial profile. The values of $\mathbf{b}(7)$, $\mathbf{b}(13)$ and $\mathbf{b}(16)$ (Figures 6.28g, 6.28m and 6.28m respectively) may be worth observing, although the former two also vary with brow and nasal shape.

Suitable candidates for observation in the facial surface are a more difficult matter, since the edge points appear to consistently dominate for all elements of b. However, $\mathbf{b}(5)$ (Figure 6.29e) appears to show stronger variation around the philtrum compared to other areas of the face, while $\mathbf{b}(6)$ (Figure 6.29f) shows stronger variation around the eyes, which is another landmark that can be used in the identification of FAS.

To show that the model can capture changes in shape, it is possible to induce specific changes and observe the level of variation that can be detected, given reasonable noise. This can be achieved by manually altering the position of points at a specific, FAS-related area and observing their impact on each element of $\mathbf{b}$. Figure 6.28r shows the effect on these values by moving the points of the facial profile along the philtrum perpendicular to the face between -1mm and 1mm, while Figure 6.29q shows the same process for the middle three philtrum nodes of the facial surface. However, in this example, the only variation in the entire face is the philtrum, and so all variables move appropriately.

Additionally, it is possible to observe the effects of moving the philtrum when there are alterations in point position across the entire face. Figure 6.28s shows the effects

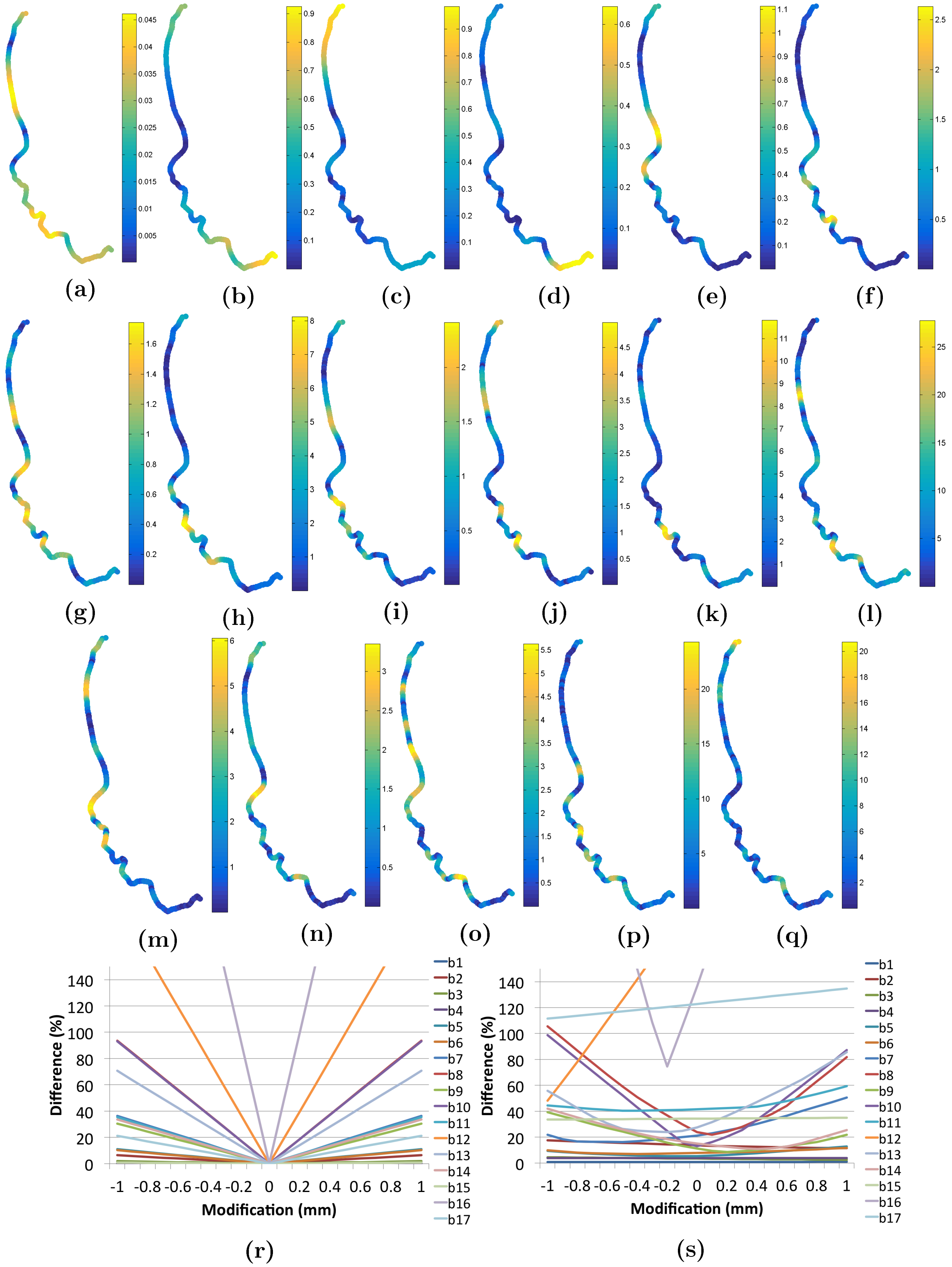


Figure 6.28: *(a)-(q) shows average variation to all model parameters across all examples when each point is perturbed perpendicular to the facial profle. Variation of all components of* **b** *with perturbation of the philtrum is shown with **(r)** and without **(s)** noise ($SNR = 10$.*

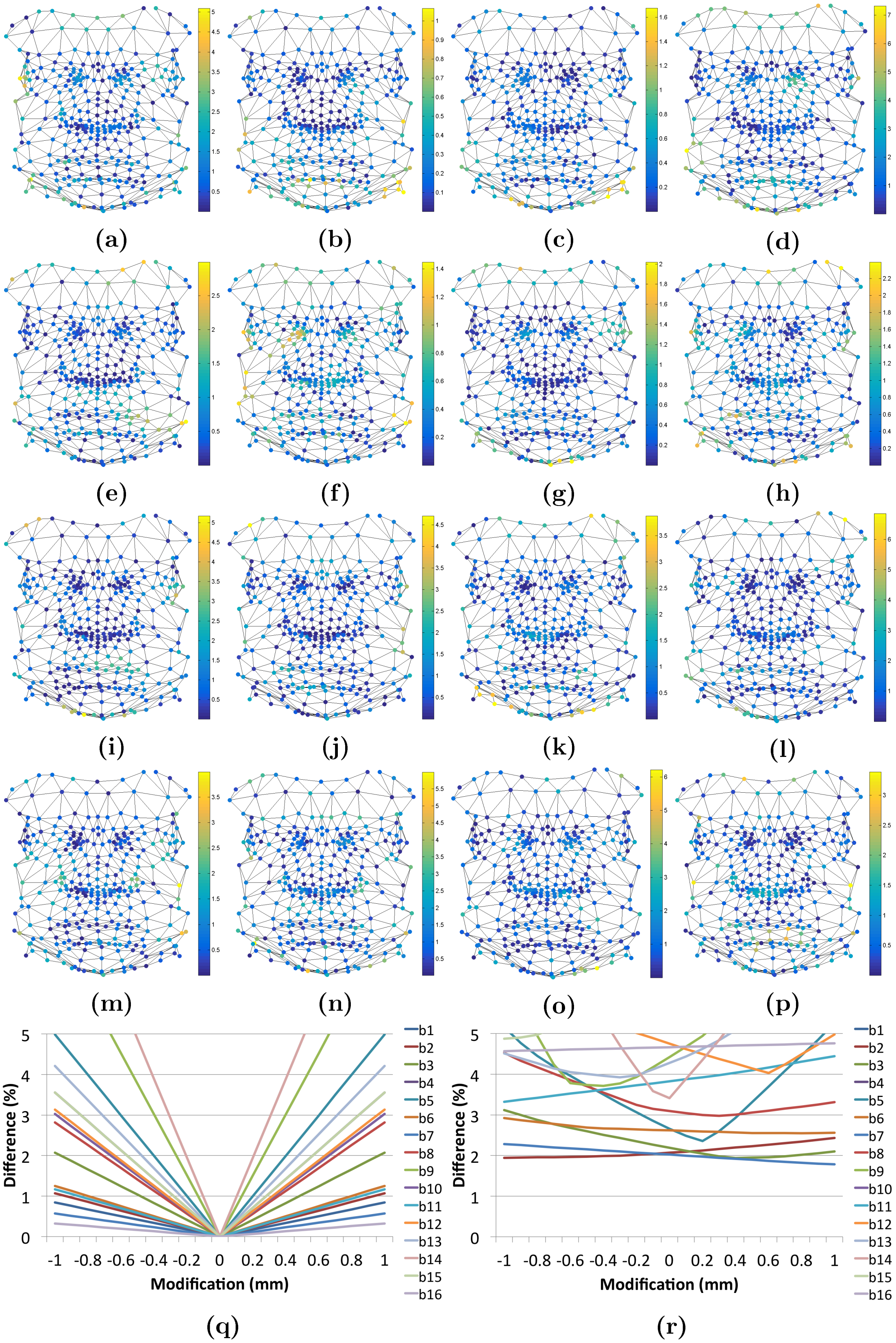
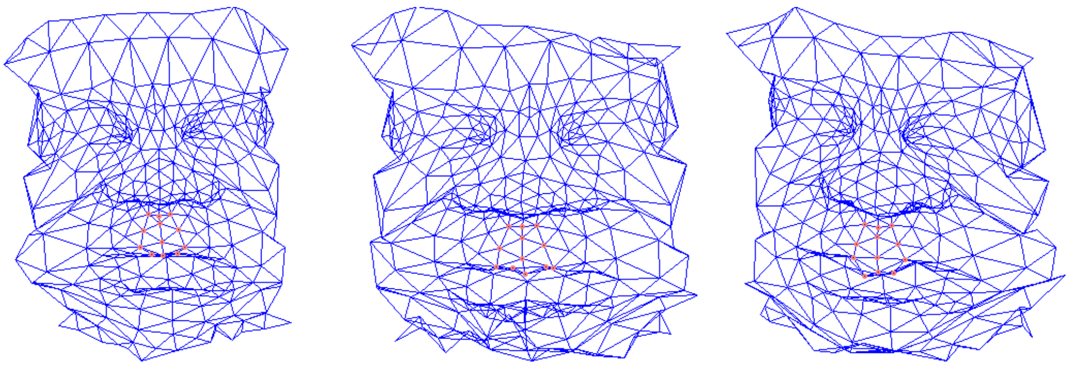


Figure 6.29: *(a)-(p) shows average variation to all model parameters across all examples when each point is perturbed perpendicular to the facial surface. Variation of all components of* **b** *with perturbation of the philtrum is shown with* ***(q)*** *and without* ***(r)*** *noise ($SNR = 20$).*



(a) (b) (c)

Figure 6.30: *The same mesh control points marked out for a set of faces to demonstrate that they all represent the same structural region of the face (in this case the philtrum) after fitting to the segmentation, and therefore could be used for direct measurement of facial biometrics.*

on **b** with additive white Gaussian noise with an SNR of 10. From this, it can be seen that certain variables including $\mathbf{b}(8)$, $\mathbf{b}(10)$ and $\mathbf{b}(16)$ maintain the v-shaped pattern of a similar gradient to the noiseless data. Similarly, Figure 6.29r shows how this is the case for $\mathbf{b}(5)$ when noise with SNR of 20 is applied to nodes on the facial surface. This demonstrates that there are a number of components to the shape model which are influenced more by philtrum shape, a useful indicator for FAS, while remaining robust to noise-related changes in other regions of the face. Consequently, it may be possible to pick out such components of the model in order to help classify FAS more robustly.

6.6.5 Direct Mesh Measurement

As has been briefly commented on, one of the advantages of using a mesh to define the fetal facial surfaces is that each control point on the mesh can be followed across all examples and should represent the same anatomical location. This means that it may not be necessary to use a shape model at all to characterise the fetal face, but rather the differences and ratios between the location of some of the anchor points may accurately simulate the measures which would be taken by a clinician. While the palpebral fissure would be an excellent example of this, as it can be used for the diagnosis of FAS, the current method used does not often contain the outer edges of the eye [176]. Scanning fetuses earlier in gestation may allow for a fuller facial model to be obtained.

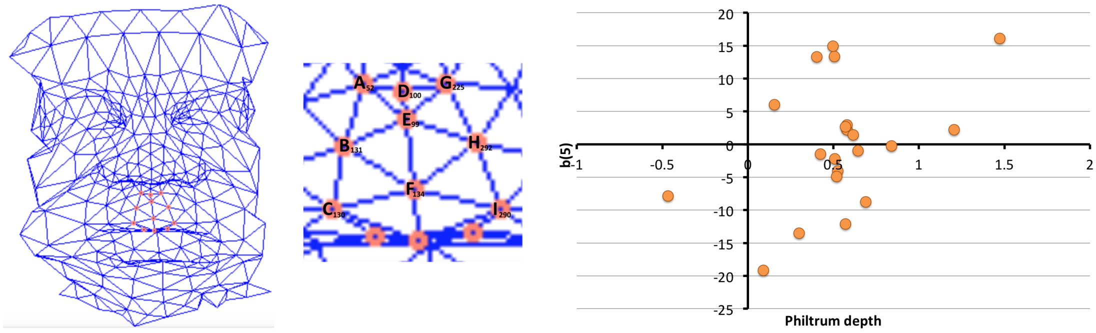
The appearance of the philtrum, however, could feasibly be acquired directly from the facial meshes. Figure 6.30 shows the control points

$$n = [52, 99, 100, 130, 131, 133, 134, 171, 172, 225, 290, 292] \tag{6.19}$$

marked on a number of examples of facial meshes, showing how they delineate the same region of the face. Using control points in this way, it could be possible to extract the philtrum surface in isolation and determine the curvature of the region, thereby providing a quantitative metric from which FAS can be assessed. Similar information could be obtained by observing the curvature of the fetal face profile. Of course, in order to improve the accuracy of such a scheme, this would have to be calculated in conjunction with other values which could characterise fetal dysmorphism, as is the case with clinical diagnosis.

Despite the potential utility of using direct measures from the mesh, pursuing a robust and exhaustive shape model may still be a more desirable outcome. This is because a shape model will account for all variation in the data that forms the training set. Therefore, the model can allow for the discovery of previously unused features in fetal face shape; it may be that the principle modes of variation between the FAS and non-FAS groups show anatomical changes that have not yet been used for diagnosis. In contrast, direct control point comparison can only be used in the context of pre-established metrics for evaluation.

It is possible to observe how indicative of philtrum shape certain aspects of the shape model are, taking into consideration the added knowledge of which variables are robust to noise-related changes from the previous section. This can be done by showing how previously identified values in $\mathbf{b}$ vary with measurements taken from the fetal face. The simplest way of measuring philtrum depth is by comparing the z-axis position of nodes on the facial surfaces. Figure 6.31a shows the philtrum nodes as well as an enlarged schematic of the philtrum, labelled with the index of each node in the matrix $\mathbf{x}$. The



(a) **(b)**

Figure 6.31: *The points which make up the philtrum on the facial surface and their position in* $\mathbf{x}$ *are indicated in* ***(a)*** *while the philtrum depth is plotted against* $\mathbf{b}(5)$ *is shown in* $\mathbf{b}$.

philtrum depth can be calculated as $\Delta z = \frac{z_B+z_C+z_H+z_I}{4} - \frac{z_E+z_F}{2}$, which can then be plotted against values in $\mathbf{b}$. Figure 6.31 shows philtrum depth against $\mathbf{b}(5)$. This appears to show an upward trend, with the value of $\mathbf{b}(5)$ increasing with philtrum depth. However, this is a relatively weak relationship and further investigation with a larger training set will be necessary in order to find out how rigidly it holds. It will also be necessary to identify how other FAS-related changes in face shape interact with the values contained in $\mathbf{b}$.

This first-pass at identifying the properties of the shape model when a small training set is used has shown that one of the main features of FAS identification can be identified from components of the shape parameter $\mathbf{b}$, and these elements appear to be robust to noise elsewhere in the fetal face. In the construction of a functional model for clinical use, it will be important to use a wide range of examples in order to capture the full variation of facial shape in the model. It may also be improved by limiting the span of the facial model to only include regions which are relevant in the diagnosis of FAS, in order to prevent extraneous variation of the model parameters.

Detecting FAS from Fetal Faces

- The sensitivity of the meshing process of the facial surface to deviation in the original segmentation has been tested, showing that the process is robust to errors.

- The utility of shape characterisation has been more difficult to assess, given that the maternal alcohol intake is unknown for this set of images. However, the sensitivity of the shape model to deviation may be a strength, detecting subtle changes in face shape that may otherwise go unnoticed.
- Alternatively, it may be more suitable to extract physical distances and ratios between control points of the facial surface mesh rather than relying on the abstracted shape model to characterise FAS.
- While it appears possible to build a representative model to characterise the shape of the fetal face surface from the position of subdivision control points, there are a number of problems that are external to the image analysing process which should be addressed in order to form a fully operational diagnostic tool.

6.7 Discussion

6.7.1 2D Profile to Detect FAS

One of the most significant problems in attempting to analyse the fetal face profile from US images is accurately acquiring the central plane of the face. This is a potential diagnostic hurdle highlighted by Kurjak *et al.* [94].

The problem of finding the central plane is of particular significance in the assessment of the fetal profile to identify FAS due to the importance of accurately determining the shape of the philtrum. In a healthy fetus, this structure should be concave in the centre and convex at the outer edges but will often be entirely convex for a fetus with FAS. Therefore, a relatively small offset (≈1-5mm) could completely change the diagnosis.

A potential use for the profile measure is to augment the full facial surface model. This is due to the fact that the full facial model may be susceptible to error in the form

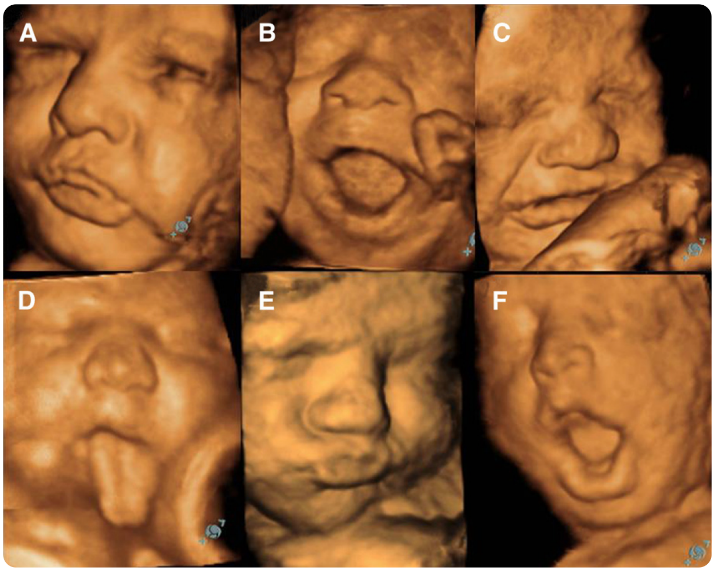


Figure 6.32: *A number of facial expressions made by fetuses during US scans, as shown by Kurjak* et al. *[94] (duplicated here with permission from the author)*

of the face being pressed against other objects, distorting its appearance. The profile may provide a set of parameters which are not affected by the distortion, acting as a second-pass on the classification. If a mesh has already been obtained, the profile can be determined accurately by finding the central points along the control mesh. This has the additional advantage that the profile could be determined even if it is off-plane from the US probe.

6.7.2 3D Surface to Detect FAS

A significant improvement resulting from the meshing of the facial surfaces is the reduction in the feature space needed to create the shape model, as the resulting mesh only contains around 300 vertices. This means that the resulting shape model cannot be dominated by small, irrelevant variations between faces but instead will respond to more significant structural changes.

The use of 3D surfaces to analyse facial dysmorphism of fetuses have a number of obstacles, irrespective of the methods and imaging modality used to acquire the surface. Figure 6.32 is taken from Kurjak *et al.*'s paper on visual analysis of the fetal face and demonstrates some of the common modes of facial expression which can be observed from

3D US images of the fetal face [94]. Unfortunately, the variation caused by expressions such as these will be taken into account when a facial shape model is constructed, potentially marking these differences as principal modes of variation, drowning out the more subtle differences in facial structure. This means that, in order to make examples directly comparable across a population, studies intending to characterise facial differences should aim to acquire the fetal face with a neutral expression. This may be difficult to achieve in fetal observations, as fetal movements can be unpredictable and expressions can only be recognised after 3D acquisitions are taken.

Similarly, the fetus's face may have a boundary between facial tissue and limbs or the wall of the uterus, rather than the far clearer boundary between facial tissue and amniotic fluid. While the segmentation methods explored here do have the ability to delineate between such boundaries, if there is any significant pressure on the face it can appear to be deformed. Soft tissue of the face, for example, can be easily displaced against other objects, resulting in a facial segmentation which does not accurately represent the 'resting' shape and potentially distorting the appearance of the resulting shape model.

As a future expansion, high curvature points on the facial surface could be used to define landmarks on the 3D facial surface, thereby entirely automating the alignment procedure. It has been observed that deviation in initialisation landmark position can result in as large a difference in the final face mesh as errors in the segmented surface. Therefore, standardising the way in which landmarks are selected could result in more accurate and comparable fetal face meshes, and subsequently more accurate shape models.

Another future expansion could be to extract the set of standard measures usually used to diagnose FAS from the control points of the mesh and compare this to manually determined measures conducted by a clinician (or, preferably, set of clinicians). The aim here would be to show if **(a)** it is possible to accurately determine the biometric values currently used to diagnose FAS and **(b)** examine if the meshing process can provide a more standardised method of assessment. This would, again, require a set of fetal face volumes with known histories of maternal alcohol intake.

6.7.3 Position in Clinical Workflow

Finally, the practicality of establishing such a diagnostic procedure in existing clinical workflow is considered. Since the segmentation of both the fetal face profile and surface have been very successful, it is likely that fetal face analysis could be incorporated into a standard prenatal scanning procedure without too much difficulty. User interaction required has been shown to be minimal, and would therefore need little training to achieve.

Major obstacles to facial characterisation are mostly caused by the behaviour of the fetus, since attenuation from the skull means the fetal face would be very difficult to observe if the fetus is facing away from the US probe. Additionally, some of their standard behaviours (such as thumb-sucking, yawning or making distinctive facial expressions) would produce a surface which may not be fit for analysis. Of course, with a large enough collection of data, it is possible that this form of variation could be taken into account, but this would require further study.

The utility of pre-term diagnosis is also an important consideration. Some studies have suggested that the physical symptoms of FAS can become more difficult to detect over time [179]. If the diagnosis can be performed as early as possible, instances of FAS could be detected more accurately. It is currently not possible to reverse the effects of FAS, but a diagnosis may prevent further damage. Additionally, early and more accurate diagnosis can allow for interventions and therapy that may prevent some of the secondary problems faced by those affected by FAS in later life, such as issues with mental health, inappropriate behaviour, substance abuse and illegal activity [177]. The low impact of diagnosis and potentially high improvement to quality of life could make screening for FAS detection a beneficial addition to clinical practice.

The 2D profile segmentation and analysis work has been presented in the following conference: T.M. Rackham, P. Hammond, J.A. Noble 'Fetal face segmentation and contour characterisation in 3D Ultrasound', MECbioeng14, London, 2014

Overview

- An automated method of segmenting the fetal face from a 2D image slice has been presented, using a combination of FS, LE and the Hough Transform to find the skull and its orientation. Then a combination of FA and LO is used to segment the facial boundary. In some cases a small amount of user interaction is required to ensure a correctly segmented boundary.
- Further to this, a method for obtaining a 3D facial surface is presented, using the initial 2D surface as initialisation to the ASP Live Wire algorithm described in Chapter 5.
- Next, an existing method of fitting a b-spline surface to the segmented point cloud is applied, in order to form a parameterised model of the fetal face surface.
- In both the 2D and 3D cases, a shape model is constructed in order to find the eigenvectors which define the modes of variation of the fetal face shape.
- Fitting the shape model back to each example of fetal face profile and surface will give a matrix **b**, which provides a quantifiable metric for the shape of each object. It is through this value that facial differences can be classified.
- While the meshing procedure and shape fitting appear to be robust to noise and deformation, the shape parameters are very sensitive to changes in profile and surface appearance.

The personal contribution in this chapter is the application of ASP Live Wire to fetal face segmentation. Additionally, the combination of modelling the segmented facial surface using a mesh and then using this simplified mesh to characterise face shape is an original contribution, although these two components were developed through independent collaborations.

Chapter 7

Discussion and Conclusions

Throughout the course of this thesis, a number of tools, specifically designed to allow clinicians to acquire a more advanced understanding of fetal development, have been designed and tested. The main objective has been to enable these tools to give the user direct control over the contour construction, keeping user-interaction intuitive and straightforward. These methods can be summarised as follows:

- Firstly, a semi-automated, interactive segmentation method has been developed for use in US, based on the Live Wire algorithm. This was then tested on a set of 2D fat images, comparing the results to a number of competing methods for completing the same task.

- This method has been expanded in order to produce a new method for evaluating a 3D volume from Live Wire contours. This has then been tested on sets of MRI and US liver images and compared against an existing approach for expanding Live Wire to 3D.

- Finally, the 2D and 3D approaches for interactive segmentation are taken into the context of fetal face segmentation and characterisation in US. Assumptions about the general appearance of the fetal face are exploited to further facilitate the segmentation process. A procedure for quantifying fetal face shape is proposed.

This final section summarises and evaluates the key features of these findings. Following this, there is a discussion on potential avenues for future work and improvements to each of the main contributions.

7.1 Fetal Fat Segmentation

This thesis has presented a method for the semi-automated segmentation of fetal fat from 2D US images by modifying the existing Live Wire algorithm. The method outlined here, FA-S Live Wire, used a combination of FA and a weak shape constraint in order to guide the segmentation process. Here, potential expansions to this work are discussed.

7.1.1 Segmentation

Chapter 4 has introduced the use of FA and a weak shape constraint to adapt the existing Live Wire algorithm and to facilitate the interactive segmentation of fat in 2D US images of fetal limb cross-sections. This was found to perform better with US than traditional intensity-based measures for Live Wire cost and remains competitive with an automated scheme that also uses FA. More importantly, the method required far less user interaction and took significantly less time than manual segmentation and the competing interactive methods and did not require specialist equipment to operate effectively.

One potential enhancement that could benefit the utility of such a tool would be the automatic selection of FA scale based on speckle appearance. This would allow FA-S Live Wire to be applied to a greater variation of images without requiring the user to pre-define a rather abstracted set of variables. Additionally, this could help further improve the ability of this method to delineate boundaries in cases for which it has already been shown to be successful.

A limitation of the study is that only 2D cross sections were used, which are prone to error for accurate quantification of fetal biometry, since errors in angle or differences in position along the limb can greatly influence results. It would therefore be useful to apply the 3D methods explored in Chapters 5 and 6 to fetal arm images.

7.1.2 Growth charts

A potential further investigation would involve constructing growth charts based explicitly on fetal fat measurements, and comparing a full fetal fat volume to other measures of fetal nutritional development, rather than just a TVol (including both fat and lean tissue). A recent study by Knight *et al.* has shown that it is possible to segment fat tissue in US, successfully segmenting 78% of captured fetal limbs from scans of 231 women from INTERGROWTH-21st data [90]. Further to this, it was discovered that segmenting fat from the fetal arm may be a more useful tool in assessing development than from the thigh, with fat representing a larger proportion of overall limb volume at this location [89].

The methods described here could facilitate the collection of such data to expand the study to observe how the fat volumes of malnourished fetuses compare. Additionally, they provide a basis from which an expansion to volumetric fat measurements could be developed. FA-S Live Wire could be combined with the ASP Live Wire method in order to form a better resolution of fat volume than fractional measures. The data set of 3D US fetal arms with corresponding GAs now available via INTERGROWTH-21st make this a very possible avenue for investigation.

7.2 Liver Volume Segmentation

In Chapter 5 potential methods for segmenting the liver in 3D from MRI and US were explored, using an expansion of the Live Wire method. In particular, a combination of LO and FA was used to estimate the position of seed points in successive slices of the image, thereby automating portions of the segmentation process.

7.2.1 ASP Live Wire

ASP Live Wire has shown great utility in being able to segment the fetal liver from e-THRIVE MR images, showing greater accuracy compared to Sparse Live Wire. In US, ASP Live Wire was found to perform reasonably but was ultimately outperformed by

Sparse Live Wire by a small margin. This could have been due to the large variation of the contour position between slices, as a result of the difficulty involved in manually segmenting the liver boundary.

As already discussed, one clear improvement would be to alter the filter scale used to calculate FA and LO depending on the imaging conditions. Since larger scales are only appropriate with a larger segmentable object, this could be determined based on the size of the objects delineated by the required user-defined segmentations. Otherwise, a multi-scale approach could allow for more accurate localisation of points in consecutive slices and a wider search window, or even allow for global point matching. The field of image registration could be drawn upon in order to improve this process further.

Another potential improvement would be to allow for the segmentation of multiple objects. Although this was not a problem that was encountered with the examples used in this thesis, it is a potential issue which should be considered. Research has already been conducted in adding this functionality to the Sparse Live Wire method, so it is likely that a similar regime could be applied directly to the point selection derived from ASP Live Wire [143].

7.2.2 Segmentation in MRI

It has been demonstrated that it is possible to assist the segmentation process, despite the challenges of low spatial resolution, specular appearance and poor contrast edges, which although not common to all MR images, are certainly problems for the e-THRIVE acquisition setting.

A major disadvantage of liver segmentation in MRI is that it is unlikely to be applied as a tool in standard clinical practice, as the modality is often only used once poor fetal health has already been established from standard US scans. In situations where pregnant women have been referred for scanning to diagnose suspected severe fetal conditions, nutritional development would be a minor aside to the health of the fetus. Introducing MRI scans as a standard procedure for pregnancy would be ill advised due to the cost and resource constraints, as well as the possible distress caused to the mother with the claustrophobic environment of the MRI scanner.

The modality is arguably more useful in investigations such as that attempted in this thesis, which could potentially provide confirmation of the measurements made in US, as these are more susceptible to error. However, this relies on appropriate US scans of fetal landmarks to compare against.

7.2.3 Segmentation in US

Segmentation in US is arguably of more utility than MRI, as this form of measurement could be included in standard protocol for maternal monitoring with very little disturbance to clinical workflow as it currently stands. However, accurately determining LVol from US images remains an incredibly difficult task, with even visual detection of boundaries presenting a challenge, let alone the automation of such a procedure.

In contrast, the appearance of the liver in MRI (when the correct settings are used) is a manageable problem. Unfortunately, MRI is often reserved for cases where the fetus is already expected to be in poor health, and malnutrition may be a less pressing issue. In order to make liver segmentation useful in clinical practice, it is essential that it can be performed in US.

To make the imaging of the liver in US a useful tool, it is essential that a set of protocols are established to ensure that the volumes obtained are suitable for boundary extraction. This may, however, limit the range of ages at which the scan can be performed, as increased size and bone density can result in severe attenuation of the signal over sections of the torso. Alternatively, it could be that further technological advances improving the quality of fetal US imaging is required before accurate liver segmentation will be viable.

7.3 Fetal Face Analysis

A further modification of Live Wire has been developed in Chapter 6, specifically designed to obtain 2D profiles and 3D surfaces from US images of the fetal face, utilising

Figure 7.1: *Example of the facial surface at the fetal stage and in early childhood for the same subject. From left to right; at 28 weeks GA, 1 month and 1 year*

FA and LO to guide the contour. This has been referred to as FA-O Live Wire. Further to this, the existing concept of subdivision surfaces has been applied to this particular problem in order to form a parameterised model by which faces can be compared. Finally, a method and workflow for obtaining quantifiable metrics to describe fetal face shape is presented, with the objective of identifying fetuses with physical signs of FAS.

The eventual goal of this research is to be able to track facial surfaces from early gestation to early childhood, and perhaps even to later in life, building a full model of facial development for various genetic conditions. Figure 7.1 shows an example of one of the facial subdivision surfaces obtained through the method described in this thesis compared to corresponding one month and one year 3D photographic images (provided by Peter Hammond).

7.3.1 2D profile segmentation and modelling

A method for obtaining profiles and processing them appropriately to form a shape model of the fetal profile has been outlined in this thesis. The model has been shown to be sensitive to small variations in appearance but it has not been possible to assess the accuracy of classification.

Although the fully-automated version of this method was found to be very successful in obtaining the correct contour, it was occasionally unable to find the correct end-points of the fetal face, resulting in an incomplete facial profile. However, a semi-automated version of the method, requiring minimal user input, was found to overcome this issue.

This accurate segmentation process has allowed for the creation of a well-defined shape model. As the alcohol exposure of the fetuses have not been available for these examples, it has not been possible to establish how accurately classification can be conducted on the shape descriptors.

The biggest consideration in obtaining a successful profile is the plane at which the initial image is taken. Only a small deviation from the central line would completely alter the shape of the segmented contour and therefore alter the values obtained to characterise shape. It is important to note that the abnormal appearance of the philtrum is a key feature in diagnosing FAS from profile images; while a normal philtrum is concave in the centre of the face, it will appear flat if the acquired slice is a few millimeters either side. Additionally, the shape of the nose will be greatly affected by such variation.

As a result, the fetal profile may need to be obtained as part of a larger, full-surface segmentation, which would be able to more accurately determine where the central plane of the face lies.

7.3.2 3D SURFACE SEGMENTATION AND MODELLING

An extension of the profile shape model has been conducted in the form of the full-facial surface model. In this case, profiles are used as the initialisation for a 3D surface segmentation, to which a mesh is fitted. The control points which define the dense surface of this mesh can then be used as the inputs for a shape model. Investigations have shown a very high level of accuracy compared to slice-by-slice semi-automated segmentation of the surface, and that the mesh fitting is robust to errors both in the initial segmentation and the landmark positioning used to initialise the mesh.

Potential obstacles to a clinically useful segmentation include fetal face expressions and deformation due to the face being pressed against limbs or the wall of the uterus. This means that even if a perfect segmentation of the fetal face is obtained, it may not be suitable to characterise the shape of the face. Ideally, images should be acquired with this in mind, avoiding such problems where possible. However, it takes time to acquire a full US volume and it may not be obvious as to whether the fetus is, for example, showing a particular expression from the views available on the clinical machine being used. A problem compounding this is that there is generally very little way of controlling the fetus's position or expression beyond sedation, which in itself may be too extreme for a standard observation.

Since a model can only capture the range of variation included in the training set, increasing the data set may allow for diagnosis under these conditions. Including examples of the fetal face subject to different expressions will allow the model to capture these standard facial deformations. The regions of the face that are associated with FAS diagnosis may still be visible with such expressions; the width of the palpberal fissure may still be apparent for a range of expressions, although its shape may change, while the shape of the philtrim may change with mouth movements but still be flatter in cases of FAS. Similarly, this may help to identify the forms of facial distortion associated with the fetal face being pressed against other structures. Exactly how the main features of FAS are affected by facial distortion and expression requires further investigation.

A potential modification to the model could be to allow for greater flexibility on certain regions of the face. For example, it is possible to make the assumption that while regions such as the cheeks (mostly soft tissue) may be highly subject to deformation, the forehead, which will have much less soft tissue, will be far less susceptible. This means that it is possible to identify regions of uncertainty on the facial mesh based on estimates of local tissue rigidity, and therefore mitigate errors caused by deformation of the face *in utero*.

Future work required to properly establish whether it is possible to accurately diagnose

FAS *in utero* would be to apply the methods described in this thesis to a far larger set of facial US images. This could then be used to define the exact shape-model values expected for facial variation exhibiting the key FAS features. It is possible that the most practical way to use this as a diagnostic tool would be to focus on a narrow band of GAs and ensure that all measurements are taken at this time. Alternatively, a more data-intensive method would be to produce a number of models for different stages of gestation.

A Composite Approach

Since the 3D segmentation tool has been so successful in obtaining full facial surfaces, this may make the 2D model obsolete, since it will not suffer from the same problem of sensitivity to deviation. The 2D profile may, however, have the advantage that it is less sensitive to facial expression, with the only main variation in shape caused by fetal movement due to whether the mouth is open or closed. In this respect, the 2D profile shape could be used as a back-up measure to the 3D surface, allowing for a 'second opinion' in cases where the facial surface is distorted by facial expression or if it is being pressed against.

Alternatively, the full facial shape model could be defined by both profile and surface measures. An additional advantage of developing a composite method is that the meshing process allows for the appropriate slice to be obtained for the facial profile very accurately.

Direct Mesh Measurement

An alternative method for assessing whether or not a fetal face is exhibiting the signs of dysmorphism associated with FAS has also been proposed. This involves taking direct measurements from the mesh control points since, assuming the initialisation landmarks have been placed accurately, these should all have comparable locations on the facial surface between examples.

Further investigations could involve comparing the results obtained from taking biometric measures directly from the fetal face surface mesh to the measures taken manually by clinicians (either taken on the corresponding segmented facial surfaces or physically measured on the neonates). This could potentially reduce the variation in diagnosis and provide more accurate, automated means of measuring standard facial biometry for the diagnosis of FAS.

Expanding the model

A final avenue for the expansion of this project would be to collect data from fetuses affected by other conditions which express themselves in the form of fetal dysmorphism. Since there is some overlap between the features observed in different conditions, it will be important to establish whether these can be accurately separated.

However, a major drawback of conducting this form of study is that it would likely require a very large amount of data with corresponding, accurate diagnosis. This could be difficult to obtain since currently, many of these conditions cannot be reliably detected before birth.

7.3.3 Establishing a protocol

One of the important issues to address in order to ensure the reliable acquisition of fetal face surfaces is that images are obtained in a standardised way. In contrast to INTERGROWTH-21st and INTERBIO-21st, the images used for the facial analysis study were not taken with any specific requirements. This has meant that the selection of images used may not have been an optimal set of examples from which to form an accurate facial model. Although accurate 3D surface segmentations have been possible, as boundaries tend to be well defined, a diagnostic model will require a greater degree of uniformity. The following should be considered when acquiring a data set for the purposes of creating a model for fetal face shape;

- The central slice of the 3D image should correspond to the centre of the face. The corpus callosum, a structure located between the two hemispheres of the brain, may be an appropriate landmark for selecting the central slice.

- Ideally, an image in which the fetus is exhibiting a neutral expression would be desirable, although this may be difficult to recognise from the standard views available during acquisition. At the very least, the mouth should be closed. While this is far simpler to identify, there are no corrective measures the clinician could practically take other than waiting and taking the volume again.

- Where possible, acquisitions which are symmetrical about the profile should be obtained. Additionally, the face should be perpendicular to the angle of the probe, since volumes obtained with one side of the face closer to the probe than the other may suffer from attenuation and result in a misleading, asymmetrical segmentation.

Through gestation, the ratio of feature size to speckle size varies, meaning that filters used for FA calculation may need to be altered with GA. Speckle size is defined by the tissue scatterer density and US probe settings. Ideally, settings could be chosen such that speckle size is an approximately known quantity, allowing for finely tuned filter scales. However, as US probes from varying manufacturers and with different properties are used in different healthcare centres, it may be inappropriate to enforce a set of standard probe settings. Additionally, factors such as maternal adipose tissue may result in varied image quality in real life applications of these methods.

As a result, the method needs to be made more robust to varying speckle-to-feature ratios. This could be done by detecting speckle size and adjusting the filter scale, or finding a suitable relationship between GA and filter scale and applying different filters to different fetuses accordingly. Alternatively, a method such as NL means could be used as a first-pass to reduce the impact of speckle, although appropriate parameters for this filter may themselves be subject to probe settings or GA.

7.4 Final Remarks

To summarise the findings of this thesis, a number of tools have been presented in order to facilitate the assessment of fetal health via biometric measurement. These efforts have focused around the adaptation of interactive segmentation methods, which give the user direct, intuitive control over where the boundaries of the object they wish to segment lie.

It is hoped that the tools outlined here can facilitate investigations into the use of more accurate measures of fetal health assessment, using a more suitable roster of fetal biometrics to determine nutritional development.

Chapter A

Abbreviations and Medical Terms

- **AAM:** Active Appearance Model
- **AC:** Abdominal Circumference
- **AGA:** Appropriate for Gestational Age
- **APAD:** Antero-Posterior Abdominal Diameter (longest dimension of abdomen)
- **ASM:** Active Shape Model
- **ASP Live Wire:** Automated Seed Point Live Wire
- **AVol:** Fractional Arm Volume
- **%BF:** Percentage Body Fat
- **BPD:** Biparietal Diameter (widest dimension of skull diameter)
- **DoG:** Difference of Gaussians
- **DSM:** Dense Shape/Surface Model
- **EFW:** Estimated Fetal Weight
- **FA:** Feature Asymmetry
- **FAS:** Fetal Alcohol Syndrome
- **FA-S Live Wire:** Feature Asymmetry and Shape guided Live Wire
- **FASP:** Fetal Anomaly Screening Programme
- **FC:** Fuzzy Connectedness
- **FGR:** Fetal Growth Restriction
- **FL:** Femur Length
- **FS:** Feature Symmetry
- **GA:** Gestational Age
- **HC:** Head Circumference
- **HELLP syndrome:** Potential complication of preeclampsia charac-

terised by **H**emolysis, **E**levated **L**iver enzymes and **L**ow **P**latelet count

- **ISUOG:** International Society of Ultrasound in Obstetrics and Gynecology
- **IUGR:** Intra Uterine Growth Restriction
- **IVUS:** Intravascular Ultrasound
- **LBP:** Local Binary Pattern
- **LE:** Local Energy
- **LO:** Local Orientation
- **LP:** Local Phase
- **LVol:** Liver Volume
- **Microcephaly:** Abnormally small head circumference
- **MRI:** Magnetic Resonance Imaging
- **N-cuts:** Normalized Cuts
- **OFD:** Occipito-Frontal Diameter (longest dimension of skull diameter)
- **Preeclampsia:** High blood pressure during pregnancy
- **RF:** Radio Frequency
- **ROC:** Receiver Operating Characteristic
- **SGA:** Small for gestational age
- **SNR:** Signal to Noise Ratio
- **TAD:** Transverse abdominal diameter (widest dimension of abdomen)
- **TVol:** Fractional Thigh Volume (also, Fractional Limb Volume)
- **UGR:** Uterine Growth Restriction
- **US:** Ultrasound

Chapter B

Subdivision Surfaces

B.1 Loop Subdivision

Loop subdivision is a method for subdivision based solely on triangular meshes [104]. New vertices are created from old vertices and the mid points of edges, their positions weighted by neighbouring points [104]. Suzuki *et al.* describe this as a process of simply "splitting and positioning" [180].

Figure B.1 shows how a new face is constructed using the Loop subdivision method from an existing face and its surrounding neighbourhood (shown in Figure B.1a). The new set of vertices used to define the new surface faces are obtained using the old vertices in addition to vertices created by splitting edges in half. This set of new vertices are repositioned based on a neighbourhood of old vertices. Those propagated from previous vertex positions are repositioned based on the connected vertices of the old mesh, as shown in Figure B.1b. Loop weights the influences of these vertices as:

$$V_{new} = \frac{5}{8}V_{old} + \frac{3}{8}Q, \tag{B.1}$$

where V_{old} and V_{new} are the old and new vertex positions respectively and Q is the average of the adjoining vertices [104]. Points created from edges are repositioned based on the location of the vertices which make up the two triangles lying on the edge in the old mesh, as shown in Figures B.1c and B.1d. Loop weights the influences of these vertices

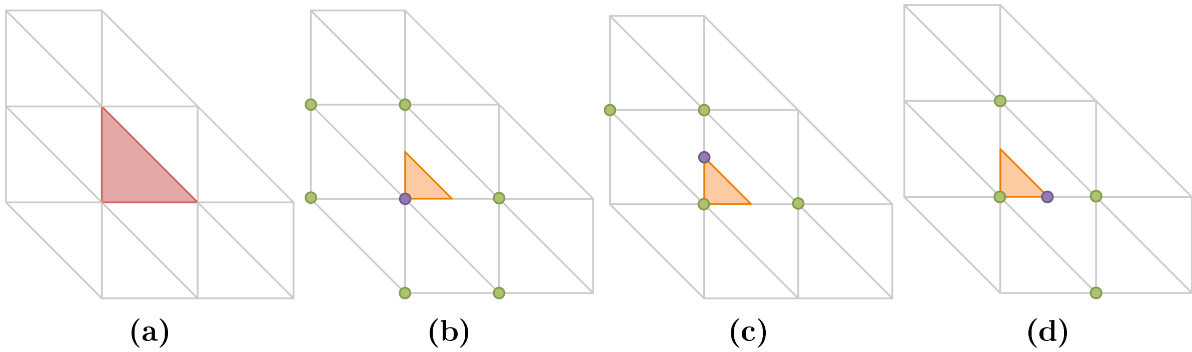


Figure B.1: *The steps involved in creating a new surface face following the Loop scheme of subdivision*

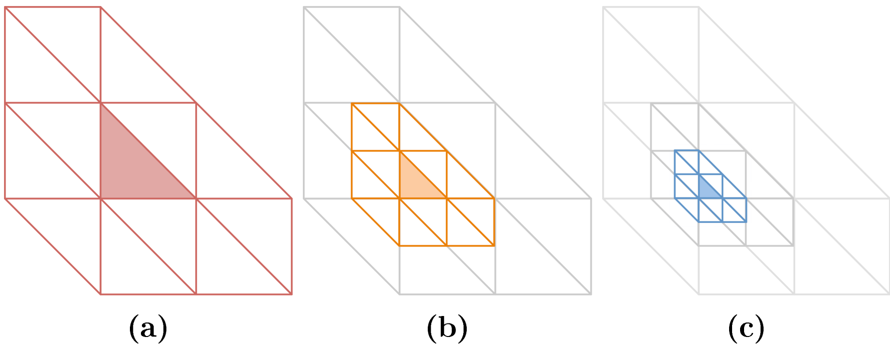


Figure B.2: *The vertex neighbourhoods required for successive subdivision iterations*

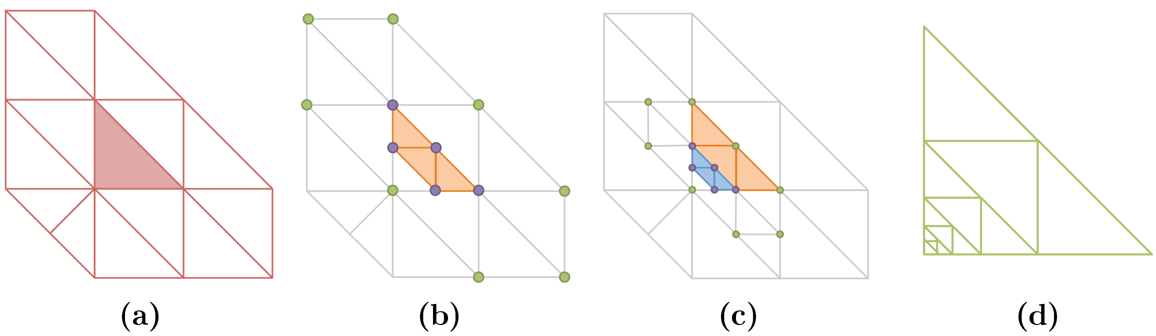


Figure B.3: ***(a)-(c)*** *Loop subdivision for an irregular surface mesh. A schematic of an infinite set of triangles which could be used to represent subdivision is shown in* ***(d)****.*

as:

$$V_{new} = \frac{3}{4}A + \frac{1}{4}O \tag{B.2}$$

where O and A are the averages of the opposite and adjacent vertices respectively [104]. By this process, each existing mesh face is refined into four new faces at each iteration.

In order to use subdivision meshes to describe an existing surface, in this case a surface created from the segmentation of the fetal face, the coarse mesh must be positioned such that the smooth subdivision surface best fits the segmentation data. Of course, this may be a difficult process if the underlying surface can only be obtained through successive subdivision of the coarse mesh after each modification of the mesh to fit the data. Stam presents a method for finding the outcome of Catmull-Clark subdivision without actually going through the process of multiple iterations of division, but instead expresses all subdivision surfaces for a given mesh in terms of a set of eigenbasis functions [173]. Stam later presented a method for evaluating the Loop subdivision surface from a coarse mesh following a similar logic [174].

Recall from Figure B.1 how a new Loop subdivision space is constructed from a neighbourhood of existing verticies. In order to fully define a face on a regular section of a triangular mesh, Stam points out that a neighbourhood of 12 vertices will be required, as shown in Figure B.2a [174]. Figure B.2b and Figure B.2c show how this process is repeated for the second and third Loop subdivision. However, as seen in Figure B.3a, it is possible for irregular cases to exist. This can be defined by an infinite set of triangles, the shape of which can be described by a set of principle modes of variation, i.e. an eigenbasis function [174].

It is this method that Stebbing has utilised for the calculation of Loop subdivision surfaces, allowing for surface points to be expressed explicitly in terms of the coarse mesh vertices [175].

Bibliography

[1] Safe passage: Pass network. *http://www.safepassagestudy.org/Default.aspx*, 2014.

[2] Interbio 21st. http://www.interbio21.org.uk/. 2010.

[3] R Achanta, A Shaji, K Smith, A Lucchi, P Fua, and S Süsstrunk. Slic superpixels. *École Polytechnique Fédéral de Lausssanne (EPFL), Tech. Rep*, 149300, 2010.

[4] S Agnihotri, H Loomba, A Gupta, and V Khandelwal. Automated segmentation of gallstones in ultrasound images. In *Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on*, pages 56–59. IEEE, 2009.

[5] T Ahonen, A Hadid, and M Pietikäinen. Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2037–2041, 2006.

[6] SJ Astley and SK Clarren. A case definition and photographic screening tool for the facial phenotype of fetal alcohol syndrome. *The Journal of Pediatrics*, 129(1):33–41, 1996.

[7] PN Baker, IR Johnson, PA Gowland, J Hykin, V Adams, P Mansfield, and BS Worthington. Measurement of fetal liver, brain and placental volumes with echo–planar magnetic resonance imaging. *BJOG – An International Journal of Obstetrics and Gynaecology*, 102(1):35–39, 1995.

[8] DJP Barker, KM Godfrey, PD Gluckman, JE Harding, JA Owens, and JS Robinson. Fetal nutrition and cardiovascular disease in adult life. *The Lancet*, 341(8850):938–941, 1993.

[9] A Belaid, D Boukerroui, Y Maingourd, and JF Lerallut. Phase based level set segmentation of ultrasound images. *ITAB 2009*, pages 1–4, 2009.

[10] ZA Bhuiyan, M Klein, P Hammond, A van Haeringen, MMAM Mannens, I Van Berckelaer-Onnes, and RCM Hennekam. Genotype–phenotype correlations of 39 patients with cornelia de lange syndrome: the dutch experience. *Journal of Medical Genetics*, 43(7):568–575, 2006.

[11] ZA Bhutta, for the International Fetal, and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). *BJOG – An International Journal of Obstetrics and Gynaecology, Special Issue: The Methodology of the INTERGROWTH–21st Project*, 120:1–142, 2013.

[12] Bitmapworld. http://www.turbosquid.com/3d–models/3d–max–little–girl–head/387777. *Turbo Squid*, 2008.

[13] JM Bland and DG Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 327(8476):307–310, 1986.

[14] V Blanz and T Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, pages 187–194. ACM Press/Addison–Wesley Publishing Co., 1999.

[15] S Boito, PC Struijk, NTC Ursem, L Fedele, and JW Wladimiroff. Fetal brain/liver volume ratio and umbilical volume flow parameters relative to normal and abnormal human development. *Ultrasound in Obstetrics and Gynecology*, 21(3):256–261, 2003.

[16] SME Boito, JAM Laudy, PC Struijk, T Stijnen, and JW Wladimiroff. Three–dimensional US assessment of hepatic volume, head circumference, and abdominal circumference in healthy and growth–restricted fetuses. *Radiology*, 223:661–665, 2002.

[17] JG Bosch, SC Mitchell, BPF Lelieveldt, F Nijland, O Kamp, M Sonka, and JHC Reiber. Automatic segmentation of echocardiographic sequences by active appearance motion models. *Medical Imaging, IEEE Transactions on*, 21(11):1374–1383, 2002.

[18] C Bottomley and T Bourne. Dating and growth in the first trimester. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 23(4):439–452, 2009.

[19] D Boukerroui, JA Noble, and M Brady. On the choice of band–pass quadrature filters. *Journal of Mathematical Imaging and Vision*, 21(1):53–80, 2004.

[20] YY Boykov and MP Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in nd images. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 105–112. IEEE, 2001.

[21] L Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[22] SM Bryan and PC Hindmarsh. Normal and abnormal fetal growth. *Hormone Research*, 65(3):19–27, 2006.

[23] A Buades, B Coll, and JM Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4(2):490–530, 2005.

[24] L Burd and JT Martsolf. Fetal alcohol syndrome: diagnosis and syndromal variability. *Physiology & Behavior*, 46(1):39–43, 1989.

[25] M Cannie, J Jani, F De Keyzer, F Van Kerkhove, J Meersschaert, L Lewi, J Deprest, and S Dymarkowski. Magnetic resonance imaging of the fetal lung: a pictorial essay. *European Radiology*, 18(7):1364–1374, 2008.

[26] PM Catalano, ED Tyzbir, SR Allen, JH McBean, and TL McAuliffe. Evaluation of fetal growth by estimation of neonatal body composition. *Obstetrics & Gynecology*, 79(1):46, 1992.

[27] E Catmull and J Clark. Recursively generated b–spline surfaces on arbitrary topological meshes. *Computer–Aided Design*, 10(6):350–355, 1978.

[28] R Chandrashekara, R Mohiaddin, RS Razavi, and D Rueckert. Nonrigid image registration with subdivision lattices: Application to cardiac mr image analysis. In *Medical Image Computing and Computer–Assisted Intervention–MICCAI 2007*, pages 335–342. 2007.

[29] CH Chang, CH Yu, FM Chang, HC Ko, and HY Chen. The assessment of normal fetal liver volume by three–dimensional ultrasound. *Ultrasound in Medicine and Biology*, 29(8):1123–1129, 2003.

[30] CH Chang, CH Yu, HC Ko, CL Chen, and FM Chang. Predicting fetal growth restriction with liver volume by three–dimensional ultrasound: Efficacy evaluation. *Ultrasound in Medicine & Biology*, 32(1):13–17, 2006.

[31] FM Chang, KF Hsu, HC Ko, BL Yao, CH Chang, CH Yu, and HY Chen. Three–dimensional ultrasound assessment of fetal liver volume in normal pregnancy: A comparison of reproducibility with two–dimensional ultrasound and a search for a volume constant. *Ultrasound in Medicine and Biology*, 23(3):381–389, 1997.

[32] Z Chang-ming, G Guo-chang, L Hai-bo, S Jing, and Y Hualong. Segmentation of ultrasound image based on texture feature and graph cut. In *Computer Science and Software Engineering, 2008 International Conference on*, volume 1, pages 795–798, 2008.

[33] CM Chen, HHS Lu, and YC Lin. An early vision–based snake model for ultrasound image segmentation. *Ultrasound in Medicine & Biology*, 26(2):273–285, 2000.

[34] HC Chen, PY Tsai, HH Huang, HH Shih, YY Wang, CH Chang, and YN Sun. Registration–based segmentation of three–dimensional ultrasound images for quantitative measurement of fetal craniofacial structure. *Ultrasound in Medicine & Biology*, 38(5):811–823, 2012.

[35] MM Chen, FV Coakley, A Kaimal, and RK Laros Jr. Guidelines for computed tomography and magnetic resonance imaging use during pregnancy and lactation. *Obstetrics & Gynecology*, 112(2, Part 1):333–340, 2008.

[36] SY Chen, HH Chang, SH Hung, and WC Chu. Breast tumor identification in ultrasound images using the normalized cuts with partial grouping constraints. In *BioMedical Engineering and Informatics, 2008. BMEI 2008. International Conference on*, volume 2, pages 28–32. IEEE, 2008.

[37] A Chodorowski, U Mattsson, M Langille, and G Hamarneh. Color lesion boundary detection using live wire. *Proceedings of SPIE, Image Processing*, pages 1589–1596, 2005.

[38] A Cifor, L Risser, D Chung, E Anderson, and J Schnabel. Hybrid feature–based diffeomorphic registration for tumour tracking in 2–d liver ultrasound images. *Medical Imaging, IEEE Transactions on*, PP(99):1–10, 2013.

[39] A Cifor, L Risser, D Chung, EM Anderson, and JA Schnabel. Hybrid feature–based log–demons registration for tumour tracking in 2–d liver ultrasound images. pages 724–727, 2012.

[40] A Ciurte, X Bresson, and MB Cuadra. A semi–supervised patch–based approach for segmentation of fetal ultrasound imaging. *Proceedings of Challenge US: Biometric Measurements from Fetal Ultrasound Images, ISBI 2012*, pages 5–7, 2012.

[41] T Cootes, ER Baldock, and J Graham. An introduction to active shape models. *Image Processing and Analysis*, pages 223–248, 2000.

[42] P Coupé, P Hellier, C Kervrann, and C Barillot. Bayesian non local means–based speckle filtering. In *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pages 1291–1294, 2008.

[43] P Coupé, P Hellier, C Kervrann, and C Barillot. Nonlocal means–based speckle filtering for ultrasound images. *Image Processing, IEEE Transactions on*, 18(10):2221–2229, 2009.

[44] P Coupé, P Yger, S Prima, P Hellier, C Kervrann, and C Barillot. An optimized blockwise nonlocal means denoising filter for 3–d magnetic resonance images. *Medical Imaging, IEEE Transactions on*, 27(4):425–441, 2008.

[45] J Cox-Brinkman, A Vedder, C Hollak, L Richfield, A Mehta, K Orteu, F Wijburg, and P Hammond. Three–dimensional face shape in fabry disease. *European Journal of Human Genetics*, 15(5):535–542, 2007.

[46] MJ Dauncey, G Gandy, and D Gairdner. Assessment of total body fat in infancy from skinfold thickness measurements. *British Medical Journal*, 52(3):223, 1977.

[47] EW Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959.

[48] D Doo and M Sabin. Behaviour of recursive division surfaces near extraordinary points. *Computer–Aided Design*, 10(6):356–360, 1978.

[49] RO Duda and PE Hart. Use of the hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11–15, 1972.

[50] S Esneault, N Hraiech, E Delabrousse, and JL Dillenseger. Graph cut liver segmentation for interstitial ultrasound therapy. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 5247–5250, 2007.

[51] AX Falcão and JK Udupa. A 3d generalization of user–steered live–wire segmentation. *Medical Image Analysis*, 4(4):389–402, 2000.

[52] AX Falcão, JK Udupa, S Samarasekera, S Sharma, BE Hirsch, and RA Lotufo. User–steered image segmentation paradigms: Live Wire and Live Lane. *Graphical Models and Image Processing*, 60(4):233–260, 1998.

[53] C Famy, AP Streissguth, and AS Unis. Mental illness in adults with fetal alcohol syndrome or fetal alcohol effects. *American Journal of Psychiatry*, 155(4):552–554, 1998.

[54] M Felsberg and G Sommer. The monogenic signal. *IEEE Transactions on Signal Processing*, 49:200–1, 2001.

[55] S Feng, SK Zhou, S Good, and D Comaniciu. Automatic fetal face detection from ultrasound volumes via learning 3d and 2d information. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2488–2495. IEEE, 2009.

[56] Y Freund and RE Schapire. Experiments with a new boosting algorithm. In *ICML*, volume 96, pages 148–156, 1996.

[57] JF Frøen, JO Gardosi, A Thurmann, A Francis, and B Stray-Pedersen. Restricted fetal growth in sudden intrauterine unexplained death. *Acta Obstetricia et Gynecologica Scandinavica*, 83(9):801–807, 2004.

[58] B Fulkerson, A Vedaldi, and S Soatto. Class segmentation and object localization with superpixel neighborhoods. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 670–677, 2009.

[59] F Gardeil, R Greene, B Stuart, and MJ Turner. Subcutaneous fat in the fetal abdomen as a predictor of growth restriction. *Obstetrics & Gynecology*, 94(2):209, 1999.

[60] C Garel, H Brisse, G Sebag, M Elmaleh, JF Oury, and M Hassan. Magnetic resonance imaging of the fetus. *Pediatric Radiology*, 28(4):201–211, 1998.

[61] H Gavin. The levenberg–marquardt method for nonlinear least squares curve–fitting problems. *Department of Civil and Environmental Engineering, Duke University*, pages 1–15, 2011.

[62] R Geva, R Eshel, Y Leitner, AF Valevski, and S Harel. Neuropsychological outcome of children with intrauterine growth restriction: a 9–year prospective study. *Pediatrics*, 118(1):91–100, 2006.

[63] Y Gielchinsky, M Zvanca, R Minekawa, N Persico, and KH Nicolaides. Liver volume in trisomy 21 and euploid fetuses at 11 to 13 weeks. *Prenatal Diagnosis*, 31(1):28–32, 2011.

[64] KM Godfrey and DJP Barker. Fetal nutrition and adult disease. *American Journal of Clinical Nutrition*, 71(5):1344S, 2000.

[65] JC Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, 1975.

[66] V Grau and JA Noble. Adaptive multiscale ultrasound compounding using phase information. *MICCAI 2005*, 3749:589–596, 2005.

[67] Me Gulmezoglu, M de Onis, and J Villar. Effectiveness of interventions to prevent or treat impaired fetal growth. *Obstetrical & Gynecological Survey*, 52(2):139–148, 1997.

[68] Z Guo, L Zhang, D Zhang, and X Mou. Hierarchical multiscale lbp for face and palmprint recognition. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 4521–4524, 2010.

[69] I Hacihaliloglu, R Abugharbieh, A Hodgson, and R Rohling. Bone segmentation and fracture detection in ultrasound using 3d local phase features. In *MICCAI 2008*, volume 5241 of *Lecture Notes in Computer Science*, pages 287–295. 2008.

[70] P Hammond, TJ Hutton, JE Allanson, B Buxton, LE Campbell, J Clayton-Smith, D Donnai, A Karmiloff-Smith, K Metcalfe, KC Murphy, M Patton, B Pober, K Prescott, P Scambler, A Shaw, ACM Smith, AF Stevens, IK Temple, R Hennekam, and M Tassabehji. Discriminating power of localized three–dimensional facial morphology. *The American Journal of Human Genetics*, 77(6):999–1010, 2005.

[71] F Hannes, P Hammond, O Quarrell, JP Fryns, K Devriendt, and JR Vermeesch. A microdeletion proximal of the critical deletion region is associated with mild wolf–hirschhorn syndrome. *American Journal of Medical Genetics Part A*, 158(5):996–1004, 2012.

[72] Z Hao, Q Wang, H Ren, K Xu, YK Seong, and J Kim. Multiscale superpixel classification for tumor segmentation in breast ultrasound images. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 2817–2820, 2012.

[73] T Hata and RL Deter. A review of fetal organ measurements obtained with ultrasound: Normal growth. *Journal of Clinical Ultrasound*, 20:155–174, 1992.

[74] I Heulens, M Suttie, A Postnov, N De Clerck, CS Perrotta, T Mattina, F Faravelli, F Forzano, RF Kooy, and P Hammond. Craniofacial characteristics of fragile x syndrome in mouse and man. *European Journal of Human Genetics*, 21(8):816–823, 2013.

[75] LM Hill, D Guzick, D Boyles, C Merolillo, A Ballone, and P Gmiter. Subcutaneous tissue thickness cannot be used to distinguish abnormalities of fetal growth. *Obstetrics & Gynecology*, 80(2):268, 1992.

[76] MC Ho, JJ Lin, YC Shu, CN Chen, KJ Chang, CC Chang, and PH Tsui. Using ultrasound nakagami imaging to assess liver fibrosis in rats. *Ultrasonics*, 52(2):215–222, 2012.

[77] H Homann, G Vesom, and JA Noble. Vasculature segmentation of ct liver images using graph cuts and graph–based analysis. In *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pages 53–56. IEEE, 2008.

[78] AM Howatson, PG Lund, and JD Todd. Engineering tables and data. *Oxford University*, page 7, 1991.

[79] DK Iakovidis, EG Keramidas, and D Maroulis. Fuzzy local binary patterns for ultrasound texture characterization. In *Image Analysis and Recognition*, pages 750–759. 2008.

[80] C Ioannou, I Sarris, LJ Salomon, and AT Papageorghiou. A review of fetal volumetry: the need for standardization and definitions in measurement methodology. *Ultrasound in Obstetrics & Gynecology*, 38(6):613–619, 2011.

[81] K Itoh, Y Yasuda, T Aihara, A Koyano, and T Konishi. Acoustic intensity histogram pattern diagnosis of liver diseases. *Journal of Clinical Ultrasound*, 13(7):449–456, 1985.

[82] SMGVB Jardim and MAT Figueiredo. Segmentation of fetal ultrasound images. *Ultrasound in Medicine & Biology*, 31(2):243–250, 2005.

[83] SS Jarvis, SV Glinianaia, MG Torrioli, MJ Platt, M Miceli, PS Jouk, A Johnson, J Hutton, K Hemming, G Hagberg, H Dolk, and J Chalmers. Cerebral palsy and intrauterine growth in single births: European collaborative study. *The Lancet*, 362(9390):1106–1111, 2003.

[84] KL Jones and DW Smith. Recognition of the fetal alcohol syndrome in early infancy. *The Lancet*, 302(7836):999–1001, 1973.

[85] JN Kapur, PK Sahoo, and AKC Wong. A new method for gray–level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing*, 29(3):273–285, 1985.

[86] M Kass, A Witkin, and D Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[87] E Keramidas, D Iakovidis, and D Maroulis. Fuzzy binary patterns for uncertainty–aware texture representation. *Electronic Letters on Computer Vision and Image Analysis*, 10(1):63–78, 2011.

[88] D. Kirwan and the NHS Fetal Anomaly Screening Programme (NHS FASP). Nhs fetal anomaly screening programme: 18+0 to 20+6 weeks fetal anomaly scan national standards and guidance for england. *fetalanomaly.screening.nhs.uk/*, 2010.

[89] CL Knight, M Ahmed, K Edwards, V Donadono, G Parry, S Rueda, JA Noble, and AT Papageorghiou. Op29. 05: Fetal arm fat: a better marker than thigh fat for nutritional status? *Ultrasound in Obstetrics & Gynecology*, 44(S1):133–134, 2014.

[90] CL Knight, M Ahmed, K Edwards, V Donadono, G Parry, S Rueda, JA Noble, and AT Papageorghiou. P13. 20: Fetal limb fat and lean volume reference ranges from an optimally healthy population. *Ultrasound in Obstetrics & Gynecology*, 44(S1):259–259, 2014.

[91] P Kovesi. Image features from phase congruency. *Technical Report 95/4*, 1995.

[92] P Kovesi. Symmetry and asymmetry from local phase. In *Tenth Australian Joint Conference on Artificial Intelligence*, volume 190, 1997.

[93] A Kuno, Y Hayashi, M Akiyama, C Yamashiro, H Tanaka, T Yanagihara, and T Hata. Three–dimensional sonographic measurement of liver volume in the small–for–gestational–age fetus. *Journal of Ultrasound in Medicine*, 21(4):361–366, 2002.

[94] A Kurjak, G Azumendi, W Andonotopo, and A Salihagic-Kadic. Three–and four–dimensional ultrasonography for the structural and functional evaluation of the fetal face. *American Journal of Obstetrics and Gynecology*, 196(1):16–28, 2007.

[95] A Larrue and JA Noble. Nakagami imaging with small windows. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pages 887–890, 2011.

[96] JAM Laudy, MMM Janssen, PC Struyk, T Stijnen, HCS Wallenburg, and JW Wladimiroff. Fetal liver volume measurement by three–dimensional ultrasonography: a preliminary study. *Ultrasound in Obstetrics and Gynecology*, 12(2):93–96, 1998.

[97] A Lee, J Deutinger, and G Bernaschek. Three dimensional ultrasound: abnormalities of the fetal face in surface and volume rendering mode. *BJOG: An International Journal of Obstetrics & Gynaecology*, 102(4):302–306, 1995.

[98] W Lee, M Balasubramaniam, RL Deter, SS Hassan, F Gotsch, JP Kusanovic, LF Gonçalves, and R Romero. Fetal growth parameters and birth weight: their relationship to neonatal body composition. *Ultrasound in Obstetrics and Gynecology*, 33(4):441–446, 2009.

[99] W Lee, M Balasubramaniam, RL Deter, SS Hassan, F Gotsch, JP Kusanovic, LF Gonçalves, and R Romero. Fractional limb volume–a soft tissue parameter of fetal body composition: validation, technical considerations and normal ranges during pregnancy. *Ultrasound in Obstetrics and Gynecology*, 33(4):427–440, 2009.

[100] W Lee, M Balasubramaniam, RL Deter, L Yeo, SS Hassan, F Gotsch, JP Kusanovic, LF Gonçalves, and R Romero. New fetal weight estimation models using fractional limb volume. *Ultrasound in Obstetrics and Gynecology*, 34(5):556–565, 2009.

[101] B Levienaise-Obadia and A Gee. Adaptive segmentation of ultrasound images. *Image and Vision Computing*, 17(8):583–588, 1999.

[102] S Liao, MWK Law, and ACS Chung. Dominant local binary patterns for texture classification. *Image Processing, IEEE Transactions on*, 18(5):1107–1118, 2009.

[103] RJ Lipinski, P Hammond, SK O'Leary-Moore, JJ Ament, SJ Pecevich, Y Jiang, F Budin, SE Parnell, M Suttie, EA Godin, JL Everson, DB Dehart, I Oguz, HT Holloway, MA Styner, GA Johnson, and KK Sulik. Ethanol–induced face–brain dysmorphology patterns are correlative and exposure–stage dependent. *PloS One*, 7(8):e43067, 2012.

[104] C Loop. Smooth subdivision surfaces based on triangles. 1987.

[105] K Lu and WE Higgins. Interactive segmentation based on the live wire for 3d ct chest image analysis. *International Journal of Computer Assisted Radiology and Surgery*, 2(3–4):151–167, 2007.

[106] R Malladi, JA Sethian, and BC Vemuri. Shape modeling with front propagation: A level set approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(2):158–175, 1995.

[107] T Maltaverne, P Delachartre, and A Basarab. Motion estimation using the monogenic signal applied to ultrasound elastography. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 33–36, 2010.

[108] L Matthews, AS Marais, HH Kay, and DL Viljoen. Oc178: Possible ultrasound markers for fetal alcohol syndrome: assessment of the fetal face and brain. *Ultrasound in Obstetrics & Gynecology*, 24(3):264–264, 2004.

[109] PA May, JP Gossage, WO Kalberg, LK Robinson, D Buckley, M Manning, and HE Hoyme. Prevalence and epidemiologic characteristics of fasd from various research methods with an emphasis on recent in–school studies. *Developmental Disabilities Research Reviews*, 15(3):176–192, 2009.

[110] C Mayer and KS Joseph. Fetal growth: a review of terms, concepts and issues relevant to obstetrics. *Ultrasound in Obstetrics and Gynecology*, 41:136–145, 2013.

[111] DJ Mellor and L Murray. Effects on the rate of increase in fetal girth of refeeding ewes after short periods of severe undernutrition during late pregnancy. *Research in Veterinary Science*, 32(3):377–382, 1982.

[112] M Mellor and M Brady. Phase mutual information as a similarity measure for registration. *Medical Image Analysis*, 9(4):330–343, 2005.

[113] M Mellor, W Zhang, R Ali, and T Szilagyi. "monogenic.m". *MATLAB Code*, 2009.

[114] I Mikic, S Krucinski, and JD Thomas. Segmentation and tracking in echocardiographic sequences: active contours guided by optical flow estimates. *Medical Imaging, IEEE Transactions on*, 17(2):274–284, 1998.

[115] SC Mitchell, JG Bosch, BPF Lelieveldt, RJ van der Geest, JHC Reiber, and M Sonka. 3–d active appearance models: segmentation of cardiac mr and ultrasound images. *Medical Imaging, IEEE Transactions on*, 21(9):1167–1178, 2002.

[116] GH Mochida and CA Walsh. Molecular genetics of human microcephaly. *Current Opinion in Neurology*, 14(2):151–156, 2001.

[117] M Mulet-Parada and JA Noble. 2D+T acoustic boundary detection in echocardiography. *Medical Image Analysis*, 4(1):21–30, 2000.

[118] RL Naeye. Placental infarction leading to fetal or neonatal death. *Obstetrics & Gynecology*, 50(5), 1977.

[119] AIL Namburete and JA Noble. Fetal cranial segmentation in 2d ultrasound images using shape properties of pixel clusters. In *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, pages 720–723. IEEE, 2013.

[120] AIL Namburete, B Rahmatullah, and JA Noble. Nakagami–based adaboost learning framework for detection of anatomical landmarks in 2d fetal neurosonograms.

[121] AIL. Namburete, RV Stebbing, and JA Noble. Cranial parametrization of the fetal head for 3d ultrasound image analysis. In *MIUA*, pages 196–201, 2013.

[122] AIL Namburete, M Yaqub, B Kemp, AT Papageorghiou, and Noble JA. Predicting fetal neurodevelopmental age from ultrasound images. *MICCAI 2014*, 2014.

[123] L Nanni, S Brahnam, and A Lumini. A simple method for improving local binary patterns by considering non–uniform patterns. *Pattern Recognition*, 45(10):3844–3852, 2012.

[124] LMM Nardozza, E Araujo Júnior, MM Barbosa, ACR Caetano, DJR Lee, and AF Moron. Fetal growth restriction: current knowledge to the general obs/gyn. *Archives of Gynecology and Obstetrics*, 286(1):1–13, 2012.

[125] TR Nelson and DH Pretorius. Three–dimensional ultrasound of fetal surface features. *Ultrasound in Obstetrics & Gynecology*, 2(3):166–174, 1992.

[126] J Ng, R Rohling, and PD Lawrence. Automatic measurement of human subcutaneous fat with ultrasound. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 56(8):1642–1653, 2009.

[127] MD Nihal al Riyami, MG Walker, LK Proctor, and Y Yinon. Utility of head/abdomen circumference ratio in the evaluation of severe early–onset intrauterine growth restriction. *J Obstet Gynaecol Can*, 33(7):715–719, 2011.

[128] JA Noble and D Boukerroui. Ultrasound image segmentation: a survey. *Medical Imaging, IEEE Transactions on*, 25(8):987–1010, 2006.

[129] L O'Donnell. Semi–automatic medical image segmentation. *Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science*, 2001.

[130] L O'Donnell, CF Westin, W Grimson, J Ruiz-Alzola, M Shenton, and R Kikinis. Phase–based user–steered image segmentation. *MICCAI 2001*, pages 1022–1030, 2001.

[131] T Ojala and M Pietikäinen. Unsupervised texture segmentation using feature distributions. *Pattern Recognition*, 32(3):477–486, 1999.

[132] T Ojala, M Pietikäinen, and D Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, 1996.

[133] T Ojala, M Pietikäinen, and T Mäenpää. Gray scale and rotation invariant texture classification with local binary patterns. volume 1842 of *Lecture Notes in Computer Science*, pages 404–420. 2000.

[134] T Ojala, M Pietikäinen, and T Maenpaa. Multiresolution gray–scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, 2002.

[135] V Ojansivu and J Heikkilä. Blur insensitive texture classification using local phase quantization. In *Image and Signal Processing*, pages 236–243. 2008.

[136] SD Olabarriaga and AWM Smeulders. Interaction in the segmentation of medical images: A survey. *Medical Image Analysis*, 5(2):127–142, 2001.

[137] F Orderud and SI Rabben. Real–time 3d segmentation of the left ventricle using deformable subdivision surfaces. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[138] S Osher and JA Sethian. Fronts propagating with curvature–dependent speed: algorithms based on hamilton–jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, 1988.

[139] AT Papageorghiou, EO Ohuma, DG Altman, T Todros, LC Ismail, A Lambert, YA Jaffer, E Bertino, MG Gravett, M Purwar, JA Noble, R Pang, CG Victora, FC Barros, M Carvalho, LJ Salomon, ZA Bhutta, SH Kennedy, and J Villar. International standards for fetal growth based on serial ultrasound measurements: the fetal growth longitudinal study of the intergrowth–21st project. *The Lancet*, 384(9946):869–879, 2014.

[140] N Paragios. A level set approach for shape–driven segmentation and tracking of the left ventricle. *Medical Imaging, IEEE Transactions on*, 22(6):773–776, 2003.

[141] N Paragios, MP Jolly, M Taron, and R Ramaraj. Active shape models and segmentation of the left ventricle in echocardiography. In *Scale Space and PDE Methods in Computer Vision*, pages 131–142. 2005.

[142] M Pietikäinen, T Ojala, and Z Xu. Rotation–invariant texture classification using feature distributions. *Pattern Recognition*, 33(1):43–52, 2000.

[143] M Poon, G Hamarneh, and R Abugharbieh. Efficient interactive 3d livewire segmentation of complex objects with arbitrary topology. *Computerized Medical Imaging and Graphics*, 32(8):639–650, 2008.

[144] O Pujol, D Rotger, P Radeva, O Rodriguez, and J Mauri. Near real–time plaque segmentation of ivus. In *Computers in Cardiology, 2003*, pages 69–72, 2003.

[145] L Quan, D Zhang, Y Yang, Y Liu, and Q Qin. Segmentation of tumor ultrasound image via region–based ncut method. *Wuhan University Journal of Natural Sciences*, 18(4):313–318, 2013.

[146] H Qureshi, O Sertel, N Rajpoot, R Wilson, and M Gurcan. Adaptive discriminant wavelet packet transform and local binary patterns for meningioma subtype classification. In *MICCAI 2008*, pages 196–204. 2008.

[147] B Rahmatullah, AT Papageorghiou, and JA Noble. Integration of local and global features for anatomical object detection in ultrasound. In *MICCAI 2012*, volume 7512 of *Lecture Notes in Computer Science*, pages 402–409. 2012.

[148] K Rajpoot, V Grau, and JA Noble. Local–phase based 3D boundary detection using monogenic signal and its application to real–time 3D echocardiography images. In *Proceedings of the Sixth IEEE International Conference on Symposium on Biomedical Imaging: From Nano to Macro*, pages 783–786. Institute of Electrical and Electronics Engineers Inc., The, 2009.

[149] K Rajpoot, JA Noble, V Grau, and N Rajpoot. Feature detection from echocardiography images using Local Phase information. *Proceedings of 12th Medical Image Understanding and Analysis*, 2008.

[150] S Ratnapalan, Y Bentur, and G Koren. Doctor, will that x–ray harm my unborn child? *Canadian Medical Association Journal*, 179(12):1293–1296, 2008.

[151] X Ren and J Malik. Learning a classification model for segmentation. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 10–17, 2003.

[152] HP Robinson and JEE Fleming. A critical evaluation of sonar crown–rump length measurements. *BJOG: An International Journal of Obstetrics & Gynaecology*, 82(9):702–710, 1975.

[153] D Rotten and JM Levaillant. Two–and three–dimensional sonographic assessment of the fetal face. 1. a systematic analysis of the normal face. *Ultrasound in Obstetrics & Gynecology*, 23(3):224–231, 2004.

[154] S Rueda. The use of ultrasound in the assessment of fetal growth and development: A survey. *(Thesis)*, 2010.

[155] S Rueda, C Knight, AT Papageorghiou, and JA Noble. Local phase–based Fuzzy Connectedness segmentation of ultrasound images. *Medical Image Understanding and Analysis, Proceedings of*, 2011.

[156] S Rueda, C Knight, AT Papageorghiou, and JA Noble. Regularised feature–based fuzzy connectedness segmentation of ultrasound images for fetal soft tissue quantification across gestation. *IEEE International Symposium on Biomedical Imaging (ISBI) 2012: From Nano to Macro*, pages 1323–1326, 2012.

[157] LJ Salomon, Z Alfirevic, V Berghella, C Bilardo, E Hernandez-Andrade, SL Johnsen, K Kalache, K-Y Leung, G Malinger, H Munoz, F Prefumo, A Toi, and W; Lee. Practice guidelines for performance of the routine mid–trimester fetal ultrasound scan. *Ultrasound in Obstetrics & Gynecology*, 37(1):116–126, 2011.

[158] LJ Salomon, Z Alfirevic, V Berghella, C Bilardo, E Hernandez-Andrade, SL Johnsen, K Kalache, KY Leung, G Malinger, and H Munoz. Practice guidelines for performance of the routine mid–trimester fetal ultrasound scan. *Ultrasound in Obstetrics and Gynecology*, 37(1):116–126, 2010.

[159] I Sarris, C Ioannou, P Chamberlain, E Ohuma, F Roseman, L Hoch, DG Altman, and AT Papageorghiou. International fetal and newborn growth consortium for the 21st century (intergrowth–21st). intra–and interobserver variability in fetal ultrasound measurements. *Ultrasound Obstet Gynecol*, 39:266–73, 2012.

[160] A Schenk, G Prause, and HO Peitgen. Efficient semiautomatic segmentation of 3d objects in medical images. In *MICCAI 2000*, pages 186–195, 2000.

[161] J Schwartz and H Galan. Ultrasound in assessment of fetal growth disorders: is there a role for subcutaneous measurements? *Ultrasound in Obstetrics and Gynecology*, 22(4):329–335, 2003.

[162] P Shah and J Kingdom. Long–term neurocognitive outcomes of sga/iugr infants. *Obstetrics, Gynaecology and Reproductive Medicine*, 21(5):142–146, 2011.

[163] PM Shankar. A model for ultrasonic scattering from tissues based on the k distribution. *Physics in Medicine and Biology*, 40(10):1633, 1995.

[164] PM Shankar. A general statistical model for ultrasonic backscattering from tissues. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 47(3):727–736, 2000.

[165] PM Shankar, VA Dumane, JM Reid, V Genis, F Forsberg, CW Piccoli, and BB Goldberg. Classification of ultrasonic b–mode images of breast masses using nakagami distribution. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 48(2):569–580, 2001.

[166] D Shen, Y Zhan, and C Davatzikos. Segmentation of prostate boundaries from ultrasound images using statistical shape model. *Medical Imaging, IEEE Transactions on*, 22(4):539–551, 2003.

[167] C Sheridan. Intrauterine growth restriction: diagnosis and management. *Australian Family Physician*, 34(9):717–723, 2005.

[168] J Shi and J Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, 2000.

[169] RJM Snijders, P Noble, N Sebire, A Souka, and KH Nicolaides. Uk multicentre project on assessment of risk of trisomy 21 by maternal age and fetal nuchal–translucency thickness at 10–14 weeks of gestation. *The Lancet*, 352(9125):343–346, 1998.

[170] JD Sonek, D McKenna, D Webb, C Croom, and K Nicolaides. Nasal bone length throughout gestation: normal ranges based on 3537 fetal ultrasound measurements. *Ultrasound in Obstetrics & Gynecology*, 21(2):152–155, 2003.

[171] L Sørensen, SB Shaker, and M De Bruijne. Quantitative analysis of pulmonary emphysema using local binary patterns. *Medical Imaging, IEEE Transactions on*, 29(2):559–569, 2010.

[172] HL Spohr, J Willms, and HC Steinhausen. Prenatal alcohol exposure and long–term developmental consequences. *The Lancet*, 341(8850):907–910, 1993.

[173] J Stam. Exact evaluation of catmull–clark subdivision surfaces at arbitrary parameter values. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, pages 395–404. ACM, 1998.

[174] J Stam. Evaluation of loop subdivision surfaces. *SIGGRAPH'99 Course Notes*, 1999.

[175] RV Stebbing, AIL Namburete, R Upton, P Leeson, and JA Noble. Data–driven shape parameterization for segmentation of the right ventricle from 3d+ t echocardiography. *Medical image analysis*, 21(1):29–39, 2015.

[176] AP Streissguth, JM Aase, SK Clarren, SP Randels, RA LaDue, and DF Smith. Fetal alcohol syndrome in adolescents and adults. *Jama*, 265(15):1961–1967, 1991.

[177] AP Streissguth, HM Barr, J Kogan, and FL Bookstein. Understanding the occurrence of secondary disabilities in clients with fetal alcohol syndrome (fas) and fetal alcohol effects (fae). *Final Report to the Centers for Disease Control and Prevention (CDC)*, pages 96–06, 1996.

[178] A Suhag and V Berghella. Intrauterine growth restriction (iugr): Etiology and diagnosis. *Current Obstetrics and Gynecology Reports*, pages 1–10, 2013.

[179] M Suttie, T Foroud, L Wetherill, JL Jacobson, CD Molteno, EM Meintjes, HE Hoyme, N Khaole, LK Robinson, EP Riley, SW Jacobson, and P Hammond. Facial dysmorphism across the fetal alcohol spectrum. *Pediatrics*, 131(3):e779–e788, 2013.

[180] H Suzuki, S Takeuchi, and T Kanai. Subdivision surface fitting to a range of points. In *Computer Graphics and Applications, 1999. Proceedings. Seventh Pacific Conference on*, pages 158–167, 1999.

[181] VW Swayze, VP Johnson, JW Hanson, J Piven, Y Sato, JN Giedd, D Mosnik, and NC Andreasen. Magnetic resonance imaging of brain anomalies in fetal alcohol syndrome. *Pediatrics*, 99(2):232–240, 1997.

[182] M Tassabehji, P Hammond, A Karmiloff-Smith, P Thompson, SS Thorgeirsson, ME Durkin, NC Popescu, T Hutton, K Metcalfe, A Rucka, H Stewart, AP Read, M Maconochie, and D Donnai. Gtf2ird1 in craniofacial development of humans and mice. *Science*, 310(5751):1184–1187, 2005.

[183] J Taylor, R Stebbing, V Ramakrishna, C Keskin, J Shotton, S Izadi, A Hertzmann, and A Fitzgibbon. User–specific hand modeling from monocular depth sequences. *Computer Vision and Pattern Recognition (CVPR)*, 2014.

[184] JL Tobin, M Di Franco, E Eichers, H May-Simera, M Garcia, J Yan, R Quinlan, MJ Justice, RC Hennekam, J Briscoe, M Tada, R Mayor, AJ Burns, JR Lupski, P Hammond, and PL Beales. Inhibition of neural crest migration underlies craniofacial dysmorphology and hirschsprung's disease in bardet–biedl syndrome. *Proceedings of the National Academy of Sciences*, 105(18):6714–6719, 2008.

[185] A Top, G Hamarneh, and R Abugharbieh. Spotlight: Automated confidence–based user guidance for increasing efficiency in interactive 3d image segmentation. In *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging*, pages 204–213. 2011.

[186] K Tunón, SH Eik-Nes, P Grø ttum, V Von Düring, and JA Kahn. Gestational age in pregnancies conceived after in vitro fertilization: a comparison between age assessed from oocyte retrieval, crown–rump length and biparietal diameter. *Ultrasound in Obstetrics and Gynecology*, 15(1):41–46, 2000.

[187] D Unay, A Ekin, M Cetin, R Jasinschi, and A Ercil. Robustness of local binary patterns in brain mr image analysis. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 2098–2101, 2007.

[188] D Unay, A Ekin, and RS Jasinschi. Local structure–based region–of–interest retrieval in brain mr images. *Information Technology in Biomedicine, IEEE Transactions on*, 14(4):897–903, 2010.

[189] A Vedaldi and S Soatto. Quick shift and kernel methods for mode seeking. In *Computer Vision–ECCV 2008*, pages 705–718. 2008.

[190] BO Verburg, EAP Steegers, M De Ridder, RJM Snijders, E Smith, A Hofman, HA Moll, VWV Jaddoe, and JCM Witteman. New charts for ultrasound dating of pregnancy and assessment of fetal growth: longitudinal data from a population–based cohort study. *Ultrasound in Obstetrics & Gynecology*, 31(4):388–396, 2008.

[191] G Vesom. *Poisson–based implicit shape space analysis with application to CT liver segmentation.* PhD thesis, Oxford University Brasenose College, 2010.

[192] OA Viegas, TJ Cole, and BA Wharton. Impaired fat deposition in pregnancy: an indicator for nutritional intervention. *American Journal of Clinical Nutrition*, 45(1):23, 1987.

[193] J Villar, LC Ismail, CG Victora, EO Ohuma, E Bertino, DG Altman, A Lambert, AT Papageorghiou, M Carvalho, YA Jaffer, MG Gravett, M Purwar, IO Frederick, JA Noble, R Pang, FC Barros, C Chumlea, ZA Bhutta, and SH Kennedy. International standards for newborn weight, length, and head circumference by gestational age and sex: the newborn cross–sectional study of the intergrowth–21st project. *The Lancet*, 384(9946):857–868, 2014.

[194] J Villar, AT Papageorghiou, R Pang, EO Ohuma, LC Ismail, FC Barros, A Lambert, M Carvalho, YA Jaffer, E Bertino, MG Gravett, DG Altman, M Purwar, IO Frederick, JA Noble, CG Victora, ZA Bhutta, and SH Kennedy. The likeness of fetal growth and newborn size across non–isolated populations in the intergrowth–21st project: the fetal growth longitudinal study and newborn cross–sectional study. *The Lancet Diabetes & Endocrinology*, 2(10):781–792, 2014.

[195] RF Wagner, MF Insana, and DG Brown. Statistical properties of radio–frequency and envelope–detected signals with applications to medical ultrasound. *JOSA A*, 4(5):910–922, 1987.

[196] RF Wagner, SW Smith, JM Sandrik, and H Lopez. Statistics of speckle in ultrasound b–scans. *IEEE Transactions on Sonics and Ultrasonics*, 30(3):156–163, 1983.

[197] J Wan and S Zhou. Features extraction based on wavelet packet transform for b–mode ultrasound liver images. In *Image and Signal Processing (CISP), 2010 3rd International Congress on*, volume 2, pages 949–955, 2010.

[198] L Wang and DC He. Texture classification using texture spectrum. *Pattern Recognition*, 23(8):905–910, 1990.

[199] N Wiest-Daesslé, S Prima, P Coupé, SP Morrissey, and C Barillot. Non–local means variants for denoising of diffusion–weighted and diffusion tensor mri. In *MICCAI 2007*, pages 344–351. 2007.

[200] N Wiest-Daesslé, S Prima, P Coupé, SP Morrissey, and C Barillot. Rician noise removal by non–local means filtering for low signal–to–noise ratio mri: Applications to dt–mri. In *MICCAI 2008*, pages 171–179. 2008.

[201] J Woo, BW Hong, CH Hu, KK Shung, CCJ Kuo, and PJ Slomka. Non–rigid ultrasound image registration based on intensity and local phase information. *Journal of Signal Processing Systems*, 54(1–3):33–43, 2009.

[202] M Yaqub, C Ioannou, A Papageorghiou, MK Javaid, C Cooper, and JA Noble. Improving boundary definition for 3D ultrasound quantification of fetal femur. *Proceedings of Medical Image Analysis and Understanding*, 2010.

[203] A Yezzi Jr, S Kichenassamy, A Kumar, P Olver, and A Tannenbaum. A geometric snake model for segmentation of medical imagery. *Medical Imaging, IEEE Transactions on*, 16(2):199–209, 1997.

[204] C Yu, C Zhang, and L Xie. A multiplicative nakagami speckle reduction algorithm for ultrasound images. *Multidimensional Systems and Signal Processing*, 23(4):499–513, 2012.

[205] W Zhang, JA Noble, and JM Brady. Real time 3–d ultrasound to mr cardiovascular image registration using a phase–based approach. In *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*, pages 666–669, 2006.

[206] M Zouqi and J Samarabandu. Prostate segmentation from 2–d ultrasound images using graph cuts and domain knowledge. In *Computer and Robot Vision, 2008. CRV'08. Canadian Conference on*, pages 359–362. IEEE, 2008.