

PROCEEDINGS OF SPIE

SPIDigitalLibrary.org/conference-proceedings-of-spie

Motion induced segmentation of stone fragments in ureteroscopy video

Soumya Gupta, Sharib Ali, Louise Goldsmith, Ben Turney, Jens Rittscher

Soumya Gupta, Sharib Ali, Louise Goldsmith, Ben Turney, Jens Rittscher, "Motion induced segmentation of stone fragments in ureteroscopy video," Proc. SPIE 11315, Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling, 1131514 (16 March 2020); doi: 10.1117/12.2549657

SPIE.

Event: SPIE Medical Imaging, 2020, Houston, Texas, United States

Motion induced segmentation of stone fragments in ureteroscopy video

Soumya Gupta^a, Sharib Ali^a, Louise Goldsmith^b, Ben Turney^b, and Jens Rittscher^a

^aInstitute of Biomedical Engineering, Old Road Campus, University of Oxford, Oxford, UK

^bDepartment of Urology, The Churchill, Oxford University Hospitals NHS Trust, Oxford, UK

ABSTRACT

Ureteroscopy is a conventional procedure used for localization and removal of kidney stones. Laser is commonly used to fragment the stones until they are small enough to be removed. Often, the surgical team faces tremendous challenge to successfully perform this task, mainly due to poor image quality, presence of floating debris and occlusions in the endoscopy video. Automated localization and segmentation can help to perform stone fragmentation efficiently. However, the automatic segmentation of kidney stones is a complex and challenging procedure due to stone heterogeneity in terms of shape, size, texture, color and position. In addition, dynamic background, motion blur, local deformations, occlusions, varying illumination conditions and visual clutter from the stone debris make the segmentation task even more challenging. In this paper, we present a novel illumination invariant optical flow based segmentation technique. We introduce a multi-frame based dense optical flow estimation in a primal-dual optimization framework embedded with a robust data-term based on normalized correlation transform descriptors. The proposed technique leverages the motion fields between multiple frames reducing the effect of blur, deformations, occlusions and debris; and the proposed descriptor makes the method robust to illumination changes and dynamic background. Both qualitative and quantitative evaluations show the efficacy of the proposed method on ureteroscopy data. Our algorithm shows an improvement of 5-8% over all evaluation metrics as compared to the previous method. Our multi-frame strategy outperforms classically used two-frame model.

Keywords: ureteroscopy, kidney stone removal, optical flow, illumination invariance, correlation-transform descriptor

1. INTRODUCTION

In the past 20 years, we have seen an increase in the number of kidney stone disease.¹ Kidney stones are formed when minerals in the urine separate and get aggregated inside the kidney or ureter. Most stones smaller than 5 mm will pass while stones larger than 5 mm may cause blockage of ureter and severe pain in the abdomen or lower back.² Abdominal X-ray, CT, Ultrasound are used for imaging of kidney stones with CT providing the most accurate result and hence is most commonly used.³ Once the stone is located, ultrasound or laser energy is used to disintegrate it into smaller fragments that are subsequently removed or further broken into smaller pieces to be flushed out in urine. The location, size and kind of stone decides the kind of therapeutic procedure required. The treatment options for kidney stones include percutaneous nephrolithotomy (PCNL), ureteroscopy, and extra-corporeal shockwave lithotripsy (ESWL).² PCNL is the most invasive technique recommended for very large stones or in case of failure from previous surgeries; ESWL is effective for small stones; and Ureteroscopy techniques are recommended for cases when size of calculi falls in the range 10-20 mm.⁴

Ureteroscopy has evolved into a routine technique for treatment of kidney stones and diagnosis of the pathology of upper urinary tract.⁵ Such procedure is performed using a long, thin, flexible tube that consists of light and a camera. The scope has a working channel through which tools like laser fibre can be inserted for stone fragmentation. Real-time video signal is produced by the endoscope and is available to the surgical team in real-time to guide their actions. Real time estimation of stone size is the most important parameter during ureteroscopy as it determines if the stone requires further fragmentation or can be directly removed. Automated segmentation is necessary for estimating the size of a given stone.

Previously, successful methods for automated detection and segmentation of kidney stones in Ultrasound (US) images have been proposed.^{6,7} To deal with noisy US images an extensive pre-processing step was first applied and then a level-set based segmentation was adopted to segment the kidney and kidney stones.⁶ Watershed segmentation has also been used for the segmentation of renal calculi.⁷ KNN and SVM classification techniques have been employed for the analysis of kidney stones from US images.⁸ Unlike US, ureteroscopic image quality is deteriorated by the kidney movement due to patient breathing, and the irrigation fluid which is continuously passed throughout the procedure.^{4,9} A region growing algorithm for segmentation of renal calculi on ureteroscopic images was proposed.⁴ However, such approaches are computationally expensive and requires the user to define a set of seed pixels, a similarity criterion, and a stopping criterion. The similarity criterion depends on difference in color values of the already found region and region to be examined. But, there exists a large variability in both texture, color and position of kidney stones. To the best of our knowledge not much work has been reported on segmentation of kidney stones in ureteroscopy images. This is due to the challenges and complexities involved in ureteroscopy data. The automatic segmentation of kidney stones is a complex and challenging task due to severe occlusions and stone heterogeneity in terms of shape, texture, color and position. Endoscopy video are often of poor quality, corrupted with noise, motion blur, and consists of stone debris obscuring the vision. Varying illumination conditions and large refractory errors from the irrigation fluid further increase the level of difficulty of the segmentation tasks.

Optical flow is a well established computer vision technique that has been successfully used in a wide variety of applications including image segmentation.¹⁰⁻¹² It is defined as an approximation of image motion based on local derivatives in a given frame sequence. Dense optical flow technique¹³ involves minimizing an energy term that consists of a data term and a regularization term. Given, a source image g_s and a target image g_t dense optical flow field $\mathbf{u} = (\mathbf{u}_x, \mathbf{u}_y)$, $\mathbf{u} \in \mathbb{R}^2$ from g_s to g_t is calculated by minimizing the energy function $E(\mathbf{u})$ given by:

$$E(\mathbf{u}) = \lambda \psi_{data}(g_s, g_t, \mathbf{u}) + \phi_{reg}(\mathbf{u}) \quad (1)$$

where ψ_{data} is the data-term that reflects the similarity of the pixels between the two images, ϕ_{reg} is the regularization term that assumes smoothness of the solution \mathbf{u} , and λ parameter controls the relative weighting of the data term and the regularization term. The data term in the original framework¹³ is based on the assumption that the brightness is constant across the frames. However, such consistency assumptions do not hold in real world applications especially in cases of medical images due to difference in imaging protocols.

Clustering of feature space formed by image intensity values and optical flow vectors has been used to provide spatio-temporal segmentation of MR image sequences.¹⁰ A hybrid technique employing gradient based optical flow and shape-based matching was proposed for detection of left ventricle motion and segmentation of a beating human heart.¹¹ A combination of active contour segmentation and dense deformation field from optical flow in a level-set framework was adopted to perform segmentation and atlas-registration.¹² Optical flow methods based on handcrafted image features have been quite successful in handling variances in illumination. Complete rank transform (CRT) based on census transform was proposed¹⁴ which involved descriptor build by counting number of pixels in a local patch that has smaller intensity w.r.t. to the center pixel. However, census transform fails to distinguish between dark and bright regions in a neighbourhood.¹⁵ A robust illumination invariant optical flow algorithm utilizing self similarity MIND descriptors was proposed.¹⁶ While it was shown that utilizing normalised correlation transform based similarity metric is more robust to illumination changes for computing deformation fields between histology brain slices and polarized light microscopy images.¹⁷

In this paper, we present a novel multi-frame total-variational optical flow based approach for segmenting stone fragments in ureteroscopy data. Such an approach benefits from relative camera motion estimation and hence can also be used to determine the motion of the stone for accurate laser targeting. Due to the presence of large illumination variation from view-point changes and non-linear reflections from debris, the underlying optical flow computation is largely affected. To deal with this problem, we propose an illumination-invariant optical flow method that is also robust to dynamic illumination and/or presence of debris. To strengthen it further, we incorporate a three frame strategy instead of conventional two-frame approach to exploit maximum temporal information between frames that can eventually handle other problems such as occlusions.

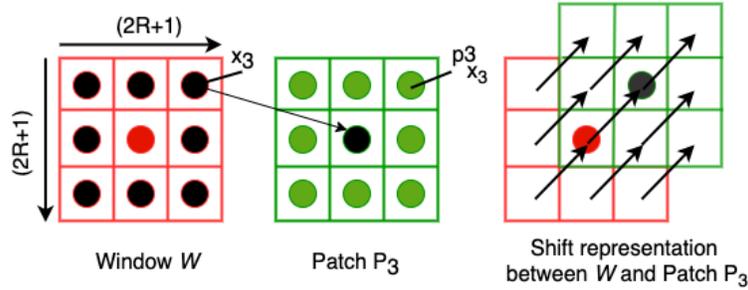


Figure 1. a) \mathbf{x}_i represented by black dots ($i \in [1,8]$) are the 8 connected neighbours of the center pixel \mathbf{x} (represented by red dot) in window W . b) A patch P_i centered at third element \mathbf{x}_3 of window W with green dots representing the neighbouring pixels $\mathbf{p}_{\mathbf{x}_3}^i$. c) Illustration of translation between corresponding pixels \mathbf{x}_i and $\mathbf{p}_{\mathbf{x}_3}^i$ ($i \in [1, 8]$) of window W and patch P_3 respectively.

2. METHOD

Our proposed algorithm employs normalised correlation transform (NCoT) based descriptors to formulate the data-term and classical bilateral filtering to enforce regularization of the flow field. The formulated data term ψ_{data} constrained with regularization ϕ_{reg} is then minimized using a first-order primal dual approach in a coarse-to-fine strategy that allows to handle large displacements. We have further extended this approach to a multi-frame model. This helps to efficiently overcome the effect of debris, occlusion and also deal with frames having insignificant or no motion at all.

Our approach utilizes the illumination invariant optical flow framework.¹⁶ To improve the accuracy and robustness of the stone segmentation we introduce the following new terms: 1) use of robust correlation transform (CoT) based descriptors as self-similarity metric for data-term, 2) extension of classically used two-frame model to a multi-frame model to deal with very small motion and occlusions, 3) normalization of CoT descriptors (NCoT) using clipped standard deviation within local neighbourhoods to eliminate large deviations and minimize the effect of debris on flow field computations, and 4) application of minimum variance quantization on computed motion field for region aggregation to provide more pronounced segmentation of stone fragments. Our experiments demonstrate the effectiveness of our proposed method compared to the previous technique¹⁶ on both a two-frame approach and a multi-frame approach.

2.1 Self-similarity descriptors

Self-similarity descriptors are vectors computed for an n -connected neighbourhood window W of radius R and size $(2R+1) \times (2R+1)$ (i.e., with $(2R+1)^2$ pixels and connectivity, $n = \{(2R+1)^2 - 1\}$, $\forall R \geq 1$) centered at pixel \mathbf{x} . Fig. 1(a) shows an example where $R=1$, then window W is of size 3×3 and $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_8\}$ are the neighbourhood pixels (represented by black dots) of center pixel \mathbf{x} (represented as a red dot), where each black dot has an associated patch P_i centered around it. Fig. 1(b) represents patch P_3 centered on \mathbf{x}_3 with its 8 neighbouring pixels $\mathbf{p}_{\mathbf{x}_3}^i$ (represented by green dots). The descriptors are computed for all pixels in window W using their corresponding patch pixels centered at each pixel. Fig. 1(c) illustrates the translation between window W centered on \mathbf{x} in image $g(\mathbf{x})$ and patch P_3 centered on \mathbf{x}_3 in $g(\mathbf{x})$.

2.1.1 Modality independent neighbourhood descriptor

MIND descriptors can be defined as sum-of-squared differences between pixels in window W centered at \mathbf{x} of an image $g(\mathbf{x})$ and its shifted versions represented by patches P_i centered at each \mathbf{x}_i in image $g(\mathbf{x})$.¹⁶ Mathematically, similarity measure of window W and patch P_i can then be written as $S_{P_i}(\mathbf{x})$:

$$S_{P_i}(\mathbf{x}) = \int_{j=0}^n (g(\mathbf{x}_j) - g(\mathbf{p}_{\mathbf{x}_i}^j))^2, \quad \text{with } \mathbf{x}_j \in W \text{ and } \mathbf{p}_{\mathbf{x}_i}^j \in P_i \quad (2)$$

where \mathbf{x}_0 and $\mathbf{p}_{\mathbf{x}_i}^0$ are the center pixels of a window W in image $g(\mathbf{x})$ and patch P_i of shifted image $g(\mathbf{x}_i)$ respectively. Similarity $S_{P_i}(\mathbf{x})$ can be normalized¹⁸ as:

$$S'_{P_i}(\mathbf{x}) = \exp^{-\frac{S_{P_i}(\mathbf{x})}{\sigma_{\mathbf{x}}^2}}, \quad \text{where } \sigma_{\mathbf{x}}^2 \text{ represents the local variance of image intensities} \quad (3)$$

The n -dimensional normalized MIND descriptor vector $\mathbf{ND}(g, \mathbf{x}, n)$ computed for pixel \mathbf{x} in image $g(\mathbf{x})$ can be expressed as :

$$\mathbf{ND}(g, \mathbf{x}, n) = (S'_{P_1}(\mathbf{x}), \dots, S'_{P_i}(\mathbf{x}), \dots, S'_{P_n}(\mathbf{x})) \quad (4)$$

2.1.2 Normalized correlation transform descriptor

Correlation transforms (CoTs) computed for each pixel in window W using local neighbourhood information can be used to obtain self-similarity neighbourhood descriptors. Computation of CoTs are based on patch mean μ_P and patch standard deviation σ_P of image intensities within a local image region and are therefore invariant to local intensity variations. This standard deviation can be clipped to avoid division by small values. In our case, this clipping technique also helps to deal with the issue of debris and dynamic background that can significantly affect the quality of segmentation. Post clipping correlation transform $NCOT_{P_i}(\mathbf{x})$ can be computed for each pixel in window W from a local patch surrounding that pixel and can be expressed as :

$$NCOT_{P_i}(\mathbf{x}) = |g(\mathbf{x}_j) - \mu_{P_i}| / clip(\sigma_{P_i}, [T_1, T_2]), \quad \text{with } \mathbf{x}_j \in W \quad (5)$$

Here, W represents the neighborhood pixels and $P_i(\mathbf{x})$ denotes shifted version of image patches centered at \mathbf{x}_i in image $g(\mathbf{x})$. μ_{P_i} and σ_{P_i} represent the mean and standard deviation in the image patch P_i . T_1 and T_2 are the minimum and maximum values for which σ_{P_i} is clipped. For our experiments we have chosen $T_1 = 0.01$ and $T_2 = 0.5$.

The n -dimensional normalized CoT descriptor vector $\mathbf{ND}(g, \mathbf{x}, n)$ computed for pixel \mathbf{x} in image $g(\mathbf{x})$ can be expressed as :

$$\mathbf{ND}(g, \mathbf{x}, n) = (NCOT_{P_1}(\mathbf{x}), \dots, NCOT_{P_i}(\mathbf{x}), \dots, NCOT_{P_n}(\mathbf{x})) \quad (6)$$

2.2 Data-term

Neighbourhood descriptors are separately computed for the source g_s and the target g_t images and then used to formulate the data-term for optimization of flow field \mathbf{u} . The coordinates of superimposed pixels are \mathbf{x} in source image g_s and $\mathbf{x} + \mathbf{u}$ in target image g_t , where \mathbf{u} is the displacement vector at $\mathbf{x} \in \mathbb{R}^2$, $\mathbf{x} = (x, y)$.

Data-term is calculated by sum of absolute differences between corresponding descriptors $\mathbf{ND}(g_s, \mathbf{x}, i)$ and $\mathbf{ND}(g_t, \mathbf{x} + \mathbf{u}_x, i)$, which is expressed as:

$$\psi'_{data} = \int_{\mathbf{x} \in \Omega} \left\{ \frac{1}{n} \sum_{i=1}^n |\mathbf{ND}(g_s, \mathbf{x}, i) - \mathbf{ND}(g_t, \mathbf{x} + \mathbf{u}_x, i)| \right\} \quad (7)$$

The data term is then linearized using a first order Taylor series expansion to obtain a convex term and can be expressed as :

$$\psi'_{data} = \int_{\mathbf{x} \in \Omega} \left(\frac{1}{n} \int_{i=1}^n |\mathbf{ND}(g_s, \mathbf{x}, i) - \mathbf{ND}(g_t, \mathbf{x} + \mathbf{u}_x^0, i) + \nabla \mathbf{ND}(g_t, \mathbf{x} + \mathbf{u}_x^0, i)(\mathbf{u}_x - \mathbf{u}_x^0)| \right) \quad (8)$$

where, \mathbf{u}_x^0 is an approximation of \mathbf{u}_x and $\nabla = [\frac{\partial}{\partial x}, \frac{\partial}{\partial y}]^T$.

To extend the classical two-frame model to three-frame approach, we calculate two separate data-terms - ψ_{data_1} (corresponding to Frame-1 and Frame-3) and ψ_{data_2} (corresponding to Frame-2 and Frame-3). The two data-terms are then averaged together to obtain the final data-term ψ_{data} :

$$\psi_{data} = \frac{1}{2} \left[\int_{\mathbf{x} \in \Omega} \left\{ \frac{1}{n} \sum_{i=1}^n |\mathbf{ND}(g_s^{frame-1}, \mathbf{x}, i) - \mathbf{ND}(g_t^{frame-3}, \mathbf{x} + \mathbf{u}_x, i)| \right\} + \int_{\mathbf{x} \in \Omega} \left\{ \frac{1}{n} \sum_{i=1}^n |\mathbf{ND}(g_s^{frame-2}, \mathbf{x}, i) - \mathbf{ND}(g_t^{frame-3}, \mathbf{x} + \mathbf{u}_x, i)| \right\} \right] \quad (9)$$

2.3 Regularization

The role of the regularizer in the flow field computation is to enforce smooth optical flow, while preserving the flow discontinuities at the edges.¹⁶ Bilateral filtering formalizes the similarity between two locations and is used in image noise removal, stereo-vision applications and lately, in optical flow.¹⁶ We have used a non-local regularization ($\phi_{reg}(\mathbf{u})$) that uses a L1 penalty to allow sharp solutions and preserve motion discontinuities,¹⁶ given by :

$$\phi_{reg}(\mathbf{u}) = \int_{\mathbf{x} \in \Omega} \int_{\forall \mathbf{x}' \neq \mathbf{x} \in W_x} w_{\mathbf{x}}^{\mathbf{x}'} |\mathbf{u}_{\mathbf{x}} - \mathbf{u}_{\mathbf{x}'}|, \quad \text{with } w_{\mathbf{x}}^{\mathbf{x}'} = e^{-\left(\frac{|\mathbf{x} - \mathbf{x}'|^2}{2\gamma_1^2} + \frac{|g(\mathbf{x}) - g(\mathbf{x}')|^2}{2\gamma_2^2}\right)}, \quad (10)$$

where (γ_1, γ_2) are normalization factors and $w_{\mathbf{x}}^{\mathbf{x}'}$ represents the weight assigned to each pixel \mathbf{x}' in the neighbourhood window W centered around pixel \mathbf{x} . This weight term is a measure of how likely the pixels \mathbf{x} and \mathbf{x}' belong to the same object. It depends on the spatial distance and the color distance between the center pixel \mathbf{x} and the neighbour pixels \mathbf{x}' . Value of the weight term ($w_{\mathbf{x}}^{\mathbf{x}'}$) will be small for pixels that are spatially distant ($|\mathbf{x} - \mathbf{x}'|$ large) from each other or those that belong to different regions (large color difference $|g(\mathbf{x}) - g(\mathbf{x}')|$), thus preventing smoothing of the edge pixels.¹⁶

2.4 Energy minimization

The flow field \mathbf{u} between the source and the target images require minimization of the energy term $E(\mathbf{u})$ (see Eq. (1)) using primal-dual optimization approach.¹⁹ Key steps of the primal-dual convex optimization for the aforementioned energy formulation in Eq. (1) are discussed below.

The smoothing error represented by $\phi_{reg}(\mathbf{u})$ in Eq. 10 is a convex function in the flow field \mathbf{u} . The flow field at each pixel \mathbf{u}_x is evaluated for all other pixels in the neighbourhood W . The spanning of smoothing error to neighbouring pixels is represented by a linear operator K . The optical flow energy term of Eq: 1 can now be re-defined as:

$$\min_{\mathbf{u}} \{\lambda \psi_{data}(\mathbf{u}) + \phi_s(K\mathbf{u})\} \quad (11)$$

The presence of a linear operator introduces complications and the flow optimization problem is therefore formulated into a saddle point min-max problem, expressed as:

$$\min_{\mathbf{u}} \max_{\mathbf{w}} \{\lambda \psi_{data}(\mathbf{u}) + \langle K\mathbf{u}, \mathbf{w} \rangle - \phi_{reg}^*(K\mathbf{w})\}, \quad (12)$$

where ϕ_{reg}^* is the complex conjugate of ϕ_{reg} , $\mathbf{w} \in \mathbb{R}^{2 \cdot |\Omega| \cdot |W_x|}$ is the dual variable in the closed convex set \mathbf{S}_w which represents the non-negative entries of the bilateral filter weights (mentioned in Eq. (10)) and $\langle \cdot \rangle$ represents the dot product. Eq. (12) is convex in \mathbf{u} and concave in \mathbf{w} . Proximal Point algorithm¹⁹ is used to solve the saddle point problem mentioned in Eq. (12). Let τ and α be positive scalars such that $\tau\alpha = \|K^T K\|$ and $(\hat{\mathbf{u}}, \hat{\mathbf{w}})$ be the initial estimates (usually set to zero). The primal-variable solution is then given by:

$$\tilde{\mathbf{u}}^k = \arg \min_{\mathbf{u} \in \mathbf{S}_u} \left\{ \lambda \psi_{data}(\mathbf{u}) + \frac{\tau}{2} \|\mathbf{u} - [\mathbf{u}_k - \frac{1}{\tau} K^T \mathbf{w}^k]\|_2^2 \right\}, \quad (13)$$

where \mathbf{S}_u is the convex set of \mathbf{u} , $\tilde{\mathbf{u}}^k$ is the current estimate, $\mathbf{u}_k = \hat{\mathbf{u}}$ is the initial estimate, \mathbf{u} represents the update values and K^T is the adjoint linear operator. The dual-variable estimate $\tilde{\mathbf{w}}^k$ is given by:

$$\tilde{\mathbf{w}}^k = \arg \min_{\mathbf{w} \in \mathbf{S}_w} \left\{ \lambda \phi_{reg}^*(\mathbf{w}) + \frac{\alpha}{2} \|\mathbf{w} - [\mathbf{w}_k - \frac{1}{\alpha} K(2\tilde{\mathbf{u}}^k - \mathbf{u}_k)]\|_2^2 \right\}, \quad (14)$$

Since Eq: (13) and (14) are both differentiable, their gradients are zero ($\frac{\partial \tilde{\mathbf{u}}^k}{\partial \mathbf{u}} = 0, \frac{\partial \tilde{\mathbf{w}}^k}{\partial \mathbf{w}} = 0$) at desired points. This gives a set of equations which are then solved to obtain the final solution.

The presented model is implemented in a coarse-to-fine approach to handle large displacements. We have employed 8 pyramid-levels in all our experiments. The minimization process starts with setting the dual variable equal to zero at the coarsest level. We have used bilinear interpolation to build the pyramids with no anti-aliasing on the fine levels and bicubic interpolation for upsampling the flow field. The flow field at finer levels of the

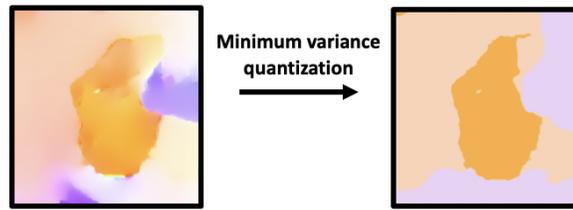


Figure 2. Optical flow image and the corresponding result after application of minimum variance quantization

Method	Frames	Sim. Metrics	Datasets					Mean
			# 1	# 2	# 3	# 4	# 5	
MIND	2	Dice C.	0.748	0.683	0.652	0.163	0.943	0.638
		Jaccard	0.597	0.518	0.484	0.088	0.892	0.516
		Corr. Coef	0.718	0.568	0.668	0.084	0.913	0.590
	3	Dice C.	0.912	0.804	0.725	0.806	0.923	0.834
		Jaccard	0.838	0.672	0.569	0.675	0.857	0.722
		Corr. Coef	0.893	0.735	0.710	0.788	0.884	0.802
NCoT	2	Dice	0.946	0.819	0.567	0.680	0.907	0.784
		Jaccard	0.898	0.693	0.396	0.515	0.831	0.666
		Corr. Coef	0.935	0.754	0.542	0.675	0.861	0.753
	3	Dice	0.889	0.806	0.902	0.852	0.933	0.876
		Jaccard	0.800	0.675	0.821	0.742	0.875	0.783
		Corr. Coef	0.866	0.760	0.897	0.838	0.905	0.853

Table 1. Quantitative evaluation of the motion induced segmentation after minimum variance quantization

pyramid are an up-scaled version of the coarser levels. The flow field obtained at each level is used to warp the source image g_s . Eq. (8) is used to calculate data-term at each level of the pyramid and Eq. (10) is used to calculate the weights required to perform non-local regularization of the computed flow field in the primal-dual minimization scheme. Each up-scaling step involves application of a 3×3 median filter to make the flow field smooth and minimize the interpolation error.

2.5 Minimum variance quantization

The flow field obtained by our algorithm is further processed using minimum variance quantization to reduce the number of colors in the segmentation result as shown in Fig: 2. It works by dividing the RGB color cube into smaller boxes of different size wherein the size of boxes depend on the distribution of colors in the given image.

3. RESULTS

We have evaluated our proposed algorithm on 4 synthetic and 1 real clinical ureteroscopy datasets. The synthetic dataset was obtained by performing in-vitro fragmentation of human kidney stone with a laser and imaged using clinical ureteroscope. 5 sets of three images each, were randomly picked from a collection of the mentioned datasets. MIND and CoT descriptors were then evaluated for each test set and analyzed quantitatively for 2 frames and 3 frames respectively.

In Table 1 it can be observed that the correlation transform based descriptors gives consistent improvements over almost all metric values. It is also evident that taking three frames into account significantly improves the segmentation accuracy. In case of three frames, the mean Dice and Jaccard are 0.876 and 0.783, respectively for CoT and 0.834 and 0.722, respectively for MIND descriptors. There is also an improvement of over 5% with CoT for correlation coefficient based evaluation metric.

For a qualitative analysis, we show optical flow outputs for random samples from one clinical dataset and three synthetic datasets in Fig. 3. The saturation and hue in the optical flow images represent the magnitude of

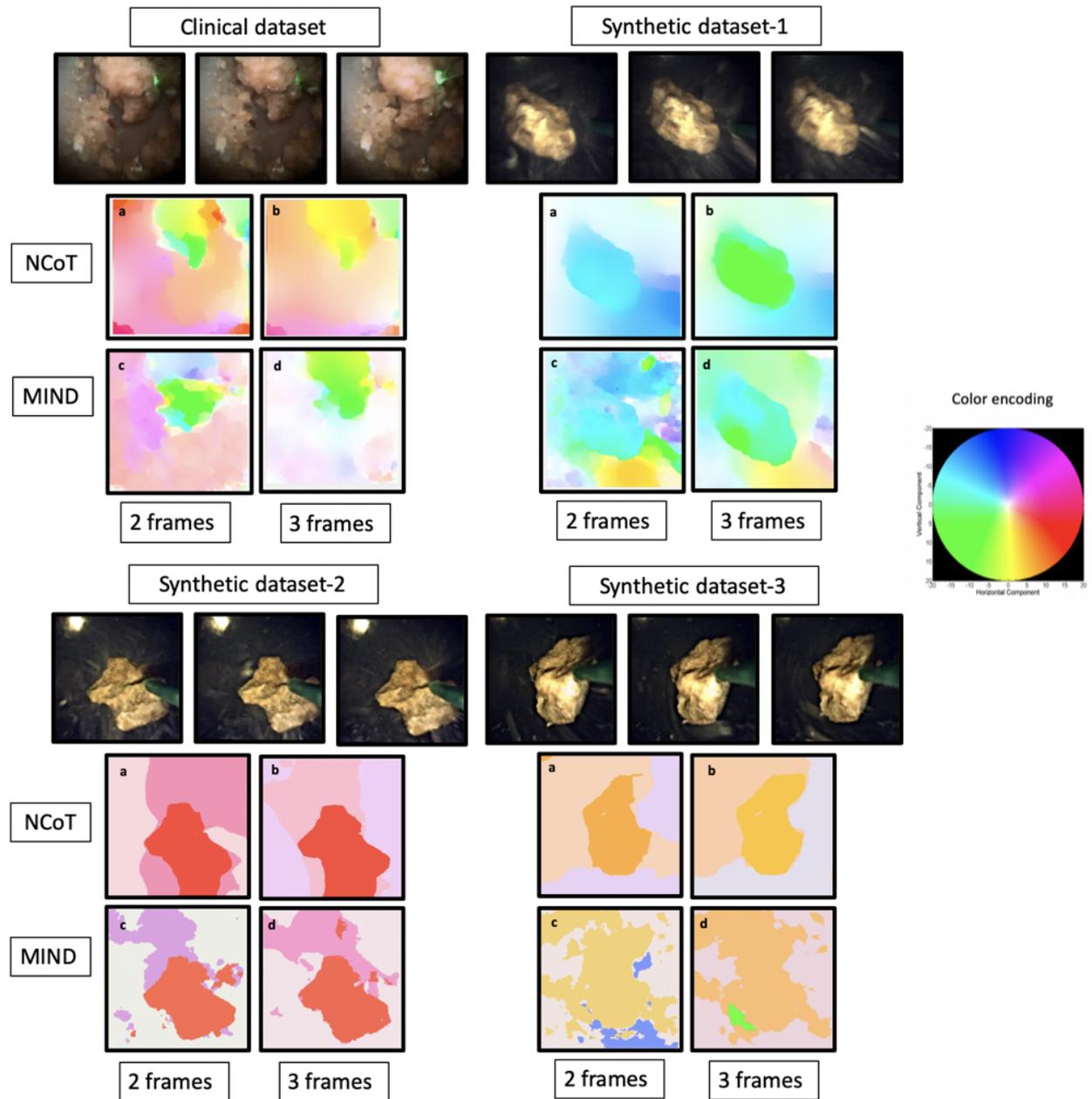


Figure 3. Qualitative evaluation of optical flow estimation with NCoT and MIND descriptors

displacement and orientation of the fragment respectively. It can be observed that our multi-frame optical flow estimation produces more accurate segmentation of large stone fragments as compared to the two-frame model. The three-frame approach involves computation of two data-terms (as mentioned in Section:2.2), thereby making the algorithm more robust to debris and other imaging artifacts like motion blur. The multi-frame approach also aids in improving the segmentation for cases where there is no significant motion between two consecutive frames. In addition, the use of CoT descriptors enables us to obtain pronounced flow field discontinuities w.r.t. MIND descriptors for the *target* stone. Although, CoT descriptor results in over-segmentation of the stone in the segmentation result of synthetic dataset-2, it is still evident that CoT descriptor delivers a consistent performance in overcoming the effect of debris to provide a more accurate segmentation of the stone fragments as compared to the MIND descriptors wherein stone debris is getting segmented as part of the stone (clearly evident in the result of synthetic dataset-3).

4. CONCLUSIONS

In this paper, we have presented a novel illumination-invariant method for segmentation of stone fragments in ureteroscopy images. To the best of our knowledge, a multi-frame dense optical flow algorithm employing a normalized correlation transform based data-term in a primal-dual optimization framework has not been proposed before. Our algorithm leverages the motion fields between multiple frames to tackle problems related to illumination variability, blur, occlusions and deformations. Qualitative and quantitative results from our experiments demonstrate that our algorithm can overcome the effect of stone debris and provide accurate segmentation of stone fragments. Although we have tested our algorithm on ureteroscopy data, it can also be applied to other video datasets. Future work will focus on combining the motion information with deep-learning based instance segmentation to further improve the segmentation quality and processing speed.

5. ACKNOWLEDGEMENTS

We would like to thank Boston Scientific for funding this project (Grant No: DFR04690). SG, LG and BT are funded by BSC, SA is supported by the NIHR Oxford BRC, and JR is funded by EPSRC EP/M013774/1 Seebibyte.

REFERENCES

1. S. Butticcè, T. E. Sener, C. Netsch, E. Emiliani, R. Pappalardo, and C. Magno, "LithoVue™: A new single-use digital flexible ureteroscope," *Central European Journal of Urology*, 2016.
2. N. L. Miller and J. E. Lingeman, "Management of kidney stones," *British Medical Journal*, 2007.
3. W. Brisbane, M. R. Bailey, and M. D. Sorensen, "An overview of kidney stone imaging techniques," *Nature Reviews Urology*, 2016.
4. B. Rosa, P. Mozer, and J. Szewczyk, "An algorithm for calculi segmentation on ureteroscopic images," *International Journal of Computer Assisted Radiology and Surgery*, 2011.
5. P. Geavlete, R. Multescu, and B. Geavlete, "Pushing the boundaries of ureteroscopy: Current status and future perspectives," *Nature Reviews Urology*, 2014.
6. K. Viswanath and R. Gunasundari, "Design and analysis performance of kidney stone detection from ultrasound image by level set segmentation and ANN classification," in *Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2014*, 2014.
7. P. Thangaraj and P. R. Tamilselvi, "A Modified Watershed Segmentation Method to Segment Renal Calculi in Ultrasound Kidney Images," *Int. J. Intell. Inf. Technol.* **8**(1), pp. 46–61, 2012.
8. J. Verma, M. Nath, P. Tripathi, and K. K. Saini, "Analysis and identification of kidney stone using Kth nearest neighbour (KNN) and support vector machine (SVM) classification techniques," *Pattern Recognition and Image Analysis* **27**, pp. 574–580, 2017.
9. J. R. Van Sörnsen De Koste, S. Senan, C. E. Kleynen, B. J. Slotman, and F. J. Lagerwaard, "Renal mobility during uncoached quiet respiration: An analysis of 4DCT scans," *International Journal of Radiation Oncology Biology Physics*, 2006.

10. S. Galic and S. Loncaric, "Spatio-temporal image segmentation using optical flow and clustering algorithm," *IWISPA 2000. Proceedings of the First International Workshop on Image and Signal Processing and Analysis. in conjunction with 22nd International Conference on Information Technology Interfaces. (IEEE)*, pp. 63–68, 2000.
11. T. Macan and S. Loncaric, "Hybrid optical flow and segmentation technique for LV motion detection," in *Medical Imaging 2001: Physiology and Function from Multidimensional Images*, C.-T. Chen and A. V. Clough, eds., **4321**, pp. 475 – 482, International Society for Optics and Photonics, SPIE, 2001.
12. V. Duay, M. Bach Cuadra, X. Bresson, and J.-P. Thiran, "Dense deformation field estimation for atlas registration using the active contour framework," *European Signal Processing Conference*, 01 2006.
13. B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.* **17**, pp. 185–203, 1980.
14. O. Demetz, D. Hafner, and J. Weickert, "The complete rank transform: A tool for accurate and morphologically invariant matching of structures," 2013.
15. L. Mei, J. Lai, X. Xie, J. Zhu, and J. Chen, "Illumination-invariance optical flow estimation using weighted regularization transform," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2019.
16. S. Ali, C. Daul, E. Galbrun, and W. Blondel, "Illumination invariant optical flow using neighborhood descriptors," *Computer Vision and Image Understanding* **145**, pp. –, 12 2015.
17. S. Ali, D. Lin, M. Axer, and K. Rohr, "Comparison of self-similarity measures for multi-modal non-rigid registration of 3d-pli brain images," in *Bildverarbeitung für die Medizin 2018*, pp. 49–54, 2018.
18. A. Buades, B. Coll, and J. . Morel, "A non-local algorithm for image denoising," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, **2**, pp. 60–65 vol. 2, June 2005.
19. A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision* **20**, pp. 89–97, 2004.