Efficient Multi-organ Segmentation in Multi-view Endoscopic Videos Using Pre-operative Priors*

Masoud S. Nosrati[1], Jean-Marc Peyrat[2], Julien Abinahed[2], Osama Al-Alao[3], Abdulla Al-Ansari[2,3], Rafeef Abugharbieh[4], and Ghassan Hamarneh[1]

[1] Medical Image Analysis Lab, Simon Fraser University, BC, Canada
[2] Qatar Robotic Surgery Centre, Qatar Science & Technology Park, Qatar
[3] Urology Department, Hamad General Hospital, Hamad Medical Corporation, Qatar
[4] BiSICL, University of British Columbia, BC, Canada

Abstract. Synergistic fusion of pre-operative (pre-op) and intra-operative (intra-op) imaging data provides surgeons with invaluable insightful information that can improve their decision-making during minimally invasive robotic surgery. In this paper, we propose an efficient technique to segment multiple objects in intra-op multi-view endoscopic videos based on priors captured from pre-op data. Our approach leverages information from 3D pre-op data into the analysis of visual cues in the 2D intra-op data by formulating the problem as one of finding the 3D pose and non-rigid deformations of tissue models driven by features from 2D images. We present a closed-form solution for our formulation and demonstrate how it allows for the inclusion of laparoscopic camera motion model. Our efficient method runs in real-time on a single core CPU making it practical even for robotic surgery systems with limited computational resources. We validate the utility of our technique on *ex vivo* data as well as *in vivo* clinical data from laparoscopic partial nephrectomy surgery and demonstrate its robustness in segmenting stereo endoscopic videos.

1 Introduction

Robotic minimally invasive surgery (MIS) systems have been gaining popularity due to their many advantages compared to traditional MIS and open surgeries including greater precision, improved dexterity and enhanced 3D immersive visualization for surgeons [9].

In robot-assisted MIS, image-guided localization and delineation of tissues, e.g. tumour and kidney in partial nephrectomy, is an important step that can significantly enhance the surgeon's perception of the surgical scene and facilitate their decision-making. However, accurate identification of various tissues in intra-op video is by no means an easy task due to the limited viewing area, presence

* This publication was made possible by NPRP Grant#4-161-2-056 from the Qatar National Research Fund (a member of the Qatar Foundation). The statements made herein are solely the responsibility of the authors.

P. Golland et al. (Eds.): MICCAI 2014, Part II, LNCS 8674, pp. 324–331, 2014.

of occluding objects (e.g. surgical tools), data acquisition noise (e.g. specular light reflection, blood and smoke), as well as similarity in the visual appearance of different tissues. To localize different tissues, surgeons rely on (mentally) combining the information they recall from pre-op scans, often computed tomography (CT), with the information they see in the intra-op stereo endoscopic video feed, a task that requires exceptional skill. Advances in intra-op imaging has introduced some other modalities into the operating room, e.g. ultrasound and X-ray [3]. However, the feasibility, quality, and information content of such data still markedly lags behind the typically high resolution 3D pre-op data, and endoscopic imaging remains the staple modality in MIS. Using efficient 2D and 3D computer vision techniques to support the analysis of endoscopic images alleviates the need for using additional equipment, e.g. fiducial markers, and helps to relax the handling of multiple intra-op data streams. Yet, the aforementioned complications related to noise and clutter pose many challenges in endoscopic video segmentation.

Some recent works in endoscopic image segmentation proposed the use of level set-based approaches [4,5] while others focused on parameter-sensitive morphological operations and thresholding techniques [2,7]. However, practical success of such methods is limited as they mostly rely on image color/intensity, do not use pre-op information, and focus on segmenting a single object. Other contributions focused on feature tracking, e.g. [11], while Mountney et al. [8] proposed a method to estimate laparoscopic camera and *periodic* organ motion. Both methods cannot handle free-form deformation of organs and the latter method assumed the camera has (on average) constant velocity, which is often not the case in MIS. In this paper, we propose a technique that imitates surgeons skill in leveraging pre-op information into the analysis of intra-op endoscopic visual cues. Our approach encodes the fused information within an energy optimization process to efficiently segment multiple objects in endoscopic videos.

Our work is inspired by the work of Prisacariu et al. [10] in the (non-medical) computer vision area where the segmentation of an object in an image was obtained by finding the six pose parameters of its 3D model. In MIS, six degrees of freedom are not enough as tissues non-rigidly deform. Unlike [10], Sandhu et al. [12] used kernel PCA to capture the shape variation and estimated the non-rigid pose of a single object. Their method, however, segments a single object in a single view image. Applying the methodology of [10] and [12] to robotic surgery applications is also not straightforward as images in endoscopic videos are highly noisy and cluttered. Moreover, [12] does not leverage stereo vision and is incapable of handling *large* occlusions (large portions of objects are occluded, e.g. by tools), both of which are common in MIS. In our formulation, we provide a *closed-form solution* (unlike [10,12]) to segment *multiple tissues* in a *multi-view* endoscopic video (here we use stereo video) based on prior knowledge extracted from pre-op data. Our approach thus simultaneously estimates the 3D pose of tissues in the pre-op domain as well as their *non-rigid* deformations from their pre-op state. Furthermore, our framework allows for the inclusion of *motion priors* on laparoscopic camera motion to stabilize the segmentation/pose

tracking in the presence of a large occlusion. Such feature is especially useful in robotic MIS as camera motion signals can be easily read using the robot's API and incorporated into our formulation to obtain even more accurate and robust results. In this work, we enforce a motion prior on the cameras only; nonetheless, our flexible mathematical formulation allows for incorporating of general non-rigid motion or temporal deformation priors on various tissues. Our method runs in *real-time on a single CPU core* which makes it suitable for robotic surgical systems as they are typically limited in computational resources.

2 Methods

For ease of exposition, we start by describing our method for segmenting a **single object in a single image**, given its 3D segmentation in the pre-op domain. Later, we show how we extend our framework to segment multiple objects in multiple 2D images (multi-view data). Let $\boldsymbol{X}_{pre,i} \in \mathbb{R}^3$ be point i of a 3D model obtained from segmented pre-op 3D data. The goal is to spatially transform and deform the model non-rigidly in 3D such that its silhouette on the 2D color image, $I : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}^3$, delineates the object of interest in I, i.e. the silhouette encapsulates the foreground. The silhouette of a 3D model is obtained by projecting the model from 3D to 2D given the projection function π and the corresponding camera's focal point f and principal point c. We represent the foreground by the level set function ϕ such that, for a pixel $\boldsymbol{x}_i \in \Omega$ in I, $\phi(\boldsymbol{x}_i) > 0$ if $\boldsymbol{x}_i$ belongs to the foreground, $\phi(\boldsymbol{x}_i) < 0$ if $\boldsymbol{x}_i$ belongs to the background, and $\phi(\boldsymbol{x}_i) = 0$ if $\boldsymbol{x}_i$ is on the object's boundary. Every 2D point $\boldsymbol{x}_i$ on the foreground is related to its corresponding 3D point $\boldsymbol{X}_{pre,i}$ by $\boldsymbol{x}_i = \pi\big(T(\boldsymbol{X}_{pre,i}; \boldsymbol{\xi})\big)$ where T transforms $\boldsymbol{X}_{pre,i}$ from the pre-op domain to $\boldsymbol{X}_{srg,i}$ in the surgical domain. $\boldsymbol{\xi} = \{\boldsymbol{\xi}_q, \boldsymbol{\xi}_w\}$ is the set of pose $\boldsymbol{\xi}_q = \{q_1, \cdots, q_n\}$ and shape parameters $\boldsymbol{\xi}_w = \{w_1, \cdots, w_m\}$. We use the weights of shape variation modes as $\boldsymbol{\xi}_w$. To segment a tissue in the 2D image I given its 3D model, we define the following residual for the i^{th} pixel, $\boldsymbol{x}_i \in \Omega$ as:

$$e_i = -g_f(\boldsymbol{x}_i)H(\phi(\boldsymbol{x}_i)) + g_b(\boldsymbol{x}_i)(1 - H(\phi(\boldsymbol{x}_i))), \tag{1}$$

where g_f and g_b are the regional terms measuring the agreement of the image pixel $\boldsymbol{x}_i$ with the foreground and background statistical models and $H(.)$ is the Heaviside step function. Ideally e_i would be zero for a perfect model-to-data fit, however, due to noise, e_i will have a distribution $p(e_i|\boldsymbol{\xi})$, which can be modelled as $\mathcal{N}(0, \sigma)$ when $\boldsymbol{\xi}$ is close to the solution (Fig. 1(a)). The residual value for all pixels $\boldsymbol{e}$ is calculated assuming that the noise is independent across pixels. The objective is to find the most likely transformation parameters $\boldsymbol{\xi}$ given the residual $\boldsymbol{e}$ by maximizing the posterior probability,

$$\boldsymbol{\xi}^* = \arg\max_{\boldsymbol{\xi}} \; p(\boldsymbol{\xi}|\boldsymbol{e}) = \arg\min_{\boldsymbol{\xi}} \; -\sum_i \log p(e_i|\boldsymbol{\xi}) - \log p(\boldsymbol{\xi}) \; . \tag{2}$$

The second term in (2) is the prior on the transformation. Here, we enforce a prior only on pose parameters ($\boldsymbol{\xi}_q$) which can also be considered as camera

motion parameters (due to their relative rigid motion); however, one can use this term to enforce a prior on the tissues' non-rigid deformation. The camera motion prior can be obtained either from the robotic surgical system or the prediction of a Kalman filter. The uncertainty in the motion estimation is modelled with a Gaussian meaned around the predicted pose parameter $\hat{\boldsymbol{\xi}}$ with covariance Σ_ξ, i.e. $p(\boldsymbol{\xi}) \sim \mathcal{N}(\hat{\boldsymbol{\xi}}, \Sigma_\xi)$. We minimize (2) by taking its derivative and setting it to zero:

$$\sum_i \frac{\partial \log p(e_i|\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} + \frac{\partial \log p(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = \sum_i \frac{-1}{\sigma} e_i \frac{\partial e_i}{\partial \boldsymbol{\xi}} - \frac{1}{\Sigma_\xi}(\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}) = 0. \tag{3}$$

As e_i is not linear in $\boldsymbol{\xi}$, to solve (3) efficiently we linearize e_i with respect to $\boldsymbol{\xi}$ using the first order Taylor approximation:

$$e_i^{lin}(\boldsymbol{\xi}) = e_i(\mathbf{0}) + \frac{\partial e_i(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}}\Big|_{\boldsymbol{\xi}=0} \Delta\boldsymbol{\xi} \;=\; e_i(\mathbf{0}) + J_i \Delta\boldsymbol{\xi}, \tag{4}$$

where J_i is the Jacobian of the i^{th} pixel's error with respect to $\boldsymbol{\xi}$. Substituting (4) into (3) and using matrix notation, we have

$$(\frac{-1}{\sigma}\boldsymbol{J}^T\boldsymbol{J} + \Sigma_\xi^{-1})\Delta\boldsymbol{\xi} = \frac{1}{\sigma}\boldsymbol{J}^T\boldsymbol{e}(0) + \Sigma_\xi^{-1}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_{t-1}) \;, \tag{5}$$

where $\boldsymbol{J}$ is the stacked matrix of all J_i pixel-wise Jacobians and $\boldsymbol{\xi}_{t-1}$ is the pose in the previous frame. At each frame of video, the linear system of equations (5) is solved efficiently for $\Delta\boldsymbol{\xi}$ (by Choleskey decomposition) with which we update the transformation/deformation parameters $\boldsymbol{\xi}$. For this linearization, we assumed $\Delta\boldsymbol{\xi}$ is small. This assumption is valid given the high video capture rate of current surgical systems, e.g. daVinci with ∼30 FPS. To handle larger transformations, one may apply a coarse-to-fine scheme. The gradient of the i^{th} pixel's error with respect to ℓ^{th} component of $\boldsymbol{\xi}$ (ℓ^{th} element of J_i) is calculated as:

$$J_i(\ell) = \frac{\partial e_i}{\partial \xi_\ell} \;=\; (-g_f - g_b)\frac{\partial H}{\partial \phi}\frac{\partial \phi}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{x}}{\partial \xi_\ell} \;=\; (-g_f - g_b)\delta(\phi)\nabla_x\phi\frac{\partial \boldsymbol{x}}{\partial \xi_\ell} \;, \tag{6}$$

where $\delta(.)$ is the Dirac delta function. Given the camera parameters f and c, every 2D point $\boldsymbol{x} = (x, y)$ in I is related to at least one corresponding 3D point $\boldsymbol{X}_{srg} = (X_{srg}, Y_{srg}, Z_{srg})$ by $x = \frac{f}{Z_{srg}}X_{srg} + c$ and $y = \frac{f}{Z_{srg}}Y_{srg} + c$. Hence,

$$\frac{\partial x}{\partial \xi_\ell} = \frac{f}{Z_{srg}^2}(Z_{srg}\frac{\partial X_{srg}}{\partial \xi_\ell} - X_{srg}\frac{\partial Z_{srg}}{\partial \xi_\ell}) \;. \tag{7}$$

We similarly calculate $\frac{\partial y}{\partial \xi_\ell}$. 3D point $\boldsymbol{X}_{srg}$ is related to $\boldsymbol{X}_{pre}$ by the transformation function T: $\boldsymbol{X}_{srg} = T(\boldsymbol{X}_{pre}; \boldsymbol{\xi}) = \boldsymbol{R}(\boldsymbol{\xi}_q)\boldsymbol{X}_{pre} + \boldsymbol{t}(\boldsymbol{\xi}_q)$, where $\boldsymbol{R}$ and $\boldsymbol{t}$ are rotation (linear) and translation matrices and can represent any rigid (linear) transformation upon the choice of pose parameters in T, $\boldsymbol{\xi}_q = \{q_1, \cdots, q_n\}$. Therefore, $\frac{\partial X_{srg}}{\partial q_i}$, $\frac{\partial Y_{srg}}{\partial q_i}$ and $\frac{\partial Z_{srg}}{\partial q_i}$ in (7) are easily calculated upon the choice of transformation function.

To handle non-rigid deformation, we generate a catalog of realistic 3D deformed shapes for the organs of interest (Fig. 1(b)) using DeformIt [6]. Having this catalog, we obtain the modes of variation through principal component analysis such that a novel 3D shape of an organ/tissue can be estimated as $\boldsymbol{X}_{pre} = \overline{\boldsymbol{X}_{pre}} + \boldsymbol{U}\boldsymbol{w}$, where $\overline{\boldsymbol{X}_{pre}}$ is the coordinates of the average shape of the organ of interest, $\boldsymbol{U} = \{u_1, \cdots, u_K\}$ are the K principal modes of variation and $\boldsymbol{w} = \boldsymbol{\xi}_q = \{w_1, \cdots, w_K\}$ are their weights. We chose the number of principal modes explaining 97% of the variance in the training set. The derivative of a 3D point $\boldsymbol{X}_{srg}$ in the surgical domain with respect to w_ℓ is calculated as $\frac{\partial \boldsymbol{X}_{srg}}{\partial w_\ell} = \boldsymbol{R}\frac{\partial \boldsymbol{X}_{pre}}{\partial w_\ell} = \boldsymbol{R} \cdot u_\ell$, where u_ℓ is the ℓ^{th} mode of variation in $\boldsymbol{U}$.

We now describe the extension to segmenting **multi-object in multi-view** images. Having N tissues of interest in the pre-op data and M views in the intra-op domain, we extend our framework to segment multiple objects in multi-view images by computing a level set function for each object. Also, all the pixels from all views contribute toward calculating the residual of each pixel. The residual of pixel $\boldsymbol{x}_i$ belonging to the n^{th} tissue is calculated as:

$$e_i = \sum_{m=1}^{M} \Big(- g^n_{f,m}\, H(\phi^n_m(\boldsymbol{x}_i)) + g^n_{b,m}\, (1 - H(\phi^n_m(\boldsymbol{x}_i))) \Big)\,, \tag{8}$$

where $g^n_{f,m}$ and $g^n_{b,m}$ are the regional terms and ϕ^n_m is the level set of the n^{th} object in the m^{th} image. Note that in the multi-view scenario, the extrinsic camera parameters $(\boldsymbol{R}^{ext}_m, \boldsymbol{t}^{ext}_m)$ have to be considered in calculating the Jacobian, i.e. $\boldsymbol{R}$ in the above equations is replaced by $\boldsymbol{R}^{ext}_m \boldsymbol{R}$.

To make our method more robust, we leverage a variety of image features (normalized RGB and YCbCr and local color histogram) to calculate the regional terms, $g^n_{f,m}$ and $g^n_{b,m}$ in (8), for different tissues in all 2D views. We estimate the probability of a given pixel $\boldsymbol{x}$ belonging to the n^{th} object (O_n), $p(\boldsymbol{x} \in O_n | I_m)$, and its background (B_n), $p(\boldsymbol{x} \in B_n | I_m)$, in the m^{th} image, I_m, by training a random forest (RF) consisting of N_b binary decision trees (here $N_b = 20$). To train a RF, we select few 20×20 patches on different tissues in 2% of all frames from the *same* patient, i.e. ∼10 frames out of ∼600 frames. In practice, surgeons may select these patches with the help of surgical tools. The regional terms are calculated as $g^n_{f,m}(\boldsymbol{x}) = -\log p(\boldsymbol{x} \in O_n | I_m)$ and $g^n_{b,m}(\boldsymbol{x}) = -\log p(\boldsymbol{x} \in B_n | I_m)$.

3 Materials, Experimental Setup, and Results

We evaluated our framework on 10 *ex vivo* lamb kidney datasets as well as three *in vivo* clinical partial nephrectomy data. We constructed the set of *ex vivo* phantoms using lamb kidneys and implanted artificial tumours outside and inside each kidney to emulate a partially exophytic and completely endophytic tumour, respectively. CT volumes and stereo video sequences of our phantoms were captured by a Somatom CT scanner (Siemens, Germany) and a daVinci S system (Intuitive Surgical, USA), respectively. We segmented the kidney and tumours in each CT using the TurtleSeg software [13]. We simulated deformations

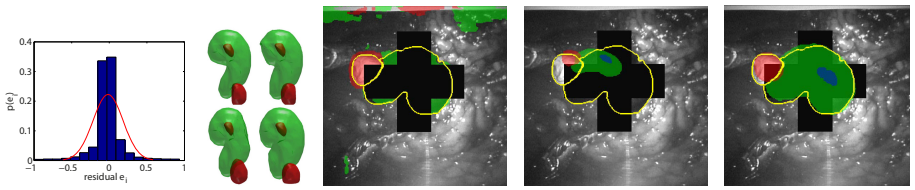


Fig. 1. (a) Residual distribution. (b) Few samples from our kidney and tumour catalog. Segmentation result in the presence of artificial tools (black cross) largely occluding the kidney and tumour phantoms using (c) ACWOE, (d) our method without any motion prior and (e) our method with motion prior. Green: kidney; Red: exophytic tumour; Blue: endophytic tumour; Yellow: ground truth.

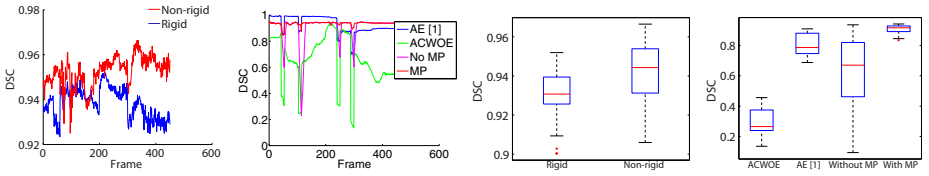
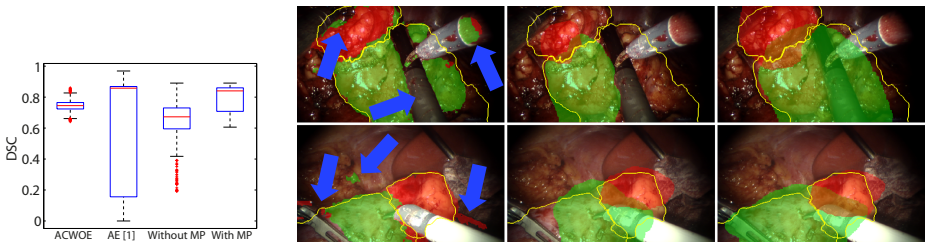


Fig. 2. DSC vs. frame of a phantom for (a) rigid vs. non-rigid transformation and, for (b) ACWOE, AE [1] and our method with and without motion prior (MP vs. No MP). Box plot representation of DSC for the whole dataset is presented in (c) for rigid vs. non-rigid transformation and in (d) for ACWOE, AE and our method with/without motion prior over the occlusion periods.

of each kidney and tumour in respectively ∼40 and ∼15 different ways using DeformIt [6] (Fig. 1(b)). To obtain ground truth segmentation for stereo videos, we used the "Rotobrush" tool of After Effect (Adobe Systems Inc., USA) as a semi-automatic segmentation tool allowing for visual inspection and correction. For initialization, we manually aligned the pre-op model with the intra-op image by choosing ∼6-8 landmarks in the CT and the first frame of the stereo video.

Our **first experiment** on *ex vivo* data focused on evaluating our method w.r.t. using a rigid vs. deformable transformation model. Fig. 2(a) shows an example Dice similarity coefficient (DSC) vs. time for one of our phantoms. It is seen in Fig. 2(a) that, as expected, incorporating the non-rigid deformation of tissues improves the final results. Note that the method in [11] does not handle non-rigid deformation and [8] can only estimate the non-rigid deformation for organs with *periodic* motion, whereas our method does not pose any such constraints. Results for all the *ex vivo* phantom datasets are shown in Fig. 2(c).

In our **second experiment**, we artificially occluded large portions of the kidney and tumour in the videos (Fig. 1(c-e)) and contrasted the performance of our method with and without the motion prior. We also compared with the state-of-the-art video segmentation method proposed by Bai et al. [1], implemented in the After Effect software (AE), and with the level-sets based



(a) (b)

Fig. 3. Clinical partial nephrectomy results. (a) DSC for three clinical cases. (b) Qualitative results of (1^{st} column) ACWOE (blue arrows indicate errors), (2^{nd} column) our method without motion prior and, (3^{rd} column) our method with motion prior. Green: kidney; Red: tumour; Yellow: ground truth.

active contours without edges (ACWOE) [14] as considered in [4,5] for endoscopic video segmentations. In this experiment, we used the constant velocity model as our motion prior. Note that there is no explicit limitation on our motion prior term and more complicated motion priors can be seamlessly deployed into our framework, e.g. using the surgical robot's API signals. To compare our method with AE, we provided AE with an accurate segmentation for the first frame and, since this software is only able to segment a single object in a single view image, we used it multiple times to segment the kidney and tumours in the left and right view channels. Fig. 2(b) illustrates the DSC vs. time when using ACWOE, AE and our method with and without motion prior for a phantom case. The dips in Fig. 2(b) are caused by the occluding objects. The figure demonstrates how incorporating a motion prior stabilizes the segmentation and pose tracking and helps overcome large tissue occlusions that may occur in MIS. Obviously, AE and ACWOE are not able to show the internal tissues, e.g. the endophytic tumours shown in blue in Fig. 1(e), as they do not use any pre-op information. They also are both fragile when a large occlusion occurs. Fig. 2(d) compares the results of ACWOE, AE and our method with/without motion prior for all the *ex vivo* phantom datasets during the occlusion periods.

In our **third experiment** we tested our method on three different clinical cases of partial nephrectomy. For each patient, we prepared a ∼20-second stereo endoscopic video from the surgical system with a frame rate of 30 FPS, i.e. ∼600 frames. Segmenting CT and preparing the ground truth was performed similar to the *ex vivo* phantom data. Each stereo video took ∼3.5 hours to segment semi-automatically using AE to create the ground truth. Quantitative and qualitative results on the real *in vivo* cases are illustrated in Fig. 3. Despite existing clutter and tool crossings, our method was able to achieve a DSC close to 0.85 on these challenging real *in vivo* cases with an average runtime of 0.045 seconds per frame using non-optimized MATLAB code on a single core 3.40 GHz CPU.

4 Discussion and Conclusions

Leveraging both pre-op data as well as endoscopic visual cues, we proposed a novel formulation with closed form solution for segmenting multiple tissues in multi-view endoscopic videos. Our formulation further incorporated a motion prior in our optimization framework to stabilize the segmentation and pose tracking. One shortcoming of our method is the use of statistical deformable model which might not represent patient-specific tissue deformations. In addition, sudden camera motion will cause our tracking to fail. However, in cases where access to the surgical robots' API is available, one can easily feed in the camera motion signals into our formulation, as a camera motion constraint, which would enable handling the complex camera motion. Although our validation on *ex vivo* phantoms and *in vivo* surgery data of partial nephrectomy confirmed the usability and great promise of the proposed framework in an augmented reality environment for MIS, more clinical experiment is needed to verify the clinical validity of the proposed method.

References

1. Bai, X., et al.: Video SnapCut: robust video object cutout using localized classifiers. ACM Trans. Graphics 28(3), 70:1–70:11 (2009)
2. Dhandra, B., et al.: Analysis of abnormality in endoscopic images using combined HSI color space and watershed segmentation. In: ICPR, vol. 4, pp. 695–698 (2006)
3. Estépar, R., et al.: Multimodality guidance in endoscopic and laparoscopic abdominal procedures. In: Intraop. Imag. Image-Guided Therapy, pp. 767–778 (2014)
4. Figueiredo, I., et al.: Variational image segmentation for endoscopic human colonic aberrant crypt foci. IEEE TMI 29(4), 998–1011 (2010)
5. Figueiredo, I., et al.: A segmentation model and application to endoscopic images. In: Image Anal. Recogn., pp. 164–171 (2012)
6. Hamarneh, G., Jassi, P., Tang, L.: Simulation of ground-truth validation data via physically- and statistically-based warps. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 459–467. Springer, Heidelberg (2008)
7. Mewes, P.W., Neumann, D., Licegevic, O., Simon, J., Juloski, A.L., Angelopoulou, E.: Automatic region-of-interest segmentation and pathology detection in magnetically guided capsule endoscopy. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 141–148. Springer, Heidelberg (2011)
8. Mountney, P., Yang, G.-Z.: Motion compensated SLAM for image guided surgery. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A., et al. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 496–504. Springer, Heidelberg (2010)
9. Pratt, P., et al.: An effective visualisation and registration system for image-guided robotic partial nephrectomy. J. Robotic Surg. 6(1), 23–31 (2012)
10. Prisacariu, V., Reid, I.: PWP3D: Real-time segmentation and tracking of 3d objects. IJCV 98(3), 335–354 (2012)
11. Puerto-Souza, G., Mariottini, G.: Toward long-term and accurate augmented-reality display for minimally-invasive surgery. In: ICRA, pp. 5384–5389 (2013)
12. Sandhu, R., et al.: A nonrigid kernel-based framework for 2D-3D pose estimation and 2D image segmentation. IEEE TPAMI 33(6), 1098–1115 (2011)
13. Top, A., Hamarneh, G., Abugharbieh, R.: Active learning for interactive 3D image segmentation. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 603–610. Springer, Heidelberg (2011)
14. Vese, L., Chan, T.: A multiphase level set framework for image segmentation using the Mumford and Shah model. IJCV 50(3), 271–293 (2002)