

# Tracking by Detection for Interactive Image Augmentation in Laparoscopy <sup>★</sup>

Jae-Hak Kim<sup>1,2</sup>, Adrien Bartoli<sup>1</sup>, Toby Collins<sup>1</sup>, and Richard Hartley<sup>3</sup>

<sup>1</sup> ISIT, Faculté de Médecine, Université d’Auvergne, France

<sup>2</sup> LASMEA, Université Blaise Pascal, France

<sup>3</sup> CECS, The Australian National University, Australia

**Abstract.** We present a system for marking, tracking and visually augmenting a deformable surgical site by the robust automatic detection of natural landmarks (image features) in laparoscopic surgery. In our system, the surgeon first selects a frame containing an organ of interest, and this is used by our system both to detect every instance of the organ in a laparoscopic video feed, and to recover the nonrigid deformations. The system then augments the video with customizable visual information such as the location of hidden or weakly visible structures (cysts, vessels, etc), or planned incision points, acquired from pre-operative or intra-operative data. Frame-rate organ detection is performed via a novel procedure that matches the current frame to the reference frame. Because laparoscopic images are known to be extremely difficult to match, we propose to use Shape-from-Shading and conformal flattening to cancel out much of the variation in appearance due to perspective foreshortening, and we then apply robust matching to the flattened surfaces. Experiments show robust tracking and detection results on a laparoscopic procedure with the uterus as target organ. As our system detects the organ in every frame, it is not impaired by target loss, contrary to most previous methods.

**Keywords:** tracking, detection, laparoscopy, conformal mapping, deformable surface

## 1 Introduction

In laparoscopic surgery, one of the difficulties for surgeons is to correctly identify the surgical site, such as the location of hidden structures and planned incision points, in the camera’s image. This is especially significant in tumor resection, when the tumor is occluded behind tissue. To resolve this, surgeons and radiologists currently use pre-operative data such as Magnetic Resonance Imaging (MRI), Computed Tomography (CT) or Ultrasound (US) images to help locate the target structure in the image.

---

<sup>★</sup> This research has been supported by Prof. Richard Hartley’s Chair of Excellence grant from Région Auvergne, France.

It is estimated that there are 1,300 to 2,700 wrong-site surgeries annually in the United States [12]. Similar problems may arise in laparoscopy because it can be difficult for the surgeons to find the target site correctly, mainly due to disorientation and the difficulty of mentally matching the laparoscopic view with the pre-operative images. Solving the medical imaging problem of automatically overlaying laparoscopic images with surgical target locations is an open, yet highly important research goal. Not only would the solution benefit surgeons, but also patients in reducing the likelihood of surgical errors and complications.

We present a new system to mark and track the surgical site in laparoscopic images. The surgical site is first marked by the surgeon in the 2D input images and in the reconstructed 3D surface video at an early stage of the laparoscopy. Then, our system automatically detects the location of the surgical site during surgery, and overlays the laparoscopic frames and reconstructed 3D surface video with visual information. Detecting the location of the surgical site requires robust feature matching and registration methods on nonrigid deformable surfaces. For this feature matching and tracking, a new approach using Shape-from-Shading (SfS) and conformal mapping is proposed and two robust methods are introduced for mapping a polygon boundary of the uterus in the reference image into the other frames by estimating a similarity transformation. Also, augmentation of the laparoscopic video using affine Moving Least Squares (affine-MLS) is proposed. We carried out experiments on in-vivo laparoscopic images captured by a Karl Storz laparoscopy system. Our method shows robust feature matching, tracking and augmentation results in laparoscopy.

***Steps of Our Proposed System.*** Our proposed system takes steps such as marking and tracking surgical sites, then augmenting the surgical target. First, surgeons and radiologists examine pre-operative data such as MRI, CT and US images and determine the surgical site. However, the surgical site is only known in this pre-operative data in this stage. When a laparoscope is placed into a patient’s body, surgeons and radiologists can see the surgical target and identify the exact location of the surgical site by referring to the pre-operative data. This identified location needs to be maintained during the surgery and it is our primary outcome to detect it at runtime. To locate and track the identified location of the target surgical site, a surgeon pauses a streaming video from the laparoscopy system, and draws a polygon around the surgical target, for instance, an ellipse surrounding a uterus. The 3D surface of the surgical target reconstructed by our system is then viewed by the surgeon in order to locate the surgical site in 3D. Next, the surgeon marks the surgical site on the 3D surface, for instance, a 3D arrow, which is stored in our system for tracking in the following frames. At this stage, anatomical landmarks can be used to register the pre-operative data to the reference laparoscopic image. This process of pausing and editing for augmentation in a laparoscopic image can be repeated by the surgeon as many times as needed during the surgical procedure. From this minor interaction by surgeons in the first image, our system computes the positions of the surgical target in the rest of the images and displays the surgical sites along with visual information such as boundaries of organs and 3D arrows in

the 2D image or 3D view. As the system remembers the location of the surgical sites and allows surgeons to find them quickly, surgeons are now free to view monitor and check a patient’s database then easily continue the surgery without referring to, or re-examining the pre-operative data to repeatedly identify the surgical site.

***Surgical Site Tracking in Laparoscopic Images.*** Feature matching is a necessary first step, however, it is not straightforward as the target organ is often deformable and many state-of-the-art feature-matching algorithms [8] in computer vision are designed only for rigid objects. Our motivation comes from the assumption that it is possible to map a reconstructed 3D shape into another representation in a space which can be easier for feature matching. For mapping to another space, a conformal mapping is applied to the known 3D shape. Conformal mapping is a popular method in computer graphics [6] as it can be used to assign a texture onto a surface. This conformal mapping flattens out the 3D shape onto a plane while preserving angles locally. This is a nice property, as given two conformally mapped flattened surfaces they should be related by a local similarity transformation. In other words, two neighbouring points on a flattened mesh can be transformed to two corresponding points on another flattened mesh by a scale, rotation and translation. (Note that the similarity transformation is valid only locally but not globally.) Therefore, feature matching is done on the flattened images instead of the original input images, and we used Pizarro and Bartoli’s feature matching algorithm [10] to incorporate the local similarity, which is based on the SIFT [7] descriptor.

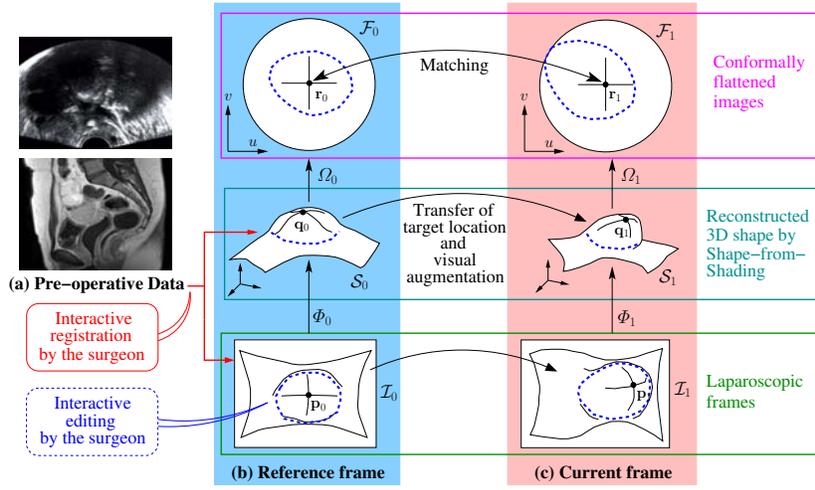
Once a set of matching points and the initial position of the organ are given, tracking becomes a problem of finding the approximation by a global similarity transformation. We assume that a set of local similarity transformations can be approximated by a global similarity transformation, which means the deformation on the surface changes only a small part of the shape of the surface. Also, the positions of the surgical targets tracked and overlaid on the original image are obtained by using affine Moving Least Squares [11].

## 2 Related work

Mountney *et al.* [8] evaluated various feature descriptors on Minimally Invasive Surgery (MIS) images and showed many feature descriptors do not work well. Although they provide an approach to selecting the best method among their evaluated algorithms and a Bayesian fusion method to increase accuracy and performance, it is preferable to have a single algorithm to find matches instead of running various feature descriptors since computing various feature descriptors for Bayesian fusion consumes most of the computation time for a practical MIS application. Su *et al.* [14] used augmented reality for stereo-laparoscopic images, however, this is limited to a rigid surface and stereo camera based laparoscopy is required to recover the 3D structure of the surface. Schaefer *et al.* [11] introduced affine-MLS for image deformation. They showed that a proper deformation can

be estimated from given key points on the surface mesh by a similarity transformation. Augmented Reality (AR) has been used in neurosurgery, otolaryngology and maxillofacial surgery [13]. Nicolau *et al.* claim that AR systems applied to MIS can increase the surgeon’s intra-operative vision by providing virtual transparency of the patient, but they also emphasized that AR systems are not robust enough since deformation of organs and human movement make registration difficult [9].

### 3 Proposed Method



**Fig. 1. System Overview.** Three columns showing (a) Pre-operative Data, (b) The Reference Frame and (c) The Current Frame. In a manual-preprocessing phase, registration is achieved interactively between (a) and (b). Then any surgical target planned from (a), for instance regions of interest or planned incision paths to be visualised in the laparoscopic images, can be transformed onto (b) via the interactive registration. The second phase involves tracking the surgical target, transferring the target location and augmenting visual information in subsequent laparoscopic images. The correspondence between (b) and (c) is achieved automatically using our robust matching method. This results in a set of robust feature matches. Determining the positions of the targets in (c) can then be achieved by mapping their locations in (b) to (c) via feature-based warping. Points  $\mathbf{p}_0$  and  $\mathbf{p}_1$  in two laparoscopic images  $\mathcal{I}_0$  and  $\mathcal{I}_1$  are back-projected by  $\Phi$  to the surfaces  $\mathcal{S}_0$  and  $\mathcal{S}_1$  as 3D points  $\mathbf{q}_0$  and  $\mathbf{q}_1$ , respectively. These 3D points are mapped to points  $\mathbf{r}_0$  and  $\mathbf{r}_1$  on planar surfaces  $\mathcal{F}_0$  and  $\mathcal{F}_1$  by a conformal mapping  $\Omega$ . Blue dotted lines indicate the boundary of the uterus, which is determined by surgeons and radiologists from pre-operative data such as MRI, CT or US in the reference image  $\mathcal{I}_0$ . The boundary and marked point in the reference frame  $\mathcal{I}_0$  are tracked to the next frame  $\mathcal{I}_1$  automatically. Our goal for augmentation is thus to determine and visualise the bounding region in  $\mathcal{I}_1$ , by performing image matching between  $\mathcal{F}_0$  and  $\mathcal{F}_1$ .

The general framework of our system is shown in Fig 1. Given a laparoscopic image  $\mathcal{I}_0$  as a reference frame, an image point  $\mathbf{p}_0$  is a projection of a 3D point  $\mathbf{q}_0$  on the surface  $\mathcal{S}_0$ . This projection can be represented by  $\Phi^{-1}$ . Therefore, the 3D point  $\mathbf{q}_0$  is  $\mathbf{q}_0 = \Phi_0(\mathbf{p}_0)$ . It can be mapped to a point  $\mathbf{r}_0$  on a flat surface  $\mathcal{F}_0$  by a conformal mapping  $\Omega_0$ . As a result,  $\mathbf{r}_0 = \Omega_0(\mathbf{q}_0) = \Omega_0(\Phi_0(\mathbf{p}_0))$ . Given another frame image  $\mathcal{I}_1$  and a point  $\mathbf{p}_1$ , in a similar way we may obtain the point  $\mathbf{r}_1$  on a planar surface  $\mathcal{F}_1$  as  $\mathbf{r}_1 = \Omega_1(\mathbf{q}_1) = \Omega_1(\Phi_1(\mathbf{p}_1))$ . Assuming that image pixel points  $\mathbf{p}_0$  and  $\mathbf{p}_1$  are a corresponding match, then the points  $\mathbf{r}_0$  and  $\mathbf{r}_1$  on the flat surfaces should be a pair of matching points in local isotropy. In other words,  $\mathbf{r}_0$  and  $\mathbf{r}_1$  are related by a similarity transformation locally. This is a key motivation that constructs our system for tracking by detection in laparoscopic images. Surgeons and radiologists use pre-operative data such as MRI, CT or US to locate the uterus and surgical site. Then the position of the surgical site is marked on the reference image  $\mathcal{I}_0$  and 3D mesh surface  $\mathcal{S}_0$ . Afterwards our system will detect and track the marked position of the uterus and surgical site in every laparoscopic image frame.

**Shape Reconstruction and Conformal Mapping.** Given an image, the 3D shape of the surface can be obtained from a monocular camera by the Shape-from-Shading (SfS) technique as summarized in [15]. In this paper, we use a real-time perspective SfS algorithm with a known light source calibration, described in [1]. In the computer graphics community, there have been many studies on surface manipulation and parametrization based on conformal mapping for texture mapping [6,4]. It is also used for 3D surface classification in computer vision [3]. With the assumption that SfS gives a correct shape reconstruction for the surface, a conformal mapping of the 3D reconstructed surface preserves angles on the flattened surface. Therefore, given two laparoscopic images and an surface shape estimated by SfS, two flattened surfaces are obtained by conformal mapping. They are related by local similarity transformations.

**Feature Matching and Outlier Removal.** For every incoming video frame, we estimate the 3D surface  $\mathcal{S}_1$  using the same SfS method as for  $\mathcal{S}_0$ , and flatten it to give us image  $\mathcal{F}_1$ . We then detect features in  $\mathcal{F}_1$  using SIFT, to give a query feature set  $\mathcal{G}_1$ . We then perform robust, nonrigid matching between  $\mathcal{G}_0$  and  $\mathcal{G}_1$  using Pizarro and Bartoli’s feature matching algorithm [10]. It works by matching features using descriptor similarity, and determines a high-probability inlier set based on spatial agreement with respect to local warp models. However, it is not completely outlier-free. We suppose that the boundary warp of the organ between  $\mathcal{F}_0$  and  $\mathcal{F}_1$  can be coarsely approximated by a similarity transform. This contrasts with the transformation of the boundary from image  $\mathcal{I}_0$  and  $\mathcal{I}_1$ , which, since these images comprise viewpoint changes, are likely to be more complex than a similarity transform. We have tested two robust similarity transformation estimation methods based on Horn’s absolute orientation [5]. The first is to use matched features with RANSAC (Random Sample Consensus) [2] and the second is to optimise a robust pseudo-huber norm cost function. Both methods allow us to detect outliers (based on their matches conflicting with the similarity

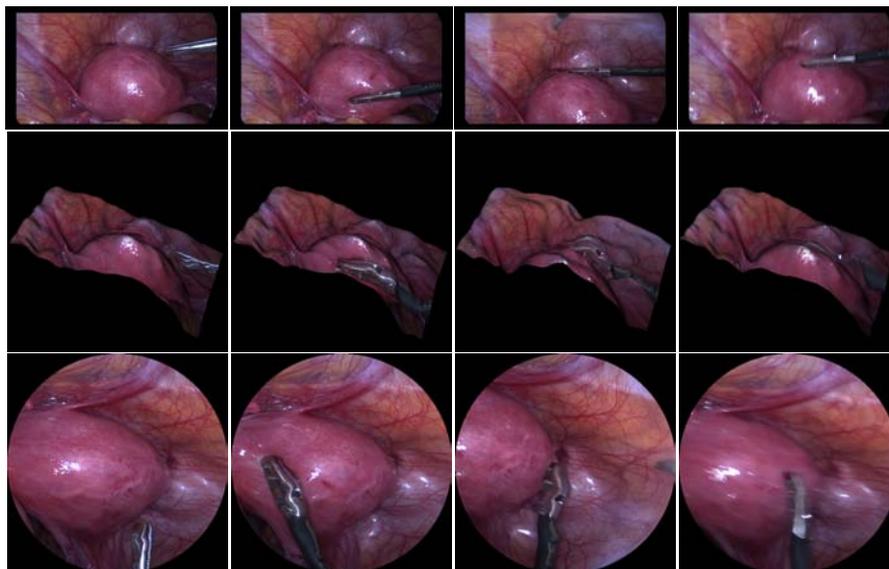
transform) which are then removed (we use a prediction error threshold). Once the boundary of the organ is estimated in  $\mathcal{F}_1$  its shape in  $\mathcal{I}_1$  is estimated as follows: first the vertices of the boundary in  $\mathcal{F}_1$  are mapped into 3D space on  $\mathcal{S}_1$  by barycentric interpolation. Then, they are projected onto  $\mathcal{I}_1$  using the camera’s projection function  $\Phi$  to give the boundary of the organ in  $\mathcal{I}_1$ .

***Surgical Target Mapping.*** Finally, we determine the positions of the surgical targets in  $\mathcal{I}_1$  as follows. First, we transform the inlier matches from  $\mathcal{F}_1$  to  $\mathcal{I}_1$ . We then use affine-MLS [11] to smoothly warp the target positions located on  $\mathcal{I}_0$  to  $\mathcal{I}_1$ , driven by the feature correspondences. This involves a single free parameter, the affine-MLS bandwidth, which we have set to  $\sigma = 150$  pixels in all of our tests. Once in position, the locations of the surgical targets are marked and presented to the surgeon, overlaid on top of the original input frame.

## 4 Experiments

We have carried out experiments on an image sequence in which a uterus is to be cut by a surgeon in order to remove a myoma (uterine fibroid) inside the uterus. As the myoma is not visible, surgeons and radiologists examine pre-operative data such as MRI and CT to determine the location of the surgical site, which will be the point of first-cut. An image sequence is captured by a Karl Storz laparoscopy system and the size of image is  $1048 \times 576$  pixels at 25 FPS. The number of captured frames is 2661 (1 min and 46 seconds). Samples of the image sequence are shown in Fig 2-(Top Row). From the input images, our SfS algorithm estimates the shape of the uterus as shown in Fig 2-(Middle Row). Then, conformal mapping is applied to the shape to obtain a flattened image at each frame as shown in Fig 2-(Bottom Row). These flattened images are used for finding feature matches. These matches are used to estimate a global similarity transformation for mapping the polygon boundary of the uterus at the reference frame to the following image frames as shown in Fig 3. Assuming that the surgeon selects frame 489 as a reference image, a polygon is drawn by the surgeon around the boundary of the uterus. The polygon is then transferred to other images when the video is resumed by the surgeon. Several frames of this result are shown in Fig 4 and Fig 5.

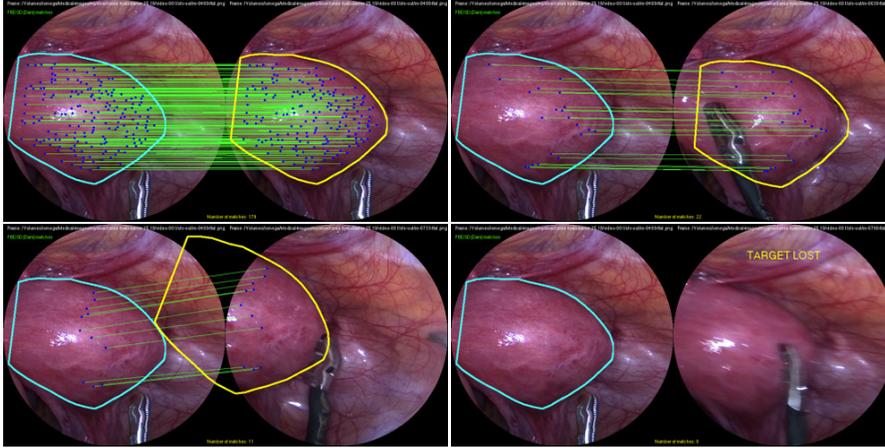
In Fig 6, the incision path and surgical site are shown as 3D arrows and a green curve, which are aligned on the surface of the uterus and augmented over all frames. The 3D arrows are orthogonal to the surface of the uterus recovered from SfS. Therefore, they provide useful information for surgeons to decide the orientation of the blade to enter the uterus. In this experiment, we used an open arc in Fig 6-(Bottom-right) as our planned incision path. This was manually marked by burning on the surface of the uterus, then overlaid in each subsequent view in the sequence by mapping using the correspondence via affine-MLS [11]. A selection of augmented frames from the sequence is shown in Fig 6. A video result is available at [http://youtu.be/LiZKmcV\\_fRg](http://youtu.be/LiZKmcV_fRg). With our system, this redundant burn mark is no longer necessary.



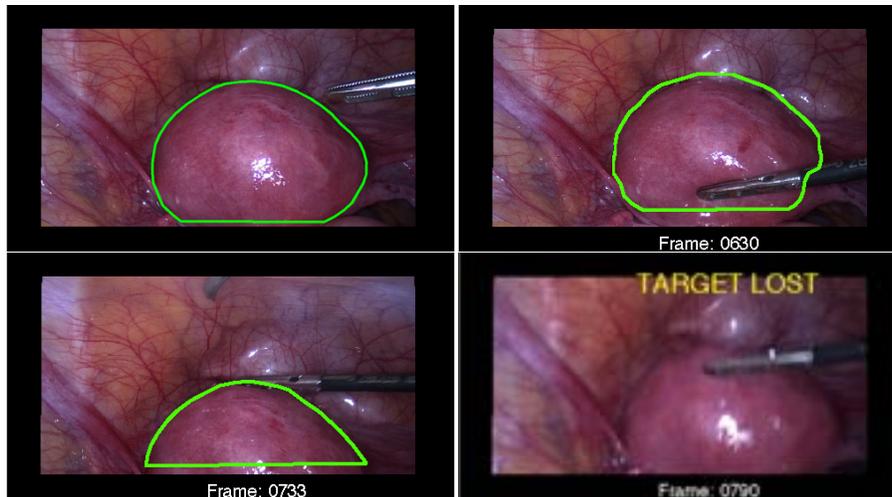
**Fig. 2. Samples of input images, 3D surface reconstruction and conformal mapping.** (Top Row) Frames number 489, 630, 733 and 790 (Left to Right). At frame 489, the uterus (an ellipsoid-like shape) is completely visible and it is a best candidate as a reference frame. At frame 630, the uterus is occluded by a surgical tool and it is deformed when pressed. At frame 733, only a half of the uterus is visible. At frame 790, the image is blurred by fast motion of the camera. (Middle Row) 3D surface shape from SfS at frame 480, 630, 730 and 790. (Bottom Row) Flattened image by conformal mapping of the 3D shape at frame 480, 630, 730 and 790.

## 5 Conclusion

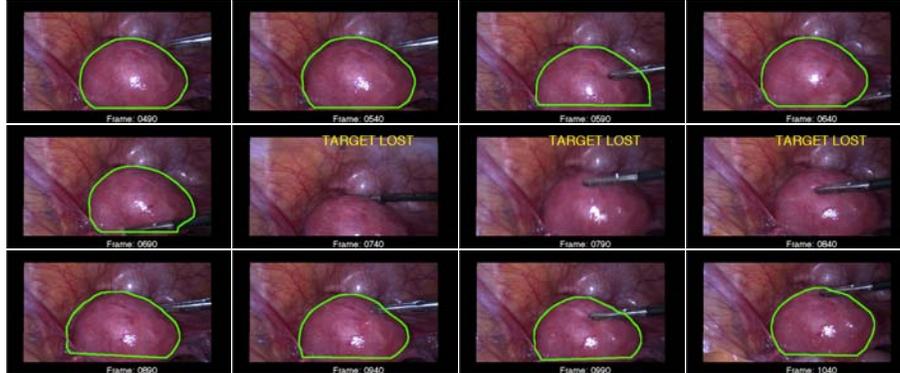
In this paper, we presented a new method for tracking and augmenting surgical targets in laparoscopic images. The system allows surgeons to mark the surgical site using pre-operative data in laparoscopic images. Our method utilizes Shape-from-Shading to recover the 3D shape of the surface, and the 3D shape is flattened by a conformal mapping which preserves angles on the surface. Feature matching is carried out on the flattened images and a global similarity transformation is applied to obtain a mapped boundary of the surgical target and outliers are removed. The surgical target registration by an affine Moving Least Squares warp made the surgical target can be localized in a laparoscopic image. Real experiments conducted on a uterus in laparoscopic images show robust tracking of the uterus and consistent surgical target augmentation. The method obviates the necessity to mark the surgical site physically on the organ surface



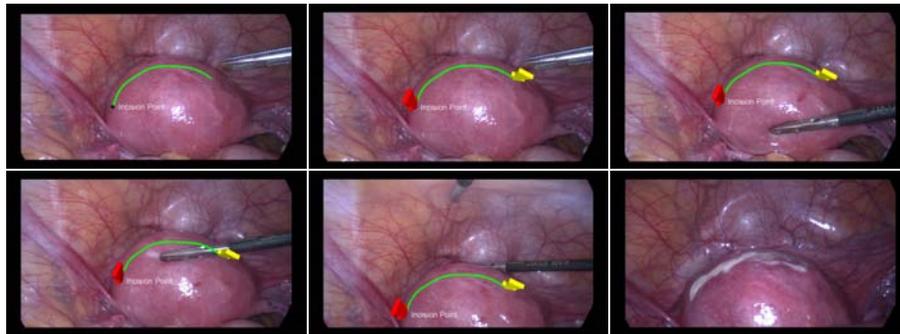
**Fig. 3.** *Feature matching and tracking between flattened images.* (Top-left) In frames 489 and 490, a total of 179 matches are found and the cyan polygon in reference frame 489 is mapped to a yellow polygon in the next frame 490. (Top-right) Matches and tracking between frame 489 and 630. In total 22 matches are found. (Bottom-left) 11 matches are found between frame 489 and 733. (Bottom-right) Tracking failed as no matches are found between frame 489 and 790.



**Fig. 4.** *Frames with uterus tracking.* (Top-Left) Reference frame 489 with a green polygon drawn on the boundary of the uterus. (Top-Right) At frame 630, tracking of the uterus is shown as a green polygon. (Bottom-Left) At frame 730, tracking is still successful for the half visible uterus. (Bottom-Right) At frame 799, tracking fails because of blurring in the image.



**Fig. 5. Tracking the uterus.** Results of tracking a uterus from laparoscopic images are shown. The boundary of the tracked uterus is indicated as a green polygon and a message (target lost) is given when the system is searching for the uterus at the current frame. Note that our system fails to track the uterus at frame 740, however tracking resumes and finds the uterus successfully at frames 890 and 1,040.



**Fig. 6. 3D augmentation in laparoscopy.** (Top row and bottom first two columns) These are sample images of 3D augmentation in laparoscopic images. A red (left) 3D arrow indicates the incision point, A yellow (right) 3D arrow shows the ending point, and a green curve shows the surgical site for the surgeon's first-cut. Our system augments this 3D visual information by tracking the uterus over all frames. (Bottom-right) At frame 2173, a surgeon made a burn mark on the surface of the uterus. This is not necessary in our system. The video is available at the link: <http://youtu.be/hvzE9VIAjPI>.

## References

1. Collins, T., Bartoli, A.: Toward live monocular 3d laparoscopy using shading and specular information. IPCAI (2012)
2. Fischler, M., Bolles, R.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24(6) (Jun 1981)
3. Gu, X., Yau, S.: Surface classification using conformal structures. *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on* pp. 701–708 vol. 1 (2003)
4. Haker, S., Angenent, S., Tannenbaum, A., Kikinis, R., Sapiro, G., Halle, M.: Conformal surface parameterization for texture mapping. *Visualization and Computer Graphics, IEEE Transactions on* 6(2), 181–189 (2000)
5. Horn, B.K.P.: Closed-form solution of absolute orientation using unit quaternions. *JOSA A* 4(4), 629–642 (1987)
6. Lévy, B., Petitjean, S., Ray, N., Maillot, J.: Least squares conformal maps for automatic texture atlas generation. *ACM SIGGRAPH conference proceedings* (Apr 2002)
7. Lowe, D.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), 91–110 (2004)
8. Mountney, P., Lo, B., Thiemjarus, S., Stoyanov, D., Zhong-Yang, G.: A probabilistic framework for tracking deformable soft tissue in minimally invasive surgery. *MICCAI - Lecture Notes in Computer Science* 4792, 34–41 (2007)
9. Nicolau, S., Solar, L., Mutter, D., Marescaux, J.: Augmented reality in laparoscopy surgical oncology. *Surgical Oncology* 20, 189–201 (2011)
10. Pizarro, D., Bartoli, A.: Feature-based deformable surface detection with self-occlusion reasoning. *International Journal of Computer Vision* (published online 27 Apr 2011) (2011)
11. Schaefer, S., McPhail, T., Warren, J.: Image deformation using moving least squares. *ACM Transactions on Graphics (TOG)* 25(3), 533–540 (2006)
12. Seiden, S.C., Barach, P.: Wrong-side/wrong-site, wrong-procedure, and wrong-patient adverse events: Are they preventable? *Archives of Surgery* 141(9), 931–939 (2006)
13. Shuhaiber, J.H.: Augmented reality in surgery. *ARCH SURG* 139, 170–174 (Jan 2004)
14. Su, L.M., Vagvolgyi, B.P., Agarwal, R., Reiley, C.E., Taylor, R.H., Hager, G.D.: Augmented reality during robot-assisted laparoscopic partial nephrectomy: Toward real-time 3d-ct to stereoscopic video registration. *Journal of Urology* 73(4), 896–900 (2009)
15. Zhang, R., Tsai, P., Cryer, J., Shah, M.: Shape-from-shading: a survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 21(8), 690–706 (1999)