

# Robust Surface Tracking Combining Features, Intensity and Illumination Compensation

Xiaofei Du, Neil Clancy, Shobhit Arya,  
George B. Hanna, John Kelly, Daniel S.  
Elson · Danail Stoyanov

Received: date / Accepted: date

## Abstract

*Purpose* Recovering tissue deformation during robotic-assisted minimally invasive surgery (MIS) procedures is important for providing intra-operative guidance, enabling *in vivo* imaging modalities and enhanced robotic control. The tissue motion can also be used to apply motion stabilization and to prescribe dynamic constraints for avoiding critical anatomical structures.

*Methods* Image-based methods based independently on salient features or on image intensity have limitations when dealing with homogeneous soft-tissues or complex reflectance. In this paper, we use a triangular geometric mesh model in order to combine the advantages of both feature and intensity information and track the tissue surface reliably and robustly.

---

Xiaofei Du  
Centre for Medical Image Computing, Department of Computer Science, University College  
London, United Kingdom  
E-mail: xiaofei.du.13@ucl.ac.uk

Neil Clancy  
Department of Surgery and Cancer, Imperial College London, United Kingdom  
E-mail: n.clancy@imperial.ac.uk

Shobhit Arya  
Department of Surgery and Cancer, Imperial College London, United Kingdom

George B. Hanna  
Department of Surgery and Cancer, Imperial College London, United Kingdom  
E-mail: g.hanna@imperial.ac.uk

John Kelly  
Division of Surgery and Interventional Science, University College London, United Kingdom

Daniel S. Elson  
Department of Surgery and Cancer, Imperial College London, United Kingdom  
E-mail: daniel.elson@imperial.ac.uk

Danail Stoyanov  
Centre for Medical Image Computing, Department of Computer Science, University College  
London, United Kingdom  
E-mail: danail.stoyanov@ucl.ac.uk

**Results** Synthetic and *in vivo* experiments are performed to provide quantitative analysis of the tracking accuracy of our method, we also show exemplar results for registering multispectral images where there is only a weak image signal.

**Conclusions** Compared to traditional methods, our hybrid tracking method is more robust and has improved convergence in the presence of larger displacements, tissue dynamics and illumination changes.

**Keywords** Non-rigid surface tracking · Multispectral imaging · Minimally invasive surgery · Illumination compensation

## 1 Introduction

Medical image computing and surgical vision can play an important role towards improving the surgeon’s operating capabilities in highly dynamic anatomical regions where tissue motion can complicate surgical dexterity and impede image-guidance or intra-operative imaging [24, 28]. In MIS, recovering *in vivo* tissue deformation in real-time by using endoscopic images has been explored predominantly for deploying robotic motion stabilization [26]. While both 2D and 3D tracking methodologies have been reported the problem of robustly tracking tissues with poor texture characteristics remains a challenging task due to the illumination complexity and variation, specular highlights and occlusions from the surgical instruments [16].

Early work on tracking tissue motion in endoscopic video focused on the use of feature-based methods in order to achieve real-time performance [17, 27]. More recently robust feature driven techniques have been developed and reported to achieve robust and long-term tracking invariant to difficult transformations [9, 20, 29]. The limitation of these approaches is that a dense region of the tissue is not recovered and rather single points of interest are detected and tracked which these can be isolated in specific regions. On the other hand, dense intensity-based methods have been reported where the tissue surface is modelled as a geometric mesh, for example, using a Thin-plate spline (TPS) or Free-form deformation (FFD) technique. Tracking is performed over the entire space covered by this model. Richa *et al.* [21] employed a TPS model to estimate the heart surface deformation using multiple visual techniques to increase robustness and spatial resolution. Braux-Zin *et al.* [4] introduced a new model of non-rigid surface registration to merge feature and intensity-based costs in a pyramidal variational approach, however, the model fails in the presence of illumination variations. Besides, additional specialized hardware can also be used for soft-tissue reconstruction [13].

In this study, we combine the advantages of both feature and intensity information to track the tissue surface robustly and in difficult conditions with poor illumination. Our method uses a triangular mesh model with geometric regularization to represent the region of interest for tracking. The energy cost function we optimize incorporates terms for feature correspondence energy and also for intensity energy. We report encouraging results and compare our algorithm with earlier works to show that performance is enhanced and the method can cope with large motions due to the feature components while also handling regions of poor textural information through the use of illumination compensated appearance. We also show preliminary results applying our method to multispectral images, where the signal can be low and difficult for tracking algorithms.

## 2 Methods

We model the tracked tissue surface as a geometric mesh which can describe the deformable tissue motion. During tracking, the locations of the mesh vertices are updated to minimize energy functions with regard to the feature correspondence alignment, dense image residuals and also illumination variations.

### 2.1 Geometric mesh model

In this work, we use the mesh model proposed by Pilet *et al.* [19]. As shown in Fig. 1, the non-rigid tissue surface  $M$  to be tracked is modelled as a 2D triangular geometric mesh with  $N$  vertices.  $\mathbf{v}_i = (v_{ix}, v_{iy})$  represents the pixel location of the  $i^{th}$  triangle vertex of the mesh. Let  $\mathbf{V}_x \in R^N$  and  $\mathbf{V}_y \in R^N$  be the vector of stacked vertex coordinates of the x-axis and y-axis respectively. To represent the shape and motion of  $M$ , we define a state vector  $\mathbf{S} = (\mathbf{V}_x, \mathbf{V}_y)^T \in R^{2N}$ .

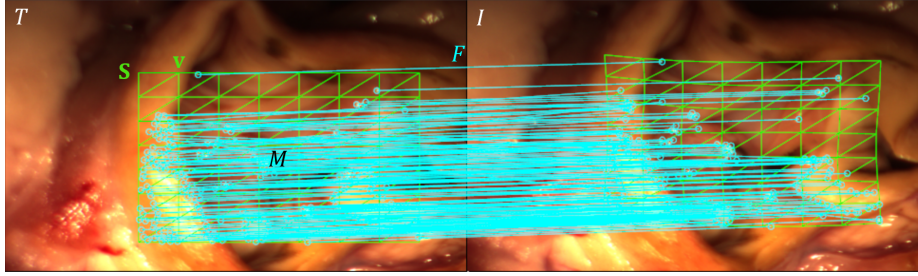


Fig. 1: The left and right images are the template image  $T$  and the input image  $I$  respectively.  $F$  is the set of feature correspondence obtained using feature matching algorithm (shown as cyan). The tissue surface  $M$  is modelled as a triangular mesh model (shown as green), so the deformation and motion of the surface  $M$  are controlled by the state vector  $\mathbf{S}$  consisting of the mesh vertices  $\mathbf{v}$ .

Any point  $\mathbf{p}$  within  $M$  can be located via the warping function  $\mathbf{W}(\mathbf{p}; \mathbf{S})$  by using its barycentric coordinate  $(b_i, b_j, b_k)$  and the vertices of the triangle it lies within  $(\mathbf{v}_i, \mathbf{v}_j, \mathbf{v}_k)$ , where the triplets  $(i, j, k)$  represent the triangle vertex indice. Barycentric vector  $\mathbf{b} \in R^N$  only contains non-zero elements at index  $(i, j, k)$ .

$$\mathbf{W}(\mathbf{p}; \mathbf{S}) = \begin{bmatrix} \mathbf{b} & \mathbf{0} \\ \mathbf{0} & \mathbf{b} \end{bmatrix} [\mathbf{V}_x \ \mathbf{V}_y]^T = \mathbf{B}\mathbf{S} \quad (1)$$

where  $b_i + b_j + b_k = 1$ . Then the task of surface tracking is to estimate the state vector  $\mathbf{S}$  of the mesh through the whole image sequence.

### 2.2 Feature-based tracking

One advantage of our framework is that it can cope with any kind of feature detection and tracking method as originally presented by [8]. In this study we

used the Speeded Up Robust Features (SURF) descriptor [2] implemented in the OpenCV library [3] to obtain feature correspondences.

Given a set of feature correspondences, to estimate the deformation of  $M$  with tissue displacement, we minimize an energy function subject to  $\mathbf{S}$  in the following equation:

$$\varepsilon_F(\mathbf{S}) = \lambda \varepsilon_R(\mathbf{S}) + \varepsilon_C(\mathbf{S}) \quad (2)$$

where  $\varepsilon_R$  represents the regularization energy,  $\varepsilon_C$  is a feature correspondence measure and  $\lambda$  controls the regularization influence [19].

No matter what kind of feature matching algorithm is used, it is usually inevitable to avoid erroneous correspondences, which we consider as outliers. The presence of outliers can severely affect deformation estimation, for instance by breaking the mesh topology. Therefore, the regularization energy  $\varepsilon_R$  is used to prevent the mesh model from overfitting the data. The mesh model can be considered as a set of hexagons, one of which is shown in green in Fig. 2.

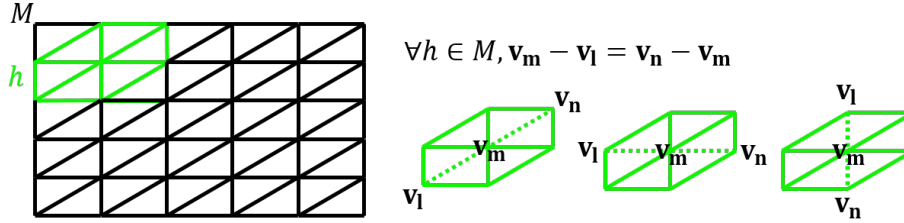


Fig. 2: A hexagonal element  $h$  in the undeformed mesh model (shown in green). The distance between co-linear vertices is equal under certain types of hexagon motion.

For each hexagon triplets  $(l, m, n)$  in the undeformed mesh model, the distances between all the co-linear vertices are equal respectively. This property can be used to preserve the regularity of the mesh. We separate the coordinates of the vertices along x and y axis, and therefore the regularization energy term is defined as in [7]:

$$\begin{aligned} \varepsilon_R(\mathbf{S}) &= \frac{1}{2} \sum_{(l,m,n) \in E} (v_{lx} - 2v_{mx} + v_{nx})^2 + (v_{ly} - 2v_{my} + v_{ny})^2 \\ &= \frac{1}{2} \sum_{(l,m,n) \in E} \sum_{i=x,y} \begin{bmatrix} v_{li} \\ v_{mi} \\ v_{ni} \end{bmatrix}^T \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}^T \begin{bmatrix} v_{li} \\ v_{mi} \\ v_{ni} \end{bmatrix} \end{aligned} \quad (3)$$

where  $E$  is composed of all the index triplets  $(l, m, n)$  for the co-linear vertices. For convenience, this can be formulated in the matrix form:

$$\begin{aligned} \varepsilon_R(\mathbf{S}) &= \frac{1}{2} (\mathbf{V}_x \mathbf{K} \mathbf{V}_x^T + \mathbf{V}_y \mathbf{K} \mathbf{V}_y^T) \\ &= \frac{1}{2} \mathbf{S}^T \begin{bmatrix} \mathbf{K} & \mathbf{0} \\ \mathbf{0} & \mathbf{K} \end{bmatrix} \mathbf{S} = \frac{1}{2} \mathbf{S}^T \mathbf{R} \mathbf{S} \end{aligned} \quad (4)$$



where  $\mathbf{R} \in R^{2N \times 2N}$  is a sparse and regular matrix which can be determined by the set of triplets  $E$ .

Let  $\mathbf{c}_t = (u, v)^T$  represent the location of one feature in the frame  $t$ , and  $F$  be the set of all correspondences we obtained after feature matching. The feature correspondence energy is defined as:

$$\varepsilon_C(\mathbf{S}) = \sum_{\mathbf{c} \in F} \rho(\|\mathbf{c}_t - \mathbf{W}(\mathbf{c}_{t-1}; \mathbf{S})\|, r) \quad (5)$$

where  $\rho$  is a robust estimator, and  $r$  is the confidence radius. The choice of the robust estimator is crucial for decreasing the effect of outliers. Various robust functions exist [4, 19, 30], in this work we follow the estimator and optimization method proposed by Zhu *et al.* [30]. Based on the modified finite Newton method [12, 14], Zhu *et al.* proposed the Progressively Finite Newton (PFN) method [30], in which the robust estimator  $\rho(\delta, r)$  uses a coarse-to-fine scheme. The initial value of  $r$  is set to 500 and is progressively reduced at a constant rate. The optimization process stops when the confidence radius  $r$  reduces to one pixel which is close to the expected precision. Because the method only needs one Newton step for each  $r$  to achieve convergence, the whole process can be solved in a fixed number of steps.

### 2.3 Deformable Lucas-Kanade method

Feature-based tracking is fast and can handle large displacements, but it has limitations because of the sparse motion field and the reliance on salient image texture. In [30], the authors employed a deformable Lucas-Kanade (DLK) method, which is a deformable variation of the intensity-based Lucas-Kanade (LK) method [1]. The energy function subject to  $\mathbf{S}$  is defined as:

$$\varepsilon_I(\mathbf{S}) = \eta \varepsilon_R(\mathbf{S}) + \varepsilon_{SSD}(\mathbf{S}) \quad (6)$$

where  $\varepsilon_R$  represents the regularization energy and  $\eta$  controls the regularization influence. The intensity energy  $\varepsilon_{SSD}$  uses the Sum of Squared Differences (SSD) to form a similarity measure between the template image  $T$  and the input image  $I$ :

$$\varepsilon_{SSD}(\mathbf{S}) = \frac{1}{2} \sum_{\mathbf{p}} [T(\mathbf{W}(\mathbf{p}; \mathbf{S})) - I(\mathbf{W}(\mathbf{p}; \mathbf{S}))]^2 \quad (7)$$

The SSD metric directly compares the illumination of every pixel  $\mathbf{p}$  in the tracked area, which makes it quite sensitive to changes in lighting. Recently, a new similarity metric called the Sum of Conditional Variance (SCV) was introduced for multi-modal medical image registration [18], and Richa *et al.* [21] used it for visual tracking. In MIS, compared to SSD, SCV is invariant to non-linear illumination variations. In this study, we improved the DLK method of Zhu *et al.* [30] by employing the SCV metric. Let  $[0, d_T]$  and  $[0, d_I]$  represent the intensity range of the template image  $T$  and the input image  $I$  respectively, the intensity energy  $\varepsilon_{SCV}$  is defined as:

$$\varepsilon_{SCV}(\mathbf{S}) = \frac{1}{2} \sum_{\mathbf{p}} [T(\mathbf{W}(\mathbf{p}; \mathbf{S})) - \hat{I}(\mathbf{W}(\mathbf{p}; \mathbf{S}))]^2 \quad (8)$$

with the SCV image

$$\hat{I}(\mathbf{W}(\mathbf{p}; \mathbf{S})) = \mathcal{E}(T(\mathbf{p})|I(\mathbf{W}(\mathbf{p}; \mathbf{S}))) \quad (9)$$

where  $\mathcal{E}(\cdot)$  is the expectation operator. This can be computed using the joint intensity distribution between  $T$  and  $I$

$$P_{ij} = \frac{1}{N_{\mathbf{p}}} \sum_{\mathbf{p}} \Phi(T(\mathbf{p}) - i) \Phi(I(\mathbf{W}(\mathbf{p}; \mathbf{S})) - j) \quad (10)$$

where  $N_{\mathbf{p}}$  represents the number of pixels,  $\Phi(x) = 1$  if and only if  $x = 0$ ,  $i \in [0, d_T]$ ,  $j \in [0, d_I]$ . The conditional expectation can then be computed as:

$$\mathcal{E}(T(\mathbf{p})|I(\mathbf{W}(\mathbf{p}; \mathbf{S}))) = \sum_i i \frac{P_{ij}}{\sum_i P_{ij}} \quad (11)$$

During tracking, the SCV image  $\hat{I}(\mathbf{W}(\mathbf{p}; \mathbf{S}))$  is computed only once for every input image  $I$  [21]. Then the optimization can be processed like the standard procedure for the LK method. After performing the first-order Taylor expansion and setting the derivative equal to zero, solving for  $\Delta \mathbf{S}$  we have:

$$\Delta \mathbf{S} = \mathbf{H}^{-1} \left( -\eta \mathbf{R} \mathbf{S} + \sum_{\mathbf{p}} (\nabla T \frac{\partial \mathbf{W}}{\partial \mathbf{S}})^T (\hat{I}(\mathbf{W}(\mathbf{p}; \mathbf{S})) - T(\mathbf{p})) \right) \quad (12)$$

with the Hessian matrix:

$$\mathbf{H} = \eta \mathbf{R} + \sum_{\mathbf{p}} (\nabla T \frac{\partial \mathbf{W}}{\partial \mathbf{S}})^T (\nabla T \frac{\partial \mathbf{W}}{\partial \mathbf{S}}) \quad (13)$$

Because any LK method is based on the assumption that the current estimate of the parameters is approximately correct. This means that by using  $\varepsilon_{SSD}$  or  $\varepsilon_{SCV}$  alone we cannot deal with significant displacement between frames [1]. We use the feature-based tracking result as initialization to fulfil this assumption and to lead the optimization toward correct convergence.

## 2.4 Template updating

After an extended sequence of tracking, it is possible that the original template will not accurately represent the tracked surface due to physiological effects such as bleeding after instrument-tissue interactions. To avoid errors caused by the appearance changes we use a template updating strategy [15]. First, at every frame we update the template image to be the tracked region of the input image. In this way, the updated template could lessen the possible appearance difference between the original template and input image. At the same time small errors will accumulate during this process and cause the template to gradually drift away. Therefore, to correct the drift we keep the original template and align the updated template with it to estimate the final update. This two-step template update with drift correction strategy can avoid local minima during optimization and prevent the drifting problem. Additionally, specular reflections create strong image gradients which are salient and can bias feature detection and appearance-based tracking metrics. We use a combination of intensity thresholding and dilation operations to remove the highlights [10, 25] (see Fig. 3).

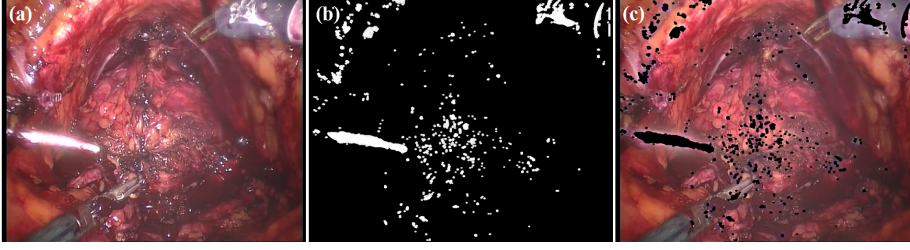


Fig. 3: Specular highlight removal procedure before tracking: (a) Before highlight removal; (b) Highlight mask; (c) After highlight removal.

### 3 Experiments and results

#### 3.1 Synthetic data experiments

As ground truth information for soft-tissue motion is not available during surgery, we used a custom simulation environment in Fig. 4 to mimic the periodic deformation of the tissue surface induced by the cardiac cycle and respiration [23]. The environment can generate synthetic image sequences by performing small but arbitrary rotations and translations to the pixels of one template image. To test the computational stability of the method, we also added different levels of Gaussian noise (noise=5%, 10% and 20%) to the synthetic sequence.

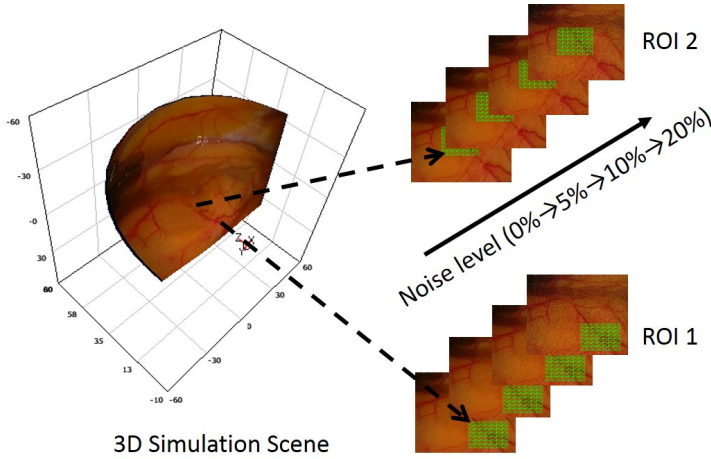


Fig. 4: Simulation environment experiment setup for generating synthetic image sequences.

To quantitatively validate the performance of our algorithm, we tracked two regions of interest (ROI) as seen in Fig. 4. The first ROI located near the right bottom corner deforms towards the center during tracking, so the displacement

is large between frames; the second ROI is located in the central area and is compressed during tracking.

Since we generated the whole synthetic sequence, we have the ground truth of the mesh vertices in each frame. We computed and compared the mean and standard deviation (STD) of the tracking error (pixel) compared to the ground truth with different methods: the feature-based PFN method, the modified intensity-based DLK using the SCV metric and our proposed hybrid PFNLK method.

The tracking results of the two ROIs with different noise levels through the sequence are shown in Fig. 5. Since the displacement between frames of this sequence is too large, the DLK method quickly loses track, while the PFN and our PFNLK method track fairly well. The PFNLK method outperforms all the other methods, but more obviously for ROI 2 than ROI 1.

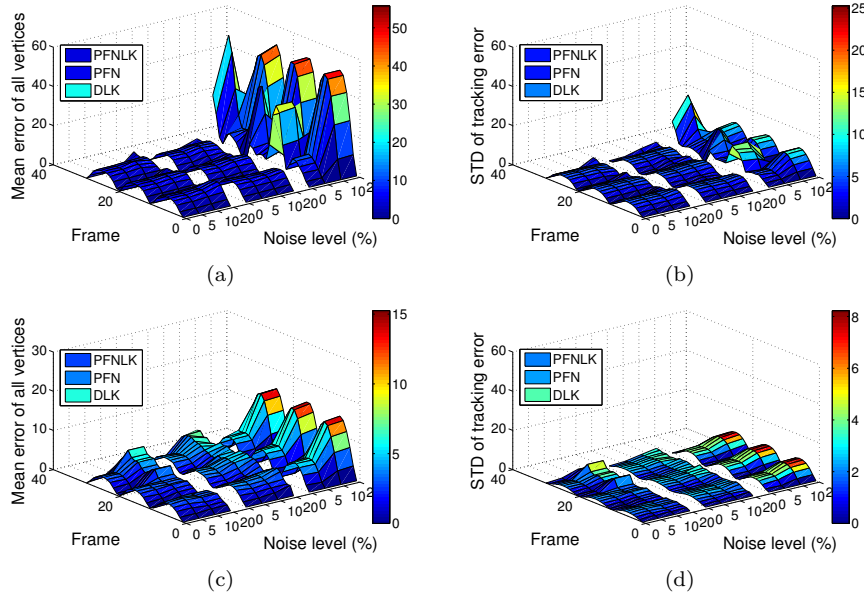


Fig. 5: Synthetic experiment results: (a) Mean error for ROI 1; (b) The standard deviation of error for ROI 1; (c) Mean error for ROI 2; (d) The standard deviation of error for ROI 2.

### 3.2 *In vivo* data experiments

For evaluating the potential clinical value of the proposed tracking algorithm, experiments on *in vivo* image sequences recorded at 25 fps using the da Vinci<sup>TM</sup> surgical robotic platform (Intuitive Surgical, Inc.) have been conducted. Since the lack of ground truth is a problem for validating tracking performance with real surgical sequences, we used the modified forward-backward tracking methodology based on even-odd frames [22]. For a given sequence, the forward tracking is made on the

even frames and the backward tracking is made on the odd frames. The assumption is that if a ROI is perfectly tracked, it should return to the initial location in the first frame. This is considered to be artificial ground truth for tracking methods. Compared to the original forward-backward tracking strategy [11], which tracked the ROI frame by frame as they move forward and backward to the beginning of the sequence, the backward tracking is decorrelated from the forward tracking by using different frames. In our experiments, we chose a robotic radical prostatectomy sequence, which is represented by  $I = (I_0, I_1, I_2, \dots, I_n)$ . Then according to the above methodology,  $FB = (I_0, I_2, I_4, \dots, I_{n-2}, I_n, I_{n-1}, \dots, I_3, I_1, I_0)$  is the corresponding modified forward-backward sequence where  $I_t$  is the frame  $t$  of the original sequence. As seen in Fig. 6, the first frame (frame 0) is the same as the last frame (frame 50) in the  $FB$  sequence.

We tracked the same ROI with the three methods, and the tracking result can be seen in Fig. 6. The DLK method loses track, but the PFN and the PFNLK methods track well. Until the last frame, the PFN method tracked back to near the original location, while the PFNLK method tracked back to the initial location. Normalized Cross-Correlation (NCC) can also be used to evaluate the tracking

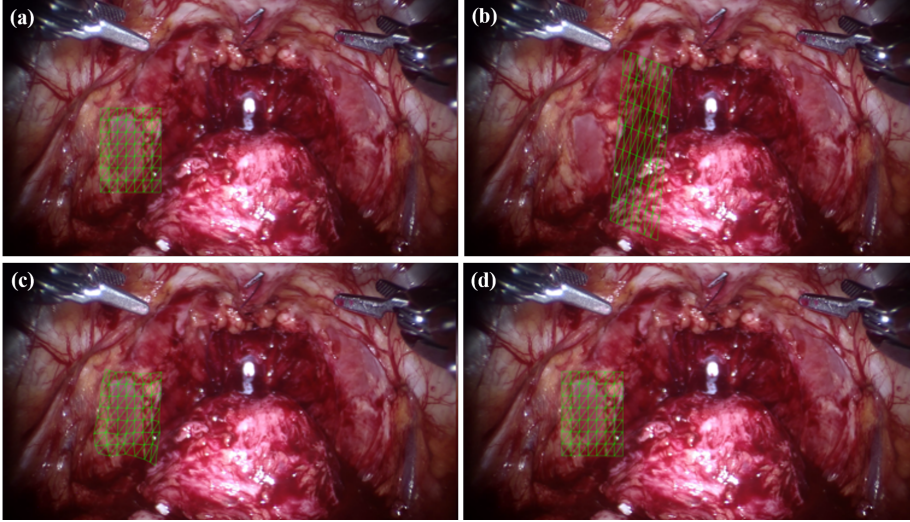


Fig. 6: Comparison of performance for a  $FB$  sequence with camera motion. The first frame (frame 0) and the last frame (frame 50) are identical, so if a ROI is perfectly tracked, it should return to the initial location in the first frame: (a) Frame 0; (b) The DLK frame 50; (c) The PFN frame 50; (d) The PFNLK frame 50.

performance quantitatively. Higher NCC value is a surrogate measure for better tracking performance as it shows close image alignment. We computed the NCC between the template ROI and the tracked ROI in Fig. 7a and computed it again after the SCV illumination mapping step shown on Fig. 7b. The similarity between Fig. 7a and Fig. 7b means the illumination changes in this sequence are not large.

As the figure indicates, the PFNLK method outperformed the PFN method which slowly drifted away during the backward tracking.

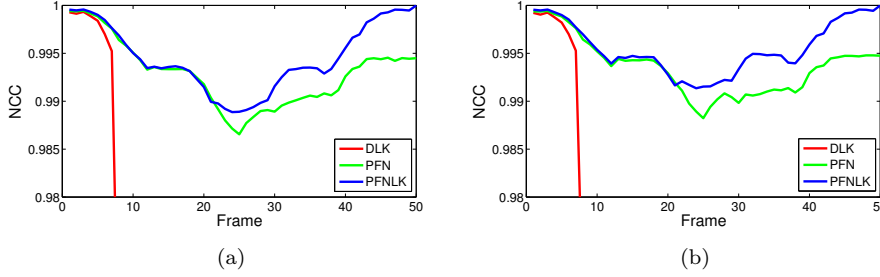


Fig. 7: The comparison of NCC and of tracked point with different tracking methods throughout the *FB* sequence: (a) NCC between the original template and tracked ROIs; (b) NCC computed after SCV illumination mapping step.

We chose another radical prostatectomy sequence<sup>1</sup> to evaluate the tracking performance in the presence of instrument occlusions. Occlusions are commonplace throughout the surgical procedure and present a significant challenge to tracking algorithms, especially if the instruments deform and manipulate the tissue of interest. This sequence consists of 600 frames and the tracked ROI is occluded by the surgical tools during certain time periods during the sequence.

We initialized each of the three methods with the same ROI and representative tracking results over the full sequence are shown in Fig. 8. Since the DLK method encountered tracking failure quickly, we left it out of the discussion below. It is possible to observe that on frame 81 (the second column), the tracked ROI is occluded by the surgical tool, and for the following frames after the occlusion the PFN optimization of the mesh is trapped in a local minima and never recovers back, as seen on frame 124 (the third column). By using the PFNLK method the tracked ROI recovered after frame 124 and continued tracking suggesting that the algorithm is more robust compared to the PFN method.

Following the same validation procedure, we computed the NCC between the template and the tracked ROIs for this sequence without and with the illumination mapping step. The results are shown in Fig. 9. A sharp drop-off and recovery can be observed between frames 172 and 308 without the illumination mapping in Fig. 9a, however, this is not reflected by visually inspecting the quality of the tracking results. In the original sequence the tracked areas were shifted to the very left side of the view during this interval, so the illumination condition changed greatly due to the camera motion. This inauthentic change reflects that the NCC metric cannot handle non-linear light changes very well. In Fig. 9b, the input images are mapped to mimic the illumination condition of the template image using the SCV metric, and we can see that the inauthentic changes of the NCC have disappeared, also the NCC went up after frame 123 for the PFNLK method, which is accordance with the visual interpretation. This means that the NCC can be trusted as a surrogate measure under similar lighting conditions. We can also

<sup>1</sup> <http://www.surgicalvision.cs.ucl.ac.uk/data>



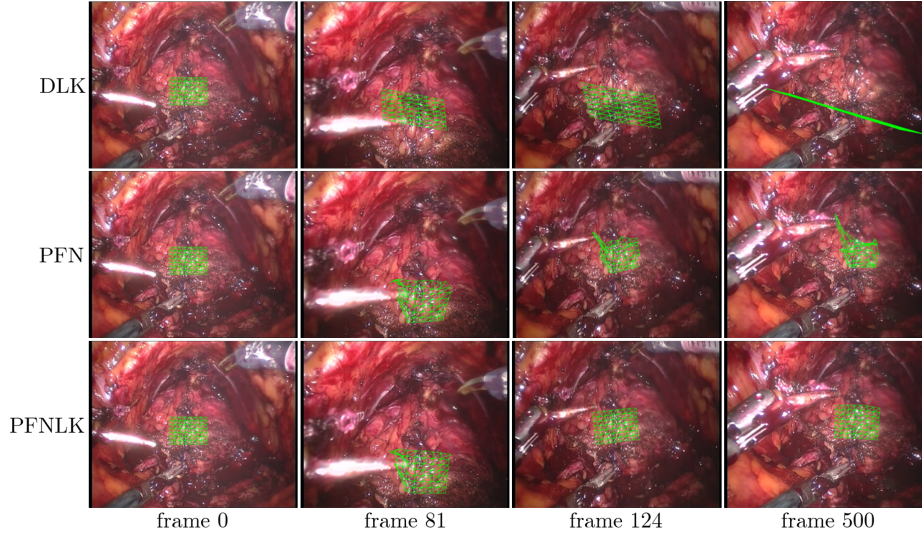


Fig. 8: Comparison of performance for occlusion sequence: (Top row) intensity-based DLK method; (Middle row) feature-based PFN method; (Bottom row) our hybrid PFNLK method.

infer from the evaluation that the SCV metric we used is necessary if there exist potential illumination variations.

Furthermore, we manually tracked one point within the ROI through the whole sequence, and use the position of the point in each frame as ground truth (GT). In Fig. 9c the tracked trajectories of different methods are illustrated with the GT, also the tracking errors are computed and shown in Fig. 9d, we can see that the tracking error of the PFNLK is the lowest, which is consistent with the analysis above.

### 3.3 Experiments with multispectral data

To illustrate the importance of using the SCV rather than an illumination sensitive metric, we provide exemplar results of registering multispectral images. In sequential multispectral images where the image stack is acquired one wavelength at a time, some images can have very low signal strength due to the camera and light-tissue interaction characteristics such as absorption and scattering as shown in Fig. 10. If the tissue under interrogation moves during acquisition, this causes misalignment of the multispectral stack and, for instance, renders spectral analysis to calculate oxygenation levels impossible [5]. We tracked the same ROI using the DLK method with and without the SCV illumination mapping step and the tracking result are shown in Fig. 10. The histogram of the SCV images in the figure is equalized because the template image is too dark.

As Fig. 10 indicates, the original DLK method in [30] loses track eventually (see top row); while our modified DLK method maps the input image  $I$  to the template image  $T$  in order to obtain the SCV image  $\hat{I}$ , after the SCV mapping

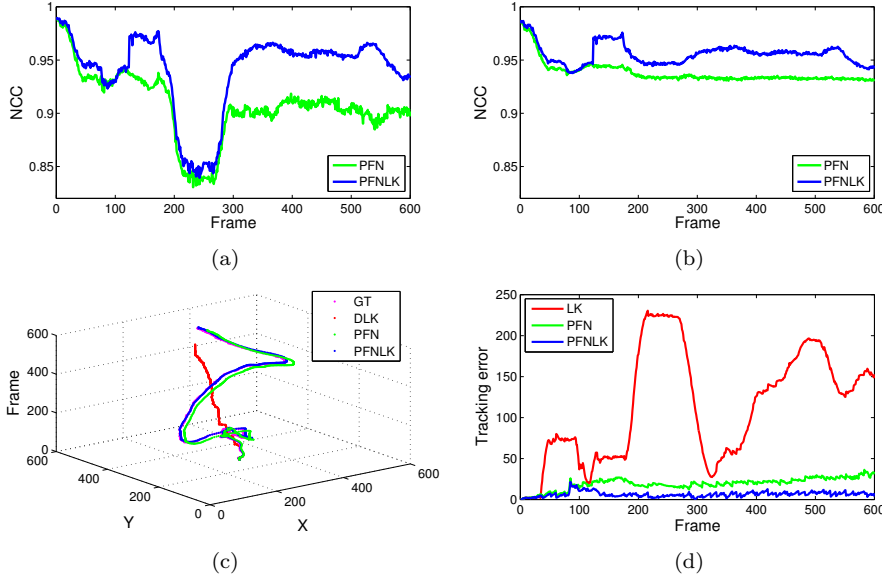


Fig. 9: The comparison of the NCC and of tracked point with different tracking methods throughout the occlusion sequence: (a) The NCC between the original template and tracked ROIs; (b) The NCC computed after SCV illumination mapping step; (c) Trajectory of the tracked point; (d) Tracking error of the tracked point.

step the images of different wavelengths are under similar illumination conditions (see middle row). The tracking result of our modified DLK method is more robust than the original DLK method (see bottom row). Since the tissue motion is quite small in the sequence, to show our tracking effect more clearly we picked images of wavelength from  $550 \sim 570 \text{ nm}$  with observable motion from the sequence and computed the absolute difference between the template and tracked ROI without and with misalignment correction. Due to the darkness, the result image is enhanced and transformed to pseudo-color image as illustrated in Fig. 11. It shows that the misalignment decreases with our method, and the spectral data can be reconstructed after the motion compensation. We also tested on other multispectral images, and showed the difference image result in Fig. 12 and Fig. 13. The misalignment of vessels on 13 is corrected using the SCV metric.

During the computation of the probability distribution functions in Equation. 10, noise will be added to the SCV images due to the impact of histogram binning [6]. If the chosen number of histogram bins is too low, the resulting SCV images will lose a lot of high frequency details; on the other hand, if the full dynamic range of the image is chosen this results in noise in the SCV images. The impact may be alleviated by using adapted histogram bins. In our experiments,  $d_T = d_I = 256$ .



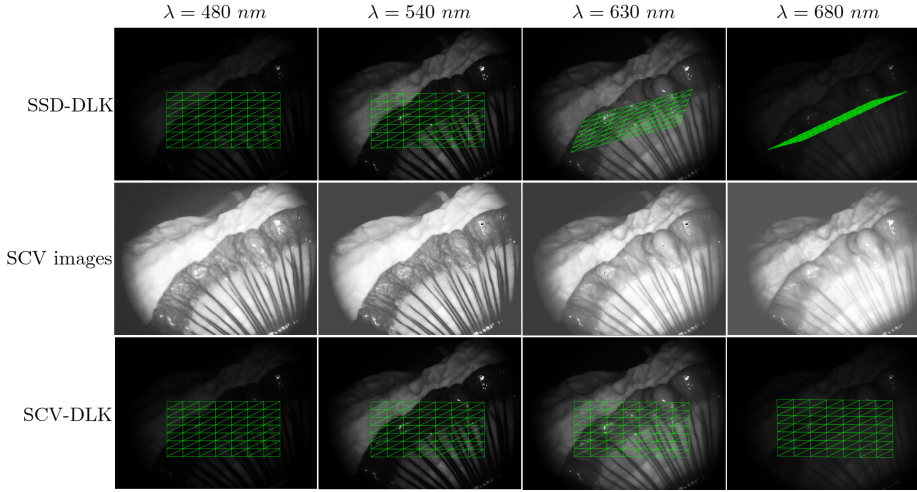


Fig. 10: The alignment of multispectral images (wavelength  $\lambda = 480 \sim 680 \text{ nm}$ ) without and with illumination compensation: (Top row) original DLK method using SSD metric; (Middle row) SCV images; (Bottom row) our modified DLK method using SCV metric.

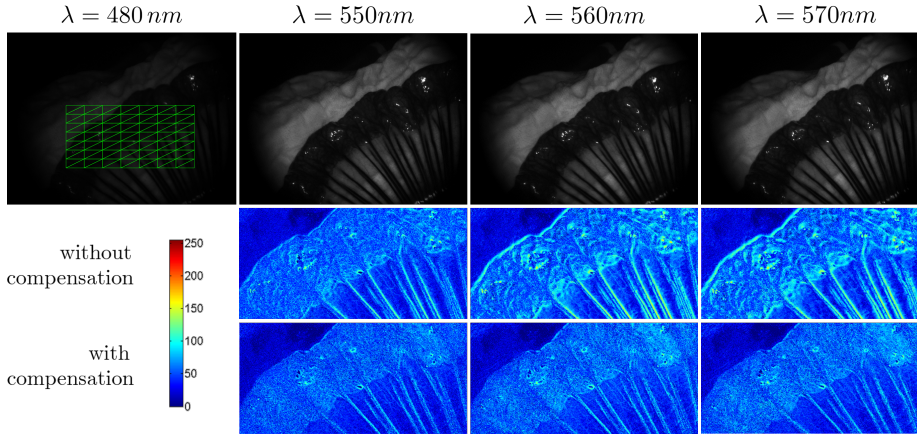


Fig. 11: The original multispectral images and the difference images without and with using SCV metric: (Top row) the template frame ( $\lambda = 480 \text{ nm}$ ) with the tracked ROI and several frames with observable motion; (Middle row) the difference ROI images without compensation; (Bottom row) the difference ROI images with compensation

#### 4 Conclusions

In this study, we presented a hybrid tracking method for estimating the deformation of soft-tissue surfaces by using a constrained geometric model combining sparse feature tracking with a modified DLK method [30]. Our algorithm uses the

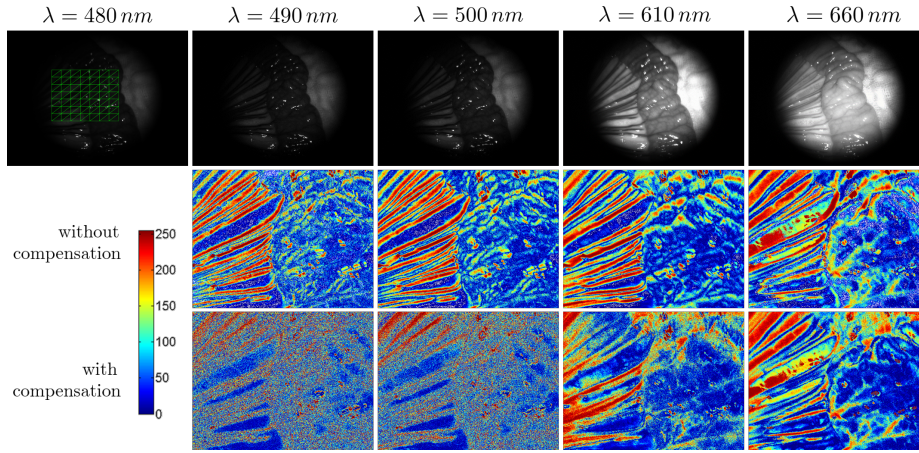


Fig. 12: The original frames and the difference images of another multispectral image sequence.

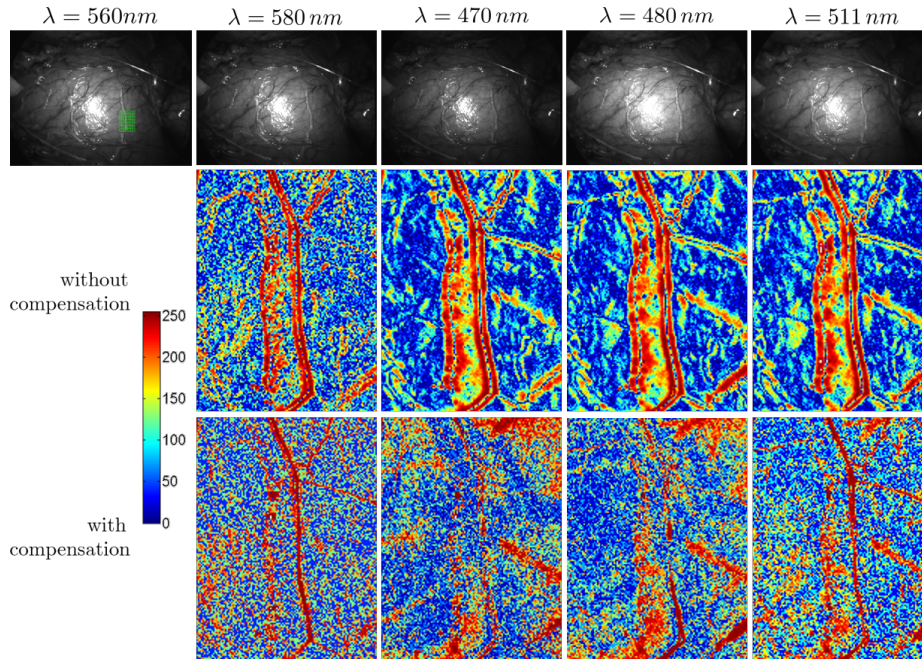


Fig. 13: Vessel misalignment correction.

SCV as the similarity metric to handle illumination variations, for example as seen in multispectral imaging, and facilitate tracking in very low light conditions where traditional approaches fail. The performance of our method on synthetic and *in vivo* datasets suggests that the hybrid approach improves the capability of correct convergence when tissue dynamics undergo large intra-frame displacements

or significant illumination change occurs. The feature tracking component of the proposed algorithm is very fast when using simple features and the mesh optimization using only the feature energy is computationally efficient allowing real-time application. We are currently investigating the GPU implementation of the improved intensity-based DLK component using SCV which we believe can also be developed to work at image-acquisition frame rates. Extending the presented algorithm to stereo images is also ongoing work, however, the regularization that we currently employ requires modification to appropriately handle deformations in 3D space.

**Acknowledgements** Xiaofei Du is supported by the China Scholarship Council scholarship. Neil Clancy is supported by an Imperial College Junior Research Fellowship. Shobhit Arya is supported by an NIHR-HTD 240 grant. The authors would like to thank the Northwick Park Institute for Medical Research (NPIMR) for their assistance with surgical arrangements.

**Conflict of interest** The authors declare no conflict of interest.

**Ethical standards** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed consent** Informed consent was obtained from all patients who were included in the study.

## References

1. Baker S, Matthews I (2004) Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision* 56(3):221–255
2. Bay H, Tuytelaars T, Van Gool L (2006) Surf: Speeded up robust features. In: *Computer Vision–ECCV 2006*, Springer, pp 404–417
3. Bradski G (2000) The opencv library. *Doctor Dobbs Journal* 25(11):120–126
4. Braux-Zin J, Dupont R, Bartoli A (2013) Combining features and intensity for wide-baseline non-rigid surface registration. In: *British Machine Vision Conference (BMVC)*, BMVA
5. Clancy NT, Stoyanov D, James DR, Di Marco A, Sauvage V, Clark J, Yang GZ, Elson DS (2012) Multispectral image alignment using a three channel endoscope in vivo during minimally invasive surgery. *Biomedical Optics Express* 3(10):2567–2578
6. Delabarre B, Marchand E (2012) Visual servoing using the sum of conditional variance. In: *Intelligent Robots and Systems (IROS)*, 2012 IEEE/RSJ International Conference on, IEEE, pp 1689–1694
7. Fua P, Brechbühler C (1996) Imposing hard constraints on soft snakes. In: *Computer Vision–ECCV 1996*, Springer, pp 495–506
8. Giannarou S, Visentini-Scarzanella M, Yang GZ (2009) Affine-invariant anisotropic detector for soft tissue tracking in minimally invasive surgery. In: *Biomedical Imaging: From Nano to Macro*, 2009. ISBI'09. IEEE International Symposium on, IEEE, pp 1059–1062

9. Giannarou S, Visentini-Scarzanella M, Yang GZ (2013) Probabilistic tracking of affine-invariant anisotropic regions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 35(1):130–143
10. Gröger M, Sepp W, Ortmaier T, Hirzinger G (2001) Reconstruction of image structure in presence of specular reflections. In: *Pattern Recognition*, Springer, pp 53–60
11. Kalal Z, Mikolajczyk K, Matas J (2010) Forward-backward error: Automatic detection of tracking failures. In: *Pattern Recognition (ICPR), 2010 20th International Conference on*, IEEE, pp 2756–2759
12. Keerthi SS, DeCoste D (2005) A modified finite newton method for fast solution of large scale linear svms. In: *Journal of Machine Learning Research*, pp 341–361
13. Maier-Hein L, Mountney P, Bartoli A, Elhawary H, Elson D, Groch A, Kolb A, Rodrigues M, Sorger J, Speidel S, Stoyanov D (2013) Optical techniques for 3d surface reconstruction in computer-assisted laparoscopic surgery. *Medical Image Analysis* 17(8):974–996
14. Mangasarian OL (2002) A finite newton method for classification. *Optimization Methods and Software* 17(5):913–929
15. Matthews I, Ishikawa T, Baker S (2004) The template update problem. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 26(6):810–815
16. Mountney P, Yang GZ (2008) Soft tissue tracking for minimally invasive surgery: Learning local deformation online. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*, Springer, pp 364–372
17. Ortmaier TJ (2003) Motion compensation in minimally invasive robotic surgery. PhD thesis, Universität München
18. Pickering MR, Muhit AA, Scarvell JM, Smith PN (2009) A new multi-modal similarity measure for fast gradient-based 2d-3d image registration. In: *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE, IEEE*, pp 5821–5824
19. Pilet J, Lepetit V, Fua P (2008) Fast non-rigid surface detection, registration and realistic augmentation. *International Journal of Computer Vision* 76(2):109–122
20. Puerto-Souza GA, Mariottini GL (2012) Hierarchical multi-affine (hma) algorithm for fast and accurate feature matching in minimally-invasive surgical images. In: *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, IEEE, pp 2007–2012
21. Richa R, Sznitman R, Taylor R, Hager G (2011) Visual tracking using the sum of conditional variance. In: *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, IEEE, pp 2953–2958
22. Selka F, Nicolau SA, Agnus V, Bessaid A, Marescaux J, Soler L (2013) Evaluation of endoscopic image enhancement for feature tracking: A new validation framework. In: *Augmented Reality Environments for Medical Imaging and Computer-Assisted Interventions*, Springer, pp 75–85
23. Stoyanov D (2012) Stereoscopic scene flow for robotic assisted minimally invasive surgery. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, Springer, pp 479–486
24. Stoyanov D (2012) Surgical vision. *Annals of biomedical engineering* 40(2):332–345

25. Stoyanov D, Yang GZ (2005) Removing specular reflection components for robotic assisted laparoscopic surgery. In: Image Processing, 2005. ICIP 2005. IEEE International Conference on, IEEE, vol 3, pp III–632
26. Stoyanov D, Darzi A, Yang GZ (2004) Dense 3d depth recovery for soft tissue deformation during robotically assisted laparoscopic surgery. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2004, Springer, pp 41–48
27. Stoyanov D, Darzi A, Yang GZ (2005) A practical approach towards accurate dense 3d depth recovery for robotic laparoscopic surgery. *Computer Aided Surgery* 10(4):199–208
28. Stoyanov D, Rayshubskiy A, Hillman E (2012) Robust registration of multi-spectral images of the cortical surface in neurosurgery. In: Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on, IEEE, pp 1643–1646
29. Yip MC, Lowe DG, Salcudean SE, Rohling RN, Ngan CY (2012) Real-time methods for long-term tissue feature tracking in endoscopic scenes. In: Information Processing in Computer-Assisted Interventions, Springer, pp 33–43
30. Zhu J, Lyu MR, Huang TS (2009) A fast 2d shape recovery approach by fusing features and appearance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 31(7):1210–1224