Motion Vector for Outlier Elimination in Feature Matching and Its Application in SLAM Based Laparoscopic Tracking

Cheng Wang[1](✉), Masahiro Oda[2], Yuichiro Hayashi[2], Kazunari Misawa[3], Holger Roth[2], and Kensaku Mori[2]

[1] Graduate School of Information Science, Nagoya University, Nagoya, Japan
{chwang,moda,yhayashi,rothhr}@mori.m.is.nagoya-u.ac.jp
[2] Graduate School of Informatics, Nagoya University, Nagoya, Japan
kensaku@is.nagoya-u.ac.jp
[3] Aichi Cancer Center Hospital, Nagoya, Japan
misawakzn@aichi-cc.jp

Abstract. This paper presents a motion vector-based method to detect and remove the outlier of the matched feature point in laparoscopic images. Feature point detected on organ surface in laparoscopic images plays an important role not only in laparoscopic tracking but also in organ surface shape reconstruction. However, many factors such as the deformation of the organ or the movement of the surgical tools result to the outliers in matched feature points, thus the feature point based tracking and reconstruction will have larger errors. Traditional methods use these points either directly (inside a RANSAC scheme) or after a prior knowledge of compensation, which may lead to larger error in tracking and reconstruction. We introduce the motion vector (MV) based method to detect outliers among the matched feature points. MV is originally used in the compression of the video streams, we exploit it to detect the movement of one feature point in different video frames. The outliers of feature point can be detected by enforcing a direction constraint with its MV. Our method had been implement under a SLAM-based framework for laparoscopic tracking, we modified the map management of SLAM for better laparoscopic tracking. The experimental results showed that our method effectively detects and removes the outliers without any prior knowledge; the average precision rate in image pairs was 95.9%.

Keywords: Laparoscopic tracking · Motion vector · SLAM

1 Introduction

In recent years, minimally invasive surgery (MIS) has became more popular due to its benefit to patients. However, MIS has drawbacks such as the limited view to surgeons. Therefore, endoscopic surgery navigation systems are used to make MIS processes safe and effective [1]. However, traditional endoscopic navigation needs additional equipment such as optical or magnetic trackers, that make

M.J. Cardoso et al. (Eds.): CARE/CLIP 2017, LNCS 10550, pp. 60–69, 2017.
DOI: 10.1007/978-3-319-67543-5_6

the endoscopic surgery navigation systems complex. To make the endoscope navigation system simple, the endoscope pose obtaining from the endoscopic videos instead of using additional equipment has been explored [2].

Visual simultaneous localization and mapping ($VSLAM$) is an approach for camera localization and 3D reconstruction. It has been introduced into laparoscopic navigation since 2006 [1,3,4]. However, even though breakthroughs have been made in recent years, many questions remain unsolved. For example, organ deformation may increase the error of laparoscopic tracking and in-vivo reconstruction.

Previous research [2] assumed rigid (or generally rigid) environments that all matched feature points are used for the estimation of laparoscope posture and 3D environments. This may result in the large difference comparing with real posture of laparoscope. Yang et al. established an online estimation of tissue deformation by exploiting a periodic motion model to estimate the organ's deformation [5]. They kept the estimation of the feature points detected on the organ surface using a filter-based SLAM. However, at least two assumptions were made: periodic organ motion, and as few other deformations as possible. A short observation for organ deformation with static camera is also needed before tracking. Clooins et al. exploited the prior information of organ shape for 3D tracking reconstruction without feature detecting in their work, and showed a good performance [6]. Mahmoud et al. extended the density of SLAM's map by enforcing cross-correlation on newly selected frames [7]. Their results showed that the reconstructed maps have higher density than the original SLAM with a higher RMSE value comparing with the segmented preoperative CT scans. Gustavo et al. exploited a HMA-based method to improve the performance of feature matching in MIS scenarios [8]. However, their work may be helpful in the reconstruction of the organ shape while it may behaves poor in real time tracking of the laparoscope.

Since previous methods use matched feature points for laparoscopic tracking, the laparoscopic tracking might fail due to the outliers existed in matched feature points. To remove these feature points, we use a motion vector-based method to judge the motion of the matched feature points. Motion vector (MV) was previously proposed [9] to detect object's movement within a SLAM solution. A voting procedure to determine the camera direction and a filtering procedure to reduce feature points were used. However, due to the difference between indoor/outdoor scenes and laparoscope scenes, the procedure we used is more strict.

Our main contribution is find a new method for the detection of the outlier existed in the matched feature points during laparoscopic tracking by the combination of the pure translational motion and the MV of the matched features. Different from the traditional method, our MV-based method detect the outliers by enforcing the direction constraint on the displacement of the feature point after the detection of eipolar constraint. The outliers in matched feature points can be detected and removed without any prior knowledge, this is quite different from the traditional method [5,9]. We explain this procedure in Sect. 2.

The performance of our method were shown in Sect. 3, the difference in tracked images and trajectories could be found.

2 Methodology

The goal of our method is to detect the outlier in matched feature points. To achieve this goal, we use three steps: (1) initial motion estimation of two images; (2) feature point selection using motion vector (MV); (3) motion refine using optimization. These three steps check the matched feature point not only with the traditional epipolar constraint, but also with the characteristic of pure transformation. Therefore, the small displacement of feature point caused by different factors can be detected and removed. We introduce the proposed feature point selection method into the ORB-SLAM to improve the tracking quality during laparoscopic tracking. The flowchart of the system implement with our method is shown in Fig. 1.

2.1 MV-Based Method for Motion Estimation

Initial Motion Estimation. Our method starts from the estimation of the motion between two images. Assume that two images at the time (m) and (n) are obtained, they are denoted as $I^{(m)}$ and $I^{(n)}$. The two images should have overlapping views so that enough feature points can be matched and used for motion estimation. Then the ORB feature is used to extract and match feature point in $I^{(m)}$ and $I^{(n)}$ [10]. After the feature matching procedure, a set of corresponding feature points $\boldsymbol{C}^{(m,n)} = \left\{\boldsymbol{x}_i^{(m)} \leftrightarrow \boldsymbol{x}_i^{(n)} \mid \boldsymbol{x}_i^{(m)} \in \boldsymbol{P}^{(m)}, \boldsymbol{x}_i^{(n)} \in \boldsymbol{P}^{(n)}\right\}$ are obtained, where $\boldsymbol{P}^{(m)}$ are the extracted feature points in $I^{(m)}$ and $\boldsymbol{P}^{(n)}$ are the extracted feature points in $I^{(n)}$, respectively.

With the use of the matched feature points $\boldsymbol{C}^{(m,n)}$, an initial fundamental matrix $\mathrm{F}^{(m,n)}$ can be computed using the eight-point algorithm with utilization

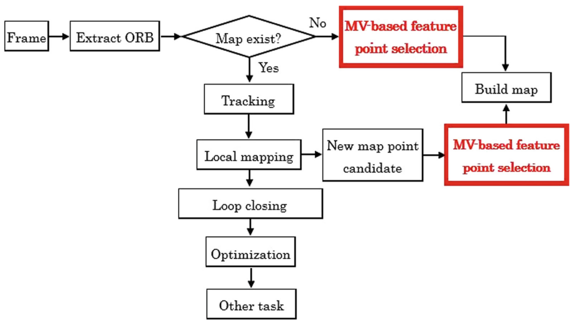


Fig. 1. ORB-SLAM flowchart implementing MV-based feature point selection procedure: tracking is used to track new frame; local mapping is used to create more map point; optimization is used for posture optimization. (Color figure online)

of a RANSAC algorithm [11] for outlier exclusion. The initial fundamental matrix $\mathrm{F}^{(m,n)}$ gives an inlier corresponding set in the matched feature points:

$$\boldsymbol{C}^{(m,n)'} = \left\{ \boldsymbol{x}_i^{(m)} \leftrightarrow \boldsymbol{x}_i^{(n)} \mid \boldsymbol{x}_i^{(m)} \leftrightarrow \boldsymbol{x}_i^{(n)} \subset \boldsymbol{C}^{(m,n)}, (\boldsymbol{x}_i^{(m)})^{\mathrm{T}} \mathrm{F}^{(m,n)} \boldsymbol{x}_i^{(n)} < \epsilon \right\} \quad (1)$$

where ϵ is an error threshold. The fundamental matrix $\mathrm{F}^{(m,n)}$ is also used to calculate the essential matrix $\mathrm{E}^{(m,n)}$ using $\mathrm{E}^{(m,n)} = \mathrm{K}^{\mathrm{T}}\mathrm{F}^{(m,n)}\mathrm{K}$, where K is the intrinsic parameter. The essential matrix $\mathrm{E}^{(m,n)}$ is decomposed to obtain the transformation between two images [11].

After this step, a subset of the corresponding feature points $\boldsymbol{C}^{(m,n)'}$ and the transformation $\mathrm{T}^{(m,n)} = \begin{bmatrix} s\mathrm{R}^{(m,n)} & \mathbf{t}^{(m,n)} \\ 0 & 1 \end{bmatrix}$ are obtained, where $\mathrm{R}^{(m,n)}$, $\mathbf{t}^{(m,n)}$ are the rotation and translation between two images, respectively. For adjacent frames in laparoscope video, the scaling s is set to 1.

MV-Based Method for Feature Point Selection. Unlike the other outdoor/indoor scenes, the in-vivo scenes are more complex and challenging due to the influence of factors such as the deformation of the organ, the movement of the forceps and so on. To improve the camera localization accuracy, only the inlier feature points are used.

To distinguish the matched feature points, we first rotate the feature points in $\mathbf{x}_i^{(m)}$ to a new position by using

$$\mathbf{x}_i^{(m)'} = s\mathrm{K}\mathrm{R}^{(m,n)}\mathrm{K}^{-1}\mathbf{x}_i^{(m)} \quad (2)$$

where $\mathbf{x}_i^{(m)'}$ is the new position of $\mathbf{x}_i^{(m)}$ after rotation $\mathrm{R}^{(m,n)}$ and scaling s. With this equation, the motion of the feature point can be changed to pure translation. Motion vectors (MVs) of feature points can be expressed as $\mathbf{x}_i^{(n)} - \boldsymbol{e}$ and $\mathbf{x}_i^{(n)} - \mathbf{x}_i^{(m)'}$, where $\boldsymbol{e}$ is the epipole on image $I^{(n)}$. For the pure translational motion, vectors $\mathbf{x}_i^{(n)} - \boldsymbol{e}$ and $\mathbf{x}_i^{(n)} - \mathbf{x}_i^{(m)'}$ are collinear [11]. However, in the abdominal cavity, due to the influence of factors mentioned above, these two vectors become non-collinear and the angle α is not 0° or 180°. Examples are shown in Figs. 2 and 3.

However, since the displacement caused by the additional motion is small, these feature point can't be detected by using constraint such as epipolar constraint or symmetric transfer error. To detect these outliers out, the angle of the motion vectors is used. The feature point is identified as outlier if

$$|cos(\alpha)| = \left| \frac{\left(\mathbf{x}_i^{(n)} - \boldsymbol{e}\right)}{\left|\mathbf{x}_i^{(n)} - \boldsymbol{e}\right|} \cdot \frac{\left(\mathbf{x}_i^{(n)} - \mathbf{x}_i^{(m)'}\right)}{\left|\mathbf{x}_i^{(n)} - \mathbf{x}_i^{(m)'}\right|} \right| < \lambda, \quad (3)$$

where λ is an threshold; otherwise, the feature point is identified as inlier. The inlier feature points are selected as the input for the motion refine procedure.

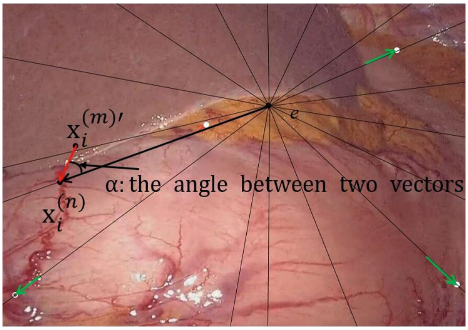


Fig. 2. An example of MV in laparoscope image. The red arrows show corresponding points $\mathbf{x}^{(m)'}$ and $\mathbf{x}_i^{(n)}$, while the black lines show $\mathbf{x}_i^{(n)}$ and the epipole $\boldsymbol{e}$. Feature points of inliers should show the correspondence as the green arrows while the outlier is the red arrows [11]. (Color figure online)

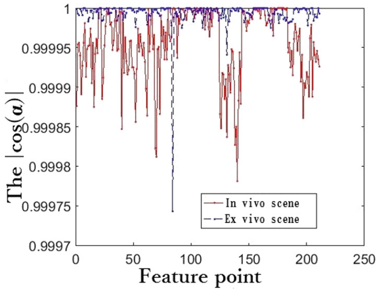


Fig. 3. The matched feature points and the cosine value of the angle α. We calculate the $|cos(\alpha)|$ in two scene: one is the ex vivo scene (blue line) and the other one is the in vivo abdominal scene exists both the non-rigid motion and rigid motion (red line). The average of $|cos(\alpha)|$ is closer to 1 in ex vivo scene while the abdominal scene is not. (Color figure online)

We can obtain a subset of $\boldsymbol{C}^{(m,n)'}$ marked

$$\boldsymbol{C}^{(m,n)''} = \left\{ \boldsymbol{x}_i^{(m)} \leftrightarrow \boldsymbol{x}_i^{(n)} \middle| \left| \frac{\left(\boldsymbol{x}_i^{(n)} - \boldsymbol{e}\right)}{\left|\boldsymbol{x}_i^{(n)} - \boldsymbol{e}\right|} \cdot \frac{\left(\boldsymbol{x}_i^{(n)} - \boldsymbol{x}_i^{\prime(m)}\right)}{\left|\boldsymbol{x}_i^{(n)} - \boldsymbol{x}_i^{\prime(m)}\right|} \right| \geq \lambda \right\}. \tag{4}$$

Motion Refine. The motion refine procedure is used to optimize the transformation of two images using the inlier feature points. The transformation is optimized by minimizing the reprojection error using

$$\mathrm{T}_r^{*(m,n)} = \underset{\mathrm{T}_r^{(m,n)}}{\arg\min} \sum_j \rho \left\| \boldsymbol{x}_j^{(n)} - proj\left(\mathrm{T}_r^{(m,n)}, \boldsymbol{X}_j^{(n)}\right) \right\|, \tag{5}$$

where $\mathbf{X}_j^{(n)}$ is the j-th recovered world coordinate of the feature point $\mathbf{x}_j^{(n)}$, $proj$ is the projection function projecting $\mathbf{X}_j^{(n)}$ onto $I^{(m)}$ [9,12], ρ is the Huber influence function [13].

Finally, we can obtain an optimized transformation $\mathrm{T}_r^{*(m,n)}$ between two images and a subset of matched feature points satisfying Eq. 4.

2.2 Application in ORB-SLAM Based Tracking

ORB-SLAM is an ORB feature-based SLAM framework and superior to other visual SLAM methods such as PTAM [12] and EKF-SLAM [3,13]. The feasibility of ORB-SLAM in endoscope surgery navigation has been proved [7].

We implemented our method on the ORB-SLAM. The flowchart of the modified ORB-SLAM is shown in Fig. 1, where the red part is the implementation of our method. We explain in details in the following parts.

Detection in Map Initialization. We modify the map initialization procedure of ORB-SLAM to initialize the map using fundamental matrix. The matched feature points should pass the test of motion model as well as our MV-based detection before they are used in map point creation.

Detection of New Map Points. New map points are created in the local mapping procedure of the ORB-SLAM. Feature points in the selected frames (called key frames) are matched and used to create new map points. New map points are created after the matched feature points pass the test of epipolar constraint and our MV-based approach.

Key Frame Management. Our MV-based method can detect the outlier of feature points especially observing the non-rigid motion. This can decrease the number of map points, and finally may resulting in the failure of laparoscopic tracking. To avoid this, we lower the threshold in the key frame selection to allow more key frames are used in building of the map. In actual implementation, we create the key frame every two frames and cull them if we have tracked enough points.

3 Experiments and Results

We validated our approach with in-vivo laparoscope videos. The videos recorded a task of exploring the abdominal cavity with a resolution of 640×480 pixels at 25 fps. The deformation in laparoscope videos were not too large [14]. We set the number of the ORB feature points in each frame to 2000, and the threshold λ was set to the mean value of $|cos(\alpha)|$ in rigid scene.

3.1 Detection Rate in Image Pairs

We saved the frames used in the map initialization during laparoscopic tracking. The ground truth is created by annotating the matched feature points manually. The matched feature points were marked true if it is found as outlier, and were marked false if it is found as inlier. The type of matched feature point is judged according to their neighbor and position. Table 1 shows the false positive (FP) rate, precision rate and recall rate of feature points of our method.

Table 1. Detection rate of proposed method in laparoscope images

Index	TP	FP	FP rate [%]	Precision rate [%]	Recall rate [%]	Miss rate [%]
1	252	17	47.1	93.7	100	0
2	129	4	13.7	96.9	100	0
3	99	6	31.6	94.3	100	0
4	88	17	5.5	98.9	100	0

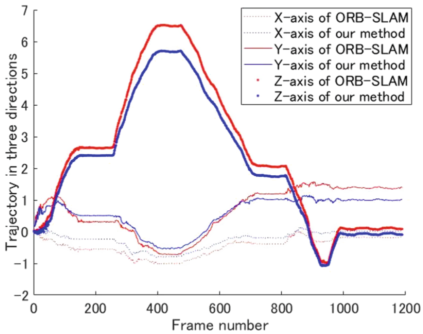


Fig. 4. Comparison of trajectory between our method and the original SLAM method. We can see the differences in two trajectories as frame changes.

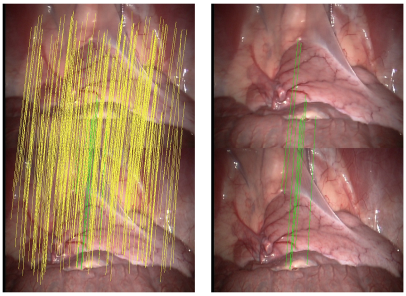
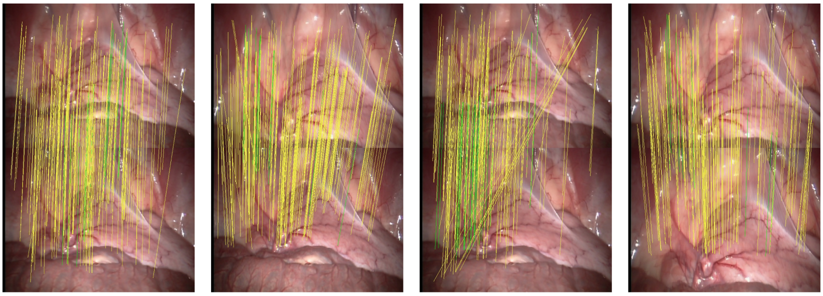


Fig. 5. Matched feature points in the key frames. Green and yellow lines show matched feature points of inlier and outlier, respectively. We have removed the yellow lines to show clearly in the right figure. (Color figure online)

3.2 Performance in Laparoscopic Video

We used the laparoscope videos mentioned above as the input of our system. The videos can be processed in real time. We obtained 5087 map points with the original ORB-SLAM and 3807 map points using SLAM implemented with our method. A comparison of trajectories between our method and the original SLAM was shown in Fig. 4. Due to no ground truth of laparoscope trajectory in real clinical scene, we only showed the difference of two methods in three directions in this figure: x-axis is the right direction, y-axis is the down direction, and z-axis is the front direction, respectively. We used the median depth of the first frame to make the trajectory under the same scale with that of the ORB-SLAM [13]. The matched feature points are shown in Figs. 5 and 6. Matched feature points are connected by yellow lines if they are judged as outliers while feature points are connected by green line if they are judged as inliers. The performance of our method in the local mapping procedure of SLAM is also validated. Figure 6 shows the matched feature points between the key frames. Some mismatched feature points together with feature points observing the non-rigid motion were detected by our method while they are poorly detected by the original ORB-SLAM.



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

Fig. 6. Matched feature points in key frames. Green and yellow lines show feature point in rigid and non-rigid motion area, respectively. (a)–(c) pairs show good results, (c) shows the mismatched feature point. (d) shows poor result caused by wrong motion estimation. (Color figure online)

4 Discussion

We confirm that the proposed method performs good in detecting and removing the outlier feature points during vision-based laparoscopic tracking. Our method exploit the characteristic of pure translation to find the outlier of the match, feature points with small displacement can be found using our method while they can't be detected by traditional method.

Figure 6(c) demonstrates that the mismatched feature points can be detected. This is because the MVs of the mismatched feature points also showing large angles. The mismatched feature points can also decrease the accuracy of the tracking.

Table 1 shows a comparison of our method and the map initialization procedure of ORB-SLAM. Our method can detect the outlier feature points even though after the test of ORB-SLAM initialization procedure. From this table, we can see our method outperform the map initialization procedure of ORB-SLAM.

However, since our method based on the estimated transformation between two images, the accuracy of our method depends on the estimated transformation. If the transformation is wrongly estimated, our method can't eliminate the outlier correctly. An undesirable result was shown in the fourth pair of Fig. 6. This result is caused by the incorrect transformation between two images.

In our experiment, the threshold λ is set according to the rigid scene. However, the value of λ can have an influence on the robust of the tracking. Too large or too small will cause the failure of the tracking, so the relationship between the λ and the tracking quality should be studied in the future.

5 Conclusion and Future Work

We proposed a motion vector-based method to detect the outlier feature points in laparoscopic video. The proposed method uses the transformation estimated between two images to distinguish feature points and achieved good performance both in the image pairs and in SLAM-based tracking. Future work contains the validation of the laparoscope posture estimated by our method, the comparison with other system such as RS-SLAM [15], and the discussion with surgeons that whether the accuracy is satisfactory for laparoscope navigation.

Acknowledgments. This work was supported by the MEXT, the JSPS KAKENHI Grant Numbers 25242047, 26108006, 26560255, 17H00869. We thanks to our lab colleagues for their help.

References

1. Baumhauer, M., Feuerstein, M., Meinzer, H.P., Rassweiler, J.: Navigation in endoscopic soft tissue surgery: perspectives and limitations. J. Endourol. **22**(4), 751–766 (2008)
2. Grasa, O.G., Bernal, E., Casado, S., Gil, I., Montiel, J.M.M.: Visual SLAM for handheld monocular endoscope. IEEE Trans. Med. Imaging **33**(1), 135–146 (2014)
3. Civera, J., Grasa, O.G., Davison, A.J., Montiel, J.M.M.: 1-Point RANSAC for extended kalman filtering: application to real-time structure from motion and visual odometry. J. Field Robot. **27**(5), 609–631 (2010)
4. Mountney, P., Stoyanov, D., Davison, A., Yang, G.-Z.: Simultaneous stereoscope localization and soft-tissue mapping for minimal invasive surgery. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 347–354. Springer, Heidelberg (2006). doi:10.1007/11866565_43
5. Mountney, P., Yang, G.-Z.: Motion compensated SLAM for image guided surgery. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6362, pp. 496–504. Springer, Heidelberg (2010). doi:10.1007/978-3-642-15745-5_61
6. Collins, T., Bartoli, A., Bourdel, N., Canis, M.: Robust, real-time, dense and deformable 3D organ tracking in laparoscopic videos. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9900, pp. 404–412. Springer, Cham (2016). doi:10.1007/978-3-319-46720-7_47
7. Mahmoud, N., Cirauqui, I., Hostettler, A., Doignon, C., Soler, L., Marescaux, J., Montiel, J.M.M.: ORBSLAM-based endoscope tracking and 3D reconstruction. arXiv preprint (2016). arXiv:1608.08149
8. Puerto-Souza, G.A., Mariottini, G.-L.: A fast and accurate feature-matching algorithm for minimally-invasive endoscopic images. IEEE Trans. Med. Imaging **32**(7), 1201–1214 (2013)
9. Li, W., Song, D.: Featureless motion vector-based simultaneous localization, planar surface extraction, and moving obstacle tracking. In: Akin, H.L., Amato, N.M., Isler, V., van der Stappen, A.F. (eds.) Algorithmic Foundations of Robotics XI. STAR, vol. 107, pp. 245–261. Springer, Cham (2015). doi:10.1007/978-3-319-16595-0_15
10. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 2564–2571 (2011)
11. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge (2003)

12. Klein, G., Murray, D.: Parallel tracking and mapping for small AR workspaces. In: 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR 2007, pp. 225–234 (2007)
13. Mur-Artal, R., Montiel, J.M.M., Tardos, J.D.: ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE Trans. Robot. **31**(5), 1147–1163 (2015)
14. Mountney, P., Stoyanov, D., Yang, G.Z.: Three-dimensional tissue deformation recovery and tracking. IEEE Signal Process. Magaz. **27**(4), 14–24 (2010)
15. Lee, G.H., Fraundorfer, F., Pollefeys, M.: RS-SLAM: RANSAC sampling for visual FastSLAM. In: International Conference on Intelligent Robots and Systems, pp. 1655–1660 (2011)