EMDQ-SLAM: Real-Time High-Resolution Reconstruction of Soft Tissue Surface from Stereo Laparoscopy Videos

Haoyin Zhou and Jagadeesan Jayender(✉)

Surgical Planning Laboratory, Brigham and Women's Hospital,
Harvard Medical School, Boston, USA
jayender@bwh.harvard.edu

Abstract. We propose a novel stereo laparoscopy video-based non-rigid SLAM method called EMDQ-SLAM, which can incrementally reconstruct thee-dimensional (3D) models of soft tissue surfaces in real-time and preserve high-resolution color textures. EMDQ-SLAM uses the expectation maximization and dual quaternion (EMDQ) algorithm combined with SURF features to track the camera motion and estimate tissue deformation between video frames. To overcome the problem of accumulative errors over time, we have integrated a g2o-based graph optimization method that combines the EMDQ mismatch removal and as-rigid-as-possible (ARAP) smoothing methods. Finally, the multi-band blending (MBB) algorithm has been used to obtain high resolution color textures with real-time performance. Experimental results demonstrate that our method outperforms two state-of-the-art non-rigid SLAM methods: MISSLAM and DefSLAM. Quantitative evaluation shows an average error in the range of 0.8–2.2 mm for different cases.

Keywords: Non-rigid SLAM · EMDQ · g2o-based graph optimization · High resolution texture · Multi-band blending · GPU parallel computation

1 Background

Three-dimensional (3D) reconstruction of tissue surfaces from intraoperative laparoscopy videos has found applications in surgical navigation [1,2] and planning [3]. One of the most important video-based 3D reconstruction methods is simultaneous localization and mapping (SLAM). Most existing SLAM methods assume that the environment is static, and the data at different time steps are aligned rigidly according to the estimated 6-DoF camera motion. This assumption is invalid for deformable soft tissues, which require much higher degrees of

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-87202-1_32) contains supplementary material, which is available to authorized users.

© Springer Nature Switzerland AG 2021
M. de Bruijne et al. (Eds.): MICCAI 2021, LNCS 12904, pp. 331–340, 2021.
https://doi.org/10.1007/978-3-030-87202-1_32

freedom to represent the non-rigid deformation and cannot be accurately recovered by the rigid SLAM methods.

Non-rigid SLAM is an emerging topic and the pioneering work is DynamicFusion [4], which first addressed the problem of incrementally building the 3D model of deformable objects in real-time. Following DynamicFusion, many non-rigid SLAM works have been proposed in the computer vision field [5,6]. However, these methods mainly focused on small regions of interest that require the object to be placed directly in front of the camera. This may not be appropriate for medical applications since the laparoscope may need to scan large areas of the tissue surface. This motivates the need for the development of large-scale non-rigid SLAM methods, which have attracted significant attention in recent years. For example, Mahmoud et al. considered the tissue deformation to be negligible and applied rigid SLAM for dense reconstruction [8,9]. Mountney et al. analyzed and compensated for the periodic motion of soft tissue caused by respiration using the extended Kalman filter [10]. These methods worked for specific situations with assumptions of the underlying tissue deformation, but cannot handle general tissue deformation. To date, only a few non-rigid SLAM works without any assumptions of the tissue deformation have been reported. For example, Collins et al. proposed to use features and tissue boundaries to track tissue deformation [11]. Song et al. proposed to combine ORB-SLAM [20] and a deformation model for tracking the motion of the laparoscope and estimating the deformation of soft tissues [15]. Recently, Lamarca et al. proposed a monocular non-rigid SLAM method called as DefSLAM for large-scale non-rigid environments, which combines the shape-from-template (SfT) and non-rigid structure-from-motion (NRSfM) methods and has obtained impressive results on laparoscopy videos [16].

In this paper, we propose a novel stereo video-based non-rigid SLAM method called as EMDQ-SLAM, which can track large camera motion and significant tissue deformation in real-time. The key algorithm of EMDQ-SLAM is the expectation maximization and dual quaternion (EMDQ) algorithm [17], which can generate dense deformation field from sparse and noisy SURF feature matches in real-time. Hence, EMDQ can efficiently track the camera motion and estimate the non-rigid tissue deformation simultaneously. However, EMDQ tracking suffers from a problem that the estimated tissue deformation may have accumulative errors. To solve this problem, we have developed a graph optimization method based on the g2o library [18], which uses the results of EMDQ as the initial values for further refinement. To preserve the high resolution color textures, we have adapted the multi-band blending (MBB) method [19] for real-time incremental applications, and have implemented GPU-based parallel computation for real-time performance.

2 Method

We first perform a GPU-based stereo matching method to estimate depths of image pixels, which was developed following Ref. [21]. Then, EMDQ-SLAM mosaics the stereo matching results at different time steps by estimating the camera motion and tissue deformation.

Without loss of generality, EMDQ-SLAM considers time $t = 0$ as the canonical frame, and estimates the camera motion and surface deformation at time steps $t = 1, 2, ...$ with respect to time 0 for non-rigid mosaicking. The 6-DoF camera motion at time t is represented using the rotation matrix $\mathbf{R}_t \in SO(3)$ and translational vector $\mathbf{t}_t \in \mathbb{R}^3$. We denote the world coordinate of a template point p at time 0 as $\mathbf{x}_{p,0}$. The tissue deformation at time t is represented by displacements of each point p, which is denoted as $\Delta\mathbf{x}_{p,t}$. And the world coordinate of point p at time t is

$$\mathbf{x}_{p,t} = \mathbf{x}_{p,0} + \Delta\mathbf{x}_{p,t}. \tag{1}$$

Hence, tissue deformation recovery is equivalent to the estimation of $\Delta\mathbf{x}_{p,t}$ for all template points p. Since the template may have millions of points, we have adapted the method from DynamicFusion [4] that uses sparse control points, or deformation nodes, to reduce the computational burden. For each template point p, its displacement $\Delta\mathbf{x}_{p,t}$ is represented by the weighted average of its neighboring deformation nodes i, which is

$$\Delta\mathbf{x}_{p,t} = \sum_{i}^{N} \left(w_i^p \Delta\mathbf{x}_{i,t} \right), \tag{2}$$

where $w_i^p = \exp(-\alpha \left\| \mathbf{x}_{i,0} - \mathbf{x}_{p,0} \right\|^2), \sum_i^N w_i^p = 1$ is the normalized weight between node i and point p, and node i is omitted if w_i^p is too small.

In summary, the parameters that need to be estimated include the 6-DoF camera motion $\mathbf{R}_t$ and $\mathbf{t}_t$, and the nodes displacements $\Delta\mathbf{x}_{i,t}$, $i = 1, 2, ..., N$. We use a two-step framework to solve this problem, which includes EMDQ tracking and g2o-based graph optimization.

2.1 EMDQ Tracking

At time t, the coordinate of node i in the camera frame is $\mathbf{x}_{i,t}^c = \mathbf{R}_t(\mathbf{x}_{i,0} + \Delta\mathbf{x}_{i,t}) + \mathbf{t}_t$. At time $t+1$, we first perform SURF [23] matching between video frame t and $t+1$, and obtain the related 3D coordinates of the SURF matches according to the pixel depths obtained by stereo matching. The number of SURF octave layers is set to one to avoid building the image pyramid and reduce the computational burden, which is reasonable because the change of image scale is small between adjacent image frames. Then, using the 3D SURF matches as the input, the EMDQ algorithm (1) obtains the displaced coordinates of nodes $\mathbf{x}_{i,t+1}^c$ from $\mathbf{x}_{i,t}^c$, and (2) removes SURF mismatches. The first output is directly used for updating the camera motion and nodes displacements at time $t+1$ in this EMDQ tracking method, and the second output will be used in the subsequent g2o-based graph optimization.

The estimated displacements of nodes between time t and $t+1$, $\mathbf{x}_{i,t+1}^c - \mathbf{x}_{i,t}^c$, include both the rigid and non-rigid components, which are caused by camera motion and tissue deformation respectively. To decompose the rigid and non-rigid components, we follow the method in Ref. [22] to estimate the rigid transformation, $\mathbf{R}_{t\to t+1}$ and $\mathbf{t}_{t\to t+1}$, between two 3D point clouds $\left\{\mathbf{x}_{i,t}^c\right\}$ and

$\{\mathbf{x}^c_{i,t+1}\}$, $i = 1, 2, ...N$. This method minimizes the sum of squared residuals and we consider the residuals as the non-rigid component, which is also the nodes displacements and we denote it as $\Delta\mathbf{x}_{i,t\to t+1} = \mathbf{x}^c_{i,t+1} - (\mathbf{R}_{t\to t+1}\mathbf{x}^c_{i,t} + \mathbf{t}_{t\to t+1})$. Finally, we update the camera motion in the world frame at time $t+1$ by

$$\mathbf{R}_{t+1} = \mathbf{R}_{t\to t+1}\mathbf{R}_t, \mathbf{t}_{t+1} = \mathbf{R}_{t\to t+1}\mathbf{t}_t + \mathbf{t}_{t\to t+1}. \tag{3}$$

The node displacements at time $t+1$ are updated by

$$\Delta\mathbf{x}_{i,t+1} = \mathbf{R}^T_t \Delta\mathbf{x}_{i,t\to t+1} + \Delta\mathbf{x}_{i,t}, i = 1, 2, ...N. \tag{4}$$

2.2 g2o-based Graph Optimization

The proposed EMDQ tracking method suffers from a problem that accumulative errors may exist in the estimated shape deformation. Specifically, Eq. (4) shows that the error of $\Delta\mathbf{x}_{i,t\to t+1}$ will result in the accumulative error of $\Delta\mathbf{x}_{i,t+1}$. Hence, in practice we found that EMDQ tracking works well for short video sequences, but may not be robust for long video sequences. The errors of $\Delta\mathbf{x}_{i,t+1}, i = 1, 2, ...N$ are mainly reflected as the differences among the neighboring deformation nodes. Hence, the deformation recovery errors can be reduced by using the as-rigid-as-possible (ARAP) [24] method, which aims to maintain the shape of deformable templates. We have developed a graph optimization method based on the g2o library [18] as the refinement step for EMDQ tracking, which integrates the EMDQ mismatch removal results and the ARAP costs. The vertices and edges of the graph are introduced in the following section.

Vertices: The graph has a total of $N+1$ vertices, which include the camera motion ($\mathbf{R}_{t+1}$ and $\mathbf{t}_{t+1}$), and displacements of N nodes ($\Delta\mathbf{x}_{i,t+1}, i = 1, 2, ...N$).

SURF Matching Edges: We denote the 3D camera coordinates of a SURF feature m at time t as $\mathbf{x}^c_{m,t}$, which can be directly obtained by the stereo matching results. Then, its world coordinate at time 0, $\mathbf{x}_{m,0}$, can be obtained according to the estimated camera motion and nodes displacements at time t. Ideally, the estimated $\mathbf{x}_{m,0}$ obtained from time t and $t+1$ should be the same, which are denoted as $\mathbf{x}^t_{m,0}$ and $\mathbf{x}^{t+1}_{m,0}$ respectively. We use the differences between $\mathbf{x}^t_{m,0}$ and $\mathbf{x}^{t+1}_{m,0}$ as the cost, that is

$$f_{\text{SURF}}(\mathbf{R}_{t+1}, \mathbf{t}_{t+1}, \Delta\mathbf{x}_{i,t+1}) = \sum_m w_m \left\|\mathbf{x}^t_{m,0} - \mathbf{x}^{t+1}_{m,0}\right\|^2, \tag{5}$$

where $w_m = 1/(e_m + 1)$ is the weight of match m, and e_m is the related error of match m in the EMDQ algorithm to distinguish inliers and outliers. We use this soft weight w_m to handle situations when EMDQ does not distinguish inliers and outliers correctly.

ARAP Edges: The basic idea of the standard ARAP method is to minimize non-rigid component after removing the rigid components [24]. Since in our method, the rigid and non-rigid components have already been decomposed, the ARAP term in EMDQ-SLAM is simplified to

$$f_{\mathrm{ARAP}}(\Delta \mathbf{x}_{i,t+1}) = \sum_{i1,i2} w_{i1}^{i2} \left\| \Delta \mathbf{x}_{i1,t+1} - \Delta \mathbf{x}_{i2,t+1} \right\|^2 , \tag{6}$$

where w_{i1}^{i2} is the weight between node $i1$ and $i2$. It is worth noting that this is a simplified ARAP method since it uses the same rigid component for all points.

2.3 GPU-Based Dense Mosaicking and MBB Texture Blending

We propose a planar TSDF [25] method to merge the template and stereo matching results at each time step. Our method reprojects each template point p to the image pixel I according to the estimated camera motion and nodes displacements, and merges the stereo matching depth of pixel I following the standard TSDF method. New template points are inserted from the stereo matching results if no existing point is reprojected to the related image pixel.

Since small misalignments are unavoidable, template point p and pixel I may have different RGB colors. Linear blending of the RGB colors may lead to blurry textures, as shown in Fig. 1(d). Hence, we employ the multi-band blending (MBB) method [19]. Traditional MBB method is offline and we refer to Ref. [19] for more details. In this paper, we mainly introduce our modifications for incremental and real-time blending. We use same notations as Ref. [19] for readers to follow easily. Our method reserves the previous information with the template points, and performs RGB color blending between the template and current image. The information at each template point p includes three spatial frequencies B_σ^p and the weight W^p. When blending with the current image, B_σ^p equals to the weighted average of that of p and I, i.e.,

$$B_\sigma^p = (1 - W_\sigma^I) B_\sigma^p + W_\sigma^I B_\sigma^I . \tag{7}$$

The weight of pixel I is obtained by Gaussian blurring $W_\sigma^I = W_{\max}^I * g_\sigma$. $W_{\max}^I = 1$ if $W^I > W^p$, otherwise $W_{\max}^I = 0$. W^p is updated by taking the max value of W^p and W^I. W^I is obtained in the same way as in Ref. [19].

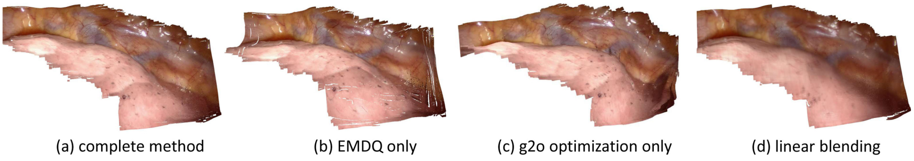

Fig. 1. Comparative study. For results in (c), g2o used estimations at the previous time step as the initial values, and Huber kernels were applied to handle outliers.

Comparative Study: As shown in Fig. 1, we conducted a comparative study to intuitively demonstrate the effects of the methods introduced in this paper. Both EMDQ tracking and g2o-based graph optimization contribute to the accuracy of EMDQ-SLAM. EMDQ tracking cannot handle long sequences robustly due to accumulative errors (Fig. 1(b)). g2o-based graph optimization suffers from the local minima problem and may be affected by SURF mismatches, which requires EMDQ results as the initial values (Fig. 1(c)). MBB blending can obtain better color textures than traditional linear blending (Fig. 1(d)).

3 Results

The source code was implemented in CUDA C++ on a desktop with an Intel Core i9 3.0 GHz CPU and NIVIDA Titan RTX GPU.

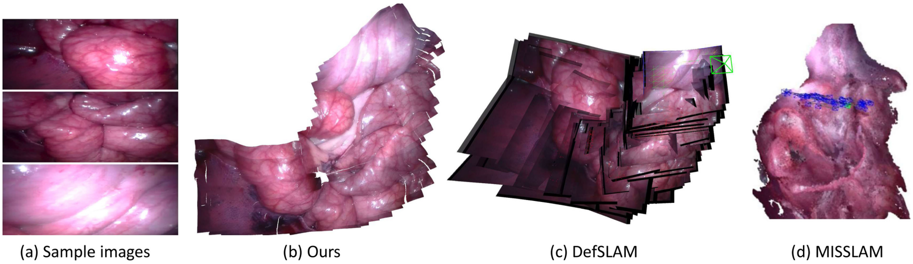

Fig. 2. Experiments on the *in vivo* porcine abdomen video from the Hamlyn dataset. (a) Ours. (b) DefSLAM result, which includes multiple overlapped local templates. (c) MISSLAM results, which is the screenshot reported in their paper.

We introduced MISSLAM [15] and DefSLAM [16] for comparison, which are recent video-based non-rigid SLAM algorithms. MISSLAM is a stereo video-based method, which is not open source software hence we used the same data reported in their paper for comparison. It may not be fair for comparing our method with DefSLAM since DefSLAM uses monocular video. Further, DefSLAM does not fully address the mosaicking problem and its mapping results are sets of overlapped local templates, rather than complete 3D point clouds or mesh models. The stereo laparoscopy videos used for validation were obtained from the Hamlyn online dataset[1]. We also captured intraoperative stereo laparoscopy videos on soft tissues at our hospital.

As shown in Fig. 2, we first conducted experiments on the Hamlyn dataset, which was obtained by scanning a porcine abdomen. The tissue had small deformation but the laparoscope motion was large. EMDQ-SLAM is able to generate a large 3D mosaic with clear textures corresponding to the area covered by the laparoscope. DefSLAM was able to track the camera motion from the monocular video, and provided multiple local templates. We also include a screenshot of the result from the MISSLAM paper for comparison. Figure 3(a) shows the results on another Hamlyn dataset, which scanned large areas of the liver surface. Due to low texture, the DefSLAM reported loss of tracking. Although it is difficult to provide quantitative comparisons since no ground truth is available, qualitative

[1] http://hamlyn.doc.ic.ac.uk/vision/.

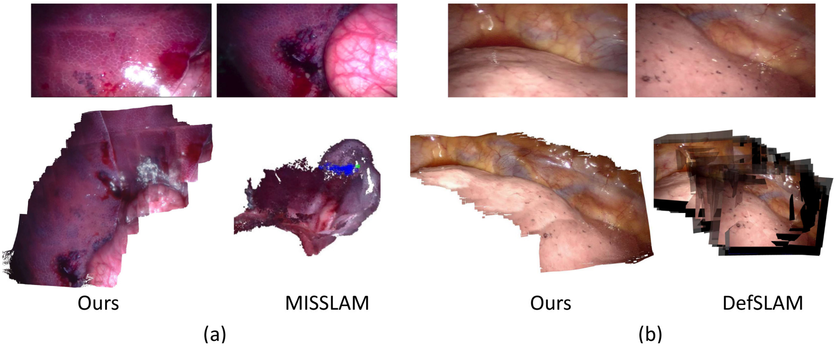

Fig. 3. (a) Experiments on the *in vivo* porcine liver video from the Hamlyn dataset. DefSLAM failed on this data. (b) Experiments on a stereo laparoscopy video captured during a minimally invasive sublobar lung surgery at our hospital.

comparisons show that our result is visually more accurate and can preserve high resolution texture.

For the experiments shown in Fig. 3(b), the video was captured during a minimally invasive sublobar lung surgery at our hospital. The surgeon was asked to move the stereo laparoscope within the patient's thoracic cavity. Due to heartbeat and respiratory motion caused by the adjacent lung, the deflated lung had significant and fast deformation. This experiment demonstrates that our method can handle highly deformable surfaces.

For the experiments shown in Fig. 4, the tissues had significant deformation due to heartbeat although the camera motion was small. Our method was able to track the deformation robustly and monitor the tissue deformation. Videos showing the results of the EMDQ-SLAM algorithm are provided as supplemental material for the paper.

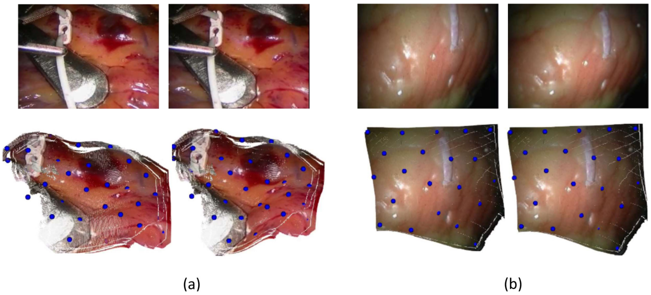

Fig. 4. Experiments on the Hamlyn dataset. (a) *In vivo* heart. (b) Heart phantom. The tissues had significant deformation and the camera motion was small.

Quantitative Experiments: The quantitative experiments were conduced using *ex vivo* porcine lungs and livers, which were placed on an Ambu bag and an anesthesia machine inflated/deflated the Ambu bag periodically to simulate the respiration, as shown in Fig. 5(a). Two electromagnetic (EM) trackers were attached to the laparoscope and on the tissue surface respectively. The laparoscope was gradually moved along the surface to create a 3D mosaic while estimating the tissue deformation, which was also measured using the EM sensor on the tissue surface. The EM coordinate frame was registered to the laparoscope coordinate frame, and our results were compared with the EM tracking results, as shown in Fig. 5(b). The errors for four cases are reported in Fig. 5(c), which were small when the EM tracker was in the field of view (FOV), but increased as the laparoscope moved far away, which is expected since the deformation of areas outside FoV cannot be monitored directly. The mean/standard deviation of errors when the EM tracker was in the FoV are 0.87/0.40, 2.1/0.58, 1.7/0.70 and 2.2/0.70 mm respectively for the four cases. The problem that areas outside FoV cannot be accurately estimated is an intrinsic drawback for large-scale non-rigid SLAM methods, since the deformation of invisible areas is obtained by extrapolation of the visible areas. However, this problem does not affect the mosaicking process because mosaicking is only performed at areas in FoV.

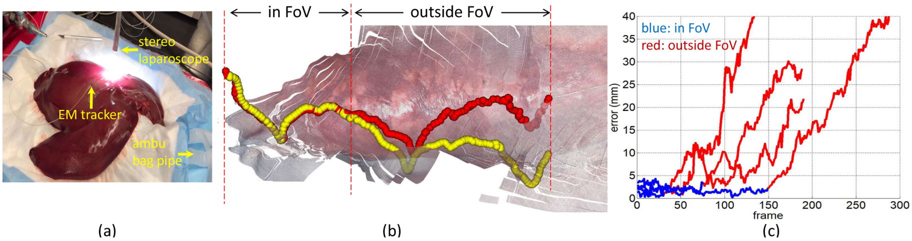

Fig. 5. Quantitative experiments. (a) Configuration. (b) Red dots are the estimated trajectory of the EM tracker by EMDQ-SLAM in the camera frame, yellows dots are the EM tracking results (ground truth). (c). Errors for four cases. Blue and red suggest the EM tracker was in and outside the field of view (FoV) respectively. (Color figure online)

Runtime: For the experiments shown in Fig. 2, Fig. 3(a)(b) and 4(a)(b), the average runtime to process one video frame was 92, 101, 97, 38 and 53 ms respectively, which included stereo matching, EMDQ-SLAM and VTK-based rendering. Hence, our method works at an update rate of around 10–26 Hz, which mostly depends on the image resolution. We use at most 1500 SURF features and 500 SURF matches at each time step to maintain the real-time performance.

4 Conclusion

The problem of large-scale non-rigid SLAM for medical applications is still an open problem. In this paper, we propose a novel non-rigid SLAM method called

EMDQ-SLAM, which uses a two-step framework to track the camera motion and estimate the tissue deformation. Although it is difficult to provide quantitative comparisons with other methods due to the lack of ground truth, qualitative comparisons shows our method can obtain visually more accurate mosaic with clear color textures. Quantitative experiments show that our method has an average error of 0.8–2.2 mm when estimating areas in the field of view.

Acknowledgments. This project was supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health through Grant Numbers K99EB027177, R01EB025964 and P41EB015898. Unrelated to this publication, Jagadeesan Jayender owns equity in Navigation Sciences, Inc. He is a co-inventor of a navigation device to assist surgeons in tumor excision that is licensed to Navigation Sciences. Dr. Jayender interests were reviewed and are managed by BWH and Partners HealthCare in accordance with their conflict of interest policies.

References

1. Maier-Hein, L., et al.: Optical techniques for 3D surface reconstruction in computer-assisted laparoscopic surgery. Med. Image Anal. **17**(8), 974–996 (2013)
2. Totz, J., Mountney, P., Stoyanov, D., Yang, G.-Z.: Dense surface reconstruction for enhanced navigation in MIS. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011. LNCS, vol. 6891, pp. 89–96. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23623-5_12
3. Lacher, R.M., et al.: Nonrigid reconstruction of 3D breast surfaces with a low-cost RGBD camera for surgical planning and aesthetic evaluation. Med. Image Anal. **53**, 11–25 (2019)
4. Newcombe, R.A., Fox, D., Seitz, S.M.: DynamicFusion: reconstruction and tracking of non-rigid scenes in real-time. In: CVPR, pp. 343–352 (2015)
5. Miroslava, S., Baust, M., Ilic, S.: Variational level set evolution for non-rigid 3D reconstruction from a single depth camera. IEEE TPAMI (2020)
6. Miroslava, S., Baust, M., Ilic, S.: SobolevFusion: 3D reconstruction of scenes undergoing free non-rigid motion. In: CVPR, pp. 2646–2655 (2018)
7. Cadena, C., et al.: Past, present, and future of simultaneous localization and mapping: toward the robust-perception age. IEEE Trans. Rob., 1309–1332 (2016)
8. Mahmoud, N., Hostettler, A., Collins, T., Soler, L., Doignon, C., Montiel, J.M.: SLAM based quasi dense reconstruction for minimally invasive surgery scenes. arXiv preprint arXiv:1705.09107 (2017)
9. Mahmoud, N., Collins, T., Hostettler, A., Soler, L., Doignon, C., Montiel, J.M.: Live tracking and dense reconstruction for handheld monocular endoscopy. IEEE Trans. Med. Imaging **13**, 38(1), 79–89 (2018)
10. 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). https://doi.org/10.1007/978-3-642-15745-5_61
11. 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). https://doi.org/10.1007/978-3-319-46720-7_47

12. Schoob, A., Kundrat, D., Kahrs, L.A., Ortmaier, T.: Stereo vision-based tracking of soft tissue motion with application to online ablation control in laser microsurgery. Med. Image Anal., 80–95 (2017)
13. Modrzejewski, R., Collins, T., Bartoli, A., Hostettler, A., Marescaux, J.: Soft-body registration of pre-operative 3D models to intra-operative RGBD partial body scans. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11073, pp. 39–46. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00937-3_5
14. Petit, A., Lippiello, V., Siciliano, B.: Real-time Tracking of 3D Elastic Objects with an RGB-D Sensor. In: IROS (2015)
15. Song, J., Wang, J., Zhao, L., Huang, S., Dissanayake, G.: MIS-SLAM: real-time large-scale dense deformable SLAM system in minimal invasive surgery based on heterogeneous computing. IEEE Rob. Autom. Lett. **3**(4), 4068–4075 (2018)
16. Lamarca, J., Parashar, S., Bartoli, A., Montiel, J.M.: DefSLAM: tracking and mapping of deforming scenes from monocular sequences. IEEE Trans. Rob. (2020)
17. Zhou, H., Jayender, J.: Smooth deformation field-based mismatch removal in real-time. arXiv preprint arXiv:2007.08553 (2020)
18. Kmmerle, R., Grisetti, G., Strasdat, H., Konolige, K., Burgard, W.: G2o: a general framework for graph optimization. In: ICRA, pp. 3607–3613 (2011)
19. Brown, M., Lowe, D.G.: Automatic panoramic image stitching using invariant features. Int. J. Comput. Vision **74**(1), 59–73 (2007)
20. Mur-Artal, R., Tard, J.D.: ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras. IEEE Trans. Rob., 1255–1262 (2017)
21. Zhou, H., Jayender, J.: Real-time dense reconstruction of tissue surface from stereo optical video. IEEE Trans. Med. Imaging **39**(2), 400–412 (2019)
22. Arun, K.S., Huang, T.S., Blostein, S.D.: Least-squares fitting of two 3-D point sets. IEEE TPAM **I**, 698–700 (1987)
23. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006). https://doi.org/10.1007/11744023_32
24. Sorkine, O., Alexa, M.: As-Rigid-As-Possible Surface Modeling. In: Symposium on Geometry Processing, vol. 4, pp. 109–116 (2007)
25. Osher, S., Fedkiw, R.: Level Set Methods and Dynamic Implicit Surfaces. AMS, vol. 153. Springer, New York (2003). https://doi.org/10.1007/b98879