



Real-Time Surface Deformation Recovery from Stereo Videos

Haoyin Zhou and Jayender Jagadeesan^(✉)

Surgical Planning Laboratory, Brigham and Women's Hospital,
Harvard Medical School, Boston, MA 02115, USA
jayender@bwh.harvard.edu

Abstract. Tissue deformation during the surgery may significantly decrease the accuracy of surgical navigation systems. In this paper, we propose an approach to estimate the deformation of tissue surface from stereo videos in real-time, which is capable of handling occlusion, smooth surface and fast deformation. We first use a stereo matching method to extract depth information from stereo video frames and generate the tissue template, and then estimate the deformation of the obtained template by minimizing ICP, ORB feature matching and as-rigid-as-possible (ARAP) costs. The main novelties are twofold: (1) Due to non-rigid deformation, feature matching outliers are difficult to be removed by traditional RANSAC methods; therefore we propose a novel 1-point RANSAC and reweighting method to preselect matching inliers, which handles smooth surfaces and fast deformations. (2) We propose a novel ARAP cost function based on dense connections between the control points to achieve better smoothing performance with limited number of iterations. Algorithms are designed and implemented for GPU parallel computing. Experiments on *ex-* and *in vivo* data showed that this approach works at an update rate of 15 Hz with an accuracy of less than 2.5 mm on a NVIDIA Titan X GPU.

Keywords: Tissue deformation recovery · Feature matching outliers · GPU parallel computation

1 Background

Tissue visualization during surgery is typically limited to the anatomical surface exposed to the surgeon through an optical imaging modality, such as laparoscope, endoscope or microscope. To identify the critical structures lying below the tissue surface, surgical navigation systems need to register the intraoperative data to preoperative MR/CT imaging before surgical resection. However, during surgery, tissue deformation caused by heartbeat, respiration and instruments interaction may make the initial registration results less accurate. The ability to compensate for tissue deformation is essential for improving the accuracy of surgical navigation. In this paper, we propose an approach to recover the deformation of tissue surface from stereo optical videos in real-time.

In recent years, several groups have investigated methods to recover tissue deformation from optical videos, and most methods are based on the minimization of non-rigid matching and smoothing costs [1]. For example, Collins et al. proposed a monocular vision-based method that first generated the tissue template and then estimated the template deformation by matching the texture and boundaries with a non-rigid iterative closet points (ICP) method [2]. In this method, the non-rigid ICP-based boundary matching algorithm significantly improves the accuracy. However, during surgery, only a small area of the target tissue may be exposed and the boundaries are often invisible, which makes it difficult to match the template. Object deformation recovery in the computer vision field is also a suitable approach to recover tissue deformation. For example, Zollhofer et al. proposed to generate the template from an RGB-D camera and then track the deformation by minimizing non-rigid ICP, color and smoothing costs [3]. Newcombe et al. have developed a novel deformation recovery method that does not require the initial template and uses sparse control points to represent the deformation [4]. Guo et al. used forward and backward L_0 regularization to refine the deformation recovery results [5]. To date, most deformation recovery methods [6,7] are based on the non-rigid ICP alignment to obtain matching information between the template and the current input, such as monocular/stereo videos or 3D point clouds from RGB-D sensors. However, non-rigid ICP suffers from a drawback that it cannot track fast tissue deformation and camera motion, and obtain accurate alignment in the tangential directions on smooth tissue surfaces. During surgery, the endo/laparoscope may move fast or even temporally out of the patient for cleaning, which makes non-rigid ICP difficult to track the tissue. In addition, smoke and blood during the surgery may cause significant occlusion and interfere with the tracking process. Hence, the ability to match the template and the input video when non-rigid deformation exists is essential for intraoperative use of deformation recovery methods.

A natural idea to obtain additional information is to match the feature points between the template and the input video. Among many types of feature descriptors, ORB [8] has been widely used in real-time applications due to its efficiency. To handle feature matching outliers, RANSAC-based methods have proven effective in rigid scenarios but are difficult to handle non-rigid deformation [9]. Another common method to address outliers is to apply robust kernels to the cost function, which cannot handle fast motion. In this paper, we propose a novel method that combines 1-point-RANSAC and reweighting methods to handle matching outliers in non-rigid scenarios. In addition, we propose a novel as-rigid-as-possible (ARAP) [10] method based on dense connections to achieve better smoothing performance with limited number of iterations.

2 Method

As shown in Fig. 1, we proposed a GPU-based stereo matching method, which includes several efficient post-processing steps to extract 3D information from stereo videos in real-time. Readers may refer to Ref. [11] for more details on this stereo matching method.

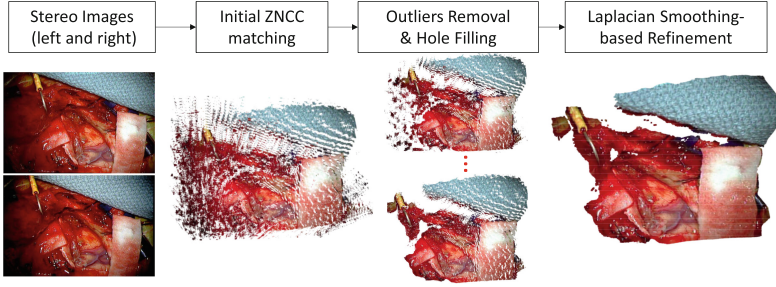


Fig. 1. The process of our stereo matching method with a pair of stereo microscopy images captured during neurosurgery.

In our system, the initial template of the tissue surface is generated by the stereo matching method, then we track the deformation of the template by representing the non-rigid deformation with sparse control points on the template, and estimating the parameters of the control points to make the deformed template match the output of the GPU-based stereo matching method. The algorithms are parallelized and run on the GPU. Similar to DynamicFusion [4], we employ dual-quaternion to represent deformation and each control point i is assigned a dual-quaternion W_i^t to represent its warp function at time t , and the template points are deformed according to the interpolation of neighboring control points. Then, the deformation recovery problem is to estimate $W_i^t, i = 1, \dots, N$, and we use the Levenberg-Marquardt algorithm to minimize the following cost function

$$f_{\text{Total}}(W_i^t) = f_{\text{ICP}} + w_{\text{ORB}}f_{\text{ORB}} + w_{\text{ARAP}}f_{\text{ARAP}}, \quad (1)$$

where f_{ICP} and f_{ORB} are based on non-rigid ICP and ORB matches between the template and the current stereo matching results respectively. The as-rigid-as-possible (ARAP) cost f_{ARAP} smoothes the estimated warp functions W_i^t , which is especially important for the estimation of occluded areas. w_{ORB} and w_{ARAP} are user defined weights. In our experiments, we use $w_{\text{ORB}} = 10.0$ and w_{ARAP} is dynamically adjusted due to the varying number of valid points in f_{ICP} and ORB matching inliers in f_{ORB} . We sum up the related weights of ICP and ORB terms for each W_i^t , and scale up or down w_{ARAP} accordingly.

A GPU-based parallel Levenberg-Marquardt (LM) algorithm was developed to minimize the cost (1). We update each W_i^t independently in the LM iterations. For the computation of the Jacobian matrix \mathbf{J} related to each W_i^t , multiple parallel GPU threads are launched to compute rows of \mathbf{J} , then we perform Cholesky decomposition to update $W_i^t, i = 1, \dots, N$.

The non-rigid ICP term f_{ICP} is determined by the distances between the deformed template and the stereo matching results. The Tukey's penalty function is employed to handle outliers. We have developed a rasterization process that re-projects the template points to the imaging plane to build correspondences between template points and the stereo matching results, which is parallelized to each template point and runs on the GPU. This rasterization step is faster than

kd-tree-based closest points search in the 3D space. Only the distance component in the normal directions are considered, which avoids the problem that non-rigid ICP is inaccurate in the tangential directions when aligning smooth surfaces.

2.1 ORB Feature Matching and Inliers Pre-selection

As shown in Fig. 2(a)–(b), standard ORB feature detection concentrates on rich texture areas, which may lead to the lack of matching information at low texture areas. Hence, we first develop a method to detect uniform ORB features to improve the accuracy of deformation recovery, which uses GPU to detect FAST corners and suppresses those if a neighboring pixel has larger corner response in parallel. Then, the ORB features of the initial template are matched to the live video frames. Two corresponding 3D point clouds are obtained, which may include incorrect matches.

Since at least three matches are needed to determine the rigid relative pose between two 3D point clouds, traditional RANSAC methods only work when the three matches are all inliers and have similar deformation [9]. Another common method to handle outliers is to apply robust kernels to the cost function, which is effective but cannot handle fast camera motion or tissue deformation. Under a reasonable assumption that local deformations at small areas of the tissue surface are approximate to rigid transforms, we propose a novel 1-point-RANSAC and reweighting method to pre-select potential matching inliers following the idea of Ref. [12], as shown in Fig. 2(c). Denoting the two sets of corresponding 3D ORB features as o_k^1 and o_k^2 , $k = 1, \dots, N$, a random match k_0 is selected as the reference, and rectify the coordinates with respect to k_0 by

$$\mathbf{S}_{k_0}^l = [o_1^l - o_{k_0}^l, \dots, o_N^l - o_{k_0}^l]_{3 \times N}, l = 1, 2. \quad (2)$$

For a reference k_0 , we denote the local rigid transform as $o_{k_0}^2 = \mathbf{R}o_{k_0}^1 + \mathbf{T}$, where $\mathbf{R} \in SO(3)$ is the rotation matrix and \mathbf{T} is the translation vector. Rigid transform for a neighboring match inlier k should satisfy

$$\mathbf{S}_{k_0}^2(k) \approx \mathbf{R}\mathbf{S}_{k_0}^1(k), \quad (3)$$

where $\mathbf{S}_{k_0}(k)$ is the k th column of \mathbf{S}_{k_0} , and \mathbf{R} can be obtained from matches that satisfy (3). We propose a reweighting method to eliminate the impacts of other matches, that is

$$d_k = \|\mathbf{S}_{k_0}^2(k) - \mathbf{R}\mathbf{S}_{k_0}^1(k)\|, w_k = \min(H/d_k, 1), \quad (4)$$

where d_k is the distance related to the k th match. w_k is the weight of the k th ORB match and if the k th match is either an outlier, or an inlier that does not satisfy (3), w_k is small. H is a predefined threshold. With a selected reference k_0 , we alternatively update \mathbf{R} from weighted $\mathbf{S}_{k_0}^1$ and $\mathbf{S}_{k_0}^2$, and update w_k according to (4). In experiments we perform 10 iterations with each k_0 . A small sum of w_k suggests that few matches satisfy (3) and k_0 may be an outlier, and we omit the

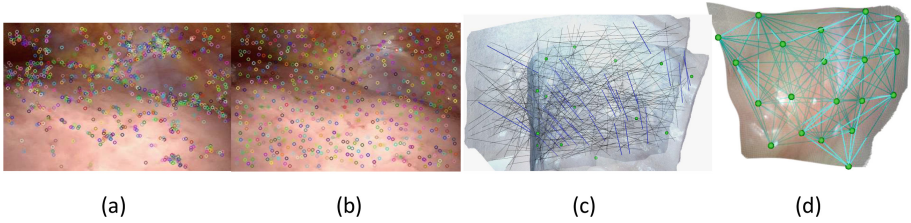


Fig. 2. (a)–(b) ORB feature detection results on laparoscopy images captured during a lung surgery using (a) OpenCV (b) Our method. (c) Matching inliers pre-selection results with a deforming phantom. The blue lines are selected inliers and black lines are identified as outliers. (d) Dense connections between control points with a silicon heart phantom. (Color figure online)

results with reference k_0 . In our experiments, we randomly select 30 different matches as the reference k_0 .

We first apply this 1-point-RANSAC + reweighting method to assign weights to ORB matches, the results of which will be used in the subsequent LM algorithm to minimize term f_{ORB} in (1). It should be clarified that we are not implying that this 1-point-RANSAC + reweighting method is able to find all inliers. To take into account all inliers, in the LM algorithm we assign the pre-selected matches the same weight as w_k , and assign other ORB matches weight according to $w_k = -1/(5H)d_k + 1$, $w_k \in [0, 1]$.

2.2 As-Rigid-As-Possible Smoothing

Traditional ARAP methods are based on sparse connections, such as triangular meshes. This type of connection is too sparse to propagate the smoothing impact fast enough, and in practice we found that it cannot perform well with the limited number of iterations in the LM algorithm. Hence, we propose to use dense connections as shown in Fig. 2(d). The weights of connections in traditional ARAP methods are sensitive and need to be specifically designed based on the angles of the triangular mesh [10], hence the ARAP cost function has to be redesigned to handle the dense connections as follows:

$$f_{\text{ARAP}} = \sum_{i_1, i_2} w_{i_1, i_2} (f_{\text{length}, i_1 i_2} + w_{\text{angle}} f_{\text{angle}, i_1 i_2} + w_{\text{rotation}} f_{\text{rotation}, i_1 i_2}) \quad (5)$$

where i_1 and i_2 are two control points. w_{i_1, i_2} is the weight of connection between i_1 and i_2 , and a smaller distance between points i_1 and i_2 at time 0 suggests larger w_{i_1, i_2} . We use $w_{\text{angle}} = 20.0$ and $w_{\text{rotation}} = 100.0$.

For control points i_1 and i_2 ,

$$\begin{aligned} f_{\text{length}, i_1 i_2} &= (\|p_{i_2}^t - p_{i_1}^t\| - \|p_{i_2}^0 - p_{i_1}^0\|)^2 \\ f_{\text{angle}, i_1 i_2} &= \text{acos}(W_{i_1}^t(p_{i_2}^0) - p_{i_1}^t, p_{i_2}^t - p_{i_1}^t) \\ f_{\text{rotation}, i_1 i_2} &= \|W_{i_1}^t(1, 2, 3, 4) - W_{i_2}^t(1, 2, 3, 4)\|^2 \end{aligned} \quad (6)$$

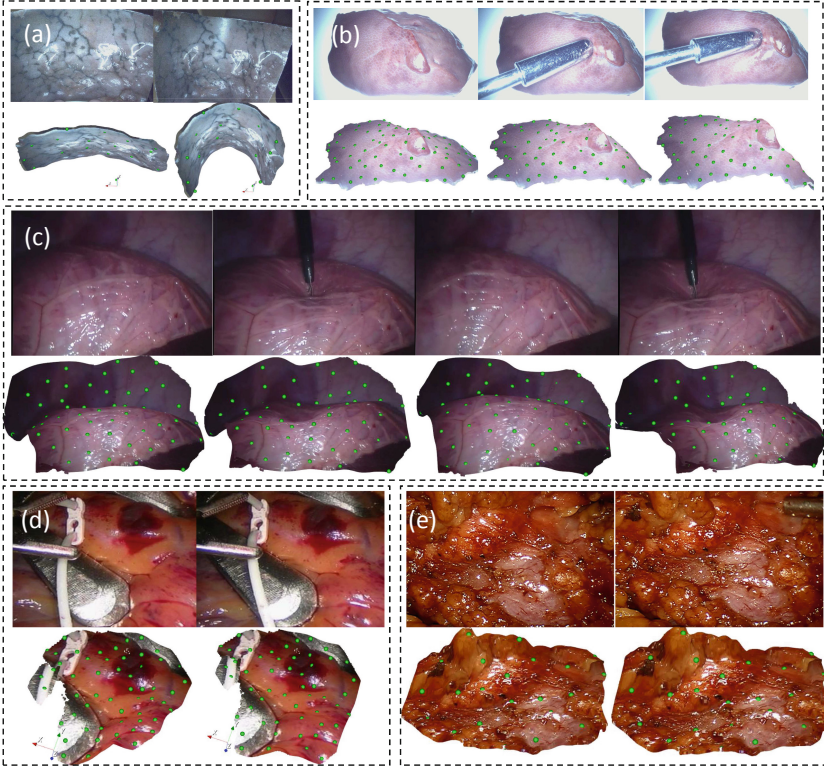


Fig. 3. Qualitative experiments. First row: input video frames. Second row: the deformed template and the control points (green dots). (a) Phantom. (b) *Ex vivo* porcine liver. (c) Hamlyn *in vivo* data with deformation caused by instrument interaction (d) Hamlyn *in vivo* data with respiration and heartbeat. (e) *In vivo* kidney data with deformation caused by respiration. (Color figure online)

where p_i^t is the coordinate of point i at time t . $f_{\text{angle},i1i2}$ equals to the angle between the normalized vectors $W_{i1}^t(p_{i2}^0) - p_{i1}^t$ and $p_{i2}^t - p_{i1}^t$, where $W_{i1}^t(p_{i2}^0)$ suggest to apply W_{i1}^t to p_{i2}^0 . $f_{\text{rotation},i1i2}$ is introduced because W_i^t has 6-DoFs, which is determined by the differences between the first four components of dual-quaternion W_{i1}^t and W_{i2}^t .

3 Experiments

Algorithms were implemented with CUDA C++ running on a desktop with Intel Xeon 3.0 GHz CPU and NVIDIA Titan X GPU. We first conducted qualitative experiments on *ex-* and *in vivo* data. As shown in Fig. 3(a), we deformed a smooth phantom with lung surface texture and captured 960×540 stereo videos with a KARL STORZ stereo laparoscope. We removed intermediate video frames between the two frames in Fig. 3(a) to simulate fast deformation, and our method

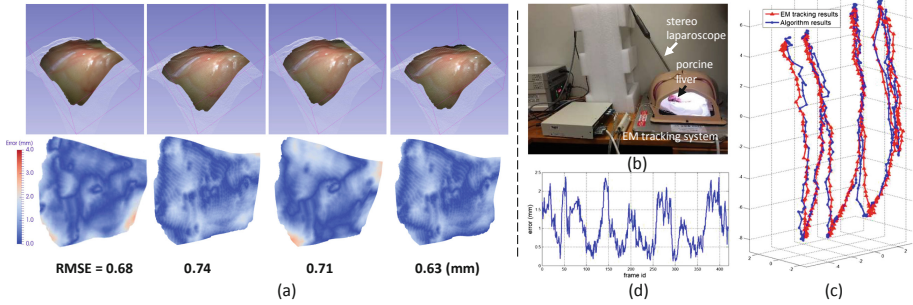


Fig. 4. Quantitative experiments. (a) Hamlyn heart Phantom data. First row: colored models are the deformed templates, white points are the ground truth. Second row: distance maps. Average runtime: stereo matching 3.8 ms, ORB feature detection and matching 10.6 ms, inliers pre-selection 4.1 ms, LM 14.2 ms. (b)–(d) Experiment with the EM tracking system. (b) Hardware. (c) 3D trajectories. (d) Errors. Average runtime: stereo matching 17.6 ms, ORB feature detection and matching 11.6 ms, inliers pre-selection 3.1 ms, LM 30.7 ms. (Color figure online)

is capable of tracking the large deformation. The second experiment was conducted with *ex vivo* porcine liver as shown in Fig. 3(b). The deformation was caused by instrument interaction, and our method is able to handle instrument occlusion. For the *in vivo* experiments shown in Fig. 3(c)–(e), we used both the Hamlyn data [13] and our data, in which the videos have camera motion and tissue deformation. We generated the tissue template before instrument interaction and then track the deformation of the template. The algorithm detected key inlier ORB features on the reconstructed surface and tracked these template features robustly in spite of respiratory and pulsatile motions, and instrument occlusions. These results highlight the robustness of tracking in spite of physiological motions and varying illumination.

We conducted two quantitative experiments. The first experiment was conducted on Hamlyn data as shown in Fig. 4(a). The Hamlyn data consists of stereo video images of a silicon phantom simulating heartbeat motion and corresponding ground truth was obtained using CT scan. The template was generated from the first video frame. Results show an RMSE of less than 1 mm and the average runtime of 32.7 ms per frame. In the second experiment, we used the EM tracking system (medSAFE Ascension Technologies Inc.) as the ground truth, as shown in Fig. 4(b)–(d). The porcine liver was placed in an abdominal phantom (The Chamberlain Group) and a medSAFE EM sensor was attached to the liver surface. We deformed the liver manually and recorded the EM sensor measurements and compared it with that of our method. Deformation estimation results on 420 video frames (Fig. 4(c)–(d)) show a mean error of 1.06 mm and standard deviation of 0.56 mm. As shown in Fig. 4(c), the maximum distance between the trajectory points is 15.7 mm. The average runtime was 63.0 ms per frame.

4 Conclusion

We propose a novel deformation recovery method that integrates the ORB feature, which is able to handle fast motion, smooth surfaces and occlusion. The limitation of this work is that it strongly relies on ORB feature matching, which may fail when the deformation is extremely large and different light reflection may make it difficult to obtain enough number of ORB matching inliers.

Acknowledgments. This project was supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health through Grant Numbers P41EB015898 and R01EB025964. Unrelated to this publication, Jayender Jagadeesan 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. Jagadeesans interests were reviewed and are managed by BWH and Partners Health-Care in accordance with their conflict of interest policies.

References

1. 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.* **40**, 80–95 (2017)
2. 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
3. Zollhfer, M., Matthias, N., Shahram, I., et al.: Real-time non-rigid reconstruction using an RGB-D camera. *ACM Trans. Graph.* **33**(4), 156 (2014)
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. Guo, K., Xu, F., Wang, Y., Liu, Y., Dai, Q.: Robust non-rigid motion tracking and surface reconstruction using L0 regularization. In: *ICCV*, pp. 3083–3091 (2015)
6. 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
7. Petit, A., Lippiello, V., Siciliano, B.: Real-time tracking of 3D elastic objects with an RGB-D sensor. In: *IROS* (2015)
8. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF, pp. 2564–2571 (2011)
9. Tran, Q.-H., Chin, T.-J., Carneiro, G., Brown, M.S., Suter, D.: In defence of RANSAC for outlier rejection in deformable registration. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) *ECCV 2012*. LNCS, vol. 7575, pp. 274–287. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33765-9_20
10. Sorkine, O., Alexa, M.: As-rigid-as-possible surface modeling. In: *Symposium on Geometry Processing*, vol. 4, pp. 109–116 (2007)

11. Zhou, H., Jagadeesan, J.: Real-time dense reconstruction of tissue surface from stereo optical video. *IEEE Trans. Med. Imaging* (2019, early access)
12. Zhou, H., Zhang, T., Jayender, J.: Re-weighting and 1-Point RANSAC-Base PnP solution to handle outliers. *IEEE TPAMI* (2018, early access)
13. Mountney, P., Stoyanov, D., Yang, G.Z.: Three-dimensional tissue deformation recovery and tracking. *IEEE Signal Process. Mag.* **27**(4), 14–24 (2010)