3D Shape Recovery of Deformable Soft-tissue with Computed Tomography and Depth Scan

Jingwei Song[1*], Jun Wang[1,2], Liang Zhao[1], Shoudong Huang[1], Gamini Dissanayake[1]
1. Centre for Autonomous Systems, University of Technology, Sydney, Australia
Jingwei.song@student.uts.edu.au {Liang.Zhao;Shoudong.Huang;Gamini.Dissanayake}@uts.edu.au
2. Institute of Remote Sensing and Digital Earth Chinese Academy of Sciences,
University of Chinese Academy of Sciences, Beijing, China. wangjun@radi.ac.cn

Abstract

Knowing the tissue environment accurately is very important in minimal invasive surgery (MIS). While, as the soft-tissues is deformable, reconstruction of the soft-tissues environment is challenging. This paper proposes a new framework for recovering the deformation of the soft-tissues by using a single depth sensor. This framework makes use of the morphology information of the soft-tissues from X-ray computed tomography, and deforms it by the embedded deformation method. Here, the key is to build a distance field function of the scan from the depth sensor, which can be used to perform accurate model-to-scan deformation together with robust non-rigid shape registration in the same go. Simulations show that soft-tissue shape in the previous step can be efficiently deformed to fit the partially observed scan in the current step by using the proposed method. And the results from the simulated sequential deformation of three different soft-tissues demonstrate the potential clinical value for MIS.

1 Introduction

Minimally Invasive Surgery (MIS), which is an indispensable tool in modern surgery, greatly benefits the patients with reduced incisions, trauma and less hospitalization time [Hu et al., 2007]. One of the most challenging task of computer assisted MIS is to build intra-operative morphology and motion of soft-tissues with stereo or ideally a monocular camera.

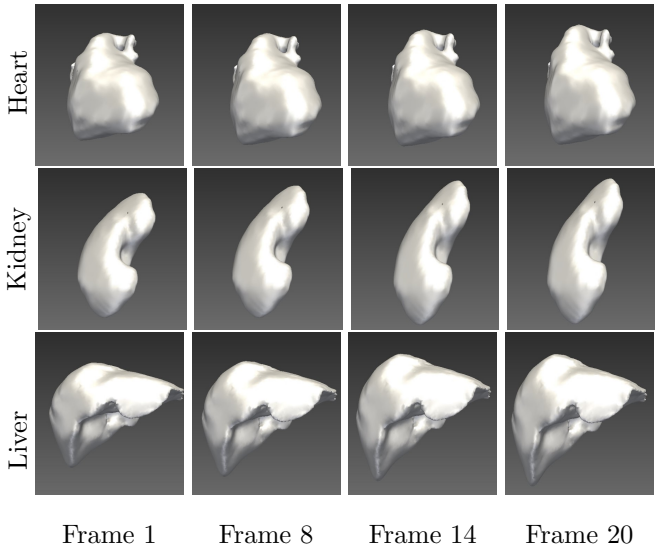


Figure 1: We propose a new framework for deforming soft-tissue organ shape to fit current observed depth scan. The recovery of the deformation of different soft-tissues at different time steps can be achieved using the proposed pipeline.

A lot of work have been devoted to 3D soft-tissues reconstruction. [Lin et al., 2015] proposed a structure from motion pipeline for partial 3D surgical scene reconstruction and localization, and in [Stoyanov, 2012] stereo images were used for extracting sparse 3D point cloud of organs. Contrary to feature based extraction, Du et al. [Du et al., 2015] employed an optical flow based approach namely Deformable LucasKanade for tracking tissue surface. All the methods described above contribute greatly to enabling implementing augmented reality or virtual reality in computer assisted interventions which will greatly promote the accuracy and efficiency of MIS. Yet all the work above focus on tracking key feature points for localization and no work has been devoted to geometry based registration and dynamic soft-tissue surface reconstruction.

Recently, more work have been reported on incremental 3D model reconstruction of deformable objects or moving human bodies. After the pioneering work of Kinect Fusion [Newcombe et al., 2011] which makes use of RGBD, efforts have been devoted on how to make full use of real-time RGBD information for judging the current shape and pose of the model. Although Kinect Fusion is basically applicable for stable and rigid object, [Zollhöfer et al., 2014] first attempted to transfer Kinect Fusion's idea in non-rigid body construction and simulation. Later on, DynamicFusion [Newcombe et al., 2015] and VolumeDeform [Innmann et al., 2016] have been proposed for more accurate 3D object reconstruction and simulation. Their template free work have achieved great success in both reference model construction and model deformation prediction. A compelling Fusion4D method is demonstrated in [Dou et al., 2016], where the topology changes are considered comparing to DynamicFusion. While, different from previous work, multi-view RGBD cameras are used instead of a single RGBD camera. These techniques like DynamicFusion may be applied to a new way of sports broadcasting or immersive telepresence in other geographic locations in the future. Despite the amazing result, their work cannot be directly applied in surgical vision due to limitations of sensor used in surgery and the high accuracy requirement in surgery. Thus, none of these methods are used in the application of computer-assisted interventions in MIS.

Inspired by the work mentioned above, similar information as point clouds can be provided from surgical vision by using shape from shading of monocular or depth from disparity of stereo camera [Maier-Hein et al., 2013]. Therefore, we propose a new framework for performing dynamic soft-tissue recovery using a single depth sensor. If the real-time pose of the organ as well as deformed shape can be recovered, more information can be applied to benefit the MIS process.

However, due to sensors and application requirements, there are two major differences between the DynamicFusion pipeline and the proposed method based on surgical vision. First, computed tomography (CT) scanning is a standard process before surgery. Thus, the pre-operative CT data provides an ideal detailed prior model for recovering the deformation. While in DynamicFusion, the current incomplete model have to be incrementally updated with noisy depth scan. As pointed out in [Dou et al., 2015], the error in propagation step accumulates, making the deformation parameter sets more unreliable. That's the reason why strategies like bundle adjustment [Dou et al., 2015] or truncated signed distance function (TSDF) surface approximation are applied to overcome it. While in MIS, as the prior reference model can be obtained from the pre-operative CT scan, this reference model can be directly used in the reconstruction of the deformed soft-tissue. Another difference is that electromagnetic (EM) tracking techniques can be applied for providing the pose of the camera in the global coordinate frame, which means not only a better initialization but also an ideal information of the camera pose can be gained as input for estimating non-rigid parameters as well as relative rigid translation between the camera and the soft-tissue. Based on the pre-operative CT data and EM tracking, we focus on reference model to scan deformation recovery to provide an efficient, robust and accurate framework for computer-assisted morphology recovering in MIS.

With respect to the differences between our scenario (MIS) and similar work in the computer vision community, in this paper we proposed an innovative framework to recover the deformed 3D structure of the soft-tissues. The deformation parameters possess the ability to deform the last update of shape to fit scan flexibly and efficiently. Thus we build a distance field function (DFF) volume as a part of the objective function and optimize the deformation parameters.

This paper is organized as follows: Section 2 provides the technical details of the deformation recovery algorithm proposed in this paper. Section 3 shows the simulation results to validate the proposed algorithm with some discussions of the pros and cons. Finally, Section 4 concludes the paper with future work.

2 Methodology

2.1 Framework overview

Our framework for recovering the deformation of the soft-tissue consists of three steps (Fig 2):

- I. Compute the DFF for the new scan.
- II. Predict visible points from the last update of the deformed model.
- III. Deform the current deformed model to fit the new scan. Here, both the deformation of the model and the non-rigid registration between the model and the new scan are accomplished simultaneously. The model is initialized by using the reference model segmented from the pre-operative CT scan at the very beginning.

Step I is an on-line pre-process. A new DFF volume is built, which records each voxel the distance to its nearest point on new scan. This DFF will not only be employed in the model-to-scan deformation and registration processes (Step III), but also in the selection of the visible points in Step II. As DFF is only built on current scan, dynamically building DFF doesn't require much computational cost. The details of generating the DFF will

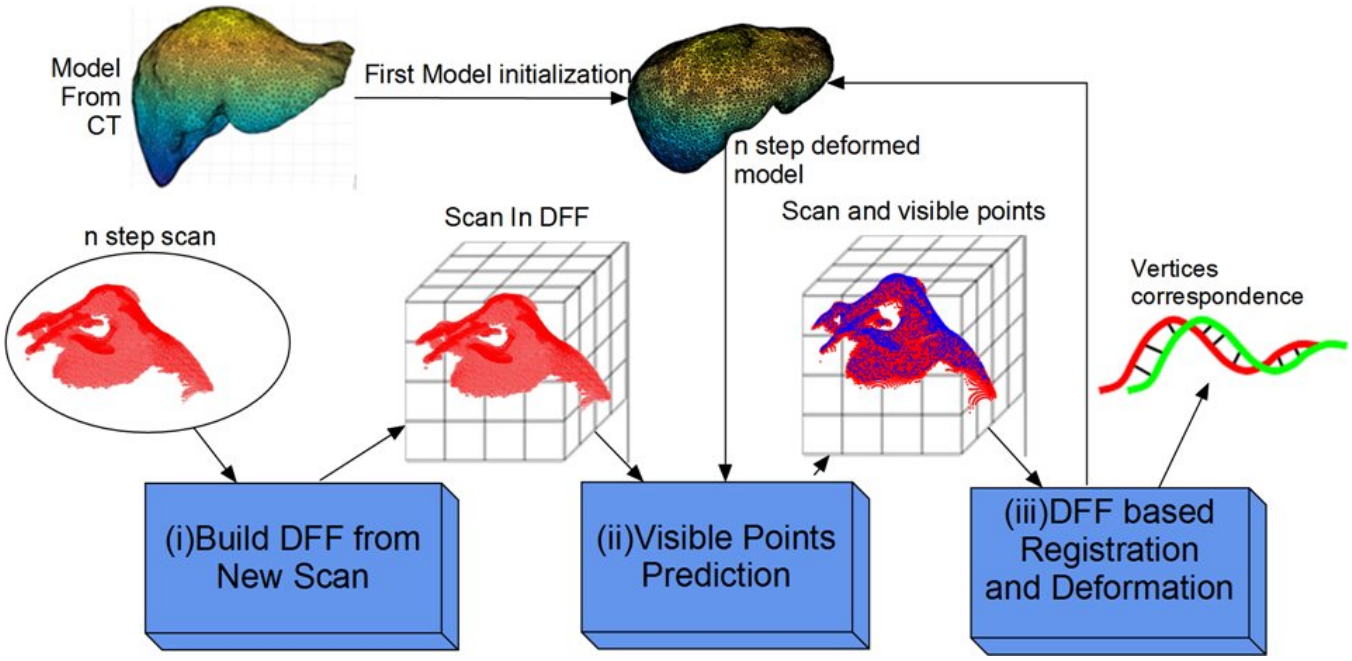


Figure 2: The framework of our deformed soft-tissue reconstruction based on DFF and pre-operative CT model.

be addressed in Section 2.4. Different from most existing approaches which traverse all the point-to-plane distances and use a threshold to decide visible points [Dou et al., 2016], in this paper we use regularized DFF volume directly by looking up the value of each vertex in DFF and the derivative functions generated from DFF and comparing with the threshold for deciding point visibility. This strategy reduces the computational cost of the visible points selection process significantly. After selecting the visible points, a cost function can be built by adjusting the deforming parameters to fit the visible points on the deformed model close to the target scan. As described in [Newcombe et al., 2015], a spatial warping field deformation [Newcombe et al., 2015] as well as source to target correspondence is built. Based on the parameters, we deformed the last updated model to the current scan which not only fits current observation but also obey As-Rigid-As-Possible principle proposed by [Sorkine and Alexa, 2007] in surface deformation. As-Rigid-As-Possible principle enables that non-visible part of the model can be inferred from current observation.

2.2 Model deformation

3D mesh models of the soft-tissue can be segmented from the pre-operative CT scans [Kenngott et al., 2015]. Here, similar to [Dou et al., 2016][Zollhöfer et al., 2014], we choose to use embedded deformation (ED) proposed by Sumner [Sumner et al., 2007] as a free form surface deformation approach. The fundamental idea of ED is to uniformly choose a set of ED nodes as the deformation elements. Each ED node is companied by an affine transformation matrix in $\mathbb{R}^{3\times3}$ and a translation vector in $\mathbb{R}^3$. Each vertex in the 3D space on the model can only be affected by several neighboring ED nodes. In this way, deformation process can be factorized by local accumulative center based affine transformation. Even though ED parameters implies global translation, we still introduce a global rigid rotation and translation as variables for the purpose of speeding up the optimization process.

The jth ED node is recorded by position $\boldsymbol{g_j}$ (3×1), a corresponding quasi rotation matrix $\boldsymbol{R_j}$ (3×3) and a translation vector $\boldsymbol{t_j}$ (3×1). For any given point $\boldsymbol{p}$, it can be mapped to a new locally deformed point $\tilde{\boldsymbol{p}}_{\boldsymbol{local}}$ and then transformed to the global coordinate frame $\tilde{\boldsymbol{p}}$.

$$\tilde{\boldsymbol{p}}_{\boldsymbol{local}} = \boldsymbol{R_j}(\boldsymbol{p} - \boldsymbol{g_j}) + \boldsymbol{g_j} + \boldsymbol{t_j} \tag{1}$$

$$\tilde{\boldsymbol{p}} = \boldsymbol{R_G}\tilde{\boldsymbol{p}}_{\boldsymbol{local}} + \boldsymbol{T_G} \tag{2}$$

where global transformation is conducted by $\boldsymbol{R_G}$ (in Euler angle form) and translation $\boldsymbol{T_G}$. This non-rigid transformation can be extended to any vertex or points mapped by k neighboring nodes:

$$\tilde{\boldsymbol{v_i}} = \boldsymbol{R_G}\sum_{j=1}^{k} w_j(\boldsymbol{v_i})[\boldsymbol{R_j}(\boldsymbol{v_i} - \boldsymbol{g_j}) + \boldsymbol{g_j} + \boldsymbol{t_j}] + \boldsymbol{T_G} \tag{3}$$

where w_j is the quantified weights for each transformation exerted by each ED node. To avoid the influence of far away ED nodes, we limit the weights of each points

by defining the w_j according to the distance:

$$w_j(\boldsymbol{v_i}) = (1 - ||\boldsymbol{v_i} - \boldsymbol{g_j}||/d_{max}). \quad (4)$$

Here d_{max} is the maximum distance of the vertex to $k+1$ nearest ED nodes. In [Sumner et al., 2007] k is set as 4.

2.3 Registration energy function

The main objective energy function of the nonlinear optimization problem consists of three terms: Rotation, Regularization and the distances between the model and the target scan. Variables in the state vector for this energy function are the $[\boldsymbol{R_j}, \boldsymbol{t_j}]$ for each ED node. The energy function is presented in the following format:

$$\min_{\boldsymbol{R_1}, \boldsymbol{t_1} \cdots \boldsymbol{R_m}, \boldsymbol{t_m}} w_{rot} E_{rot} + w_{reg} E_{reg} + w_{\mathcal{D}} E_{\mathcal{D}} \quad (5)$$

where m is the number of ED nodes.

In order to prevent the optimized parameters leading deformation to an unreasonable way, here we follow the method proposed in [Sumner et al., 2007] of constraining model with Rotation and Regularization.

Rotation:

$$\begin{aligned} Rot(\boldsymbol{R}) = (\boldsymbol{c_1} \cdot \boldsymbol{c_2})^2 + (\boldsymbol{c_1} \cdot \boldsymbol{c_3})^2 + (\boldsymbol{c_2} \cdot \boldsymbol{c_3})^2 + \\ (\boldsymbol{c_1} \cdot \boldsymbol{c_1} - 1)^2 + (\boldsymbol{c_2} \cdot \boldsymbol{c_2} - 1)^2 + (\boldsymbol{c_3} \cdot \boldsymbol{c_3} - 1)^2 \end{aligned} \quad (6)$$

where $\boldsymbol{c_1}$, $\boldsymbol{c_2}$ and $\boldsymbol{c_3}$ are the column vectors of the matrix $\boldsymbol{R}$. The term E_{rot} sums the rotation error of all the matrix.

$$E_{rot} = \sum_{j=1}^{m} Rot(\boldsymbol{R_j}). \quad (7)$$

Regularization. For each ED node, the deformation exerts on itself and from other ED nodes should be almost the same. Otherwise, the surface will not be smooth. Therefore, we introduce the term E_{reg} to sum the transformation errors from each ED node. Note that a huge weight of regularization makes the non-rigid transformation deteriorate to the rigid transformation.

$$E_{reg} = \sum_{j=1}^{m} \sum_{k \in \mathbb{N}(j)} \alpha_{jk} ||\boldsymbol{R_j}(\boldsymbol{g_k} - \boldsymbol{g_j}) + \boldsymbol{g_j} + \boldsymbol{t_j} - (\boldsymbol{g_k} + \boldsymbol{t_k})||^2 \quad (8)$$

where α_{jk} is the weights calculated by the Euclidean distance of the two ED nodes. In [Sumner et al., 2007], α is uniformly set to 1. $\mathbb{N}(j)$ are neighboring nodes to the node j. And for each vertex, there are k neighbors.

Distances to the target scan. After deciding the rotation matrix and transformation vector of ED nodes, all the vertex in the mesh can be transformed to their new positions and the distances between these vertices on the deformed model to the target scan needs to be minimized. These distances can be easily looked up from a predefined loss function DFF. The lower the value is, the closer the deformed vertex to the target surface (not necessarily to the correct correspondences but at least close to the surface). Details of DFF definition will be described in Section 2.4. The positions of vertex are calculated by Eq 3 and compared in Eq 9. Minimizing this term means deforming keeping the transformed model close to the target surface of the scan.

$$E_{\mathcal{D}} = \sum_{i \in \mathbb{L}} ||\mathcal{D}(\tilde{\boldsymbol{v}}_{\boldsymbol{i}})||^2 \quad (9)$$

where $\tilde{\boldsymbol{v}}_{\boldsymbol{i}}$ is the deformed position of $\boldsymbol{v_i}$. $\mathcal{D}(\cdot)$ is the corresponding voxel value recorded in DFF. $\mathbb{L}$ defines the set for all the visible points for calculating object function.

2.4 Distance Field Function

The key ingredient in the non-rigid deformation is the registration process, which decides how to deform the model so that it can best fit the target scan. Current work employ back-projection as the registration method ([Dou et al., 2016], [Newcombe et al., 2015], [Innmann et al., 2016]), in each iteration back-projection keeps projecting and lifting current points back and forth. In this paper, we modified the Directional Distance Function proposed by Dou in [Dou et al., 2013] as a DFF by ignoring the directions. Unified volume based distance function provides a robust and efficient target loss function for surface matching. Results (Table 1) show that unified 3D object function makes model fit better to scan than back-projection method. Back-projection process keeps projection current point to image and then back to 3D space which is computationally inefficient and is not easy to implement especially in its complex form of Jacobian.

Similar to the Directional Distance Function proposed in [Dou et al., 2013], we also record at each voxel its distance $\mathcal{D}(\cdot)$ to the closest point on the surface of the target scan.

The difference is, in [Dou et al., 2013], they record the pointing vector of each voxel to closest surface point and use it for calculating Jacobians for faster speed. While, we find out that this method makes the spatial distribution of Jacobian not uniformly scattered which will increase points misalignment. Therefore, we applied the traditional numerical Jacobian calculation with 3D version of Robert's operator [Davis, 1975]. This rigorous Jacobian calculation strategy makes optimization more accurate.

2.5 Optimization

Both registration and deformation parameter estimation processes are carried out simultaneously by minimizing

the energy functios. Here, we use Levenberg-Marquardt (LM) algorithm to solve the nonlinear optimization problem [Madsen et al., 2004].

$$(\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{J} + \mu\boldsymbol{I})\boldsymbol{h} = -\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{f} \tag{10}$$

where $\boldsymbol{J}$ is the corresponding Jacobian of the energy function and $\boldsymbol{f}$ is the energy function. Different from conventional Gauss-Newton (GN) optimization method, LM introduces an extra term $\mu\boldsymbol{I}$ which controls the aggressiveness of GN approach. If the step is over confident, the damping factor μ will be increased, otherwise it will be lowered. Another key point is that solving global and local transformation together will lead to insufficient of information (to be more specific, singularity in solving the linear equation) which is caused by the fact that ED parameters also contains information about global rotation and transformation. Through LM algorithm, this numerical problem can be avoided.

We would like to point out that one benefit of volume based object registration is that the corresponding Jacobian is smoothly and evenly distributed along the volume. This avoids abrupt changes in the back-projection. Thus the unified distance field function provides better alignments and efficient computation.

3 Results and Discussion

3.1 Simulation Setup

Simulated datasets were generated from the different real soft-tissue models to demonstrate the effectiveness of the deformation recovery algorithm proposed in this paper. Three different soft-tissue models (heart, liver and right kidney) were downloaded from OpenHELP [Kenngott et al., 2015], which were segmented from a CT scan of a healthy, young male undergoing shock room diagnostics. In the simulation, each model was randomly deformed as the ground truth by using the embedded deformation approach [Sumner et al., 2007]. The deformation of the soft-tissue is simulated by randomly exerting 2-3 mm movement on a random vertex on the model with respect to the status of the deformed model from the last frame. Then, camera poses with trajectories looped around the model were simulated to generate the point cloud scan from the randomly deformed model. Gaussian noises were added to the perfect camera poses to simulate the data from the EM tracking system. The distance from the camera center to the model is around 200mm. Pinhole model is use to simulate the stereo camera with the camera intrinsic parameters as:

$$\begin{bmatrix} 520 & 0 & 640 \\ 0 & 520 & 320 \\ 0 & 0 & 1 \end{bmatrix}$$

Fig 3 is an example observation of a liver model. In each frame the camera only observes part of the deformed model which makes the recovery of the soft-tissue even more challenging.

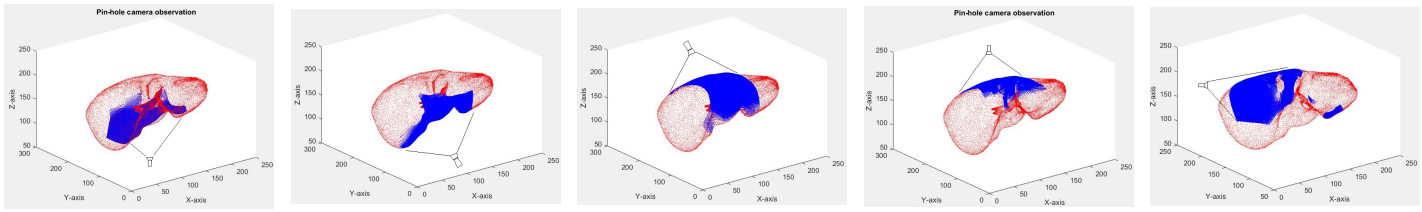
3.2 Simulation Results

In the model-to-scan deformation and registration process, the size of the downsampling grid is set to 20mm to obtain the ED nodes, and the number of neighboring points is set to 4. These are the default parameters in the embedded deformation in [Sumner et al., 2007]. The weights used in the optimization proposed in Eq 5 are set to 1, 20000 and 100 for rotation, smoothness and distance error respectively, which is proposed in [Newcombe et al., 2015] as a practical weight combination.

Fig 4 are the visible points-to-scan registration error map which is generated by taking corresponding value in the voxel of DFF. Results show that most points are correctly fit to reference model and the maximum error is about 4 mm. Some abrupt big error results from small details which can not be fit by sparse ED nodes. To solve this problem, ED nodes should be sampled denser which will increase computations at cost. There is a tradeoff between computation and accuracy.

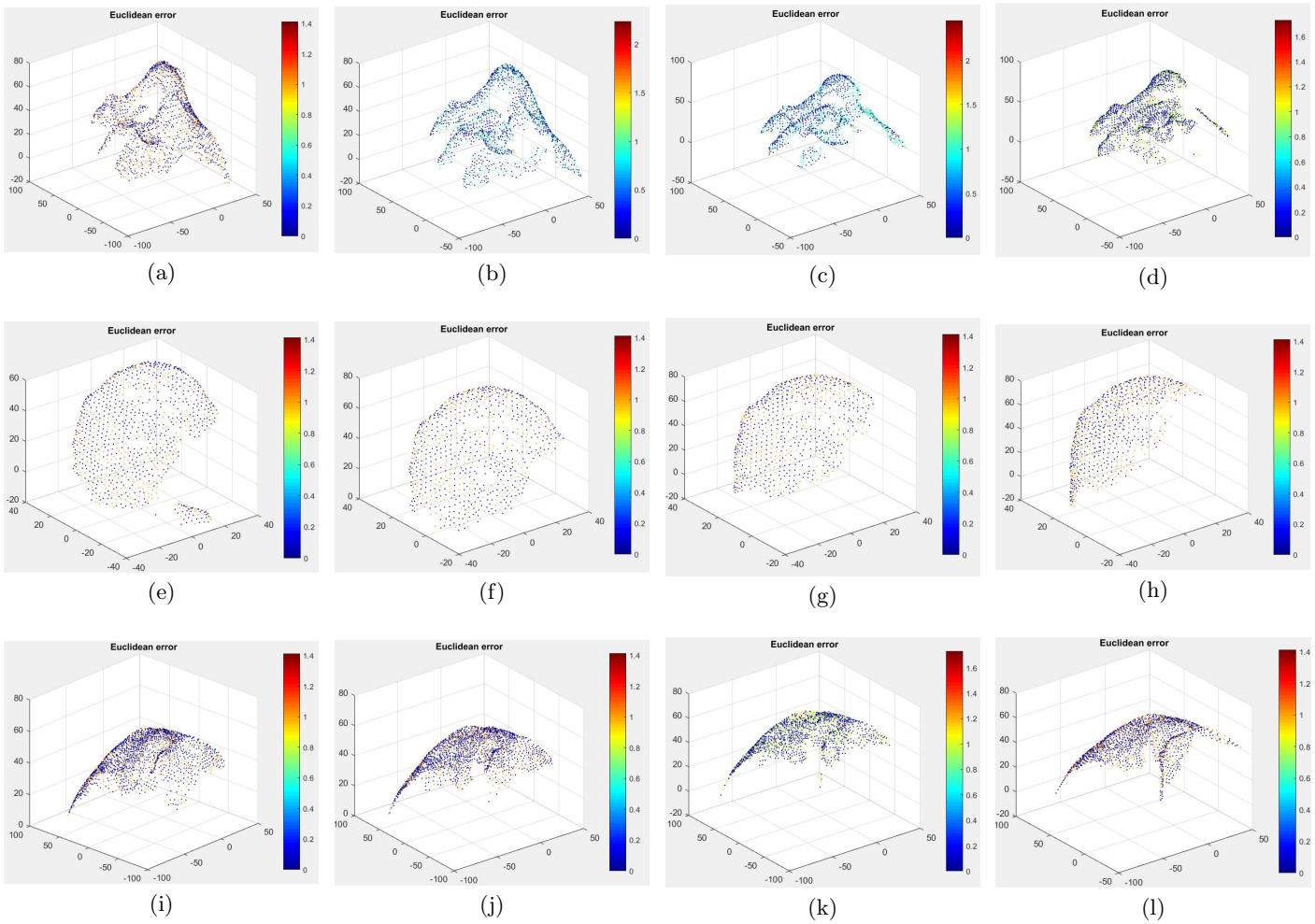
To illustrate the accuracy of DFF based registration process proposed in this paper, as comparison, the back-projection approach used in [Newcombe et al., 2015] and [Dou et al., 2016] is also implemented using the same datasets. Similar to the proposed algorithm, we define the error from the back-projection approach to be the minimal distance from a transformed point to the closest point from the scan. The mean errors were used to compare the effectiveness of these two methods and the quantitative comparison of the accuracy is shown in Table 1.

Fig 5 shows the comparisons between the models generated from the proposed algorithm and the corresponding ground truth which are used to generate the scans from the camera. It is clear that the deformed models are close to the ground truth at the area where the model was observed. On the contrary, the farther the point on the model is away from the observation, the larger error it could be. This is due to the lack of information and the smoothness in our energy function exerting on the unobserved part of the model. In other words, these unobserved points are predicted through the minimization of our energy function. Even though the prediction could not be that accurate, as more parts are observed, the accuracy will be increased significantly. Fig 6 shows the last frame of the deformed model which is presented in the form of Axial, Coronal and Sagittal map. All the results demonstrated that the deformed models get quite close to scan but areas far away to the observation shows obvious errors.



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

Figure 3: (a) to (e) are the simulation of generating the depth scan observation from the deformed liver model. The blue points are the simulated depth observations. The point cloud in red is the deformed model.



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

(e) (f) (g) (h)

(i) (j) (k) (l)

Figure 4: The results of model-to-scan registration colored by the matching error (in mm) which is directly obtained from the DFF. (a)-(d) are selected error map from the heart model;(e)-(h) are selected error map from the right kidney model;(i)-(l) are selected error map from the liver model.

In the optimization process of all the experiments, using the DFF makes the LM algorithm converged very fast within about 3-8 iterations. And issues of singularity, divergence or bad fitting never happened.

There is a limitation in the framework proposed in this paper that we need CT scan as the initial model and EM sensor to provide global pose of the camera. Different from DynamicFusion, in the minimal invasive surgery scenario, the scope is quite close to the object (it is set to 200-300 mm in our simulation) which limits

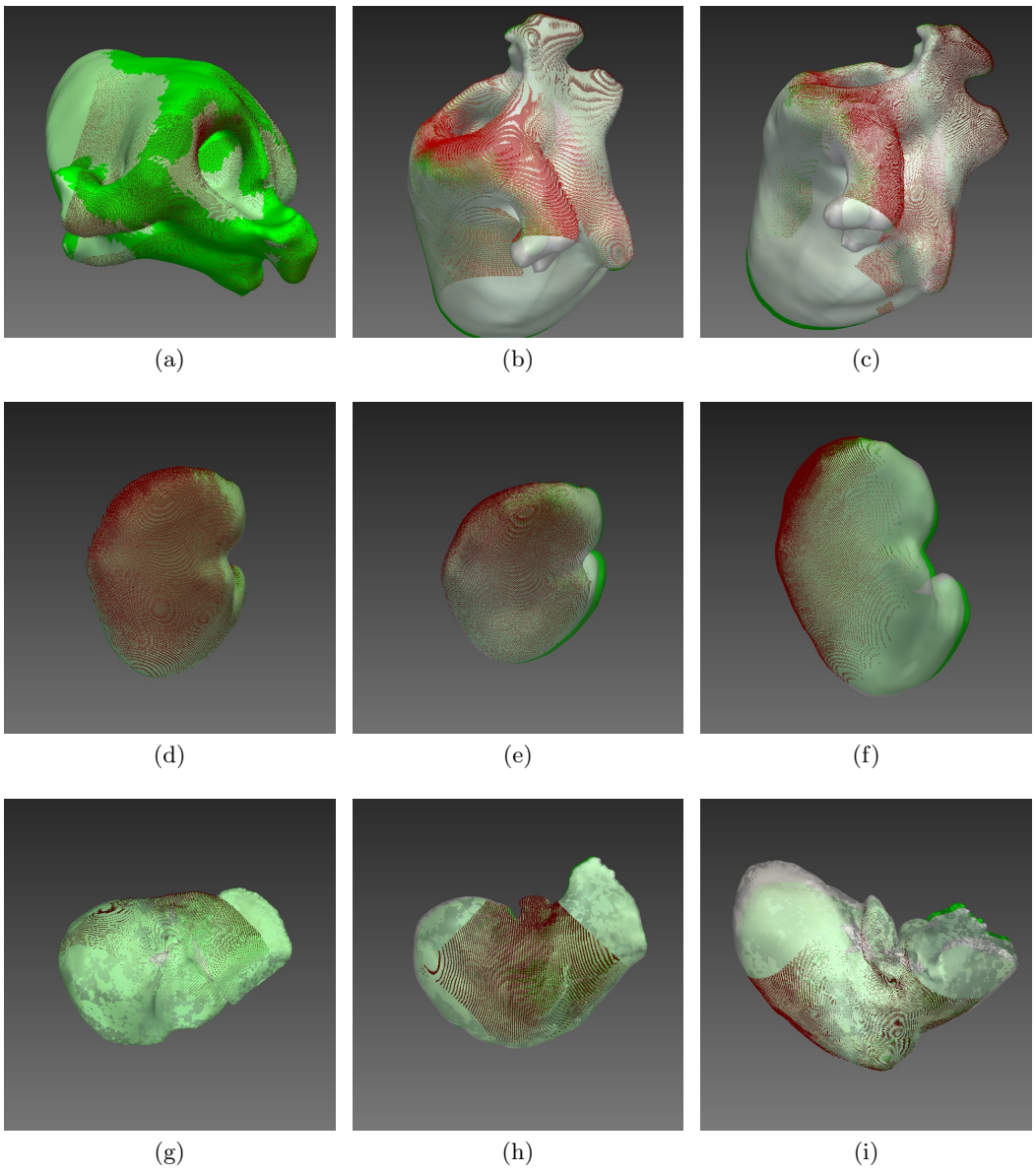


Figure 5: The comparison between the deformed models recovered from the proposed algorithm and the ground truth used for generating the depth observations, by using the heart model (a)-(c), the kidney model (d)-(f) and the liver model (g) to (i) respectively. The models in green are the ground truth, while the models in white are the recovered soft-tissues.

the field of view. If only small part of model is observed (Fig 3), scan could be easily initialized to a different area thus fused to a wrong shape. Considering the easy access to CT and EM, we make full use of them for better accuracy.

4 Conclusions

The contribution of this paper is a deformation recovery framework for the 3D reconstruction of the deformable soft-tissue in the scenario of MIS based on the pre-operative CT data and real-time depth sensing. The distance field function is proposed for robust, efficient and accurate optimization and the model-to-scan registration and model deformation can be solved simultaneously in the proposed framework. Simulations results using three public available soft-tissue models show that different deformations were recovered accurately using the proposed algorithm with very good convergence which is promising for real-time implementation.

Future work will focus on 3 parts. 1. Investigating

Table 1: Accuracy comparison between our approach and Back-projection approach (mm). Each value is calculated by averaging all the points of all the frames.

	DFF based approach	Back-projection based approach
Heart	0.36	0.91
Liver	0.30	0.60
Right Kidney	0.35	0.76

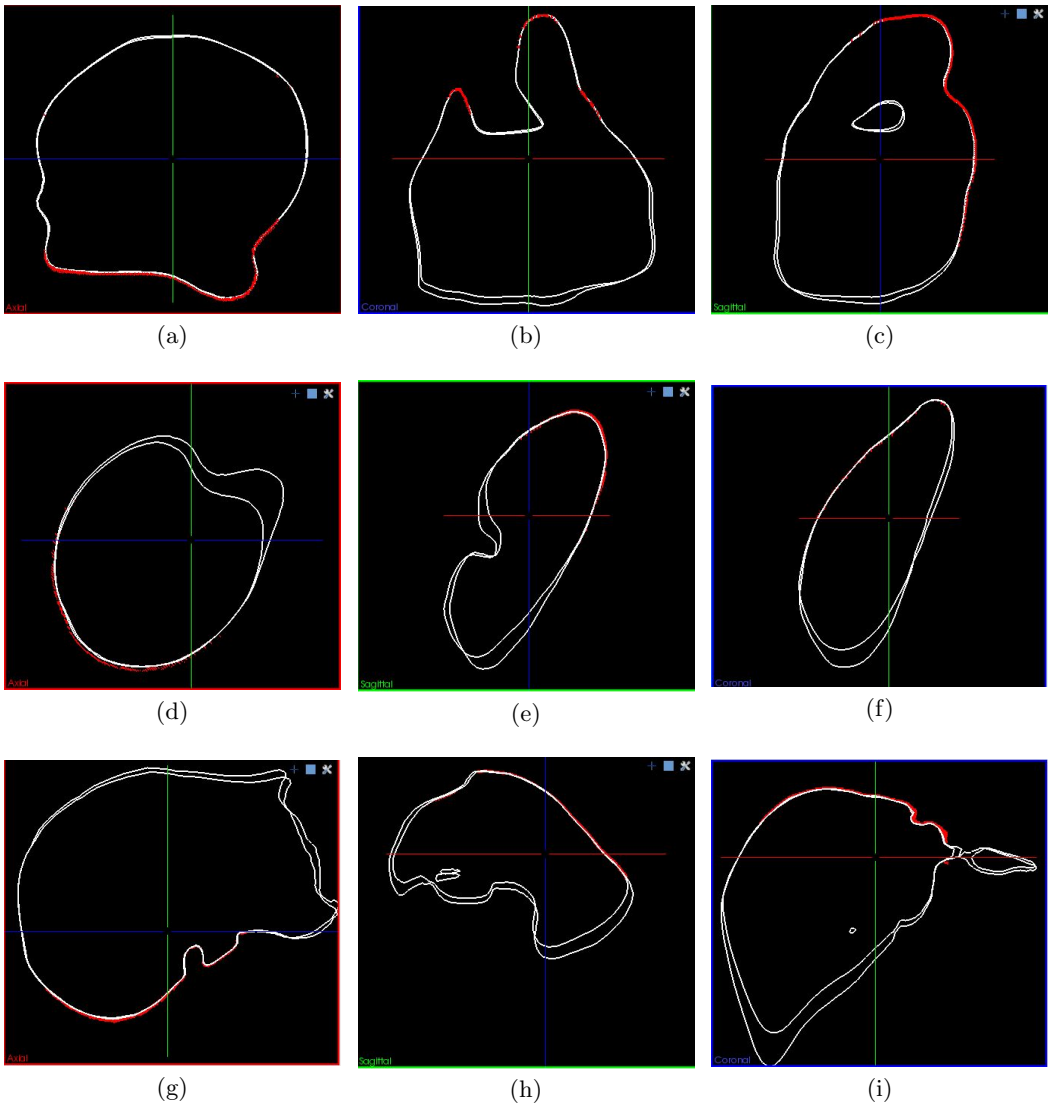


Figure 6: The Axial ((a)-(c)), Coronal ((d)-(f)) and Sagittal ((g)-(i)) views of the deformed model and ground truth at the last frame. The red points denotes the scan of the last frame.

further about the achievable accuracy of the recovered soft-tissue model with respect to the scan accuracy and the limited field of view of the camera. 2. Implementing our method and test it on real-time clinical experiment. Utilizing stereo camera with both RGB and depth information for more robust shape recovery. 3. Investigate the more complicated deformation scenario like fusing new observed data with topology changes like a cut on the soft-tissue.

Acknowledgments

This work is supported in part by the Center of Autonomous System, University of Technology, Sydney, the Australian Governments International Postgraduate Research Scholarship and Endeavour Research Fellowship and China Scholarship Council. We would also like to thank OpenHELP (Heidelberg laparoscopy phantom) for providing organ models.

References

[Davis, 1975] Davis, L. S. (1975). A survey of edge detection techniques. *Computer graphics and image processing*, 4(3):248–270.

[Dou et al., 2013] Dou, M., Fuchs, H., and Frahm, J.-M. (2013). Scanning and tracking dynamic objects with commodity depth cameras. In *Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on*, pages 99–106. IEEE.

[Dou et al., 2016] Dou, M., Khamis, S., Degtyarev, Y., Davidson, P., Fanello, S. R., Kowdle, A., Escolano, S. O., Rhemann, C., Kim, D., Taylor, J., et al. (2016). Fusion4d: real-time performance capture of challenging scenes. *ACM Transactions on Graphics (TOG)*, 35(4):114.

[Dou et al., 2015] Dou, M., Taylor, J., Fuchs, H., Fitzgibbon, A., and Izadi, S. (2015). 3d scanning deformable objects with a single rgbd sensor. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 493–501. IEEE.

[Du et al., 2015] Du, X., Clancy, N., Arya, S., Hanna, G. B., Kelly, J., Elson, D. S., and Stoyanov, D. (2015). Robust surface tracking combining features, intensity and illumination compensation. *International journal of computer assisted radiology and surgery*, 10(12):1915–1926.

[Hu et al., 2007] Hu, M., Penney, G., Edwards, P., Figl, M., and Hawkes, D. J. (2007). 3d reconstruction of internal organ surfaces for minimal invasive surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 68–77. Springer.

[Innmann et al., 2016] Innmann, M., Zollhöfer, M., Nießner, M., Theobalt, C., and Stamminger, M. (2016). Volumedeform: Real-time volumetric non-rigid reconstruction. *arXiv preprint arXiv:1603.08161*.

[Kenngott et al., 2015] Kenngott, H., Wünscher, J., Wagner, M., Preukschas, A., Wekerle, A., Neher, P., Suwelack, S., Speidel, S., Nickel, F., Oladokun, D., et al. (2015). Openhelp (heidelberg laparoscopy phantom): development of an open-source surgical evaluation and training tool. *Surgical endoscopy*, 29(11):3338–3347.

[Lin et al., 2015] Lin, B., Sun, Y., Qian, X., Goldgof, D., Gitlin, R., and You, Y. (2015). Video-based 3d reconstruction, laparoscope localization and deformation recovery for abdominal minimally invasive surgery: a survey. *The International Journal of Medical Robotics and Computer Assisted Surgery*.

[Madsen et al., 2004] Madsen, K., Nielsen, H. B., and Tingleff, O. (2004). Methods for non-linear least squares problems.

[Maier-Hein et al., 2013] Maier-Hein, L., Mountney, P., Bartoli, A., Elhawary, H., Elson, D., Groch, A., Kolb, A., Rodrigues, M., Sorger, J., Speidel, S., et al. (2013). Optical techniques for 3d surface reconstruction in computer-assisted laparoscopic surgery. *Medical image analysis*, 17(8):974–996.

[Newcombe et al., 2015] Newcombe, R. A., Fox, D., and Seitz, S. M. (2015). Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 343–352.

[Newcombe et al., 2011] Newcombe, R. A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A. J., Kohi, P., Shotton, J., Hodges, S., and Fitzgibbon, A. (2011). Kinectfusion: Real-time dense surface mapping and tracking. In *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on*, pages 127–136. IEEE.

[Sorkine and Alexa, 2007] Sorkine, O. and Alexa, M. (2007). As-rigid-as-possible surface modeling. In *Symposium on Geometry processing*, volume 4.

[Stoyanov, 2012] Stoyanov, D. (2012). Stereoscopic scene flow for robotic assisted minimally invasive surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 479–486. Springer.

[Sumner et al., 2007] Sumner, R. W., Schmid, J., and Pauly, M. (2007). Embedded deformation for shape manipulation. *ACM Transactions on Graphics (TOG)*, 26(3):80.

[Zollhöfer et al., 2014] Zollhöfer, M., Nießner, M., Izadi, S., Rehmann, C., Zach, C., Fisher, M., Wu, C., Fitzgibbon, A., Loop, C., Theobalt, C., et al. (2014). Real-time non-rigid reconstruction using an rgb-d camera. *ACM Transactions on Graphics (TOG)*, 33(4):156.