Deformable Shape-From-Motion in Laparoscopy using a Rigid Sliding Window

Toby Collins
http://isit.u-clermont1.fr/content/Toby-Collins
Benoît Compte
http://isit.u-clermont1.fr/content/Benoît-Compte
Adrien Bartoli
http://isit.u-clermont1.fr/~ab

ALCoV-ISIT
Université d'Auvergne
Clermont-Ferrand
France

Abstract

We present an approach for deformable 3D surface reconstruction using laparoscopic monocular video data that is based on breaking down the nonrigid reconstruction problem into a sequence of temporally-localised rigid reconstructions. Our main contribution, the Rigid Sliding Window algorithm provides a means to achieve this whereby Rigid Shape-from-Motion is performed over a sliding window that adapts according to the amount of deformation detected in the scene. Our method is particularly applicable for laparoscopy, where, by adapting the window, it can exploit low-frequency deformations caused by for example breathing. Our algorithm is demonstrated on both synthetic and in-vivo image sequences.

1 Introduction and Background

An important computer vision task in Minimally Invasive Surgery is to recover the 3D structure of deformable tissues from endoscopic images. Solutions to this have several important applications, including intra-operative surgical guidance, motion estimation and compensation and pre-operative data registration. Currently, state-of-the-art methods differ along two main axes; the information used to infer 3D structure, and the imaging hardware. In vivo 3D reconstruction has been attempted previously using stereo endoscopes [4, 7, 10] and active 3D methods based on structured sensors [1]. These simplify the 3D reconstruction problem when compared with methods based on standard monocular endoscopes, yet come at the price of considerable hardware investment and operational training. By contrast, monocular methods require no hardware modification. The 3D reconstruction problem is considerably more difficult however and remains an open challenge.

In this paper we study monocular laparoscopic 3D reconstruction based on Shape-from-Motion (SfM.) SfM has been attempted previously, including the reconstruction of the abdominal cavity [9] and heart [6]. Several of the recent SfM methods are based on Simultaneous Localisation And Mapping (SLAM) [3, 5, 8], which can provide a framework for 3D reconstruction in real time. However SLAM necessarily assumes the 3D scene is rigid, which is mostly an unrealistic requirement in laparoscopic data. Nonrigid deformation is caused for example by breathing and surgical intervention, and these effects will defeat rigid-based SfM methods. Recently, a deformable SfM method has been applied to surface reconstruction based on low-rank shape models [6]. However, selecting the number of shape bases is usually non-trivial. An extension of SLAM to handle periodic deformations was presented in [8].

We present a new nonrigid SfM approach which specifically exploits the fact that the nature of deformation in laparoscopic data often is of low temporal frequency. That is, although over a full video sequence the scene is deformable, over a small temporal window of the order of a few frames the scene motion may be approximated rigidly. We present an algorithm which breaks down the deformable 3D reconstruction problem into a sequence of temporally-localised rigid reconstructions. A naive implementation using fixed-sized windows may produce unstable results however. This is because (*i*) the degree of nonrigid motion may vary over time and (*ii*) there may be insufficient image motion with which to reliably recover 3D shape. Our method automatically tunes the temporal windows to facilitate a good trade-off between increasing the amount of rigid baseline and reducing the degree of nonrigid deformation over the window. We call our algorithm a Rigid Sliding Window (RSW) approach to deformable laparoscopic reconstruction. The RSW is particularly suitable for laparoscopic data. In contrast to arbitrary imaging conditions, the degree of rigidity between image frames is determined by the relative speed of the camera with respect to the surface's rate of change due to deformation.

The structure of this paper is broken down as follows. In Section 2 we present the RSW algorithm. An overview of the algorithm's principles is given in §2.1 and the complete algorithm is given in §2.2. In §3 we present results using synthetic data of a deforming kidney model, and provide empirical analysis of the performance of the RSW algorithm, in particular with respect to varying camera speeds. We then present some experimental results on real in vivo data of a deforming pig liver and in §4 we present concluding remarks and directions for future research.

2 The Rigid Sliding Window

2.1 Principle

Our Rigid Sliding Windows reconstruction algorithm processes sequentially the laparoscopic video by finding contiguous windows of frames for which rigid SfM is able to perform well. The duration of these windows is determined by the amount of nonrigidity in the scene. For example, given a highly deforming scene the rigid assumption may only be valid over a couple of frames, whereas for an approximately rigid scene the assumption holds over the whole video sequence. The task of the RSW algorithm is to determine the optimal window size and to perform an approximate rigid 3D reconstruction over these windows. The output of the algorithm is therefore not a single 3D reconstruction, but a sequence of 3D reconstructions capturing the different deformed states of the 3D scene.

Critical to our approach is how to optimally determine the rigid window size. There is an inherent tradeoff between having larger windows, which increases the camera baseline and provides better stability for depth and camera pose estimation, yet remaining small enough such that the 3D motion within the window is approximately rigid. We base our algorithm on testing the hypothesis of rigid motion, and increasing the window size up to the point where this hypothesis breaks down. More precisely, suppose we have a temporal window spanning frames captured between times t_0 and t_1. At some time $t_0 \leq i \leq t_1$, suppose that the endoscope's projection matrix is given by $\mathbf{P}_i = \begin{bmatrix} \mathbf{K} & 0_3 \end{bmatrix} \begin{bmatrix} \mathbf{R}_i & \mathbf{t}_i \\ 0_3 & 1 \end{bmatrix}$ with rotation $\mathbf{R}_i$, translation $\mathbf{t}_i$ and projection intrinsics $\mathbf{K}$. We assume $\mathbf{K}$ to be known and radial distortion undone, which can be determined easily with an offline calibration procedure [11]. Suppose for a 3D point $\mathbf{x}_p$ in the scene, we have its 2D position $\mathbf{y}_p^i$ in the camera's image at time i. $\mathbf{x}_p$ respects the rigid motion model over $t_0 \leq i \leq t_1$ if the residual error of its projection can be explained by image noise. Assuming additive gaussian noise, we have:

$$\mathbf{r}_{p,i} = \left(\psi\left(\mathbf{x}_p; \mathbf{K}, \mathbf{R}_i, \mathbf{t}_i\right) - \mathbf{y}_p^i\right) \sim N\left(\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \sigma^2 \mathbf{I}_2\right) \tag{1}$$

with $\psi(\mathbf{x}_p;\mathbf{K},\mathbf{R}_i,\mathbf{t}_i) = \begin{bmatrix} u/w \\ v/w \end{bmatrix}$ and $[u,v,w]^T = \mathbf{P}_i \begin{bmatrix} \mathbf{x}_p \\ 1 \end{bmatrix}$. We can test whether the point violates the rigid assumption at time i if $\mathbf{r}_{p,i}$ is an outlier with respect to the noise model. We chose to use the z-score statistic. Here, $\mathbf{r}_{p,i}$ is marked an outlier if $z = \mathbf{r}_{p,i}^{\top}\Sigma^{-1}\mathbf{r}_{p,i} > \tau$ for some tolerance τ. Returning back to the RSW concept, we can infer that a window between t_0 and t_1 supports rigid motion if the ratio of the number of outliers to the number of inliers is small. Conversely, if it is large then the window is too large to support rigid motion. The idea of testing the rigid hypothesis in this way, and adjusting the sliding window accordingly as it advances through the frame sequence is at the core of our RSW algorithm.

Algorithm 1 RSW 3D Reconstruction

$t_0 \leftarrow 1, t_1 \leftarrow 1$ //Sliding window initialised to start of sequence
Mode 1: Window Expansion
repeat
  $t_1 \leftarrow t_1 + 1$, call $rigidSfM(t_0,t_1)$
until ratio of inliers $< \rho$
//Rigid assumption is now violated
if baseline $\geq b$ **then**
  Output reconstruction for window t_0 to $(t_1 - 1)$
  goto Mode 2
else
  //Reconstruction unachievable:
  $t_0 = t_0 + 1, t_1 = \max(t_0, t_1 - 1)$
  goto Mode 1
end if
Mode 2: Window Contraction
repeat
  $t_0 \leftarrow t_0 + 1$, call $rigidSfM(t_0,t_1)$
until ratio of inliers $\geq \rho$ or $t_0 == t_1$
//Rigid assumption is now satisfied
goto Mode 1

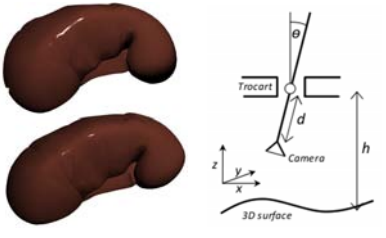
2.2 Algorithm Details

We present pseudo-code for our RSW in Algorithm 1. The window is parameterised by start and end frames t_0 and t_1 respectively. These are initially set at the first frame, and are then modified according to two distinct operation modes. In Mode 1: Window Expansion, t_1 is incremented up to the point where the rigid motion model is violated (*i.e.* the inlier ratio becomes less than a tolerance ρ.) At each iteration a call to $rigidSfM(t_0,t_1)$ is made which denotes calling a rigid SfM algorithm using frames in the window t_0 to t_1. This involves the following general processes (1) tracking points over the window and (2) solving for the camera poses and 3D points. Usually this is done with RANSAC model estimation followed by bundle adjustment. If the rigid model becomes violated at time t_1, we test whether there has been a sufficient motion baseline to stably estimate 3D structure between times t_0 to $(t_1 - 1)$. A simple strategy we take is to estimate the baseline using mean disparity of the inliers over the window. In all our tests we have set $b = 5$ pixels. If the baseline is insufficient it means that there is high deformable motion at time t_0, and a rigid reconstruction is unachievable. Here the sliding window advances without outputting a reconstruction. If the baseline was sufficient, the algorithm outputs the reconstruction and proceeds to Mode 2: Window Contraction. In this mode the sliding window is contracted by incrementing t_0, until the window satisfies the rigid model again. At this point we return to Mode 1, and the process of expanding the window using t_1 repeats again.

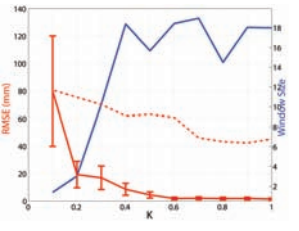
3 Experimental Results

Synthetic data. We designed a synthetic experimental setup to show empirically the performance of the RSW algorithm by simulating the motion of a laparoscope viewing a deforming organ. The setup was as follows. A 3D kidney model (comprising 1820 vertices

bounded by a volume of 120mm^2) was deformed via free-form deformation into a plausible deformed state 1(a). The mean vertex displacement between these states was 7.6mm. A 60-frame sequence (2 seconds at 30fps) was then generated by linearly interpolating between the undeformed and deformed states. A laparoscope camera was synthesised with intrinsics given by a calibration of a Karl Storz 2.9mm zero degree laparoscope. We synthesised the motion by pivoting the laparoscope about a virtual trocar 1(b). Defining the world coordinate frame at the trocar pivot, the camera's position is given by $\mathbf{M} = \mathbf{R}_x(\theta)\mathbf{T}_z(d)$, where $\mathbf{T}_z$ denotes a translation along the negative z axis by $d$mm, and $\mathbf{R}_x(\theta)$ denotes a rotation about the x axis by an angle of θ radians. The performance of the RSW algorithm is directly influenced by the amount of deformation exhibited by the 3D scene relative to the amount of rigid motion between the camera and the scene. To test this, we simulated the motion of the camera with varying speed whilst observing the deforming kidney, which we position at a distance of $h = 65$mm from the trocar. Specifically, at frame i, the camera's position is given by: $\theta = ki\pi/500$ and $d = 20 + ki/3$, where k denotes a speed factor that we vary from 0 to 1. Figure 1(c) shows, for each camera speed the distance between the camera's position in the 1^{st} and 60^{th} frames (second column) and the average optic flow magnitude between consecutive frames (third column.) It is important to note that at $k = 0.0$, the optic flow is entirely generated by nonrigid deformation, yet as the camera speed increases the optic flow becomes dominated by the rigid motion. We ran the RSW algorithm for each camera speed as follows. We synthesised image correspondences by randomly selecting 90 vertices which remained visible to the camera and perturbed their image positions with noise $\sigma = 1.0$. We set the RSW's free parameters to $\rho = 0.7$ and $\tau = 1.96$. Figure 1(d) shows the reconstruction performance as a function of k. The Root Mean Squares Error of the points' 3D positions (solid red) clearly reduces with increased camera speed. We also note that the average window size increases, showing that the RSW algorithm is exploiting more frames with increased camera speeds. We also compared against the RMSE error obtained by performing rigid SfM (including bundle adjustment) over the whole 60 frames (dashed red). Since this does not handle deformable motion, this clearly results in less accurate reconstructions.



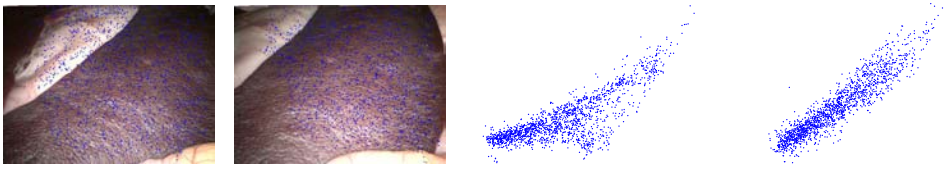
k	End-point Distance (mm)	Mean OF Magnitude (pixels)
0.00	0.00	1.6
0.10	2.35	1.90
0.20	4.74	2.70
0.30	7.15	3.70
0.40	9.60	4.50
0.50	12.08	5.40
0.60	14.59	6.80
0.70	17.13	7.70
0.80	19.70	8.90
0.90	22.30	10.20
1.00	24.79	11.50



(a) Kidney model in undeformed (top) and deformed (bottom) states (b) Laparoscope model (c) Motion information (d) Synthetic reconstruction results

Figure 1: Experiments with a synthetic kidney model.

In-vivo liver sequence. A second set of experiments was conducted on real in-vivo data. The data comprises a sequence spanning 120 frames capturing a pig liver which deforms due to respiration. The data spans just over one breathing cycle. Two example frames are shown in Figure 2(a,b). We ran the RSW algorithm with the same values for ρ and τ, but with $\sigma = 1.5$. We used optic flow [7] to estimate the image motion between t_0 and t_1. Flow estimates failing left/right constancy were removed, and 4,000 random samples from the flow field were kept with which to reconstruct the point clouds (shown in blue.) The point clouds for the two example frames are shown in 2(c-d). In the absence of ground truth data, no quantitative performance measures are available for this sequence, however qualitatively the reconstructions look reasonable, and capture the slant and gradual curve of the liver surface.



(a) Frame 24 (b) Frame 76 (c) Point cloud (Frame 24) (d) Point cloud (Frame 76)

Figure 2: In-vivo experiments with a pig liver.

4 Conclusion and Future Work

In this paper we have presented an approach for reconstructing deformable 3D surface from image sequences by posing the task as solving multiple, temporally local rigid reconstructions. We have provided the Rigid Sliding Window, an algorithm for sequentially processing an input video and determining subwindows for which the rigid hypothesis holds. In future work we aim to conduct more experiments to further validate the method, exploit smooth camera motion to help stabilise results in very narrow windows, and aim to reconstruct a complete 3D surface model using the set of multiple rigid reconstructions.

References

[1] J. D. Ackerman, K. Keller, and H. Fuchs. Surface reconstruction of abdominal organs using laparoscopic structured light for augmented reality. *Three-Dimensional Image Capture and Applications*, 2002.

[2] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. *ECCV*, 2004.

[3] D. Burschka, M. Li, M. Ishii, R. Taylor, and G. Hager. Scale-invariant registration of monocular endoscopic images to CT-scans for sinus surgery. *MICCAI*, 2004.

[4] F. Devernay, F. Mourgues, and È. Coste-Manière. Towards endoscopic augmented reality for robotically assisted minimally invasive cardiac surgery. *International Workshop on Medical Imaging and Augmented Reality*, 2001.

[5] M. Hu, G. Penney, P. Edwards, M. Figl, and D. Hawkes. 3D reconstruction of internal organ surfaces for minimal invasive surgery. *MICCAI*. 2007.

[6] M. Hu, G. P. Penney, D. Rueckert, P. J. Edwards, F. Bello, R. Casula, M. Figl, and D. J. Hawkes. Non-rigid reconstruction of the beating heart surface for minimally invasive cardiac surgery. *MICCAI*, 2009.

[7] W. W. Lau, N. A. Ramey, J. J. Corso, N. V. Thakor, and G. D. Hager. Stereo-based endoscopic tracking of cardiac surface deformation. *MICCAI*, 2004.

[8] P. Mountney and G.-Z. Yang. Motion compensated SLAM for image guided surgery. *MICCAI*, 2010.

[9] J. M. M. Montiel Óscar G. Grasa, Javier Civera. EKF monocular SLAM with relocalization for laparoscopic sequences. *ICRA*, 2011.

[10] D. Stoyanov, A. Darzi, and G.-Z. Yang. 3D depth recovery for soft tissue deformation during robotically assisted laparoscopic surgery. *MICCAI*, 2004.

[11] Z Zhang. A flexible new technique for camera calibration. *PAMI*, 2000.