Automatic Verification of Laparoscopic 3D Reconstructions with Stereo Cross-Validation

Richard MODRZEJEWSKI [1,2,3]***, Toby COLLINS*** [2]***, Adrien BARTOLI*** [1]***, Alexandre HOSTETTLER*** [2]***, Luc SOLER*** [2] ***and Jacques MARESCAUX*** [2]

EnCoV [1] *- UMR 6602 - CNRS/UBP/SIGMA, Tel : +33(0)4 73 17 81 23*
Institut Pascal, 63000 Clermont Ferrand
IRCAD [2] *and IHU-Strasbourg* [3]*, Tel : +33(0)3 90 41 36 00*
1 Place de l'Hopital, 67000 Strasbourg
Contact: richard.modrzejewski@ihu-strasbourg.eu

An important task in computer-assisted laparosurgery is to automatically and densely reconstruct the surgical environment in 3D. The most successful approaches are Simultaneous Localization and Mapping (SLAM) and Structure-from-Motion (SfM). Nevertheless SLAM and SfM can both fail for several reasons. The ability to automatically and quickly detect reconstruction failures is therefore a crucial component in a robust system. We present stereoscopic cross-validation as the first solution for this task. This is designed for SfM reconstruction, works with calibrated stereo laparoscopes, and requires no additional sensors or interventional imaging.

1 Introduction and Contributions

Automatic, accurate and reliable 3D reconstruction of anatomical structures from laparoscopic images is important in many surgical navigation approaches [8, 6, 2, 9, 16, 14, 13, 5]. The main motivation is to enable image-based registration between detailed 3D organ models built from CT or MR data, which in turn enables *Augmented Reality-based surgical guidance* [11, 7, 3, 4, 12, 1]. This guidance allows the surgeon to see where hidden sub-surface structures such as vessels and tumours are located in the laparoscopic images, by fusing the registered organ model with the laparoscopic video. The most successful reconstruction approaches that handle 'large' scenes are SLAM and SfM. These work by registrating points on the target's surface on multiple laparoscopic images and reconstructing them in 3D space. They are compatible with current surgical workflows, require no modification to current laparoscopic hardware, require no tracking sensors, and are applicable for both monocular and stereo laparoscopes. The key difference between SLAM and SfM is that SLAM reconstructs the scene *live*, whereas SfM reconstructs it offline after a short *exploration phase* typically lasting under a minute. The advantage of SfM is that it does not have real-time constraints, which allows more powerful methods to be used to achieve greater robustness [17]. Nevertheless, both approaches can fail because of poor texture, poor image quality or very strong illumination change. Our main contribution is to provide an automatic, robust solution for detecting SfM reconstruction failure using what we call *stereo cross-validation*.

2 Method and Results

State-of-the-art monocular SfM implementations such as Theia [15] take as input a set of 2D images (often called keyframes), and returns a 3D model of the environnement and a set of camera poses associated with the keyframes. SfM reconstruction fails when there is significant errors in the camera poses and/or 3D model.

We propose to detect these failures using an approach inspired by *cross-validation*, which is a technique for assessing how properties of a fitted model stay the same even if we use an other approach to generate them. In our case the model is the SfM reconstruction. We propose to *withhold* image data from the SfM process, and

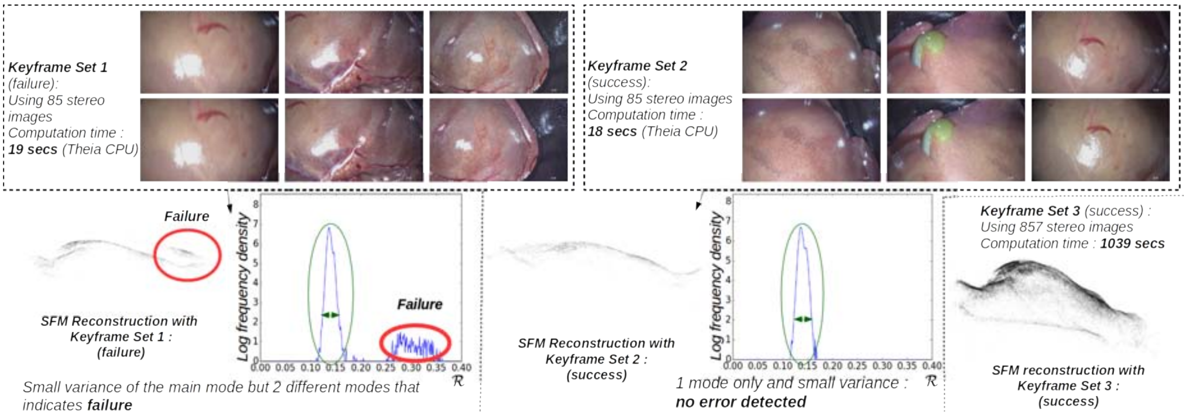

Figure 1: *Our method applied to an exvivo porcine liver dataset using different keyframe sets*

then use this withheld data to cross-validate the reconstruction. For this to work effectively we must minimize correlations between the withheld and non-withheld data. Our strategy is to use a stereo laparoscope and perform SfM with one of the camera's images, then validate the reconstruction with the second camera's images. The main idea is to compute for each point of the reconstruction, the ratio between its depth and its correponding depth value computed with a semi dense stereo algorithm [18]. We then compute the distribution $\mathcal{R}$ of these ratios over all keyframes. A good reconstruction generates a unimodal distribution with low variance. For the stereo, an algorithm providing sparse but reliable depthmap is prefered to avoid false positives. From this depth ratio distribution, we can:

1. Reject or accept a reconstruction
2. Compute a reconstruction confidence score
3. Reject incorrect surface regions or keyframes corresponding to non principal modes in the depth ratio distribution
4. Indicate where more exploration is needed (when surface regions or keyframes have been rejected)

We focused on option 1 which we implemented in Algorithm 1. This assumes the reconstructed model contains a set of 3D points $\mathcal{P}$ (called a pointcloud).

Algorithm 1 SfM Verifier

function IsCorrect(*Keyframes* $\mathcal{K}$, *PointCloud* $\mathcal{P}$)
  $\mathcal{R} \leftarrow \{\}$
  for $t \in [1, T]$ **do**
    $\mathbf{D} \leftarrow$ stereo depthmap from $\mathbf{K}_t \in \mathcal{K}$
    for $\boldsymbol{p}_j \in \mathcal{P}$ **do**
      Project $\boldsymbol{p}_j$ onto $\mathbf{D}$ and measure the depth value d using bilinear interpolation
      if (d is defined) **then**
        Add $\frac{\boldsymbol{p}_j}{d}$ to $\mathcal{R}$
  n $\leftarrow$ number of modes of $\mathcal{R}$
  $\sigma \leftarrow$ variance of the primary mode
  return (n == 1) $\wedge$ ($\sigma < \tau$)

A decision threshold τ is needed and set to 0.07 empirically. In the future, τ should be learned from data gathered from the camera used.

We used for our experiment a challenging ex-vivo dataset which has been tested in a preliminary work by running state-of-the art SLAM [10]. Both monocular and stereo methods failed. We then tested Theia SfM with keyframe sets sampled uniformly from an exploration video (2 minutes in length). In Fig.1 some example images are shown. Three different reconstructions were computed from three keyframe sets. In the first two sets, 85 keyframes were used. For the first reconstruction, there is a misalignment between a region on one of the lobes, which was caused by errors in the laparoscope's estimated 3D poses due to strongly homogeneous texture. If left undetected, this type of error can cause significant errors through a surgical guidance pipeline. For the second reconstruction no misalignement appeared. The third reconstruction used the 857 images from the video as keyframes. Nevertheless, while the two first reconstructions took around 20 seconds to compute, the third reconstruction is about 50 times slower. Real time results are not always needed, yet an overly slow reconstruction can be problematic. Associated with reconstructions, we plot the histogram showing the logarithmic frequency density distribution of $\mathcal{R}$. In the failure case, we see two modes which indicates the reconstruction falsely estimates the liver's scale in some regions while on the successful case, only one mode appears.

3 Conclusion and future work

Our method has been tested with Theia but it can also be used with any monocular SfM algorithm if stereo data are available. We presented a novel reconstruction verification method using stereo information to detect 3D reconstruction failures. In future works, we will evaluate this method in vivo.

Acknowledgements. This work is part of a project of IHU Strasbourg called LASAR2. We would like to thank the surgical team of Prof. Didier Mutter and especially Prof. Patrick Pessaux.

References

[1] Sylvain Bernhardt, Stéphane A Nicolau, Luc Soler, and Christophe Doignon. The status of augmented reality in laparoscopic surgery as of 2016. *Medical image analysis*, 37:66–90, 2017.

[2] Toby Collins, Daniel Pizarro, Adrien Bartoli, Michel Canis, and Nicolas Bourdel. Real-time wide-baseline registration of the uterus in monocular laparoscopic videos. In *MICCAI conference, Nagoya, Japan*, 2013.

[3] Toby Collins, Daniel Pizarro, Adrien Bartoli, Michel Canis, and Nicolas Bourdel. Computer-assisted laparoscopic myomectomy by augmenting the uterus with pre-operative mri data. In *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*, pages 243–248. IEEE, 2014.

[4] Nazim Haouchine, Jeremie Dequidt, Igor Peterlik, Erwan Kerrien, Marie-Odile Berger, and Stephane Cotin. Image-guided simulation of heterogeneous tissue deformation for augmented reality during hepatic surgery. In *Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on*, pages 199–208. IEEE, 2013.

[5] N. Mahmoud, A. Hostettler, T. Collins, L. Soler, C. Doignon, and J. M. M. Montiel. SLAM based Quasi Dense Reconstruction For Minimally Invasive Surgery Scenes. *ArXiv e-prints*, May 2017.

[6] Nader Mahmoud, Inigo Cirauqui, Alexandre Hostettler, Christophe Doignon, Luc Soler, Jacques Marescaux, and JMM Montiel. Orbslam-based endoscope tracking and 3d reconstruction. *arXiv preprint arXiv:1608.08149*, 2016.

[7] Nader Mahmoud, Óscar G Grasa, Stéphane A Nicolau, Christophe Doignon, Luc Soler, Jacques Marescaux, and JMM Montiel. On-patient see-through augmented reality based on visual slam. *International Journal of Computer Assisted Radiology and Surgery*, 12(1):1–11, 2017.

[8] Lena Maier-Hein, Peter Mountney, Adrien Bartoli, Haytham Elhawary, D Elson, Anja Groch, Andreas Kolb, Marcos Rodrigues, J Sorger, Stefanie Speidel, et al. Optical techniques for 3d surface reconstruction in computer-assisted laparoscopic surgery. *Medical image analysis*, 17(8):974–996, 2013.

[9] Abed Malti, Adrien Bartoli, and Toby Collins. Template-based conformal shape-from-motion-and-shading for laparoscopy. In *International Conference on Information Processing in Computer-Assisted Interventions*, pages 1–10. Springer, 2012.

[10] Raúl Mur-Artal and Juan D. Tardós. ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras. *arXiv preprint arXiv:1610.06475*, 2016.

[11] Stéphane Nicolau, Luc Soler, Didier Mutter, and Jacques Marescaux. Augmented reality in laparoscopic surgical oncology. *Surgical oncology*, 20(3):189–201, 2011.

[12] Gustavo A Puerto-Souza, Jeffrey A Cadeddu, and Gian-Luca Mariottini. Toward long-term and accurate augmented-reality for monocular endoscopic videos. *IEEE Transactions on Biomedical Engineering*, 61(10):2609–2620, 2014.

[13] Danail Stoyanov, Marco Visentini Scarzanella, Philip Pratt, and Guang-Zhong Yang. Real-time stereo reconstruction in robotically assisted minimally invasive surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 275–282. Springer, 2010.

[14] Deyu Sun, Jiquan Liu, Cristian A Linte, Huilong Duan, and Richard A Robb. Surface reconstruction from tracked endoscopic video using the structure from motion approach. In *Augmented Reality Environments for Medical Imaging and Computer-Assisted Interventions*, pages 127–135. Springer, 2013.

[15] Christopher Sweeney, Tobias Hollerer, and Matthew Turk. Theia: A fast and scalable structure-from-motion library. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 693–696. ACM, 2015.

[16] Johannes Totz, Stephen Thompson, Danail Stoyanov, Kurinchi Gurusamy, Brian R Davidson, David J Hawkes, and Matthew J Clarkson. Fast semi-dense surface reconstruction from stereoscopic video in laparoscopic surgery. In *International Conference on Information Processing in Computer-Assisted Interventions*, pages 206–215. Springer, 2014.

[17] Bill Triggs, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. Bundle adjustmenta modern synthesis. In *International workshop on vision algorithms*, pages 298–372. Springer, 1999.

[18] Koichiro Yamaguchi, David McAllester, and Raquel Urtasun. Efficient joint segmentation, occlusion labeling, stereo and flow estimation. In *European Conference on Computer Vision*, pages 756–771. Springer, 2014.