Endoscopic Stereo Reconstruction: a Comparative Study

Mostafa Parchami[1], Jeffrey A. Cadeddu[2], and Gian-Luca Mariottini[1]

Abstract— **Advances in robotic surgery especially in minimally-invasive surgery (MIS) has increased the need for translating computer-vision algorithms in endoscopic imagery to support surgical decisions. While methods for stereo reconstruction have been extensively investigated for man-made environments, such an extensive and detailed study on the pros and cons of stereo reconstruction for endoscopic images. In this paper, we extensively compare several state-of-the-art methods on both simulated as well as real endoscopic images over controlled in-lab and phantom models observed by a daVinci stereo endoscope. The advantages and disadvantages of each compared method over the major steps of a stereo-reconstruction pipeline are discussed and supported by exhaustive experiments and discussions.**

I. Introduction

Recent advances in endoscopic technology promoted a novel surgical paradigm called minimally invasive surgery (MIS). In MIS, long thin surgical tools are inserted into the patient's body through tiny incisions, and the surgical site is made visible to the surgeon by means of an endoscopic camera. Despite the benefits for the patient (such as reduced trauma and hospitalization time), MIS is challenging for the surgeon who still experiences a reduced awareness of the patient's anatomy due to the limited field of view of the endoscopic camera.

As such, computer-aided navigation systems have been developed in the past years that promise to enhance the surgeon's view about high-risk anatomical targets by fusing pre-operative radiological data onto the live endoscopic video. At the core of these systems is the capacity for the computer to accurately perceive in real time the tri-dimensional (3-D) dynamic structure of the soft-tissue surgical scene.

Achieving both sub-millimeter accuracy and real-time performance in estimating the 3-D tissue geometry from the stereo images have been a Holy Grail for a long time. While stereo reconstruction in man-made environments has been thoroughly investigated [15], [16], no work exists in the endoscopic-vision arena that extensively compares state-of-the-art stereo-reconstruction methods for endoscopic imagery. Stereo reconstruction in endoscopic images is dramatically challenging when compared to man-made environments, because of the large lens distortion, the many texture-less areas, the occlusions introduced by the surgical tools, specular highlights, smoke and blood [4]. Some recent publications target stereo analysis and compare the disparity between single pairs of images from endoscopic images [5]. While these papers provide a survey over the current methods, they do not present any quantitative comparison of stereo reconstruction accuracy, and in particular at each stage of the reconstruction pipeline. Therefore, a comparison between available methods for each individual step of a stereo-reconstruction pipeline is extremely important and it will greatly help the community to increase awareness over those stages the most critical in a stereo endoscopic reconstruction.

In our previous workshop work [17], we provided an initial study of different stereo-reconstruction methods with a particular focus on the image-processing stages necessary to improve image quality. In this work we greatly extend over that initial study, and adopt a controlled in-lab model to perform an extensive comparison between a larger number of state-of-the-art stereo-reconstruction methods from endoscopic images. We also specifically focus this study on comparing two key aspects of a stereo-reconstruction pipeline: disparity and stereo-triangulation stages, which have been shown to critically affect the overall final accuracy. The in-lab model (with known dimensions) is used for the first time to conduct precise tests on all the different stages of a stereo pipeline. Finally, results on endoscopic imagery of a realistic organ's phantom model are also presented. Based on our results, we present in depth discussion on the most suitable algorithms for each stage of the stereo reconstruction pipeline in terms of both accuracy and computational time.

The rest of the paper is organized as follows: Sect. II introduces the stereo reconstruction pipeline with a particular focus on the compared disparity and stereo-triangulation methods. Sect. III provides details about the experimental setup and the results from simulated and real endoscopic images. Finally, Sect. IV discusses the presented results and describes future direction of investigations.

II. Stereo Reconstruction

Stereo reconstruction aims at obtaining a metric 3-D reconstruction of a scene as it is being observed by two (left $\{L\}$ and right $\{R\}$) endoscopic cameras. Stereo reconstruction consists of a sequence of steps, as illustrated in the flowchart of Fig. 1.

The first step consists of de-interlacing and filtering each image to improve the search for similarities (or correspondences). The second step uses image undistortion to remove the image effects caused by the endoscope lenses. Third, the pair of filtered and undistorted images are rectified and

[1]M. Parchami and G.L. Mariottini are with the Computer Science and Engineering Dept., The University of Texas at Arlington, 500 UTA Boulevard, Arlington, TX, USA. Email: mostafa.parchami@mavs.uta.edu, gianluca@uta.edu

[2]J. Cadeddu is with the Dept. of Urology, The University of Texas Southwestern, Harry Hines Blvd., Dallas, TX, USA. Email: jeffrey.cadeddu@utsouthwestern.edu

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

input to a disparity calculation method that obtains a dense disparity map, which is finally used as input to a triangulation phase that estimates the 3-D coordinates of points.

Calibration: To start the reconstruction process, a preliminary camera-calibration phase is required to estimate the intrinsic (${}^{L}\mathbf{K}$, ${}^{R}\mathbf{K}$), extrinsic (${}^{R}_{L}\mathbf{R}$ and ${}^{R}_{L}\mathbf{t}$), as well as the lens-distortion parameters. Since calibration parameters are used in several steps within the stereo reconstruction pipeline, it is important to obtain accurate calibration parameters. For this purpose, we used the MATLAB Camera Calibration Toolbox [1] and we were able to calibrate the stereo endoscope of the daVinci surgical platform with sub-pixel accuracy (max 0.6 pixels of re-projection error). We used a 3×3 cm. calibration checkerboard to achieve this result.

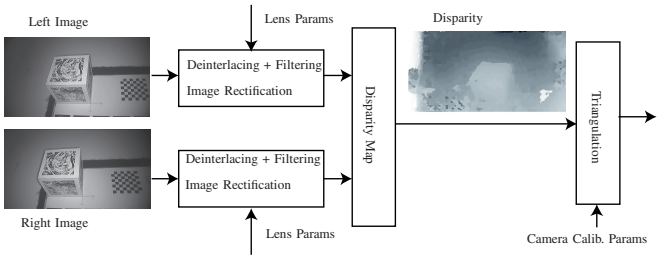


Fig. 1. *3*-D reconstruction framework.

Deinterlacing and Filtering: Interlacing is commonly used in video streaming to enhance motion perception to the viewer. However, it can introduce undesired image artifacts that will negatively alter the quality of the reconstructed anatomical scene. To remove interlacing, a linear interpolation algorithm [14] gives comparatively better results among other deinterlacing algorithms on endoscopic images and hence this algorithm is employed here.

Endoscopic images are also subject to noise, and a noise removal step was adopted here to enhance image quality. We used a Gaussian filter [13] with $\sigma = 3$ which showed encouraging results. In order to have a fair comparison between all the proposed stereo-reconstruction methods, in this work we applied the same image pre-processing before the disparity calculation stage.

Image Undistortion and Rectification: Image undistortion algorithms have been used to correct images for the (radial) lens distortion effects. Rectification uses the camera calibration parameters to align the left and right image to have horizontal epipolar line, so the correspondence search space can be effectively reduced from two dimensions (image-to-image) to a 1-D search (image to epipolar line). In this preliminary study we didn't consider the effect of rectification.

A. Disparity Calculation

A disparity map is a matrix containing the distance (in pixels) from each point in the left to the corresponding point in the right image [3]. Disparity calculation algorithms take a pair of undistorted and rectified images, as well as the camera-calibration parameters, to estimate the disparity.

There exist three different approaches to calculate the disparity map: sparse, dense, and semi-dense methods. Sparse methods only compute disparity at a limited number of image points (feature matches). Dense methods, instead, measure similarity in a sliding window to find the most similar in the other image. Most recent state-of-the-art methods have an additional global optimization combined with block matching [4]. Since semi-dense and sparse stereo reconstruction algorithms do not work well in texture-less environments, we focus on dense disparity algorithms.

Based on the results given in [5], this work will focus on comparing three dense disparity calculation methods: *Stereo Block Matching* (SBM), *Stereo Semi-Global Block Matching* (SSGBM), and *Variational* (SVar) methods. The OpenCV implementations of these methods are available at [6].

SBM is a real-time algorithm (can process a 1920×1080 in milliseconds) that uses moving average sliding-window correlation and then finds extrema among different windows. First of all, it creates a feature-image from left and right pairs and applies correlation to find disparity to approximate Laplacian of Gaussian method [7].

SSGBM [8] is based on the idea of pixel-wise matching of mutual information and approximating a global, 2-D smoothness constraint by combining 1-D constraints. A constraint is added to support smoothness by penalizing changes of neighboring disparities. Minimization is narrowed to minimizing cost function along a single row of an image. After calculating disparity, several steps are applied to refine the map (such as removing peaks and Intensity consistent disparity selection) [8].

SVar is based on a variational method that uses a combination of multilevel adaptive technique and multi-grid approach to achieve a real-time performance. This method adapts the regularizer based on the current state thus improving convergence speed during optimization [9].

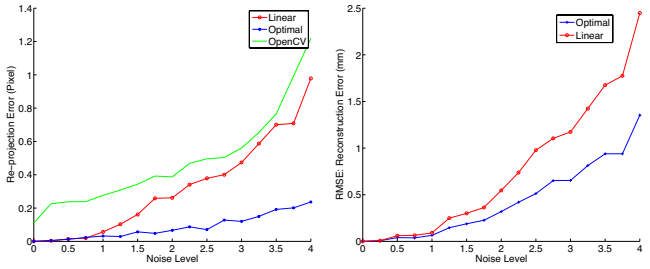
B. Triangulation

Triangulation estimates the position of a 3-D point by reprojecting the corresponding pixels points found in the left and the right images in the 3-D space [10]. During the past years, several triangulation algorithms have been proposed. Among all of them, in this paper we present a comparison between the most popular: the *OpenCV method*, the *Linear algorithm* [10], and the *Optimal-triangulation method* [10].

While the OpenCV method is fast (it only uses a matrix-vector multiplication), the accuracy is one of its major drawbacks. In fact, the OpenCV method assumes that the left and right cameras are collinear and that the camera intrinsic parameters are the same for both cameras, which is not true in practice.

On the other hand, the Linear method [10] does not impose any restrictive assumption about the camera calibration and relative cameras' pose, and uses two linear constraints for each corresponding pair of pixels. SVD decomposition is used to solve these linear equations [6].

The Optimal triangulation has a preliminary phase that tries to minimize the pixel noise in the image plane by



(a) (b)

Fig. 2. Simulation Results: Comparison between stereo-triangulation methods (Since the OpenCV triagulation has a high 3-D error, it is not shown in the plots to make other methods comparable). *(a)* Re-projection error for increasing pixel noise level. *(b)* Reconstruction error for increasing pixel noise level.

shifting points towards the epipolar lines. After this phase, a linear algorithm is applied to triangulate points. A detailed description of this method is available in [10].

III. Experiments

In this section we present an extensive comparison between the different combinations of disparity and stereo triangulation methods described in Sect. II. The goal is to compare both accuracy and speed for each individual steps and of the overall reconstruction pipeline. For this purpose, we designed three different set of tests. The first test is a simulation in MATLAB to assess the triangulation error by assuming full knowledge of the calibration and of the corresponding points. The second and third parts consist of using real images from the daVinci stereo endoscope of both an in-lab model as well as of a laparoscopic sequence. This test was useful to determine the overall accuracy of the pipeline under different assumptions (of III).

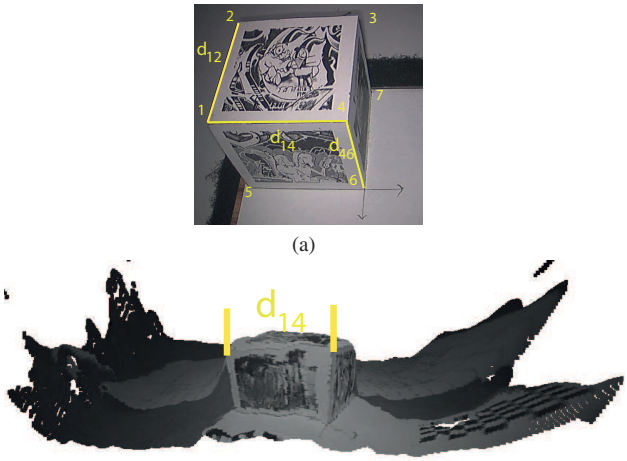
A. Simulation

We assumed no noise in the calibration parameters and used MATLAB to simulate the stereo endoscope, with the same parameters than the real daVinci endoscope. A set of 1000 randomly-generated points were projected onto the image planes, and a pixel noise with an increasing power was added to both correspondences. The error calculated is the distance between the reprojection in the left image of both the ground-truth and each reconstructed point. The results shown in Fig. 2 are an average over 100 iterations for each noise level and the noise level is increased from zero to four pixels. As shown in this graph, error raises by increasing noise level for all methods. However, the Optimal method always outperforms the other methods. The OpenCV method does not consider rotation between cameras and assumes that both cameras have the same intrinsic parameters: as such, it exhibits considerable error even with zero image noise.

B. Real Endoscopic Sequences

In this section, we present the comparison between several 3-D stereo reconstruction methods and evaluate the accuracy of the overall pipeline on real endoscopic images. The aim is to compare the performance of the nine combinations between the three disparity and the three triangulation algorithms (cf., Sect. II).

In-lab model: The first video sequence consists of an accurate model built in lab (cf., Fig. 3(a)) so that we could accurately compare the distances between 3d points (e.g., corners). The ground-truth distances between corners, d^*_{ij}, are known from the model, where i and j are indexes of specific corners. The error $\varepsilon_{ij} = \widehat{d_{ij}} - d^*_{ij}$ is the distance between the estimated (3) and ground-truth distances in mm. The RMSE was calculated according to these error values. In this experiment, we used 10 frames from a video taken



(a)

(b)

Fig. 3. *(a)* Error metrics with in-lab model. *(b)* Reconstructed point cloud.

by the daVinci endoscope. For each image, we selected 30 line-segments with known length, d^*_{ij}. Fig. 4 shows the results of this experiment. As observed, the combination of StereoSGBM with Optimal triangulation gives the best results however, it is not real-time. Confidence level for the mean reconstruction error using this sequence of image pairs is 96%. Time consumption in ms. for each method is listed in the Table under Fig. 4. The values presented represent the time taken to estimate the 3-D point cloud from a pair of endoscopic images (resolution:1920×1080) on a core i7 3GHz laptop.

Laparascopic Sequence: Validation of a surgical-vision system is an important step towards establishing the use of such a system for clinical use. In order to validate the results in a real-world scenario, we used the stereo videos from phantom heart model from the Hamlyn Centre Laparoscopic/Endoscopic Video Dataset [12]. The dataset provides intrinsic, extrinsic and lens distortion parameters as well as the ground-truth point cloud from a CT scan.

We first used the stereo images and the calibration parameter to reconstruct the 3-D point cloud. Second, the two point clouds are aligned using the Iterative Closest Point algorithm implemented by [18]. Third, the ground-truth CT scan was used to calculate the error as the distance between each corresponding points after the alignment. Fig. 5 illustrates the error after registration for SSGBM, SBM and SVar using Optimal triangulation. The box plot shows error for one pair

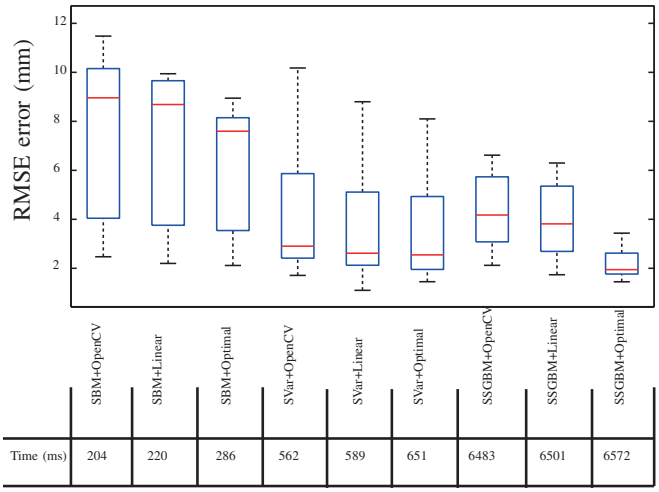


Fig. 4. Reconstruction error and time consumption for 9 different combinations of disparity and triangulation algorithms.

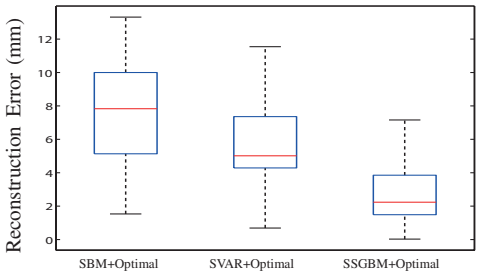


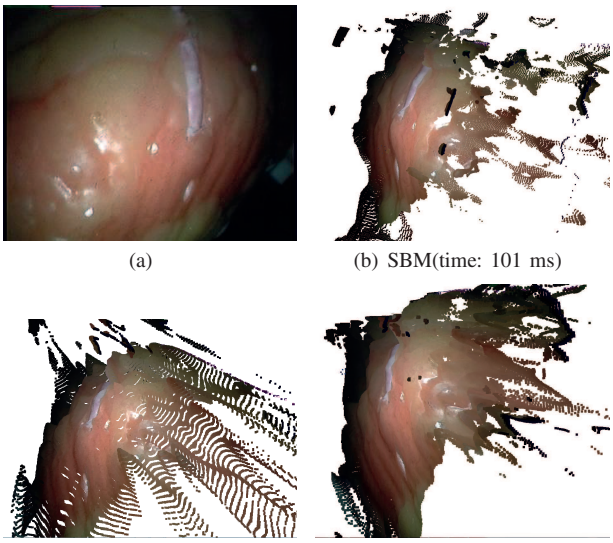
Fig. 5. Reconstruction error for SBM, SSGBM and SVar on heart model.

of images. As shown in this figure, the SSGBM outperforms the other methods and has smaller error.

Fig. 6 illustrates three different disparity methods on a phantom heart model (resolution: 360×288). SSGBM outperforms other methods and has a better visual output compared to SBM and SVar. While SVar and SBM have less accuracy for further objects, they still result in a promising 3-D reconstruction output and in faster computation time.

IV. Discussion and Conclusions

The medical imaging literature lacks a rigorous and extensive comparative study of stereo-reconstruction methods from endoscopic imagery. In this paper, we presented a comparative study to illustrate the performance of each algorithm used in the stereo reconstruction pipeline. We compared 9 different combinations of 3 disparity and stereo-triangulation methods on controlled and real endoscopic images. As our experiments illustrated, there is a trade-off between accuracy and speed. SSGBM along with the Optimal triangulation method has the best performance but, it cannot be used for real-time reconstruction applications. On the other hand, SBM is fast but it is not as accurate as the SSGBM. However the SVar stands somewhere in between them. Therefore, we suggest to use SSGBM for non real-time applications that demand accuracy, and SBM for realtime purposes. In the future, we will compare both GPU implementations of these methods, as well as in real scenario (ex-vivo and in-vivo).



(a) (b) SBM(time: 101 ms) (c) SVar(time: 163 ms) (d) SSGBM(time: 456 ms)

Fig. 6. Comparison between triangulation methods. *(a)* An image from left camera of endoscope [12]. *(b), (c), (d)* Reconstructed point cloud using StereoBM, StereoVar and StereoSGBM respectively, using Optimal triangulation.

References

[1] J.-Y. Bouguet. Camera Calibration Toolbox for MATLAB. $http : //www.vision.caltech.edu/bouguetj/calib_doc/$. Accessed: 2014-02-15.

[2] Z. Zhang. A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell., 22(11):13301334, Nov. 2000.

[3] N. Atzpadin, P. Kauff, and O. Schreer. Stereo analysis by hybrid recursive matching for real-time immersive video conferencing. IEEE Trans. Circuits and Sys. for Video Tech., 14(3):321334, Mar. 2004.

[4] D. Stoyanov. Surgical vision. Annals of Biomedical Engineering, 40(2):332345, 2012.

[5] S. Roehl, S. Bodenstedt, S. Suwelack, H. Kenngott, B. P. Muller-Stich, R. Dillmann, and S. Speidel. Dense gpu-enhanced surface reconstruction from stereo endoscopic images for intraoperative registration. Med. phys., 39(3):16321645, 2012.

[6] G. Bradski. OpenCV library in C++. Dobbs Journ. Soft. Tool,2000.

[7] K. Konolidge. Realtime stereo and motion analysis on passive video images using an efficient image-to-image comparison algorithm requiring minimal buffering, 2007. US Patent 7,194,126.

[8] H. Hirschmuller. Stereo processing by semiglobal matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(2):328341, 2008.

[9] S. Kosov, T. Thormählen, and H.P. Seidel, Accurate real-time disparity estimation with variational methods, ISVC, 2009.

[10] R. Hartley and A. Zisserman. Multiple view geometry in computer vision. Cambridge university press, 2003.

[11] Point cloud library. $http : //pointclouds.org$. Accessed: Feb. 2014.

[12] S. Giannarou, M. Visentini-Scarzanella, .G-Z Yang, "Probabilistic Tracking of Affine-Invariant Anisotropic Regions," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 99, 2012.

[13] D.A. Forsyth and J. Ponce, Computer Vision: A Modern Approach, Prentice Hall Professional Technical Reference, 2002.

[14] L. Chulhee, and J. Lee. "Deinterlacing with motion adaptive vertical temporal filtering." IEEE Tran. on Cons. Electr., 55(2): 636-643, 2009.

[15] P., Musialski, P. Wonka, D. Aliaga, and M. Wimmer, L. Gool, W. Purgathofer, A survey of urban reconstruction, Computer Graphics Forum, 32(6):146-177, 2013.

[16] G. Guidi, M. Russo, D. Angheleddu, 3D survey and virtual reconstruction of archeological sites, Digital App. in Archaeology and Cultural Heritage, Elsevier, 2014.

[17] M. Parchami, J.A. Caddedu, G.L. Mariottini, A Comparative Study on 3-D Stereo Reconstruction from Endoscopic Images, PETRA, 2014.

[18] D.J. Kroon, Iterative Closest Point using finite difference optimization to register 3D point clouds affine, 2009.