

Recovering 3D information of human soft tissue using stereo endoscopic images

Ziyang Chen¹, Davide Alberti¹, Aldo Marzullo², Giancarlo Ferrigno¹, and Elena De Momi¹

¹Department of Electronics, Information and Bioengineering, Politecnico di Milano

²Department of Mathematics and Computer Science, University of Calabria

INTRODUCTION

Robot assisted minimally invasive surgery (RAMIS) has been widely applied in various clinical treatment, and da Vinci surgical system is the most typical representative because of its special advantages, like hand-eye coordination and 3D vision. However, the movement of the stereo endoscope inside human body is limited. Augmented reality (AR) is considered to be integrated into RAMIS, since it can provide more visualization. Recovering 3D information of surgical scene directly affects the performance of AR. Many researchers have implemented 3D reconstruction based on stereo images, such as the semi-global block matching (SGBM) [1]. It has high disparity search efficiency for common stereo images, while the effect in medical field remains to be explored. Hence, an enhanced semi-global block matching approach with preprocessing (P-SGBM) is developed to recover the 3D information of endoscopic images in this paper.

METHODOLOGY

Stereo vision based 3D reconstruction can be divided into several parts, including stereo camera calibration, image rectification, stereo correspondence and depth mapping, as shown in Fig.1. In particular, stereo correspondence and depth mapping are crucial for recovering the space information, and they are implemented by means of the following steps,

1) Image Preprocessing To enhance the model performance and robustness, image preprocessing is a necessary way before the disparity generation. Here, Gaussian filtering and image equalization are combined to perform the image preprocessing, since it improves the quality on medical images because of removing noise and enlarging global contrast.

2) Matching cost computation The left and right pixels in the scanning line of processed images are matched horizontally, and BT cost [2] is adopted to measure their dissimilarity based on the linear intensity interpolation,

$$\hat{C}(p_k, q_k, I_L, I_R) = \min_{q_k - \frac{1}{2} \leq q \leq q_k + \frac{1}{2}} |I_L(p_k) - \tilde{I}_R(q)| \quad (1)$$

Where p_k, q_k denote the pixels to be measured in stereo images. I_L is the intensity function of left scanline, while \tilde{I}_R is the interpolated function of right scanline. The matching cost with disparity value d_k can be defined as,

$$C(p_k, d_k) = \min \{ \hat{C}(p_k, q_k, I_L, I_R), \hat{C}(q_k, p_k, I_R, I_L) \} \quad (2)$$

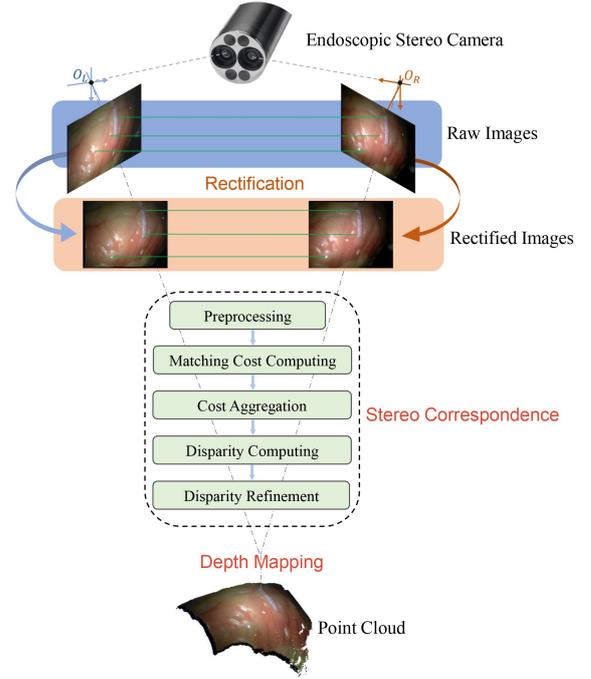


Fig. 1: The reconstruction process using stereo images.

3) Cost aggregation Considering the possibility of pixel mismatching, the smoothness of neighboring disparity is also regarded as a constraint. The cost S'_{r_l} is defined recursively as [1],

$$S'_{r_l}(p_k, d_k) = C(p_k, d_k) + \min(S'_{r_l}(p_k - r_l, d_k), S'_{r_l}(p_k - r_l, d_k \pm 1) + L_1, \min_{n \in [d_1, d_2]} S'_{r_l}(p_k - r_l, n) + L_2) - \min_{i \in [d_1, d_2]} S'_{r_l}(p_k - r_l, i) \quad (3)$$

Where $p_k - r_l$ denotes the left pixel near p_k , and L_1, L_2 are the different penalty factors. n represents the remaining values within the specific disparity range, while i is potential values with the same range. Then, the cost values in different directions are summed to calculate the final matching cost S_r . Empirically, we set L_1, L_2 as 144, 576 respectively, d_1 is -32 and d_2 is 96.

4) Disparity computation The final disparity map consists of the optimal disparity value corresponding to each pixel. Winner Takes All is adopted to extract the disparity value d_k by minimizing the aggregate cost S_r . Fig. 2 presents the searching process of optimal disparity.

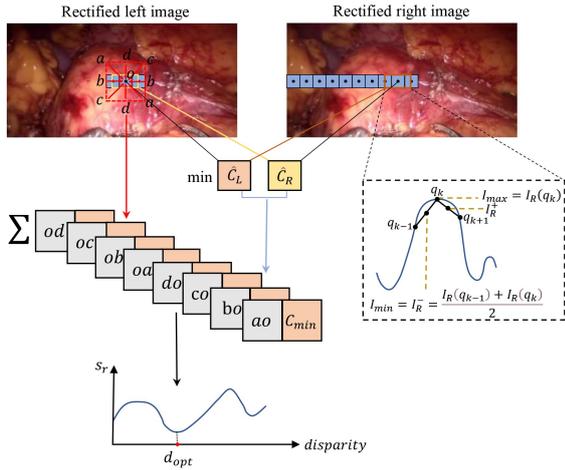


Fig. 2: Schematic diagram of disparity generation.

5) Disparity refinement By following [1], three optimization methods are utilized to improve the quality: (i) Uniqueness test is implemented to delete error values to avoid falling into local optimal solution; (ii) Subpixel interpolation aims to enhance the smoothness of disparity map; (iii) Left-right consistency check can eliminate disparity errors caused by possible object occlusion.

6) Depth Matching Following the camera projection transformation, the depth value z is denoted as,

$$Z = f * b / (p_k - q_k) \quad (4)$$

Where f is the camera focus, and b represents the baseline horizontal distance of the stereo camera. Lastly, global point cloud traversal is applied to remove potential singular points and black background points.

EXPERIMENTAL EVALUATION

To evaluate the practical performance of the proposed approach, we used the laparoscopic images from EndoAbs dataset [3]. Also, three different metrics [4] were adopted to evaluate the reconstruction performance,

1. *Accuracy*. The average Euclidean distance between the reconstructed point cloud and the ground truth.

2. *Computing Time*. Time from image preprocessing to point cloud generation.

3. *The percentage of matched pixels*. Ratio of reconstruction points of interest to the total pixels in the image.

Fig. 3 presents the preprocessed images of different organs and the final generated point clouds. Then, the quantitative evaluation is completed to verify the model performance, and the corresponding box plot is shown in Fig. 4. It can be seen that P-SGBM can have an apparent improvement in accuracy and save little computing time, although the percentage of matched pixels decreases slightly. In the meanwhile, the Wilcoxon rank-sum test ($p < 0.05$) is performed, and it shows that there are significant differences in Accuracy and Computing Time.

CONCLUSION

This paper proposes a 3D reconstruction approach named P-SGBM, and the experimental result shows that it

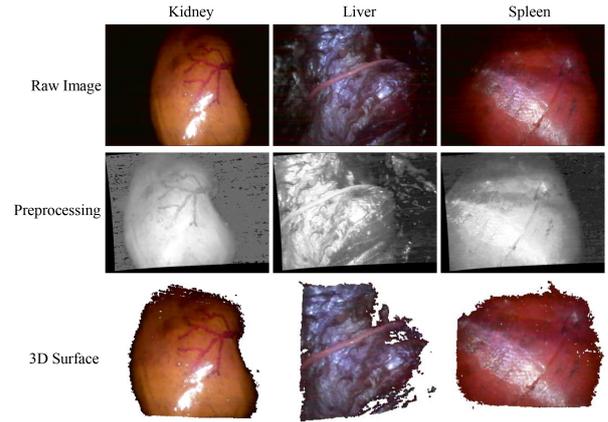


Fig. 3: Qualitative evaluation on Kidney, Liver and Spleen.

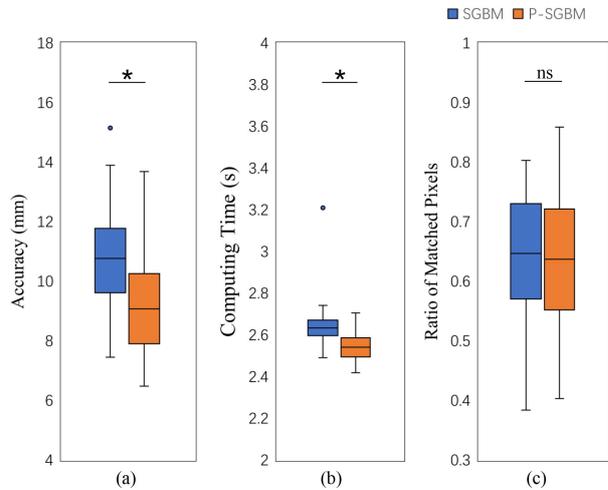


Fig. 4: Box plot representation of the quantitative performance with SGBM [1] and P-SGBM. (a)-(c) presents the result of accuracy, computing time and the percentage of matched pixels sequentially.

can improve the performance of accuracy and computing time compared with the traditional one. In the future, deep learning based approaches [5] will be explored to develop a real-time implementation.

REFERENCES

- [1] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 30, no. 2, pp. 328–341, 2007.
- [2] S. Birchfield and C. Tomasi, "Depth discontinuities by pixel-to-pixel stereo," *International Journal of Computer Vision*, vol. 35, no. 3, pp. 269–293, 1999.
- [3] V. Penza, A. S. Ciullo, S. Moccia, L. S. Mattos, and E. De Momi, "Endoabs dataset: endoscopic abdominal stereo image dataset for benchmarking 3d stereo reconstruction algorithms," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 14, no. 5, p. e1926, 2018.
- [4] V. Penza, J. Ortiz, L. S. Mattos, A. Forgione, and E. De Momi, "Dense soft tissue 3d reconstruction refined with super-pixel segmentation for robotic abdominal surgery," *International journal of computer assisted radiology and surgery*, vol. 11, no. 2, pp. 197–206, 2016.
- [5] G. Yang, J. Manela, M. Happold, and D. Ramanan, "Hierarchical deep stereo matching on high-resolution images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5515–5524.