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

Ziyang Chen[1], Davide Alberti[1], Aldo Marzullo[2], Giancarlo Ferrigno[1], and Elena De Momi[1]

[1] *Department of Electronics, Information and Bioengineering, Politecnico di Milano*

[2] *Department 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,

$$\hat{C}\left(p_k, q_k, I_{\mathrm{L}}, I_{\mathrm{R}}\right) = \min_{q_k-\frac{1}{2} \le q \le q_k+\frac{1}{2}} \left| I_{\mathrm{L}}\left(p_k\right) - \widetilde{I}_{\mathrm{R}}(q) \right| \tag{1}$$

Where p_k, q_k denote the pixels to be measured in stereo images. I_{L} is the intensity function of left scanline, while $\widetilde{I}_{\mathrm{R}}$ is the interpolated function of right scanline. The matching cost with disparity value d_k can be defined as,

$$C\left(p_k, d_k\right) = \min\left\{\hat{C}\left(p_k, q_k, I_{\mathrm{L}}, I_{\mathrm{R}}\right), \hat{C}\left(q_k, p_k, I_{\mathrm{R}}, I_{\mathrm{L}}\right)\right\} \tag{2}$$

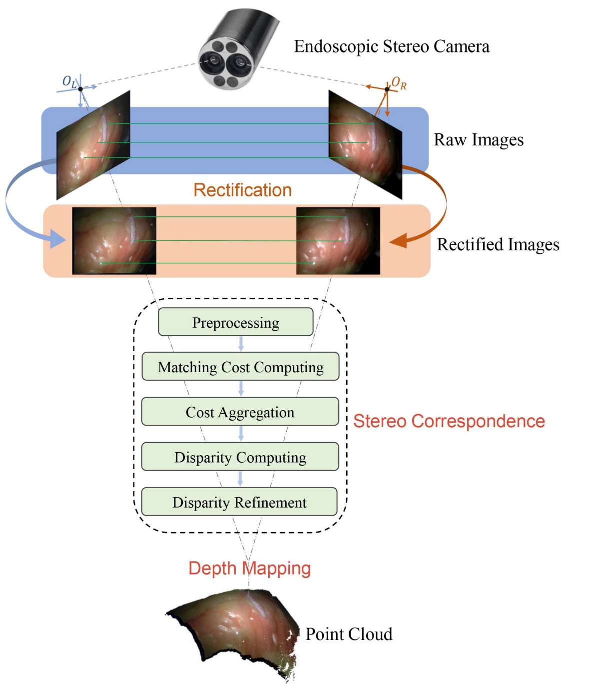

Fig. 1: The reconstruction process using stereo images.

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],

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

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.