H-047

Endoscope based gastric 3D reconstruction and surface mosaic

Mengyao Duan† Rong Xu† Jun Ohya†

1. Introduction

Gastric endoscope is a very essential tool used for examining any gastric abnormalities or growths. It is a flexible fiber-optic scope with a light that helps a physician see inside certain internal organs. Since endoscopic microsurgery could observe the internal structure via a minor wound, which can also reduce complications and shorten the recovery time of patients.

Though endoscope based microsurgery has many outstanding advantages, it has some weak points. 1. It has difficulty to obtain the precise location information and the 3D structure information of lesions. 2. Gastric endoscope only provides a narrow view angle of the internal organs.

Structure-from-motion (SFM) [1] is a significant paradigm of 3D reconstruction in various conditions, which may be coupled with local motion signals. The motion trajectory of the camera could be estimated straightforwardly, and also reconstructs the 3D coordinates of feature points. However, the limitation is that it performs robust only within a certain limit of view angle and rotation. Though it does not have much influence on small object or general 3D reconstruction, if on the occasion of large object or interior scenes 3D reconstruction, it may fail to reconstruct whole structure or perform in high accuracy. Therefore, in order to achieve 3D reconstruction in case of wide-angle or large object, 3D point cloud stitching is an intuitive and essential process.

In order to expand the view angle and describe the details of internal organ structures, this research aims at achieving dense 3D reconstruction from a video sequence acquired by a moving endoscope.

2. SIFT Based 3D Point Clouds Stitching for SFM

SFM is an efficient method to obtain the motion of camera without calibration beforehand. Since SIFT features utilized by SFM is reliable and allows wide baseline matching with much rotation, scaling, etc., we propose a method of 3D point cloud stitching for SFM reconstruction based on SIFT feature.

Since SFM reconstruction method is mainly used for image sequences and SIFT algorithm is utilized for feature extraction, according to the basic principle described above, the process of the proposed method is as follows:

(1) We assume three frames to be reconstructed. First, we divide them into two groups, and each group contains two successive frames. Group1 contains frame1 and frame2, group2 contains frame2 and frame3, so that frame2 is shared by two groups.

(2) Find the matching pairs of feature points in group1 and group2 independently by SIFT algorithm, and then record the coordinates of the matched features in the overlapping frame (frame2) in group1 and group2 independently.

(3) Self-calibrate the camera to calculate the intrinsic and extrinsic parameters, so that the camera motion is obtained, and then perform 3D reconstruction of each group by the principle of the epipolar equation.

(4) Find all the 3 pairs of corresponding feature points from frame2 between two groups; here we assume that there are n pairs of corresponding feature points.

(5) We match each two pairs of points then make the 3 pairs of points coplanar in the same 3D world coordinate system. Each three pairs of feature points from group1 $(\mathbf{A_1}, \mathbf{A_2}, \mathbf{A_3})$ and group2 $(\mathbf{B_1}, \mathbf{B_2}, \mathbf{B_3})$ may define two planes: palne1 and plane2. Let $\mathbf{A_1}$ be the original of XYZ axes, and $\mathbf{A_1 A_2}$ be Y-axis, $\mathbf{A_1}, \mathbf{A_2}$ and $\mathbf{A_3}$ form YZ-plane. Similarly, $(\mathbf{B_1}, \mathbf{B_2}, \mathbf{B_3})$ are defined for group2. Then we zoom $\mathbf{B_1\,B_2}$ to the same scale of $\mathbf{A_1 A_2}$ by the following equation:

$$\mathbf{B}'_{(\mathbf{x,y,z})} = \mathbf{B}_{(\mathbf{x,y,z})} \cdot \mathbf{M_{sc}} \tag{1}$$

where $\mathbf{M_{sc}}$ is the scaling matrix of the transformation from line $\mathbf{B_1\,B_2}$ to line $\mathbf{A_1 A_2}$. Then we match these two lines by the following equation, so that the two lines are collinear.

$$\mathbf{B}''_{(\mathbf{x,y,z})} = \mathbf{B}'_{(\mathbf{x,y,z})} \cdot \mathbf{M_{tran}} \cdot \mathbf{M_{rot1}} \cdot \mathbf{M_{rot2}} \tag{2}$$

where $\mathbf{M_{tran}}$ is the transformation matrix from $\mathbf{B_1}$ to $\mathbf{A_1}$, $\mathbf{M_{rot1}}$ and $\mathbf{M_{rot2}}$ are the rotation matrices about Z-axis and Y-axis, and then we rotate $\mathbf{B_3}$ by the following equation, until the two groups are coplanar.

$$\mathbf{B}'''_{(\mathbf{x,y,z})} = \mathbf{B}''_{(\mathbf{x,y,z})} \cdot \mathbf{M_{rot3}} \tag{3}$$

where $\mathbf{M_{rot3}}$ is the rotation matrix about Y-axis.

(6) The transformation matrix of each of 3 pairs of corresponding feature points is represented by:

$$\mathbf{M_{(i)}} = \mathbf{M_{sc}} \cdot \mathbf{M_{tran}} \cdot \mathbf{M_{rot1}} \cdot \mathbf{M_{rot2}} \cdot \mathbf{M_{rot3}} \quad \mathbf{(0<i<=n)} \tag{4}$$

(7) Then repeat step (5) and step (6), until all the 3 pairs of corresponding feature points are matched for the transformation matrices.

(8) Calculate the total Euclidean distance between the corresponding feature points in group 1 and group 2 respectively by different transformation matrix. Finally, the least squares method is utilized to estimate the transformation matrix $\mathbf{M_{(j)}}\mathbf{(0<j<=n)}$ based on two groups of 3D point cloud.

3. Dense Point Cloud Generation

Since the reconstruction by SFM is done only for the feature points, it suffers from lack of maps and textures. In addition, the amount of feature points is limited in quantity, which may result in sparse 3D reconstruction. Therefore, a method to obtain a dense 3D reconstruction result is necessary. However, in any case to obtain dense reconstruction, the lack of feature points should be dealt with first. So we propose a method for generating dense point clouds.

3.1 Region-growing Algorithm for Stereo Matching

In this research, since we utilize SIFT algorithm to extract features for SFM based 3D reconstruction and 3D point cloud stitching, here we also utilize robust SIFT feature points as seed points. In this process, if the number of seed points is more than one point, then the program is continued.

If we iterate for all of the pixels in the images, it must be quite

† Graduate School of Global Information and Telecommunication Studies, Waseda University

time consuming and redundant. In fact, in order to avoid reconstructing too much dense 3D points in a certain area, we divide the reference image by 3×3 block, and project the 3D points to the reference image, and if there are more than one point in each block, keep it as the seed point, and others are removed. After we select all the seed points, we consider the points in the center of each 3×3 block as the points to be matched, which is in the four-neighborhood of the seed points' block. Then, the depth of each point to be matched is obtained by the depth testing method described in the next section.

3.2 Depth Testing Method by Furukawa [2]

In this section, we compute the depth of each point to be matched. Here we utilize depth testing method. The process to determine the depth of one point to be matched is as follows:

We assume that one seed point is projected to $\mathbf{M_1}$ in image1 and $\mathbf{N_1}$ in image2, and we consider image1 as the reference image. $\mathbf{M_2}$ is one point to be matched. Figure 1 shows the process of depth testing method by Furukawa.

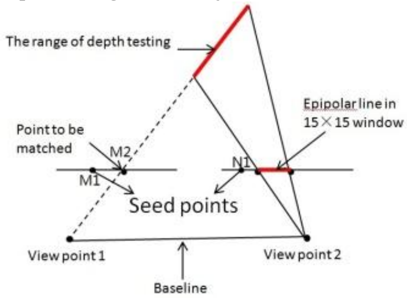

Fig. 1 The process of depth testing method

(1) We calculate the epipolar line $\mathbf{L}$ that passes through $\mathbf{M_2}$ in image2, and we consider $\mathbf{N_1}$ as the center of a 15×15 window that line $\mathbf{L}$ intersects the window at point $\mathbf{C_1}$ and $\mathbf{C_2}$.
(2) Reconstruct the point $\mathbf{N_1}$ with the point $\mathbf{C_1}$ and $\mathbf{C_2}$ respectively, so that we can obtain two 3D points $\mathbf{D_1}$ and $\mathbf{D_2}$. So the depth range of $\mathbf{N_1}$ is $[\mathbf{min}(\mathbf{D_1}, \mathbf{D_2}), \mathbf{max}\,(\mathbf{D_1}, \mathbf{D_2})]$.
(3) Project the points in the range of depth to image2, and then find the point whose feature is the most similar to $\mathbf{N_1}$.
(4) If the distance between the new point and the seed point is in a certain value range, keep it, otherwise remove it.
Regard the new point as the seed point, and repeat the step 1-4, until no new 3D point to be created.

In the end, in order to show the result more intuitively, we obtain the RGB information of each 3D dense point from the 2D images, and then directly show the points having a size expanded by three times from their original sizes.

4. Experiments and Results

In the experiment, we reconstruct the gastric wall. Here we captured an image sequence, and obtained three consequent images for reconstruction and stitching. We divide the images into two groups, where group1 contains image1 and image2, group2 contains image2 and image3, as shown in Figure 2.

Fig. 2 Two groups to be reconstructed and stitched

Then we reconstruct the images of these five groups by SFM. Next, we realize 3D point stitching for each two successive groups by our proposed method. Here we stitch each adjacent two groups, such as group1 and group2; group2 and group3, respectively. Figure 3 shows the results of the sparse 3D point cloud before stitching and after stitching.

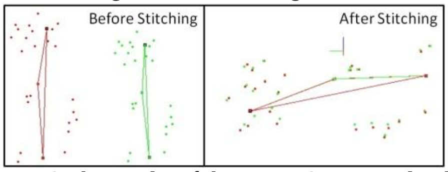

Fig. 3 The results of the sparse 3D point cloud

After the stitching of the sparse 3D point cloud, we generate the dense point cloud by utilizing region-growing algorithm and depth testing method proposed by Furukawa. Here we first generate the dense point clouds from the three results after stitching shown in Figure 3, where the dense point clouds of the stitching result of group1 and group2, as shown in Figure 4

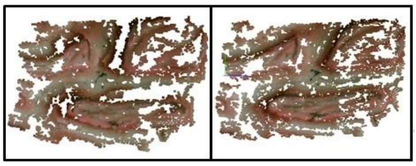
Fig. 4 The result of dense 3D point cloud after stitching

Finally, we reconstruct the continuous surface from the uniform dense point cloud by the method of Poisson surface reconstruction [3], as shown in Figure 5.

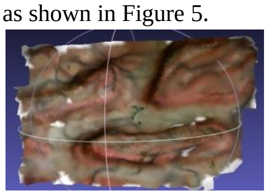
Fig. 5 The final result of Poisson surface reconstruction

5. Conclusion and Future Work

In this study, we have proposed a method for SFM-based 3D reconstruction and surface mosaic. SIFT-based feature extraction is utilized in three major steps: 3D reconstruction, 3D point cloud stitching and dense point cloud generation. The experimental results presented in Chapter 6 demonstrate that the proposed method achieves a high accuracy. The basic 3D shapes of the model and colors are reconstructed precisely.

Though SIFT algorithm is robust, it is time consuming and does not work well for non-rigid objects. In addition, some algorithm without feature extraction will be introduced to deal with 3D reconstruction in a high accuracy and efficiency. Therefore, the future work is to find a better method for feature extraction, or a method that is not based on features. The accuracy of 3D point cloud stitching will also be improved.

Reference

[1] Richard Hartley and Andrew Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2003
[2] Yasutaka Furukawa and Jean Ponce, "Accurate, Dense, and Robust Multiview Stereopsis", *IEEE Trans. Pattern Anal. Mach. Intell. 32(8)*: 1362-1376, 2010
[3] Michael Kazhdan, Matthew Bolitho and Hugues Hoppe, "Poisson surface reconstruction", *Symposium on Geometry Processing*, 2006