Three-Dimensional Endoscopy

Thorsten Thormählen, Hellward Broszio, Peter N. Meier ⋆

Information Technology Laboratory, University of Hannover, Germany
email: thormae@tnt.uni-hannover.de, broszio@tnt.uni-hannover.de
⋆ Dept. of Gastroenterology and Hepatology, Medical School of Hannover, Germany
email: meier.peter@mh-hannover.de

ABSTRACT

This contribution presents a robust estimation technique for computation of camera motion and 3D reconstruction of a scene from an endoscopic image sequence. The technique offers new possibilities for more objective diagnoses and intervention planning. Every ordinary flexible monocular video endoscope can be used without modifications.

1. INTRODUCTION

Diagnosis and intervention in endoscopy is based on image sequences of a video camera. But an image sequence provides no depth information. If the usual stereoscopic human sight is missing, navigation of instruments is difficult. Recognition and evaluation of pathological structures and the estimation of their spatial extension can only be achieved by experience. Diagnosis is very subjective. Measurement of calibrated data for colour and texture and a three-dimensional (3D) reconstruction of pathological structures can make endoscopic diagnosis more objective and reproducible. Therefore this paper presents an approach that allows 3D reconstruction from endoscopic image sequences.

The described approach estimates the camera motion and uses the estimated camera motion to reconstruct the 3D scene. The camera motion must be estimated from the image sequence because it cannot be controlled or accurately measured. In computer vision this problem is called the *structure from motion* problem. Since the structure from motion approach only requires an image sequence as input data an ordinary flexible monocular video endoscope can be used without any modifications.

In most known approaches which generate an endoscopic 3D scene a special endoscope is developed or an additional device is inserted into the working channel of the endoscope. Schubert and Müller [8] presented a structured light method for bronchoscopy which can be used for measurement of hollow biological organs. The 3D scene is estimated by the deformation of a projected ring of laser light. A laser fibre probe is inserted into the instrument channel. Armbruster and Scheffler [1] developed a prototype of an endoscope that measures the 3D scene by deformation of several projected lines. The accuracy of this structured light method is increased by a phase shift of the lines. Kemper et al. [5] have examined the deformation of a 3D scene with an electronic-speckle-pattern interferometer. A stiff minimal inversive endoscope is extended with an external holographic camera. Some endoscope manufactures offer special stereo endoscopes for 3D reconstruction or 3D visualization. Since stereo endoscopes have two optical channels they have a larger diameter than usual endoscopes and are not often used in practice.

Other approaches exist that only evaluate the endoscopic image sequence to reconstruct the 3D scene. Yeung et al. [10] proposed a shape from shading algorithm. Deguchi et al. [2] have used a structure from motion approach for endoscopic image sequences, where camera motion and 3D scene is estimated by the factorization method.

The approach presented here consists of two parts. First the very robust estimation of camera motion. Based on conventional feature point detection and correspondence analysis, the robustness of the estimation is achieved by using a highly efficient outlier detection of misaligned correspondences. The second part consists in the reconstruction of the 3D scene by using the known camera motion.

The next section describes the proposed estimation algorithm for the motion of the camera. Section 3 explains the reconstruction of the 3D scene. Section 4 gives some results. The paper ends with conclusions.

2. ROBUST ESTIMATION OF CAMERA MOTION

The estimation method illustrated in Fig. 1 consists of three processing steps: detection of feature points, determination of correspondences and robust estimation of the camera motion parameters.

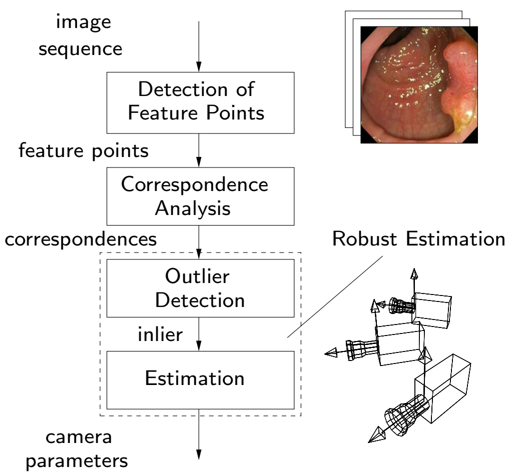

Figure 1: Steps of the proposed algorithm to estimate camera motion parameters.

2.1. Detection of Feature Points

The feature points are detected with subpixel accuracy using the Harris feature point detector [4]. For each image of the sequence a list of feature point coordinates $L = \{\mathbf{p}_1, \ldots, \mathbf{p}_i, \ldots, \mathbf{p}_M\}$ is extracted, where $\mathbf{p}_i = (x_i, y_i)^\top$ is the image coordinate of a feature point.

2.2. Correspondence Analysis

The feature points in list L and L' of two successive images are assigned by measuring normalised cross-correlation between 15×15 windows surrounding the feature points. The correspondences are established for those feature points which have the highest cross-correlation. This results in a list of correspondences $L_c = \{q_1, \ldots, q_i, \ldots, q_N\}$, where $q_i = (\mathbf{p}_i, \mathbf{p}'_i)$ is a correspondence.

2.3. Estimation of Camera Translation

For the estimation of camera motion parameters from corresponding feature points, the real camera must be represented by a mathematical camera model. Required relation between corresponding feature points in two images follows from the motion model.

2.3.1. Camera Model

The perspective camera model with the focal length f describes the projection of a 3D scene coordinate $\mathbf{P} = (X, Y, Z)^\top$ into the image plane (Fig. 2). The projected 2D image point $\mathbf{p} = (x, y)^\top$ can be calculated as follows

$$\mathbf{p} = \begin{pmatrix} x \\ y \end{pmatrix} = \frac{f}{Z} \cdot \begin{pmatrix} X \\ Y \end{pmatrix} \qquad (1)$$

A shift of the optical centre and the third-order radial lense distortions respect to the image centre (barrel and cushion) are also taken into account. For simplification these camera parameters are only mentioned briefly. A detailed camera model can be found in [9].

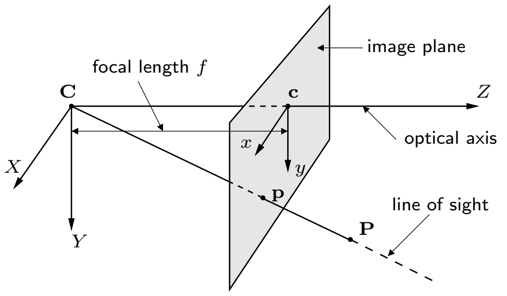

Figure 2: Camera model.

2.3.2. Motion Model

The proposed algorithm to estimate the translation of the camera and the scene geometry is based on a motion model, which is illustrated in Fig. 3. The general camera operation is given by the position of the camera centre $\mathbf{C}' = (C_X, C_Y, C_Z)^\top$ relative to the coordinate frame (X, Y, Z) and the rotation matrix $\mathbf{R}$, which can be specified as three successive rotations $\mathbf{R}_\varphi$, $\mathbf{R}_\vartheta$ and $\mathbf{R}_\rho$ around the Y-, X- and Z-axes, respectively. The rotations are named throughout this paper as φ for pan, ϑ for tilt and ρ for roll.

If $\mathbf{P}$ is a 3D point represented in the camera coordinate frame (X, Y, Z) and $\mathbf{P}'$ represents the same point in the coordinate frame (X', Y', Z') after the camera operation, then the transformation is given by

$$\mathbf{P}' = \mathbf{R}(\mathbf{P} - \mathbf{C}') \qquad (2)$$

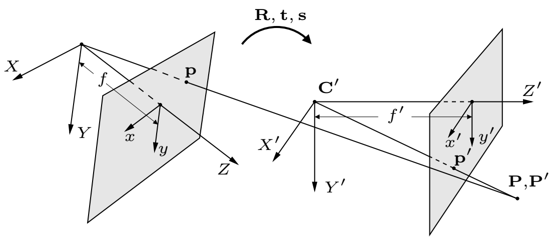

Figure 3: Projection of 3D point $\mathbf{P}$ for a moved camera.

with

$$\begin{aligned}\mathbf{R} &= \mathbf{R}_\rho \mathbf{R}_\vartheta \mathbf{R}_\varphi \\ &= \begin{bmatrix} cos\rho & sin\rho & 0 \\ -sin\rho & cos\rho & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\vartheta & sin\vartheta \\ 0 & -sin\vartheta & cos\vartheta \end{bmatrix} \\ &\quad \begin{bmatrix} cos\varphi & 0 & -sin\varphi \\ 0 & 1 & 0 \\ sin\varphi & 0 & cos\varphi \end{bmatrix} \\ &= \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \end{aligned} \tag{3}$$

It is convenient not to make the camera translation explicit, and instead to represent the transformation in equation (2) with

$$\mathbf{t} = (t_X, t_Y, t_Z)^\top = -\mathbf{R}\ \mathbf{C}' \tag{4}$$

Equation (2) can be rewritten as

$$\mathbf{P}' = \mathbf{R}\mathbf{P} + \mathbf{t} \tag{5}$$

Combining equation (1) and (5) the relationship between the image coordinates $\mathbf{p}_i = (x_i, y_i)^\top$ and $\mathbf{p}'_i = (x'_i, y'_i)^\top$ of a projected 3D point $\mathbf{P}_i = (X_i, Y_i, Z_i)^\top$ can be expressed as

$$\begin{aligned} x'_i &= f' \frac{r_{11}x_i + r_{12}y_i + r_{13}f + f\frac{t_X}{Z_i}}{r_{31}x_i + r_{32}y_i + r_{33}f + f\frac{t_Z}{Z_i}} \\ y'_i &= f' \frac{r_{21}x_i + r_{22}y_i + r_{23}f + f\frac{t_Y}{Z_i}}{r_{31}x_i + r_{32}y_i + r_{33}f + f\frac{t_Z}{Z_i}} \end{aligned} \tag{6}$$

where the focal length after the camera operation is $f' = sf$.

Assuming pure translation and the absence of zoom, which is valid for most endoscopic image sequences, the motion model can ignore rotation and zoom operations. Therefore the equation (6) can be simplified

$$\begin{aligned} x'_i &= \frac{x_i + f\frac{t_X}{Z_i}}{1 + \frac{t_Z}{Z_i}} \\ y'_i &= \frac{y_i + f\frac{t_Y}{Z_i}}{1 + \frac{t_Z}{Z_i}} \end{aligned} \tag{7}$$

Elimination of Z_i between these two equations gives

$$\begin{aligned} (y'_i - y_i)\frac{t_X}{t_Z} - (x'_i - x_i)\frac{t_Y}{t_Z} \\ +(x'_i y_i - x_i y'_i)/f = 0 \end{aligned} \tag{8}$$

With the substitution

$$\frac{t_X}{t_Z} = \tilde{t}_X \qquad \frac{t_Y}{t_Z} = \tilde{t}_Y \tag{9}$$

and two correspondences the following linear equation system can be build

$$\begin{aligned} \begin{bmatrix} (y'_1 - y_1) & (x_1 - x'_1) \\ (y'_2 - y_2) & (x_2 - x'_2) \end{bmatrix} \begin{pmatrix} \tilde{t}_X \\ \tilde{t}_Y \end{pmatrix} = \\ \begin{pmatrix} (x_1 y'_1 - x'_1 y_1)/f \\ (x_2 y'_2 - x'_2 y_2)/f \end{pmatrix} \end{aligned} \tag{10}$$

This equation system can be solved explicitly for $\tilde{t}_X$ and $\tilde{t}_Y$.

The translation of the camera can only be determined up to an arbitrary scale factor k, which leads into the following constraint

$$\sqrt{t_X^2 + t_Y^2 + t_Z^2} = k \tag{11}$$

This constraint implies the third motion parameter t_Z with

$$t_Z = \frac{k}{\sqrt{\tilde{t}_X^2 + \tilde{t}_Y^2 + 1}} \tag{12}$$

2.3.3. Outlier Detection

Due to misalignments, usually some of the correspondences are incorrect. To achieve a robust estimation of camera motion parameters, a random sampling algorithm [3] for outlier detection is employed. Two correspondences are sufficient to calculate the camera motion parameters using equations (10) and (12). Because equations (10) can be solve explicitly this calculation is very fast. Support of this parameter set is measured by evaluating how good the displacements of the other correspondences can be expressed by this parameter set. If it does not match the majority of the correspondences, a new parameter set is chosen randomly until a satisfying solution is found. The correspondences, which support the satisfying solution are called inliers - the others outliers. Only the inliers are applied to estimate the camera motion parameters and 3D scene in the following processing step.

2.3.4. Estimation of Camera Motion

The camera motion parameters are estimated using a least-squares technique applied to all inlier correspondences by minimizing the residual error of the cost function (13) over $\tilde{t}_X$ and $\tilde{t}_Y$.

$$\sum_i \epsilon_i(\tilde{t}_X, \tilde{t}_Y)^2 \rightarrow \min \tag{13}$$

with the residuals

$$\begin{aligned} \epsilon_i(\tilde{t}_X, \tilde{t}_Y) = (y'_i - y_i)\tilde{t}_X - (x'_i - x_i)\tilde{t}_Y \\ + (x'_i y_i - x_i y'_i)/f \end{aligned} \tag{14}$$

The minimum of the cost function can be calculated by linear regression.

3. 3D RECONSTRUCTION

The scheme for automatic reconstruction of the 3D scene from an image sequence is illustrated in Fig. 4. The reconstruction process consists of three steps:

- Calculation of 3D coordinates for each detected inlier feature point giving a cloud of 3D feature points.
- Generation of a 2D triangle mesh in the image plane with inlier corners as vertices, which converts the 3D point cloud into a triangle surface description.
- Final texturing to assign a natural look to the 3D reconstruction.

3.1. Calculation of 3D Coordinates

Known camera motion enables the calculation of 3D point coordinates belonging to each inlier correspondence. The triangulation of two lines of sight from two different cameras gives the 3D coordinate for each correspondence. Due to erroneous detection of feature points, the lines of sight do not intersect in most cases (Fig. 5).

Therefore, a correspondence of two 3D points $\tilde{\mathbf{P}}_i$ and $\tilde{\mathbf{P}}'_i$ can be determined for each feature point separately. The 3D points are located where the lines of sight have their smallest distance. The arithmetic mean of $\tilde{\mathbf{P}}_i$ and $\tilde{\mathbf{P}}'_i$ gives final 3D coordinate $\mathbf{P}_i$.

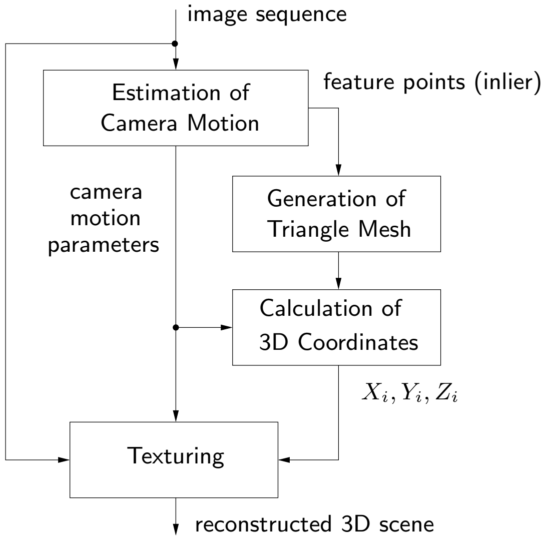

Figure 4: Scheme for reconstruction of the 3D scene.

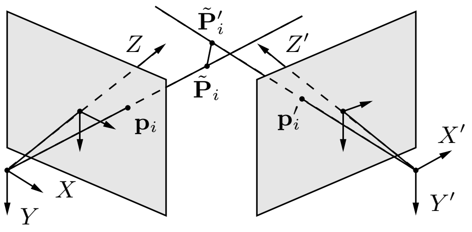

Figure 5: Triangulation of lines of sight for correspondences.

3.2. Generation of a Triangle Mesh

The shape representation of the 3D reconstruction by a textured triangle mesh is common practice. For recovering the shape, the cloud of 3D points has to be converted into a triangle mesh. Diverse possibilities to triangulate a 3D point cloud make the recovering of the shape problematic. Therefore, an approach is used to triangulate the feature points in the image plane by using Delaunay triangulation [6]. Depending on the 2D triangle meshes in the image planes the 3D points are connected to a 3D triangle mesh (Fig. 6).

The characteristics of the Delaunay triangulation and the known correspondences of feature points between successive images enables

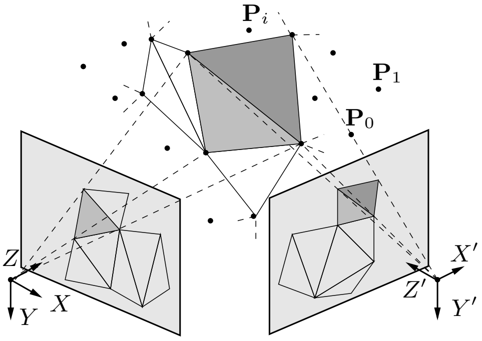

Figure 6: Generation of 2D triangle meshes in the image planes. Connecting 3D points advised by the 2D triangle meshes.

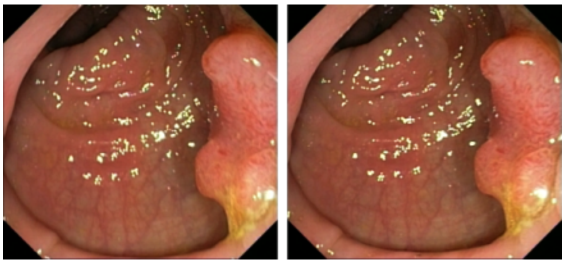

Figure 7: First two images of the sequence "polyp in colon".

the generation of a closed and coherent 3D triangle mesh. The feature points of each image of sequence belonging to inlier correspondences, provide a partition of the 3D triangle mesh.

3.3. Texturing

For a realistic-looking 3D model, a texturing algorithm is used, which estimates the texture for each surface triangle from the endoscopic image sequence. The principle of binding texture information from an image sequence to a single surface triangle was introduced by [7]. The vertices of a surface triangle are projected into the image plane by using the known camera parameters. The projected triangle contains the image part which defines the texture map. Due to many possibilities of images from which the texture information can be taken, each triangle is assigned to that image of the sequence, which ensures the highest texture resolution defined by the ratio of texture elements per surface area.

4. RESULTS

The method has been applied to an image sequence that was recorded on a real intervention.

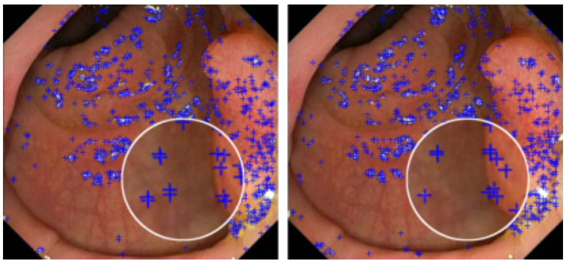

Figure 8: Extracted feature points from the first two images of the sequence.

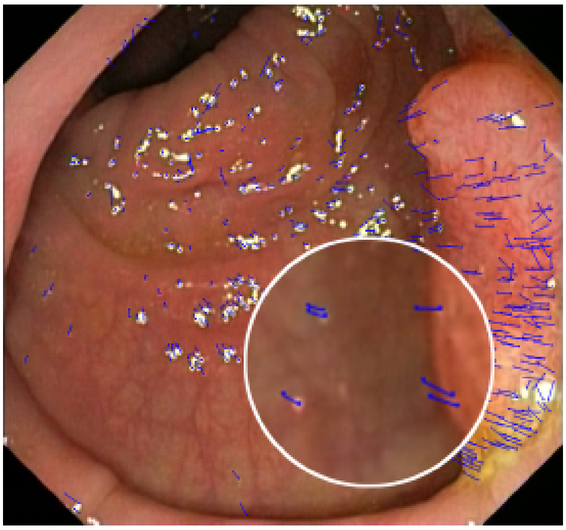

Figure 9: Displacement vectors of corresponding corners between the first two images.

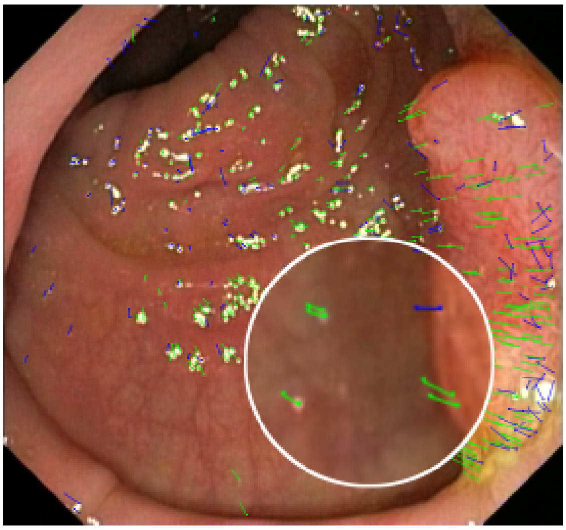

Figure 10: Displacement vectors of the remaining correspondences after outlier elimination (inlier = green displacement, outlier = blue displacement).

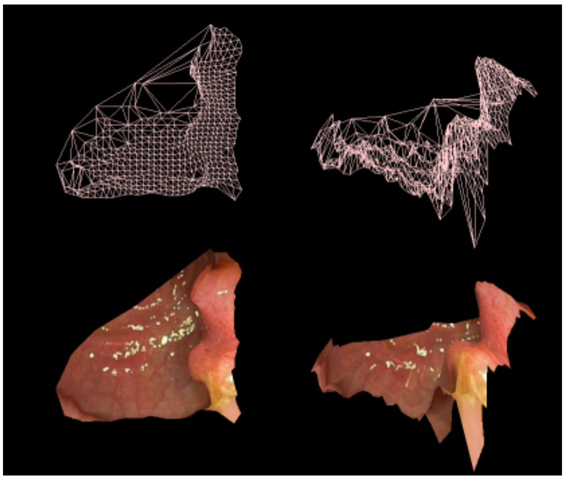

Figure 11: Reconstructed 3D surface model “polyp in colon” from two different views with and without textures.

In the image sequence an polyp in the colon can be seen (Fig. 7). Fig. 8 shows the result of feature point detection for the first two images of the sequence. The feature points detector finds 930 features points in the first image and 898 in the second. With this feature point set 528 correspondences are calculated (Fig. 9). From these, the applied random sampling algorithm selects 348 supported correspondences (Fig. 10), that are used to estimate camera motion. With the estimated camera motion a 3D reconstruction of the polyp is performed. Fig. 11 exhibits the reconstructed 3D scene from two different views.

5. CONCLUSIONS

A robust estimation technique for computation of camera translation was presented. Based on conventional feature detection and correspondence analysis, the robustness of the approach is achieved by the use of a highly efficient outlier detection of misaligned correspondences. The camera translation estimator is based on a very fast linear regression technique that computes accurate parameters.

These accurate parameters are used to reconstruct the observed 3D scene. Though the reconstruction method is still very basic the potential of the approach is obvious. In future work the accuracy of the method should be improved and the estimation error should be analyzed.

Visualization of the reconstructed 3D scene offer new possibilities for diagnosis and intervention planning. Furthermore it will be possible to visualize a virtual endoscope in the 3D scene, which informs the physician about the current position of the endoscope. This 3D representation of an operation can support the navigation or can help to teach medical students in an illustrative manner.

6. REFERENCES

[1] Armbruster K, Scheffler M. Messendes 3D-Endoskop. horizonte. 12; p 15-6; January 1998.

[2] Deguchi K, Sasano T, Arai H, Yoshikawa H. 3-D Shape Reconstruction from Endoscope Image Sequences by the Factorization Method. IEICE Transactions on Information and Systems. Vol. E79-D; No. 9; p 1329-36; 1996.

[3] Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. Communications of the Association for Computing Machinery (ACM). Vol. 24; p 381-95; 1981.

[4] Harris C, Stephens M. A combined corner and edge detector. Proceedings of 4th Alvey Vision Conference. p 147-51; 1988.

[5] Kemper B, Merker A, Lai A, von Bally G. A Novel Electronic-Speckle-Pattern Interferometer (ESPI) for Dynamic Holographic Endoscopy. Fifth International Conference on Optics Within Life Sciences Volume V, Optics and Lasers in Biomedicine and Culture. 1998; Heraklion, Crete, Greece; Springer 1999.

[6] Lischinski D. Incremental Delaunay Triangulation. In: Heckbert PS, editor. Graphics Gems IV. p 43-59; Academic Press; 1994.

[7] Niem W, Broszio H. Mapping texture from multiple camera views onto 3D-object models for computer animation. International Workshop on Stereoscopic and Three-Dimensional Imaging. Santorini, Greece; September 6-8; 1995.

[8] Schubert M, Müller A. Evaluation of endoscopic images for 3-dimensional measurement of hollow organs. Biomedical Engineering. 43(1); p 32-3; 1998.

[9] Tsai RY. A Versatile Camera Calibration Technique For High-accuracy 3-D Machine Vision Metrology Using Off-the-Shelf Cameras and Lenses. IEEE Transactions on Robotics and Automation. Vol. 3; 1987.

[10] Yeung SY, Tsui HT, Yim A. Global Shape from Shading for an Endoscope Image. Medical Image and Computer Assisted Intervention, Second International Conference Proceedings MICCAI’99. p 328-32; 1999.