RECTIFICATION AND ROBUST MATCHING USING ORIENTED IMAGE TRIPLETS FOR MINIMALLY INVASIVE SURGERY

N. Conen [a], C. Jepping [a], T. Luhmann [a], H.-G. Maas [b]

[a] Jade University of Applied Sciences, Institute for Applied Photogrammetry and Geoinformatics (IAPG),
Ofener Str. 16/19, 26121 Oldenburg, Germany - { niklas.conen, christian.jepping, luhmann }@jade-hs.de
[b] Technische Universität Dresden, Institute of Photogrammetry and Remote Sensing,
01062 Dresden, Germany - hans-gerd.maas@tu-dresden.de

Commission III, WG III/1

KEY WORDS: Image triplet, rectification, epipolar lines, semi-global optimisation, reliability, real-time, endoscopy

ABSTRACT:

Stereo endoscopes for minimally invasive surgery have been available on the market for several years and are well established in some areas. In practice, they offer a stereoscopic view to the surgeon but are not yet intended for 3D measurements. However, using current knowledge about the camera system and the difficult conditions in object space, it is possible to reconstruct a highly accurate surface model of the current endoscopic view. In particular for medical interventions, a highly reliable point cloud and real-time computation are required. To obtain good reliability, a miniaturised trinocular camera system is introduced that reduces the amount of outliers. To reduce computation time, an approach for generation of rectified image triplets and their corresponding interior and exterior camera parameters has been developed. With these modified and parameterised images it is possible to directly process 3D measurements in object space. Accordingly, an efficient semi-global optimisation is implemented by the authors. In this paper the special camera system, the rectification approach and the applied methodology of matching in rectified image triplets are explained. Finally, first results are presented. In conclusion, the trinocular camera system provides more reliable point clouds than a binocular one, especially for areas with repetitive or poor texture. Currently, the benefit of the third camera is not as great as desired.

1. INTRODUCTION

Endoscopic images are currently only used for visual purposes, but not for measurements in medical applications. Stereo endoscopes offer a real-time 3D view of the treatment area and would thus fulfil requirements for 3D measurement purposes. Surgeons would greatly benefit from such a support during a medical intervention.

One task herein is the metrically correct surface reconstruction of the tissue that should be treated. This task includes many difficulties due to complex conditions in object space:

- The tissue is soft and deforms continuously
- The surface is usually not well textured
- The scene can contain specular highlights on the surface, which are caused by liquids
- There may be smoke present in the scene, which is caused by special surgical procedures
- The relative position between the endoscope and the surface is not fixed
- There are moving obstacles like surgical instruments which cause occlusions

The algorithm for surface reconstruction should perform in real-time to allow continuous observations of the permanently changing treatment area, it has to be robust against low structured textures and reflections, and it should be able to deal with obstacles. In combination with the position of surgical instruments, it is possible to constantly measure the distance between the surface and the instrument tip, as well as the penetration of the instrument into the tissue. The instrument position can be tracked by attached target points and an external navigation system, or by stereo images of the endoscope, or both.

Once the surface is reconstructed, it can be registered to other 3D models e.g. from pre-operative MRT data. With the pre-operative data, the surgeon is able to plan the entire intervention beforehand and select e.g. areas of risk, healthy tissue and treatment areas. If this model is properly superimposed to the real situation during intervention, the surgeon would greatly benefit because it is hardly possible to distinguish exactly between these areas in real endoscopic images.

Besides, specific measurement tasks must be related to specific surgeries due to a variety of different methods of treatment, which use different kinds of endoscopes and instruments. Hence, the size of working area varies too. Endoscopes can be flexible or rigid and can have different diameters. The presented investigations focus on laparoscopic surgeries, which mean operations in the abdominal or pelvic cavities. Basically, rigid endoscopes with a diameter of about 10 mm are employed for this treatment. The working depth is about 40 mm to 100 mm.

In general, measurements in surgery would be very helpful, but due to the difficult circumstances inside the human body it is very challenging. Within this paper, which concentrates on robust matching, a calibrated three-camera system based on an endoscopic component arrangement is chosen. Including the third camera offers an improvement in the reliability and precision of the resulting point cloud. In order to simplify the matching algorithm, a rectification approach is developed that allows for horizontal and vertical epipolar lines and provides the related interior and relative orientation parameters. In a next step, an efficient semi-global (SG) optimisation is applied. Finally, two representative results are shown and compared to reference models.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
doi:10.5194/isprsannals-III-3-27-2016

2. RELATED WORK

In practice, the use of endoscopic images for measurement purposes is not well established. However, in research several approaches for special kinds of endoscopic interventions have been published.

Röhl et al. (2011) published an approach for real-time surface reconstruction from stereo endoscopic images. The interior and relative orientation is determined by a common computer vision approach using a checkerboard pattern. For dense image matching a modified census similarity measure and the so called 'hybrid recursive matching' is applied. This algorithm requires the previously computed and the current disparity map as well as the stereo image pair to generate robust 3D point clouds in real-time. The authors reach framerates of 40 Hz for images with a resolution of 320x240 pixels. The resulting average depth difference with respect to a reference model is determined to less than one millimetre.

An overview about different techniques for 3D reconstruction in laparoscopy is shown in Maier-Hein et al. (2013). Regarding stereoscopy the authors explain that computational stereo is a mature method for 3D reconstruction, because there are already many well-established algorithms. Nevertheless, they also emphasize the main challenge to make it robust and reliable for practical applications. These requirements are not properly resolved so far. The authors finally suggest that one solution could be a combination of different sensors and algorithms. However, a three camera solution is not considered.

The advantage of a three-camera system has already been published by Maas (1992). The author introduced a third camera to reduce ambiguities on the epipolar lines in a particle tracking application. With two cameras there are frequently multiple possible matching candidates along these lines. Whereas, with a third camera in a not collinear arrangement, epipolar lines calculated from a point in one image and a corresponding candidate in another image intersect in the third image (Figure 4). These three corresponding candidate points offers the opportunity for mutual comparative measurements. This redundancy should bring more robustness and reliability for endoscopic surface reconstruction and is thus applied in this work. This procedure is also explained in Hartley et al. (2003) and is called 'epipolar transfer'.

Investigations regarding orientation, calibration and matching with image triplets are published by Mayer (2003). The author estimates the trifocal tensor for the orientation of the cameras and calculation of the third corresponding point. The points in the third image are just computed for the best stereo matches. These stereo matching costs are finally weighted by the costs between the first and third image point. This procedure improves the results slightly for some image triplets, but in relation to the benefit, the computation time is too high.

Heinrichs et al. (2007) proposed an efficient semi-global matching (SGM) procedure for image triplets. During cost calculation in an image pair, the authors compute the corresponding point in the third image and calculate the similarity measures regarding the two other images. The following semi-global (SG) optimisation is applied once to the mean of the three corresponding cost values. This is more efficient than applying the SG optimisation three times to the costs between each pair. This idea is also considered in this paper. In a final comparison with stereo analysis, the authors reach slightly better results (about 7.5 % more correct matches and 33 % better standard deviation) with image triplets while the computation time is about 13 % higher.

Furthermore, the authors rectify the image triplets to optimise the matching computation time (Heinrichs et al., 2006). The transformation leads to horizontal and vertical epipolar lines. However, their method requires three uncalibrated images and does not provide the interior and exterior orientation parameters of the resulting rectified images. Hence, a customised solution is developed, which is similar to Ayache et al. (1988) and described in section 4.1.

3. CAMERA SYSTEM AND CALIBRATION

Firstly, the development of the trinocular camera system and its system calibration are described. Then an a priori estimation of the systems resolution and its measurement accuracy in object space is given.

Concerning the investigations of a real stereo endoscope and its suitability for measurements by Conen et al. (2014), the requirements for the trinocular demonstrator are specified:

- Miniature: minimal outer diameter for realistic laparoscopic scales and a maximal baseline between the cameras for best possible intersection angles of the image rays
- Image quality: low noise in bad light conditions, quality is more important than resolution
- Stability: interior and relative orientation parameters remain stable during measurements
- Connectivity: simple connection to a computer and easy capture of the image data

As a result, three separate 'Awaiba NanEyeGS' cameras (Figure 1 left) are chosen, each with a diameter of 6 mm, a resolution of 640x640 px² and a pixel size of 3.6 µm. The cameras have a CMOS-RGB-sensor with global shutter and reach framerates up to 100 fps, according to the manufacturer. The interface works via USB 3.0 with image access using software of the manufacturer.

Figure 1. Awaiba NanEyeGS and attachment

The relative orientation is fixed by a cylindrical mounting, the focus is adjusted to the average working distance (~70 mm) and fixed by tape (Figure 1 right). To accomplish the miniature constraint, the cameras are arranged in an equilateral triangle, which provides, according to Maas (1992), the optimum configuration for minimal ambiguities. The outer diameter of the trinocular head (without mounting) is about 14 mm, which is already close to conventional stereo endoscopes and let assume that a later product development will lead to an even smaller instrument. If the optical axis of the cameras would additionally be arranged like a tetrahedron, with a height of the average acquisition distance, the overlap of the images could be optimised. However, this could not be realised yet. The cameras

are rotated such that all resulting images are approximately aligned to the same horizon and two images are side by side with a minimum y-parallax. The two parallel cameras ((1) left, (2) right) provide a stereoscopic view, while the third camera ((3) top) offers additional information for image matching. The cylindrical mounting is attached to a variable friction arm to enable stable recordings from different viewpoints.

The interior camera parameters are computed by a bundle adjustment (BA) over a 60x60x15 mm³ three-dimensional test field with 43 circular and coded targets of 2 mm diameter each (Figure 3). The obtained parameters are the principal distance, principal point, radial-symmetric lens distortion, decentring distortion, affinity and shear. The relative orientation, described by three rotation matrices and translation vectors, is estimated by space resection using the images and interior parameters followed by the BA. The resulting image measurement precision for calibration is determined to 1/32 of a pixel due to the use of circular targets.

The maximum acquisition distance for calibration is about 100 mm, which results in a target diameter of about 32 px in image. Consequently, the pixel size in object space is 0.063 mm. Thus, point clouds derived from dense image matching maintain at least the same point density, which enables a very detailed surface reconstruction.

A preliminary accuracy estimation for forward intersection is conducted by a Monte-Carlo simulation (MCS). A 3D point in 100 mm distance perpendicular to the centre of image 1 is computed 10^5 times considering an image measuring accuracy of a half pixel. The interior and relative orientation parameters are varied by their standard deviations of the calibration. The variations are normally distributed. The trinocular forward intersection yields deviations of 0.038 mm and 0.032 mm for XY position and 0.692 mm for depth in Z. Whereas for the stereo case (camera 1 and 2) 0.050 mm and 0.035 mm for position and 0.934 mm for depth are obtained. Due to the poor height-to-base ratio the deviation in depth is in both cases about 20 times higher than in lateral position (the base distances are determined by calibration to 7.521 mm, 6.955 mm and 6.939 mm). However, the key finding is that the trinocular arrangement leads to approximately one-third better results than the binocular one.

4. METHODOLOGY

In this section the rectification of the image triplets, dense matching with SG optimisation, subsequent sub-pixel interpolation and the calculation of 3D point clouds will be described.

4.1 Rectification of oriented image triplets

As already mentioned in section 2, image triplets can be rectified such that the epipolar lines are horizontal between image 1 and 2 and vertical between 1 and 3. When adding the line between image 2 and 3, this allows for matching in corresponding image rows and columns which simplifies and accelerates the algorithm.

A further improvement would be the direct processing of these images for measurements in object space. Thus, an approach to calculate the interior and relative orientation parameters of the rectified cameras has been developed. The initial orientation parameters must be determined beforehand or have to be known. The basic steps and the impact on epipolar lines are described briefly:

1. Distortion correction → straight epipolar lines
2. Parallel alignment of the three z-axes → parallel epipolar lines; epipoles are at infinity
3. Rotation around the z-axis of camera 1 and 2 → horizontal epipolar lines
4. Rotation around the z-axis of camera 3 and distortion (affinity and shear in x direction) of camera 1 → vertical epipolar lines

The principal distance has to be unified to reach the same scale. In a last step, an affine homography from the initial to the rectified camera is computed. The detailed procedure is described in the next three subsections.

4.1.1 Exterior orientation: To achieve parallel epipolar lines, the three images have to be aligned parallel to the trifocal plane. This plane is described by the projection centres $\mathbf{x}_{01}$, $\mathbf{x}_{02}$ and $\mathbf{x}_{03}$ (Figure 2).

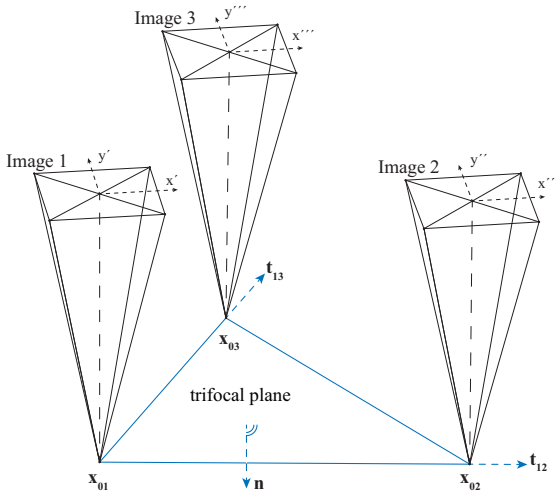

Figure 2. Camera arrangement and trifocal plane

Consequently, the new camera´s line of sight axis (z-axis) must be perpendicular to the trifocal plane and hence is equal to the normal vector $\mathbf{n}$. The normal vector can be calculated by the cross product of the vectors $\mathbf{t}_{12}$ and $\mathbf{t}_{13}$ (1).

$$\begin{aligned} \mathbf{t}_{12} &= \mathbf{x}_{02} - \mathbf{x}_{01} \\ \mathbf{t}_{13} &= \mathbf{x}_{03} - \mathbf{x}_{01} \\ \mathbf{n} &= \mathbf{t}_{12} \times \mathbf{t}_{13} \end{aligned} \tag{1}$$

To achieve horizontal epipolar lines between image 1 and 2, their new x-axis has to be equal to the vector $\mathbf{t}_{12}$. Likewise, to gain vertical epipolar lines in image 3, its new y-axis has to be equal to $\mathbf{t}_{13}$. Considering these constraints, the new rotation matrices can be constructed by horizontally concatenating the normalised direction vectors of the new axes. This is possible because the unit vectors $\mathbf{r}_x$, $\mathbf{r}_y$, $\mathbf{r}_z$, which describe the new axes, contain the direction cosines with respect to the original axes. These direction cosines are also part of the 3x3 rotation matrices. Thus, the new rotation matrices are described as follows:

$$\mathbf{R} = \begin{bmatrix} \mathbf{r}_x & \mathbf{r}_y & \mathbf{r}_z \end{bmatrix} \quad \text{… basic form} \tag{2}$$

$$\mathbf{R}_1 = \mathbf{R}_2 = \begin{bmatrix} \frac{\mathbf{t}_{12}}{|\mathbf{t}_{12}|} & -\left(\frac{\mathbf{t}_{12} \times \mathbf{n}}{|\mathbf{t}_{12} \times \mathbf{n}|}\right) & \frac{\mathbf{n}}{|\mathbf{n}|} \end{bmatrix} \tag{3}$$

$$\mathbf{R}_3 = \begin{bmatrix} \frac{\mathbf{t}_{13} \times \mathbf{n}}{|\mathbf{t}_{13} \times \mathbf{n}|} & \frac{\mathbf{t}_{13}}{|\mathbf{t}_{13}|} & \frac{\mathbf{n}}{|\mathbf{n}|} \end{bmatrix} \tag{4}$$

Note that the columns of the resulting rotation matrices have to be normalised to unit vectors. The vector $\mathbf{r}_y$ in matrix $\mathbf{R}_1$ and $\mathbf{R}_2$ as well as the vector $\mathbf{r}_x$ in matrix $\mathbf{R}_3$ are consequently resulting by the orthogonality constraint of the rotation matrix. As an interim result, the new relative orientation parameters for the rectified images are given by the rotation matrices $\mathbf{R}_1$, $\mathbf{R}_2$ and $\mathbf{R}_3$ and the projection centres $\mathbf{x}_{01}$, $\mathbf{x}_{02}$ and $\mathbf{x}_{03}$, which must remain unmodified. The epipolar lines between image 1 and 2 are horizontal and the lines in image 3 are vertical. The epipolar lines in image 1, which are assigned to points in image 3, still remain diagonal.

4.1.2 Interior orientation: The remaining diagonal lines in image 1 are aligned vertically by a shear factor in x-direction, which is applied by the appropriate distortion parameter. The distortion for affinity and shear is described by equation 5.

$$\Delta x'_{aff} = C_1 x' + C_2 y' \qquad \Delta y'_{aff} = 0 \tag{5}$$

Initially, the scale factor for x-direction (C_1) can be neglected and is set to zero to rearrange equation 5 as follows:

$$C_2 = \frac{\Delta x'_{aff}}{y'} \cdot \tag{6}$$

In a next step, the values for $\Delta x'_{aff}$ and y' are calculated by the projection of a 3D point $\mathbf{x}_a$ into image 1, while $\mathbf{x}_a$ is orthogonal to the new image 3 and not equal to its projection centre.

$$\mathbf{x}_a = \mathbf{x}_{03} + \mathbf{n} \qquad \text{arbitrary 3D point orthogonal to image 3 and not equal to } \mathbf{x}_{03} \tag{7}$$

$$\begin{pmatrix} x'_a \\ y'_a \\ 1 \end{pmatrix} = \mathbf{x}'_a = \mathbf{P}_1 \cdot \mathbf{x}_a \qquad \text{projection in image 1} \tag{8}$$

The projection matrix $\mathbf{P}_1 = \mathbf{K}_1 \cdot (\mathbf{R}_1 \mid -\mathbf{x}_{01})$ requires the new camera matrix $\mathbf{K}_1$ with the initial principal distance from image 1 (c_1) and assigning the principal point to the centre of the image.

Since $\mathbf{x}_a$ is orthogonal to the centre of image 3, the projection $\mathbf{x}_a'$ describes the offset to this centre in image 1. Hence, the shear factor can be calculated by (9).

$$C_2 = \frac{-x'_a}{y'_a} \tag{9}$$

In the last step, the vertical epipolar lines between image 1 and 3 have to be aligned to the same column. This is done by a scale factor, which is determined by comparing corresponding distances in both images. The distance in image 1 is defined by an arbitrary 2D point on the x-axis (e.g. $(1, 0, c_1)^T$) and the image centre. In order to get the corresponding distance in image 3, the 2D image point is transformed into object space and projected to image 3.

$$\mathbf{x}_b = \mathbf{x}_{01} + \mathbf{R}_1^T \begin{pmatrix} 1 \\ 0 \\ c_1 \end{pmatrix} \qquad \text{transformation into object space using the principal distance } c_1 \tag{10}$$

$$\begin{pmatrix} x'''_b \\ y'''_b \\ 1 \end{pmatrix} = \mathbf{x}'''_b = \mathbf{P}_3 \cdot \mathbf{x}_b \qquad \text{projection to image 3} \tag{11}$$

$$m = \frac{x'''_b}{1} \qquad \text{scale factor} \tag{12}$$

$$C_1 = m - 1 \qquad C_2 = \frac{-x'_a}{y'_a} m \tag{13}$$

The resulting scale factor m (12) is applied as distortion parameter C_1 of image 1 (13), which scales the x-direction. The shear factor C_2 has to be adjusted accordingly (13). In conclusion, the new relative and interior camera parameters for the rectified images are given as follows:

- Camera 1: $\mathbf{x}_{01}$, $\mathbf{R}_1$, c_1, C_1,C_2
- Camera 2: $\mathbf{x}_{02}$, $\mathbf{R}_2$, c_1
- Camera 3: $\mathbf{x}_{03}$, $\mathbf{R}_3$, c_1

By adjusting the principle points, the image content can be shifted in order to get as less as possible empty areas within the rectified images.

4.1.3 Image transformation: Finally, three affine homography matrices are generated by a 2D direct linear transformation between the initial and rectified corner points. The latter ones are determined by projecting the initial 3D corner points into the new cameras. The resulting homographies enables the conversion of image points from unrectified to rectified images and vice versa. It has to be considered that, when applying to the original images, a distortion correction is necessary beforehand. Therefore lookup tables are computed in advance to achieve a shorter computation time. An exemplary result of an undistorted and rectified image triplet is shown in Figure 3.

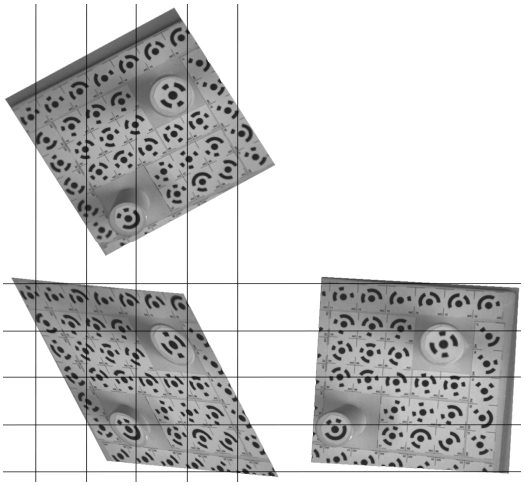

Figure 3. Rectified image triplet and epipolar lines. Three dimensional test field for calibration

4.2 Matching in rectified and oriented image triplets

As illustrated in Figure 3, the image contents are rotated, sheared and scaled, and cannot be properly matched using local cost functions. Alternatively, the matching windows could be transformed appropriately using least-squares-matching (LSM) (Förstner, 1982; Bethmann & Luhmann, 2010), which would entail an interpolation on grey values by affine or higher-order geometric transformation functions. Pixel based cost functions, e.g. mutual information (MI) as in Hirschmüller (2005), could be applied too, but are not investigated yet in this application.

A further alternative is given by a reduced rectification approach, which still leads to horizontal epipolar lines, but does not apply shear and scale to image 1. Additionally, image 3 is aligned parallel to the stereo pair. Consequently, the image content of the resulting triplet is exactly aligned to the same horizon without remaining scale differences. On the one hand, this allows for quadric cost windows, but on the other hand the epipolar lines of image 3 remain diagonal.

4.2.1 Epipolar constraint: To avoid the cost calculation for each pixel along the diagonal lines, points in image 3 are directly calculated by epipolar transfer, as mentioned in section 2. Figure 4 demonstrates the procedure of epipolar line intersection. It also indicates that the reference point or template is defined in image 1 and the search area in image 2 along the horizontal epipolar line in a predefined disparity range.

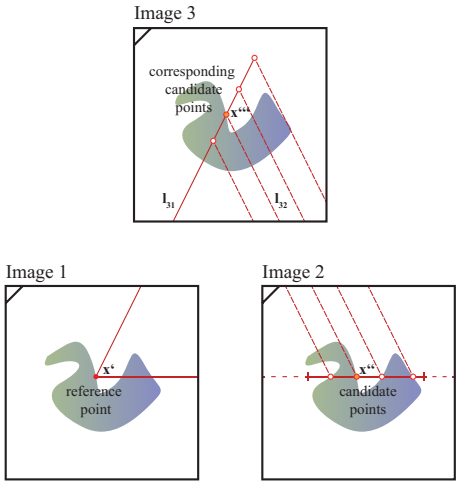

Figure 4. Epipolar line intersection in image 3

The corresponding candidate points in image 3 are directly calculated by the intersection of epipolar lines using fundamental matrices (14). Due to the rectification the fundamental matrices $\mathbf{F_{31}}$, $\mathbf{F_{32}}$ have a special form, which simplifies the calculation of $\mathbf{l_{31}}$ and $\mathbf{l_{32}}$ (15).

$$\mathbf{x'''} = \mathbf{l_{31}} \times \mathbf{l_{32}} = (\mathbf{F_{31}}\mathbf{x'}) \times (\mathbf{F_{32}}\mathbf{x''}) \tag{14}$$

$$\begin{pmatrix} l_a \\ l_b \\ l_c \end{pmatrix} = \begin{pmatrix} 0 & 0 & f_{13} \\ 0 & 0 & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f_{13} \\ f_{23} \\ f_{31}x + f_{32}y + f_{33} \end{pmatrix} \tag{15}$$

Thus, for each epipolar line only l_c has to be calculated while l_a and l_b remain stable. Consequently, the third element w of the resulting vector of the cross product (16) is constant. Even the terms $l_{31c}l_{32b}$ and $l_{31c}l_{32a}$ can be calculated beforehand if the reference point $\mathbf{x'}$ is given.

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} l_{31a} \\ l_{31b} \\ l_{31c} \end{pmatrix} \times \begin{pmatrix} l_{32a} \\ l_{32b} \\ l_{32c} \end{pmatrix} = \begin{pmatrix} l_{31b}l_{32c} - l_{31c}l_{32b} \\ l_{31c}l_{32a} - l_{31a}l_{32c} \\ l_{31a}l_{32b} - l_{31b}l_{32a} \end{pmatrix}, \begin{pmatrix} x''' \\ y''' \\ 1 \end{pmatrix} = \begin{pmatrix} u/w \\ v/w \\ w/w \end{pmatrix} \tag{16}$$

As a result, for each reference point in image 1 a list of candidate pairs along the disparity range of the epipolar line in image 2 is achieved. To construct the disparity space image (DSI), a similarity measure is calculated pairwise for each point triplet. This leads to three cost values (c_{12}, c_{13}, c_{23}), which are assigned to the images 1-2, 1-3 and 2-3.

4.2.2 Cost calculation and aggregation: The normalised correlation coefficient (NCC) is applied as local cost function because it provides reliable results. Short computation times are gained by a formula for the correlation coefficient (17), which is free of mean values, so that just one loop is needed.

$$c_{fg} = \frac{\sum_{i=1}^{n} g_i f_i - \frac{1}{n}\sum_{i=1}^{n} g_i \sum_{i=1}^{n} f_i}{\sqrt{\left(\sum_{i=1}^{n} g_i^{\,2} - \frac{1}{n}\left(\sum_{i=1}^{n} g_i\right)^2\right)\left(\sum_{i=1}^{n} f_i^{\,2} - \frac{1}{n}\left(\sum_{i=1}^{n} f_i\right)^2\right)}} \tag{17}$$

g_i, f_i: grey value i of the reference (g) or test (f) window

A further acceleration is achieved by integral images (Viola et al., 2001) that allow for the computation of the sum or the sum of squares of an arbitrary rectangular area in an image by only four storage accesses. Therefore, six integral images are generated beforehand (three for the sum and three for the sum of squares) using OpenCV. Only the sum of $g_i f_i$ has to be calculated in the loop, whereby the storage access to each grey value is most time consuming. As a result, all cost triplets and their image coordinates are stored in one DSI data structure.

4.2.3 Optimisation and disparity computation: In a next step, the SG optimisation according to Hirschmüller (2005) is applied to the DSI. To reduce computation time, the approach of Heinrichs et al. (2007) is considered and slightly modified. Due to the trinocular camera arrangement and the epipolar constraint a left-right check is not implemented. For the search of the best neighbouring disparity, the costs as well as its previously attached penalties are summarised. The maximum sum wins. The penalties are attached to each cost value separately because the single costs are needed for the sub-pixel interpolation. The filter is applied in eight directions and aggregated in one resulting DSI_{SG}. After the SG optimisation, the best matches are determined by a simple winner-takes-all approach whereby the corresponding costs are summarised.

4.2.4 Disparity refinement: Before computing 3D points, a sub-pixel interpolation is applied to the corresponding points in image 2 and 3; their reference point remains fix. The interpolated point is determined by the maximal turning point of a parabola which is fitted to the neighbouring cost values of the best match. The neighbouring costs are the mean values of the corresponding costs c_{12} and c_{23} for the sub-pixel in image 2, and c_{13} and c_{23} for the sub-pixel in image 3. Thus, by weighting with c_{23} the interpolation is influenced by the epipolar constraint. However, this procedure succeeds if there are no occlusions between the point triplets.

4.3 3D point computation

The last step is the computation of 3D points using the matched correspondences and the orientation parameters of the rectified images, according to section 4.1. A threshold for good matches is defined on the basis of the cost function. If all three costs lie above the threshold, a 3D point is calculated. In case of occlusion it is possible to reduce to a stereo case, but due to the robustness and computation time this is not applied. Consequently, 3D points are always computed from three corresponding image points. This is achieved by a forward intersection in terms of a least squares adjustment.

5. RESULTS AND ANALYSIS

In this section, experimental results of real test objects are shown and analysed. The objects are captured by the camera system described in section 3. Only static cases have been investigated so far due to limitations in system synchronisation. To evaluate the measurement accuracy, the results are compared against reference models. Furthermore, the influence of the third camera is demonstrated by reducing the camera system and algorithm to a stereo case. Finally, the computation times of single steps of the algorithm are determined.

5.1 Comparison with reference models

Two test objects are investigated for the evaluation, on the one hand a rigid and well textured statue (Figure 5 top left) and on the other hand raw pieces of meat, which are deformable and contain reflective spots (Figure 5 top right). The median acquisition distances are approximately 105 mm and 140 mm. The dense trinocular matching results in 227,548 points for the statue (Figure 5 centre left) and in 262,829 points for the meat (Figure 5 centre right) using a window size of 11x11 px² and the same cost threshold. For comparison against a reference model, the data is cleaned from 2,774 and 455 disconnected points (Table 1). Concerning the statue, the disconnected points mainly belong to the poorly textured background. The final point clouds are almost free of visible outliers.

The reference models are generated by the 3D fringe projection system 'Vialux zSnapper® portable', which provides a measuring accuracy of 20-50 µm for the depth. With the available equipment, this is an accurate and easy way to capture reference data. Selecting the same area as for the test measurement the statue is captured with about 60,000 points and the meat with about 25,000 points. Due to the statue's rigid body it is measured by multiple views and stitched together using reference points. Thus, a higher point density is achieved. In contrast the meat is soft and continuously deforming so that one single scan is captured. To achieve minimum deformations between the test and reference measurement, the scenario is observed within a narrow time frame.

Regarding the test measurements the reference scans are of higher precision, whereas the point resolution is much lower. Thus, for an area-based comparison the reference points are meshed. Applying a best-fit alignment between the reference meshes and the cleaned test point clouds, the comparisons are computed and visualised with a colour code (Figure 5 bottom).

The colour code indicates larger variations for the meat as for the statue. Furthermore, the deviations of the statue seems to be distributed much smoother whereas the colour code for the meat looks speckled and uneven.

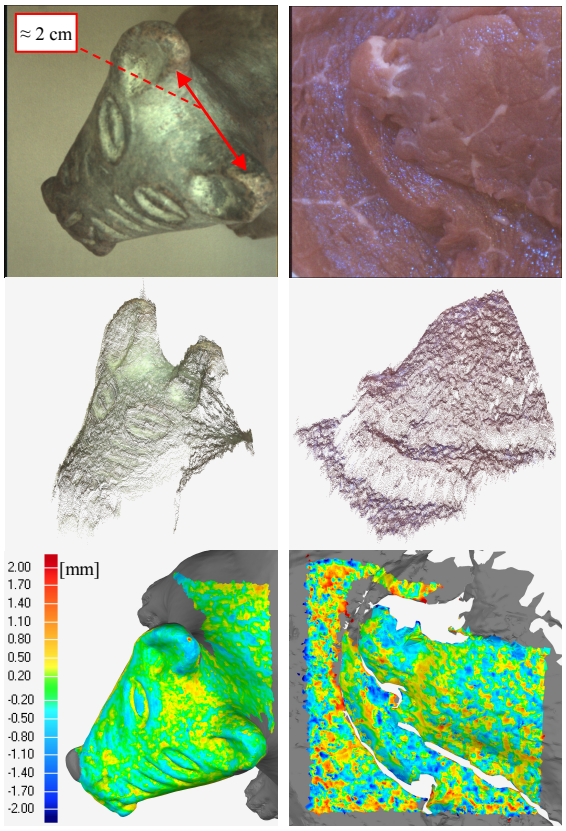

Figure 5. Test objects (top), resulting point clouds (centre) comparisons with reference model (bottom)

		Statue	Meat
Point cloud		points	points
	Ref.	60,000	25,000
	Test orig.	227,548	262,829
	Test filt.	224,774	262,374
Approx. median acquisition distance		mm	mm
	h	105	140
Median of real standard deviations for forward intersection		mm	mm
	$\bar{s}_x$, $\bar{s}_y$	0.004	0.007
	$\bar{s}_z$	0.033	0.068
Simulated (MCS) standard deviations for forward intersection		mm	mm
	s_x, s_y	0.062	0.080
	s_z	0.769	1.336
3D deviation after best fit alignment		mm	mm
	s_0	0.240	0.664

Table 1. 3D comparison and deviations

In comparison to the 3D deviation between the reference and test dataset (s_0), the median standard deviations of the real forward intersections ($\bar{s}_x$, $\bar{s}_y$, $\bar{s}_z$) are much better. However, these values are far too optimistic due to the epipolar constraint and sub-pixel interpolation and are therefore not well suited for comparison. Hence, a MCS is computed using the same parameters described in section 3. For the statue the simulated points are between the eyes and for the meat in the centre of the surface. The resulting deviations are more realistic and closer to the 3D deviations (s_0). The remaining discrepancies can be explained by the best-fit alignment over all points.

The 3D deviations as well as the results of the MCS for the statue are better than for the meat. This is primarily caused by the approximately 35 mm larger acquisition distance. Secondly,

the lack of good texture and the presence of small reflections on the meat surface are affecting the result. Nevertheless, both point clouds reaches a standard deviation s_0 with respect to their reference surface of less than one millimetre.

5.2 Binocular vs. Trinocular

In comparison to a stereo analysis (left and right image) the trinocular matching steadily leads to better results when the SG optimisation is not applied. There are more correct matches with higher similarity values and still less outliers. Figure 6 (red point clouds) demonstrates the impact on the matching without SG optimisation of the statue introducing the third view. The trinocular analysis (right) leads to roughly 14 % more correct matches using the same filter threshold.

In fact, the results show no significant differences when the SG optimisation is applied (Figure 6 green point clouds, Figure 5 centre left). Both point clouds can be fitted to the reference model with similar deviations listed in Table 1. The trinocular analysis seems to be marginally better, but this has to be investigated in more detail. In particular real scenarios with bad measuring conditions have to be taken into account. As can be shown with the latest test measurements, these special cases significantly benefit from the third camera.

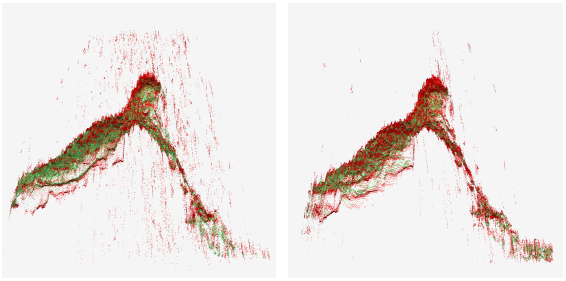

Figure 6. Side view of the statue analysis: binocular (left) and trinocular (right) with (green) and without (red) optimisation

5.3 Computation time

The current implementation needs about 20,000 ms in total for the statue example using an Intel Core i5 quad core processor. The current implementation uses multiple threads, however it is not entirely optimised for speed. Less than 20 ms are required for the rectification with the lookup table, the DSI is computed within 5,380 ms, the SG optimisation takes 11,700 ms and the disparity selection with 3D point calculation needs 2,900 ms. The result is a very dense point cloud with a ground sample distance (GSD) of less than 0.1 mm, which is probably not needed in practice. Thus, if every tenth pixel is considered the GSD decreased ten times (< 1.0 mm), but the computation time is about 100 times shorter (215 ms).

6. DISCUSSION

Currently, the real-time constraint can be fulfilled if the point resolution in object space, the disparity range or the size of the matching window is reduced. However, there are many other approaches to accelerate the matching without taking a loss in quality or quantity.

The matching process might be accelerated by a hierarchical approach using image pyramids, first applying the matching on the coarsest level and then refining in each larger level.

The most time consuming parts of the algorithm are the DSI calculation and the SG optimisation. These procedures are highly parallelisable and can be implemented on special hardware, such as a GPU (Ernst et al., 2008) or a FPGA (Banz et al., 2010), which leads to framerates of up to 30 Hz. Recent studies of Spangenberg et al. (2014) results in framerates of more than 16 Hz even on a CPU of a standard PC. The authors investigated the standard stereo approach, but the trinocular solution described in this contribution is in the basic steps similar to it such that these optimisations could also be applied.

The similarity measures the sum of absolute difference and the census transform (Zabih, 1994) have already been investigated with an older trinocular camera system. On the one hand these tests result in a faster computation, while on the other hand the NCC generates the most reliable results. Thus the NCC is chosen. The combination of SGM and MI, which is proposed by Hirschmüller (2005), is not considered yet. However, this could be a great benefit because MI is also robust and a pixel-by-pixel solution (Hirschmüller et al., 2007). Hence, the complete rectification approach with horizontal and vertical epipolar lines (section 4.1) can be applied because there is no matching window to be transformed.

In addition, the rectified images could be shifted using the principal point, as mentioned in section 4.1.2, such that the disparities in the right and top images are equal for corresponding points. According to Heinrichs et al. (2006) this could be realised by the condition that the slope of the epipolar lines between the right and top image must be one. This would circumvent the intersection of epipolar lines.

If not every pixel is considered for the test point cloud, the smoothing effect of the SG optimisation becomes particularly evident. By considering every fourth pixel, the resolution in object space is about 0.4 mm, but the smoothing effect bridges a wider range. Thus, the point cloud consists of less high frequently variations and the comparison with the reference surface yields statistically to slightly better results.

7. CONCLUSION AND OUTLOOK

In this paper, a miniature trinocular camera system with the dimensions of a real endoscope for laparoscopy is presented. The demonstrator is calibrated using standard photogrammetric approaches and enables extensive measurements with a high point resolution on the entire object surface. The close camera arrangement leads approximately to a 20 times higher standard deviation in depth than in lateral direction. Furthermore, an approach for rectification of image triplets, which also provide the new relative and interior orientation parameters of the rectified cameras is introduced. Thus, 3D measurements are done directly in the rectified images. The subsequent matching requires a slightly different rectification approach, such that the windows of the local similarity measure (NCC) need not to be transformed. Due to the rectification, the corresponding point triplets are calculated very efficiently. The most exhaustive part is the SG optimisation which is applied once during the process using the sum of the three cost values. Finally, the resulting point clouds are fitted to their reference models with a 1-sigma standard deviation (68.3 %) of less than one millimetre. This should also be reached for the 3-sigma standard deviation such that 99.7 % of the points are below 1 mm. This seems to be sufficient for laparoscopic applications. The results also point out that trinocular analyses are in general more reliable than binocular analyses, especially when the SG optimisation is not

applied. However, if the optimisation is performed, the point clouds of the stereo analysis are comparable to the trinocular results, when the surface is well textured.

Future work focus on the optimisation of the algorithm in order to achieve more reliable point clouds in particular for real scenarios with fluids and reflections. Moreover, image sequences are usually available in practice and should be exploited to accelerate the matching. One idea starts with an initial reconstruction of the surface with the full disparity range, which must be done once. With the assumption of marginal changes between two frames, the following image sequence can be analysed using a strongly shortened and dynamic disparity range for each pixel separately.

According to section 4.2.3, the costs are combined by summarising, which basically yield to the same results as the arithmetic mean like in Heinrichs et al. (2007). In future needs to be investigated whether other operators leads to better and more reliable results. By considering matches in all three images, the minimum cost value might be an alternative. If the minimum has a high cost value, the two corresponding costs must be even better. If occlusions are present and it is required to switch to the stereo case, the median might be a good choice. In contrast to the mean the median neglects single outliers completely.

An encapsulated work would be the proof of the thesis of Smirnov et al. (2010). The authors ascertained when using a local matching approach from a certain window size a trinocular analysis provides a higher quality and a shorter computation time than a binocular analysis.

REFERENCES

Ayache, N. and Hansen, C., 1988. Rectification of Images for Binocular and Trinocular Stereovision. In: *9th International Conference on Pattern Recognition*, Rome, Italy, Vol. 1, pp. 11-16.

Banz, C., Hesselbarth, S., Flatt, H., Blume, H. and Pirsch, P., 2010. Real-Time Stereo Vision System using Semi-Global Matching Disparity Estimation: Architecture and FPGA-Implementation. In: *International Conference on Embedded Computer Systems*, Samos, Greece.

Bethmann, F. and Luhmann, T., 2010. Least-squares matching with advanced geometric transformation models. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 38, Part 5, pp. 86-91.

Conen, N. and Luhmann, L., 2015. Kalibrierung und 3D-Messung mit einem medizinischen Stereoendoskop. In: *Photogrammetrie – Laserscanning – Optische 3D-Messtechnik, Beiträge der Oldenburger 3D-Tage 2015*, Wichmann, Berlin, pp. 186-195.

Ernst, I. and Hirschmüller, H., 2008. Mutual Information based Semi-Global Stereo Matching on the GPU. In: *4th International Symposium on Visual Computing*, Las Vegas, USA, pp. 228-239.

Förstner, W., 1982. On the geometric precision of digital correlation. In: *International Archives of Photogrammetry*, Vol. 24(3), Finland, pp. 176-189.

Hartley, R. and Zisserman, A., 2003. *Multiple View Geometry in Computer Vision*. University Press, Cambridge, UK.

Heinrichs, M., Rodehorst, V. and Hellwich, O., 2007. Efficient Semi-Global Matching for Trinocular Stereo. In: *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Science*, Munich, Germany, Vol. 36, Part 3/W49A, pp. 185-190.

Heinrichs, M. and Rodehorst, V., 2006. Trinocular Rectification for Various Camera Setups. In: *Symposium of ISPRS Commission III - Photogrammetric Computer Vision*, Bonn, Germany, pp. 43-48.

Hirschmüller, H., 2005. Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. In: *Conference on Computer Vision and Pattern Recognition*, IEEE, San Diego, USA.

Hirschmüller, H. and Scharstein, D., 2007. Evaluation of Cost Functions for Stereo Matching. In: *Conference on Computer Vision and Pattern Recognition*, IEEE, Minneapolis, USA, pp. 1-8.

Maas, H.-G., 1992. Complexity analysis for the determination of image correspondences in dense spatial target fields. In: *The International Archives of Photogrammetry and Remote Sensing*, Vol. 29, Part B5, pp.102-107.

Maier-Hein, L., Mountney, P., Bartoli, A., Elhawary, H., Elson, D., Groch, A., Kolb, A., Rodrigues, M., Sorger, J., Speidel, S. and Stoyanov, D., 2013. Optical Techniques for 3D Surface Reconstruction in Computerassisted Laparoscopic Surgery. In: *Medical Image Analysis*, Vol. 17, Issue 8, pp. 974-996.

Mayer, H., 2003. Robust Orientation, Calibration, and Disparity Estimation of Image Triplets. In: *Pattern Recognition*, Proceedings of the 25th DAGM-Symposium, Magdeburg, Germany, pp. 281-288.

Röhl, S., Bodenstedt, S., Suwelack, S., Kenngott, H., Müller-Stich, B. P., Dillmann, R. and Speidel, S., 2011. Real-time surface reconstruction from stereo endoscopic images for intraoperative registration. In: *Medical Imaging 2011: Visualization, Image-Guided Procedures, and Modeling*, *Proceedings of SPIE*, Vol. 7964.

Smirnov, S., Gotchev, A. P. and Hannuksela, M., 2010. Comparative analysis of local binocular and trinocular depth estimation approaches. In: *Real-Time Image and Video Processing*, *Proceedings of SPIE*, Vol. 7724.

Spangenberg, R., Langner, T., Adfeldt, S. and Rojas, R., 2014. Large Scale Semi-Global Matching on the CPU. In: *Proceedings of Intelligent Vehicles Symposium*, IEEE, pp. 195-201.

Viola, P. and Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. In: *Proceedings on Computer Vision and Pattern Recognition*, IEEE, Vol. 1, pp. 511-518.

Zabih, R. and Woodfill, J., 1994. Non-parametric Local Transforms for Computing Visual Correspondence. In: *Proceedings of European Conference on Computer Vision*, Stockholm, Sweden, pp. 151-158.