ToF/RGB Sensor Fusion for 3-D Endoscopy

Sven Haase[1], Christoph Forman[1,2], Thomas Kilgus[3], Roland Bammer[4], Lena Maier-Hein[3], Joachim Hornegger[1,2]

[1]Department of Computer Science, Pattern Recognition Lab, Friedrich-Alexander University of Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany
[2]Erlangen Graduate School in Advanced Optical Technologies (SAOT)
[3]Division of Medical and Biological Informatics, DKFZ Heidelberg
[4]Department of Radiology, Stanford University, Stanford, California, USA

E-mail: Sven.Haase@informatik.uni-erlangen.de

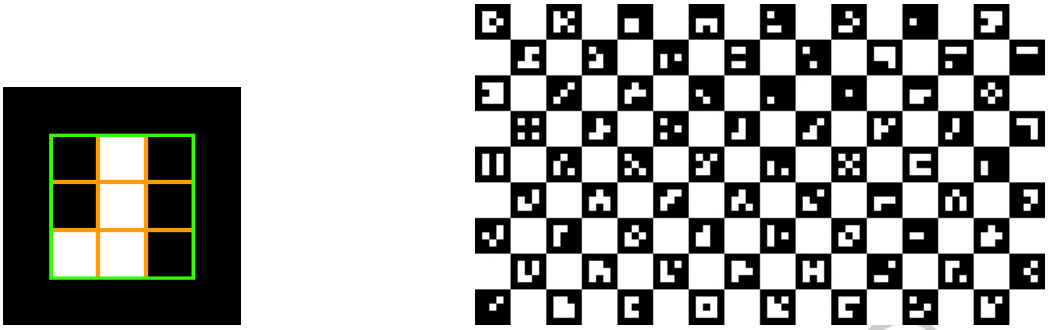
Abstract.

Acquisition of 3-D anatomical structure in minimally invasive surgery is an important step towards intra-operative guidance. In this context, the first prototype of a Time-of-Flight/RGB endoscope was engineered for simultaneous range and color data acquisition. Intrinsic and stereo camera calibration are essential to achieve an intuitive visualization of colored surfaces. Due to the early prototype stage, inhomogeneous illumination and low resolution (64×50 px) complicate the calibration significantly. To overcome these challenges, we propose a fully automatic multiscale calibration framework using a self-encoded marker for checkerboard detection. A first application demonstrates the feasibility of intra-operative measurement. Using our calibration scheme, we achieved a reprojection error of less than 0.7 px for the Time-of-Flight camera and 0.5 px of the RGB camera. Our framework eases calibration and enables future applications to use combined range and colored data.

1. Introduction

Minimally invasive procedures have become popular in the community of abdominal surgery [1]. Compared to conventional open surgery, minimally invasive procedures reduce post-operative trauma, scars and recovery time and thereby shorten hospital stays. Augmenting the conventional 2-D information with 3-D surface data enables novel medical applications, e.g. registration of intra-operative data with pre-operative information [2], recognition of risk situations in 3-D [3] or simple metric measurements [4]. Three different approaches have been proposed for endoscopy to estimate 3-D surface information: Stereo vision, structured light and Time-of-Flight (ToF). In stereo endoscopy, the displacement of corresponding features in the images of two cameras is used to calculate depth information based on a disparity map [5]. However, this approach requires a set of corresponding features in both images and is computationally expensive. Recently, a first prototype of a 3-D endoscope using structured light was proposed by Schmalz et al. [6]. This system provides no color information and thereby lacks an intuitive visualization to assist surgeons. ToF technology provides the ability to acquire a dense 3-D surface in real-time. In ToF imaging the scene is illuminated by modulated light. Range information is calculated by the phase shift between emitted and measured light on the chip. Additionally, the amplitude values of the measured signal provide a grey scale representation of the observed scene. In an initial approach, Penne et al. [7] replaced the conventional video camera of an endoscope with a ToF camera. To use the additional surface information in surgery a combination with conventional 2-D color information is required. This sensor fusion has been demonstrated for two individual devices [8, 9] and in particular for ToF endoscopy by Penne et al. [10]. Compared to those setups, our ToF and RGB sensor acquire data through a common optical system. This improvement allows to use a robust homographic mapping instead of exploiting the error-prone range information. Both mapping techniques require a calibration of the RGB and ToF sensor beforehand.

Conventionally, the corners of neighboring patches of a checkerboard pattern are used as features for camera calibration. Those feature points can be detected automatically with established calibration frameworks. For ToF endoscopy calibration is a highly recurrent task as besides an initial calibration a recalibration of the system must be performed each time the endoscope optics are changed. Due to inhomogeneous illumination in the ToF images, conventional checkerboard detection algorithms result in a high error rate. Since the RGB and the ToF chip do not share the same geometries both sensors need to be calibrated separately. Therefore, we use 2-D barcodes for feature point identification as proposed by Fiala et al. [11]. Conventional checkerboard detection systems have to identify the complete checkerboard in each image. Our approach uses the entire field of view at all distances even if the checkerboard is only partially visible. For higher robustness in low resolution ToF images (64×50 px) we adopted the marker of Forman et al. [12] with a reduced barcode size of 3×3. The detected corner points are used for camera calibration to estimate the intrinsic and extrinsic parameters.



(a) (b)

Figure 1. Detailed illustration of the self-encoded marker. (a) A single patch with green lines separating the barcode and the border, and orange lines separating the barcode boxes. (b) The set of all barcodes used for calibration.

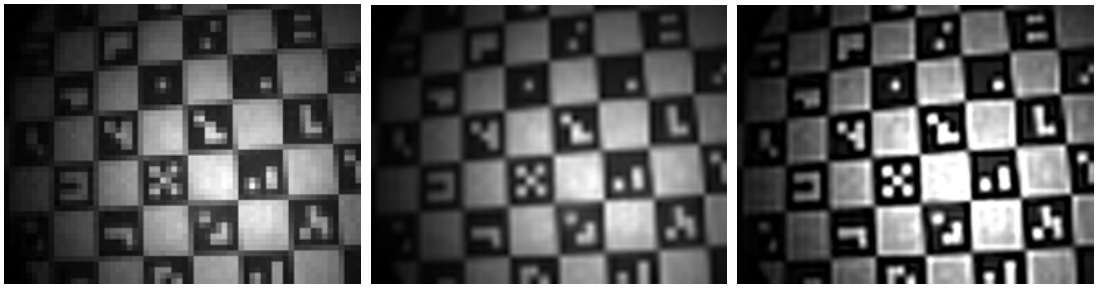
In this paper, we propose a fully automatic camera calibration scheme for fusing 2-D color information with 3-D ToF information for 3-D endoscopy. We demonstrate that the resulting data can be used for metric measurement within abdominal surgery in an initial application. The paper is organized as follows: Section 2 describes the methods used for camera calibration and sensor fusion. Experiments and results are shown in section 3. Section 4 summarizes the whole paper and gives an outlook about future work.

2. Camera Calibration and Sensor Fusion

This section details our approach to ToF/RGB sensor fusion in detail. The chapter is split into three major aspects. Initially, we will describe the camera calibration using the self-encoded marker. Therefore, the marker detection and barcode identification process is proposed in detail. Once the cameras are calibrated, the sensor fusion is illustrated describing the generic ToF/RGB sensor fusion and the endoscope specific homographic mapping. The final part of this chapter gives a small outlook on using the fused data for metric measurements within the human body.

2.1. Camera Calibration

Self-encoded Marker As proposed by Fiala et al. [11] barcodes can be used for feature point identification, which are also required for camera calibration. In our approach, we use a checkerboard marker with unique barcodes embedded in the checkerboard patches for a recognition of the feature points independently of the rotation of the entire marker [12]. The 2-D barcode is described by 3×3 blocks as depicted in Fig. 1(a). Those small barcodes are recognized robustly even in low resolution images. All barcodes are embedded into a checkerboard patch and thus surrounded by a black border. The feature points for calibration are the checkerboard corners identified by the barcodes.



(a) (b) (c)

Figure 2. Illustration of the image enhancement pipeline: (a) Original amplitude image. (b) Upsampled data generated with bicubic interpolation. (c) Image after applying an unsharp mask.

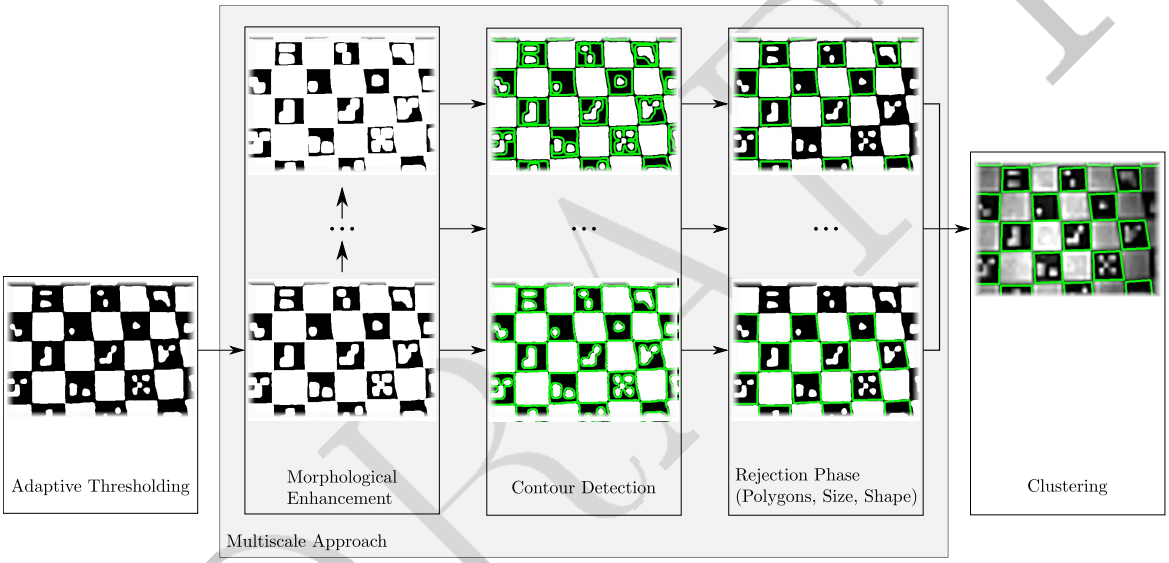


Figure 3. Flowchart of the marker detection process. The different scales of the patch recognition phase vary in the number of erosions and dilations applied on the binary image. The examples show an image section of a Time-of-Flight amplitude image. Green pixels denote the detected contours.

Time-of-Flight Image Enhancement As we use an implementation of a pixel-accurate framework, upsampling the ToF images is the initial step. Next, a preprocessing of the amplitude images as shown in Fig. 2(a) is required to compensate for inhomogeneous illumination. After resampling, we perform unsharp masking for local contrast enhancement [13]. As illustrated in Fig. 2(c) the contrast was obviously improved after applying our preprocessing pipeline.

Marker Detection The marker detection process is depicted in Fig. 3 and demonstrated for a ToF amplitude image. For marker detection we use a binarized version of the ToF amplitude and RGB input image. Therefore, an adaptive thresholding technique is performed calculating the threshold individually for each pixel depending

Figure 4. Flowchart of the marker identification process describing the phases leading from barcode calculation to checkerboard corner detection.

on its neighborhood. In order to retrieve the contours of each checkerboard patch, morphological operators are applied on the binary images to separate the blurred patches. Enhancing the detection algorithm proposed by Haase et al. [14], we cope with inhomogeneous illumination in the acquired images using a multiscale approach. Within each scale an opening is performed, which is a dilation of the eroded image. Across multiple scales the number of erosions applied on the image before applying the same number of dilations defines the individual scale. The output of the morphological enhancement is used for contour detection as proposed by Freeman [15] to find the checkerboard patches. Subsequently, a shape analysis is performed on all contours. First, the contours are approximated by polygons [16] and rejected if an approximation by four points is not achieved within ten iterations. Then, the contours are analyzed by their shape and length. Contours with a non-square shape or with an unexpected size are rejected. Finally, a clustering, similar to [17], is performed across all scales to combine contours of different scales describing the same checkerboard patch.

Marker Identification The marker identification process is depicted in Fig. 4. First, all detected patches are identified by their barcode. The barcode is calculated by dividing each patch in 5×5 blocks and analyzing the inner 3×3 blocks. The barcode is represented by a unique hash value calculated by the number of black blocks and their position. Subsequently, these identified barcodes are associated in a common structure describing their neighboring patches. The same structure is constructed for the ground truth image, respectively. To verify the identified barcodes each associated neighbor of an identified patch is compared to the ground truth neighbor. A score calculated by the validity of all four neighbors of each patch is finally used to reject a incorrect identified barcode. For the upcoming calibration process all identified checkerboard corners are then corrected with subpixel accuracy by gradient analysis.

Camera Calibration The previously identified checkerboard corners are utilized for camera calibration, subsequently. For estimating the intrinsic parameters the corners in all views are associated to their real world coordinates using prior knowledge about the checkerboard geometry. Following the approach of Zhang et al. [18] the focal lengths $(f_x,\ f_y)$ and the principal point $(c_x,\ c_y)$ assembling the camera matrix $\boldsymbol{K} \in \mathbb{R}^{3\times3}$ are estimated. The matrix is calculated by minimizing the reprojection error using a Levenberg-Marquardt optimization.

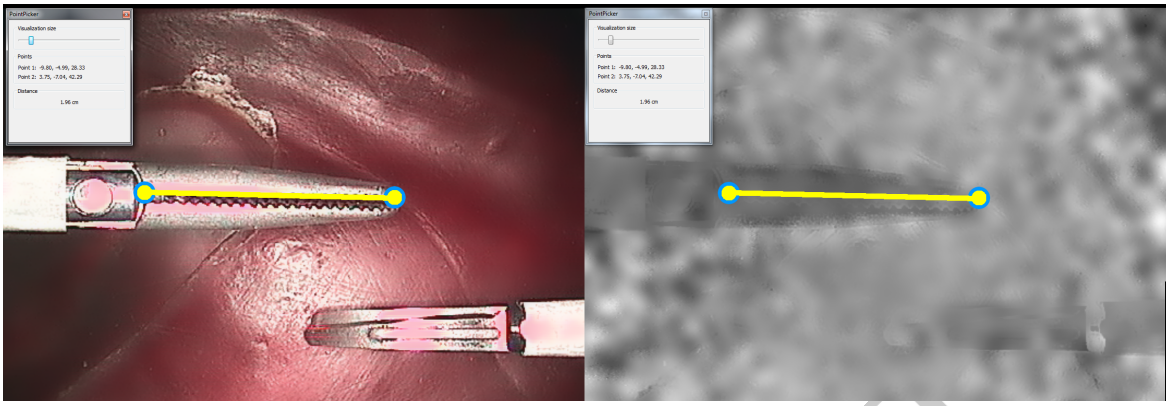


Figure 5. Measuring the length of the endoscopic tool in a realistic scenario. The yellow line denotes the measured length. Left: Mapped RGB image. Right: Time-of-Flight range image.

2.2. Sensor Fusion

Previously proposed approaches for ToF/RGB image fusion estimate the relative transformation using the extrinsic parameters of both sensors. First, the 3-D world coordinates are calculated with the ToF range data. Second, all 3-D points are transformed into the RGB sensor coordinate system and projected onto the RGB image plane. The relative transformation is described as:

$$\boldsymbol{R} = \boldsymbol{R}_{\mathrm{RGB}} \left(\boldsymbol{R}_{\mathrm{ToF}}\right)^{\mathsf{T}}, \quad \boldsymbol{t} = \boldsymbol{t}_{\mathrm{RGB}} - \boldsymbol{R}\,\boldsymbol{t}_{\mathrm{ToF}}, \tag{1}$$

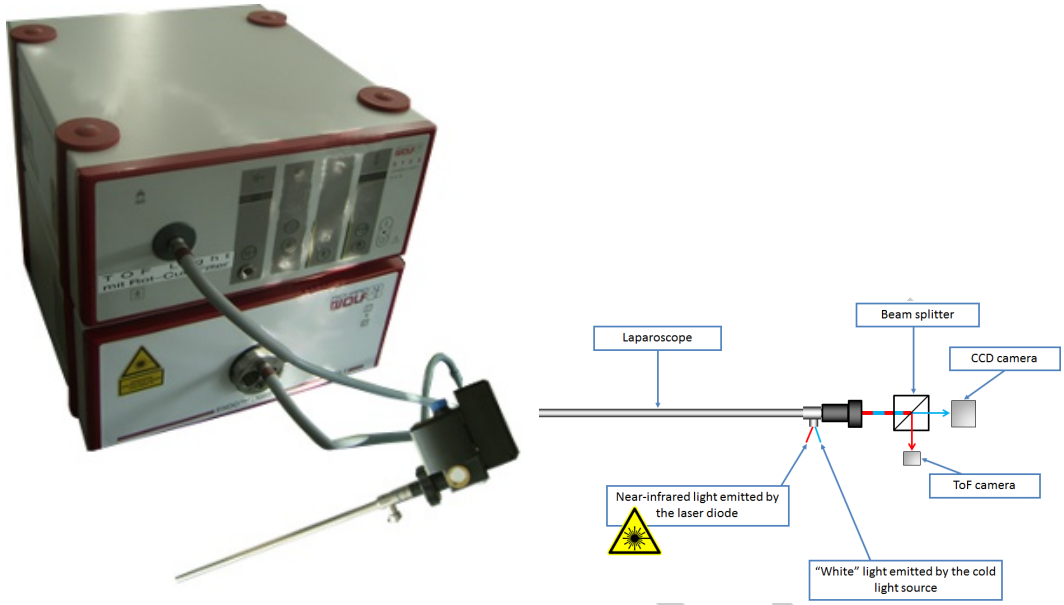
where $\boldsymbol{R} \in \mathbb{R}^{3\times3}$ denotes a rotation matrix and $\boldsymbol{t} \in \mathbb{R}^3$ a translation vector and the index denotes the modality. In our prototype both sensors acquire the scene through the same optical system. A beam splitter separates the incoming signal into near-infrared light for the ToF chip and the residual for the RGB chip (see Fig. 6(b)). Since ToF and RGB images share the same center of projection [19], this allows us to use a homographic mapping for transforming a 2-D RGB pixel $\boldsymbol{x}_{\mathrm{RGB}}$ onto the ToF chip following the equation:

$$\boldsymbol{x}_{\mathrm{ToF}} = \boldsymbol{H}\boldsymbol{x}_{\mathrm{RGB}}, \tag{2}$$

where $\boldsymbol{H} \in \mathbb{R}^{3\times3}$ denotes the homography between both sensor planes. This homography is estimated within the calibration process using the extrinsic parameters of both sensors.

2.3. Measuring in 3-D

Fusing both ToF and RGB sensor data allows us to perform metric 3-D measurements within the human body more intuitively. The surgeon is now capable of picking points either in the 3-D mesh representation or in the color domain and calculate distances as depicted in Fig. 5. Compared to Field et al. [4] the ToF surface mesh does not rely on features and therefore provides measurable points in a dense manner all over the observed scene. The intrinsic parameters of the ToF camera combined with the



(a) Light sources and rigid endoscopic optic. (b) Time-of-Flight/RGB endoscope system using a beam splitter to acquire data through one optical system.

Figure 6. Prototype of a 3-D Time-of-Flight/RGB endoscope.

mapping described in Sect. 2.2 allow to transform 2-D RGB pixels $\boldsymbol{x}_{\mathrm{RGB}}$ into 3-D world coordinates $\boldsymbol{X}_{\mathrm{World}}$ for calculating distances. This calculation is given by:

$$\boldsymbol{X}_{\mathrm{World}} = \boldsymbol{K}_{\mathrm{ToF}}^{-1}(\boldsymbol{H}\boldsymbol{x}_{\mathrm{RGB}}), \tag{3}$$

where $\boldsymbol{K}_{\mathrm{ToF}}$ denotes the intrinsic camera matrix of the ToF sensor.

3. Experiments and Results

The following chapter describes the evaluation of our sensor fusion in detail. First, the experiments and the setup is specified. Second, all results of our evaluation are shown and discussed.

3.1. Experiments

Time-of-Flight/RGB Endoscope All experiments were performed using a 3-D endoscope prototype (Richard Wolf GmbH, Knittlingen, Germany). The prototype acquires ToF (64×50 px) and RGB (640×480 px) data simultaneously through one optical system at a frame-rate of 30 fps. ToF technology utilizes modulated light and is based on a phase measurement [20]. The phase shift between emitted and measured signals is then utilized to calculate a range image as well as a gray scale amplitude image of the observed scene as depicted in Fig. 2. Furthermore, a flag image is delivered to indicate for each pixel if its measured range is reliable or erroneous. Nevertheless, most valid pixels show a low signal-to-noise ratio due to several error sources, e.g. temperature

related or amplitude related offsets [21]. Since our endoscopic system is in an early prototype stage, all experiments for metric measurements were performed in a phantom study with real instruments and a realistic liver phantom.

Calibration Data To enable a reliable evaluation the calibration pattern was observed from 100 different views. The views were shifted in all directions and acquired from different angles and thereby contain a varying amount of checkerboard corners. For evaluating the robustness of our calibration algorithm the reprojection error for a setup using 70 different views was calculated for both sensors in 30 repetitions. Furthermore, the focal lengths and the principal point were calculated for different numbers of views for 30 repetitions each. The views were chosen randomly from all 100 views without using any view twice within one repetition. Evaluating the barcode identification process was performed by labeling all images by an expert for ground truth data.

Sensor Fusion For sensor fusion evaluation we constructed a realistic medical scenario and used the generic mapping according to Eq. 1 as well as the homographic mapping according to Eq. 2. In order to obtain a quantitative comparison for both techniques the normalized mutual information (NMI) [22] was calculated as a similarity measurement using the RGB image and the amplitude image. A checkerboard representation of both input images as depicted in Fig. 8 shows qualitative improvements.

Measuring in 3-D The metric measurements were evaluated by virtually calculating the length of the tool tip using a point picker and comparing it to the ground truth data measured on the real tool, subsequently. This measurement was repeated for 30 successive frames acquired in a realistic medical scenario as depicted in Fig. 5.

3.2. Results and Discussion

Calibration Data The advantage of our multiscale approach is noticeable in the rejection phase of Fig. 3. On the last scale our algorithm was able to detect different patches compared to the first. In terms of marker identification we achieved an identification rate of 92.7% for the RGB images. The small improvement for the RGB images is due to the fact that almost all completely visible barcodes were identified using the conventional approach. The residual were expert labeled barcodes that were only partially visible and thereby not identified by our algorithm. In terms of the ToF data, we improved the identification rate of the barcodes from 92.0% [14] to 96.4% using our multiscale approach. Partially visible barcodes are less an issue in the ToF images due to the fact that the expert was not able to identify those barcodes either. Note that all identified barcodes are verified beforehand. Therefore, no erroneous identified barcodes are retained for calibration. As shown in Fig. 7(a) increasing the number of different views and thereby increasing the amount of checkerboard corners for calibration improves the robustness of the intrinsic parameters. We want to point out that a number

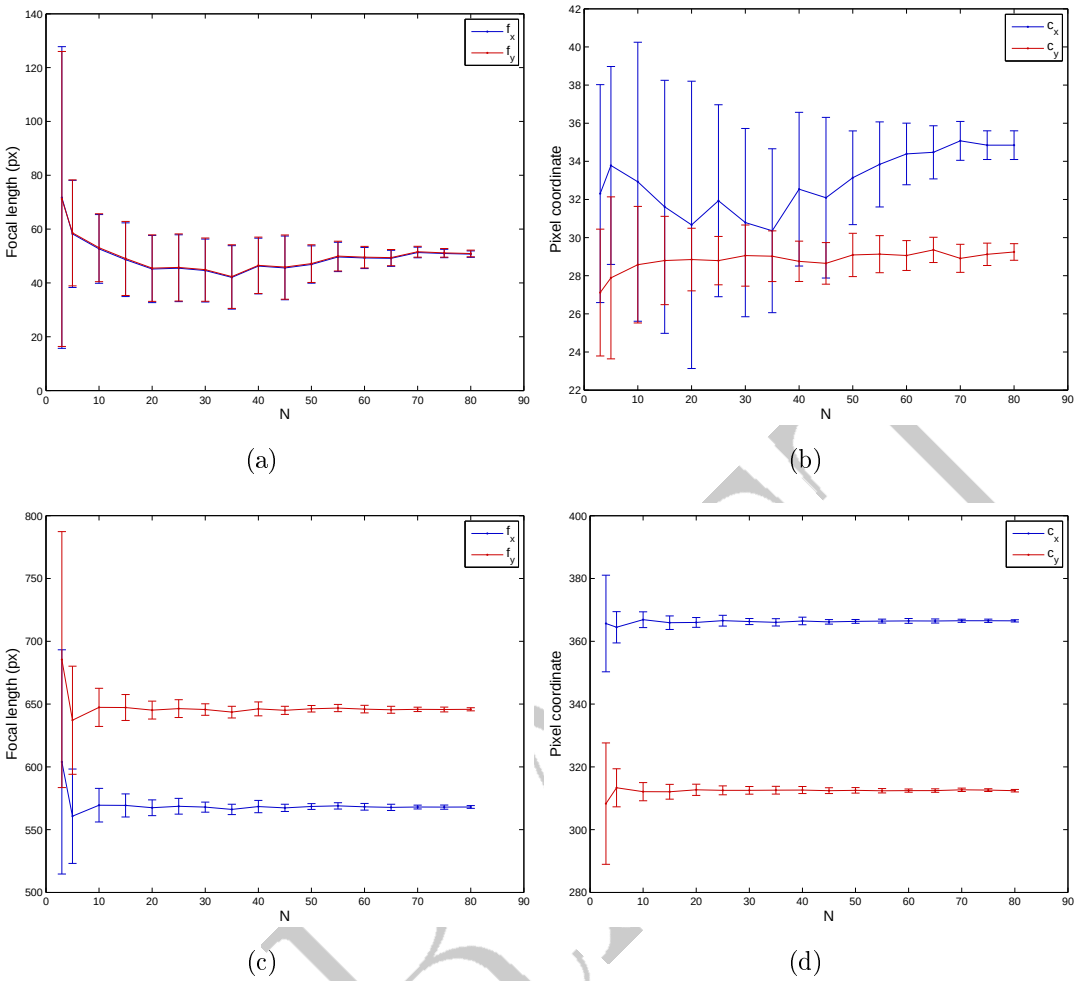
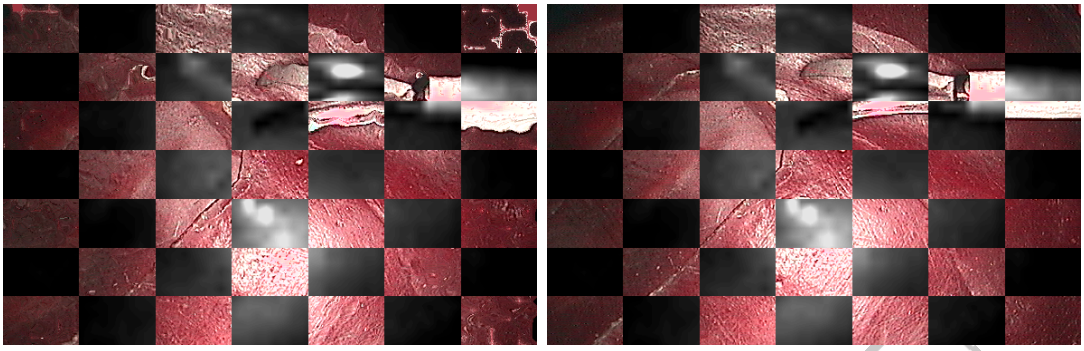


Figure 7. Plots of the mean and the standard deviation of the focal lengths (f_x, f_y) and the principal point (c_x, c_y) for different number of checkerboard views N. The first row represents the Time-of-Flight data. The second row represents the RGB data.

of 70 different views seems sufficient to result in a reliable calibration output as all relative standard deviations result in less than 5%. For 30 repetitions using 70 images for calibration the mean reprojection error resulted in 0.63 px for the ToF sensor and 0.49 px for the RGB sensor. The reprojection error allows a first interpretation of the quality of the estimated parameters, where a high reprojection error indicates that the parameters fit poorly to the input data. On the other hand a low reprojection error combined with a huge collection of input data indicates high quality parameters. We managed to keep the reprojection error in subpixel domain for all input data. Comparing Fig. 7(a) and Fig. 7(c) leads to the conclusion that the RGB sensor allows less calibration images for robust results due to its better signal-to-noise-ratio and higher resolution.

Sensor Fusion As shown in Fig. 8, the homographic mapping is independent of the error-prone ToF range values and, therefore, is more robust. The improved results are noticeable along the border of the observed tool and at the corners of the image where



(a) (b)

Figure 8. Two checkerboard views of the Time-of-Flight/RGB fusion result. (a) Naive mapping by projecting the world coordinates on the RGB chip. (b) Improved homographic mapping with more robust results noticeable at the border of the endoscopic tool.

range data is usually less reliable. The NMI between the amplitude image and the raw RGB image resulted in 0.92. After naive mapping it is improved to 0.93. With our homographic transformation we achieved an NMI of 0.95. This leads to the conclusion that the initial misalignment of both input images is not huge, but can be improved further by our approach.

Measuring in 3-D The tool tip measured as depicted in Fig. 5 has a length of 20.0 mm. Averaging the values using the range data for calculating this length resulted in a mean length of 18.0 mm with a standard deviation of 3.3 mm. This accuracy enables rough estimations within the human body. But due to the early development stage of this ToF/RGB endoscope, the accuracy is not yet sufficient enough for assisting surgeons during minimal invasive procedures. Though, it demonstrates the feasibility of measurements with improved hardware.

4. Summary and Outlook

In this paper we presented an easy-to-use calibration technique for ToF/RGB sensor fusion demonstrated for a 3-D endoscope. We assembled a powerful framework that identifies a self-encoded marker even on low resolution endoscope images. The framework identified more than 96% of all barcodes in the ToF images. Furthermore, the reprojection error of the ToF camera using our identified subpixel accurate checkerboard corners resulted in 0.63 px using 70 different views containing more than 1500 feature points. The fused output provides an intuitive visualization of the acquired data for the surgeon. Furthermore, it enables various applications where range and corresponding color information is needed. We proposed a metric measurement tool for calculating 3-D distances on the 2-D color and the 3-D range image. However, the data acquired by the ToF chip are still error-prone which shows the need for further hardware improvements.

Compared to Haase et al. [14] we added three major contributions. First, we enhanced the detection pipeline by a multiscale approach. Second, the conventional ToF/RGB mapping was replaced by a homographic mapping. Third, we showed the feasibility to measure metric distances with the fused data.

Henceforth, further applications benefit from both range and color information to achieve more robust or more reliable results. Medical applications in particular, including 3-D instrument tracking or reconstruction of the whole abdomen by stitching different 3-D views, could be improved significantly by using both color and surface data.

Acknowledgments / Disclaimer

We gratefully acknowledge the support by the Deutsche Forschungsgemeinschaft (DFG) under Grant No. HO 1791/7-1. This research was funded/ supported by the Graduate School of Information Science in Health (GSISH) and the TUM Graduate School. The authors gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the DFG in the framework of the German excellence initiative.

References

[1] T. Heikkinen, S. Msika, G. Desvignes, et al. Laparoscopic surgery versus open surgery for colon cancer: Short-term outcomes of a randomised trial. *Lancet Oncol*, 6(7):477–484, 2005.

[2] S. Röhl, S. Bodenstedt, S. Suwelack, et al. Real-time surface reconstruction from stereo endoscopic images for intraoperative registration. In *Proc SPIE*, volume 7964, page 796414, 2011.

[3] S. Speidel, G. Sudra, J. Senemaud, M. Drentschew, B. P. Müller-Stich, C. Gutt, and R. Dillmann. Recognition of risk situations based on endoscopic instrument tracking and knowledge based situation modeling. In *Proc SPIE*, volume 6918, page 69180X, 2008.

[4] M. Field, D. Clarke, S. Strup, and W. Seales. Stereo endoscopy as a 3-D measurement tool. *Conf Proc IEEE Eng Med Biol Soc*, 1:5748–51, 2009.

[5] U. D. A. Mueller-Richter, A. Limberger, P. Weber, K. W. Ruprecht, W. Spitzer, and M. Schilling. Possibilities and limitations of current stereo-endoscopy. *Surg Endosc*, 18:942–947, 2004.

[6] C. Schmalz, F. Forster, A. Schick, and E. Angelopoulou. An endoscopic 3D scanner based on structured light. *Med Image Anal*, 16(5):1063 – 1072, 2012.

[7] J. Penne, K. Höller, M. Stürmer, et al. Time-of-Flight 3-D Endoscopy. In *MICCAI 2009, Part I*, pages 467–474, 2009.

[8] M. Lindner and A. Kolb. Data-Fusion of PMD-Based Distance-Information and High-Resolution RGB-Images. In *Proc ISSCS*, volume 1, pages 121–124, 2007.

[9] S. Á. Gudmundsson and J. R. Sveinsson. TOF-CCD image fusion using complex wavelets. In *Proc ICASSP*, pages 1557–1560, 2011.

[10] J. Penne, C. Schaller, R. Engelbrecht, L. Maier-Hein, B. Schmauss, H.-P. Meinzer, and J. Hornegger. Laparoscopic Quantitative 3D Endoscopy for Image Guided Surgery. In *Bildverarbeitung für die Medizin*, pages 16–20. Springer, 2010.

[11] M. Fiala and C. Shu. Self-identifying patterns for plane-based camera calibration. *Mach Vis Appl*, 19(4):209–216, May 2008.

[12] C. Forman, M. Aksoy, J. Hornegger, and R. Bammer. Self-encoded marker for optical prospective head motion correction in MRI. *Med Image Anal*, 15(5):708–719, 2011.

[13] D. F. Malin. Unsharp Masking. *AAS Photo-Bulletin*, (16):10–13, 1977.

[14] S. Haase, C. Forman, T. Kilgus, R. Bammer, L. Maier-Hein, and J. Hornegger. ToF/RGB Sensor Fusion for Augmented 3D Endoscopy using a Fully Automatic Calibration Scheme. In T. Tolxdorff, T. M. Deserno, H. Handels, and H.-P. Meinzer, editors, *Bildverarbeitung für die Medizin*, Informatik Aktuell, pages 111–116. Springer, 2012.

[15] H. Freeman. Boundary encoding and processing. *Picture Processing and Psychopictorics*, pages 241–266, 1970.

[16] D. H. Douglas and T. K. Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(2):112–122, Oct. 1973.

[17] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. Data structures for disjoint sets. In *Introduction to Algorithms*. The MIT Press, 2 edition, 2001.

[18] Q. Zhang. Extrinsic calibration of a camera and laser range finder. In *In IEEE International Conference on Intelligent Robots and Systems (IROS*, page 2004, 2004.

[19] M. Ben-Ezra. Segmentation with Invisible Keying Signal. In *CVPR*, pages 1032–1037, 2000.

[20] A. Kolb, E. Barth, R. Koch, and R. Larsen. Time-of-Flight Sensors in Computer Graphics. In *Eurographics (State-of-the-Art Report)*, pages 119–134. The Eurographics Association, 2009.

[21] A. Kolb, E. Barth, R. Koch, and R. Larsen. Time-of-Flight Cameras in Computer Graphics. *Computer Graphics Forum*, 29(1):141–159, 2010.

[22] C. Studholme, D. Hill, and D. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognit*, 32(1):71–86, 1999.