# ToF/RGB Sensor Fusion for Augmented 3-D Endoscopy using a Fully Automatic Calibration Scheme

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

[1]Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg
[2]Department of Radiology, Stanford University, Stanford, California, USA
[3]Division of Medical and Biological Informatics, DKFZ Heidelberg
[4]Erlangen Graduate School in Advanced Optical Technologies (SAOT)
sven.haase@informatik.uni-erlangen.de

**Abstract.** 3-D Endoscopy is an evolving field of research and offers great benefits for minimally invasive procedures. Besides the pure topology, color texture is an inevitable feature to provide an optimal visualization. Therefore, in this paper, we propose a sensor fusion of a Time-of-Flight (ToF) and an RGB sensor. This requires an intrinsic and extrinsic calibration of both cameras. In particular, the low resolution of the ToF camera ($64{\times}50$ px) and inhomogeneous illumination precludes the use of standard calibration techniques. By enhancing the image data the use of self-encoded markers for automatic checkerboard detection, a reprojection error of less than 0.23 px for the ToF camera was achieved. The relative transformation of both sensors for data fusion was calculated in an automatic manner.

## 1  Introduction

The benefits of 3-D data in medicine to speed up and improve quality of surgeries are described in several publications [1,2]. In comparison to conventional techniques, Time-of-Flight (ToF) cameras are a popular modality due to their markerless and non-invasive data acquisition. They measure 3-D scenes in real-time and hold potential to provide a surgeon with up to date surface data during surgery. Especially 3-D endoscopy gained lots of attention recently [3]. It is generally expected that the incorporation of photometric information into geometric data is able to improve segmentation, classification and registration in a significant way. Furthermore, it eases the interpretation of the data for surgeons. As the grayscale information delivered by the ToF sensor is insufficient for texture information due to its low resolution ($64{\times}50$ px), ToF/RGB data fusion has already been proposed in [4], [5] and for 3-D endoscopy in [3]. The 3-D endoscope used in this paper is a prototype utilizing a ToF sensor in combination with an RGB chip and is described in Sect. 2.3. Besides an initial calibration of the

system, each time the endoscope optics are changed a recalibration needs to be performed. In practice, we experienced that due to inhomogeneous illumination and the low resolution of the ToF sensor, illustrated in Fig. 1, the user has to identify the checkerboard in each acquisition manually, which is a time consuming and tedious procedure. Consequently, data enhancement is an inevitable preprocessing step for application of an automatic marker detection framework. For checkerboard detection we propose a method that does not have to recognize the whole checkerboard in each image and that does not have to make sure that both sensors cover all corners during measurement.

In comparison to Penne et al., in this paper, we propose the use of self-encoded markers [6] in order to establish a fully-automatic calibration routine for low-resolution sensor fusion.

## 2  Materials and Methods

The following sections describe our approach to ToF/RGB sensor fusion in detail. This process can be split into two major aspects. First, the camera calibration and the required image enhancement for the ToF amplitude image and second, the fusion of ToF and RGB data.

### 2.1  Camera Calibration

In order to apply a pixel-accurate framework for self-encoded marker detection [6] a bicubic upsampling of the ToF amplitude image data is performed. This technique enables subpixel-accuracy in low-resolution ToF data. In order to compensate for illumination inhomogeneities, we performed the well established technique of unsharp masking for local contrast enhancement [7]. Fig. 1 shows qualitative results of our preprocessing pipeline.

The calibration is split into two major stages. First, the corners of the observed checkerboard are detected and identified. Second, these corners serve as input for intrinsic and extrinsic calibration.
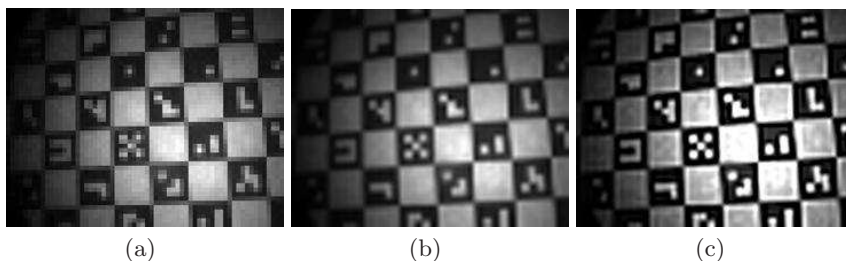


     (a)          (b)          (c)

**Fig. 1.** Illustration of the image enhancement pipeline: (a) shows the original amplitude image, (b) the upsampled data using bicubic interpolation and (c) the image after applying an unsharp mask.

For automatic detection of the corners, self-encoded markers are utilized. The process is described precisely as follows: The contours of the checkerboard squares are extracted using an adaptive thresholding scheme and are verified using contour shape analysis. Next, the markers are identified according to their unique bar codes using a nearest neighbor classification based on a template database. For eliminating false detection a sanity check of the four diagonal neighbors is performed. Finally, the corners of the verified markers serve as input for subpixel corner detection based on gradient analysis.

Based on the detected corners we apply an established optical calibration method [8] that estimates the intrinsic parameters and the distortion coefficients for both cameras separately. According to [8], the detected corners are re-used to calculate the extrinsic parameters for all image pairs that were acquired for calibration. Next, the poses (position and orientation) of the two cameras w.r.t. each other are calculated by using the extrinsic parameters (rotation and translation) estimated for each view:

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

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. Due to the fact that for each view a slightly different transformation will be estimated, averaging all results weighted by their certainty is necessary. The certainty of each transformation is calculated depending on the amount of detected corners in the ToF image of this view. Note that the normalized quaternion representation [9] is used for averaging all rotations.

## 2.2   Sensor Fusion

In order to merge the data of both sensors, the 3-D position $\boldsymbol{X}_{\mathrm{ToF}}$ is calculated for each rectified ToF pixel $\boldsymbol{x}_{\mathrm{ToF}}^{\mathrm{rect}}$ by utilizing the intrinsic camera matrix $\boldsymbol{K}_{\mathrm{ToF}}$ of the ToF camera $\boldsymbol{X}_{\mathrm{ToF}} = \boldsymbol{K}_{\mathrm{ToF}}^{-1}\,\boldsymbol{x}_{\mathrm{ToF}}^{\mathrm{rect}}$. Next, $\boldsymbol{X}_{\mathrm{ToF}}$ is transformed from ToF camera coordinates into RGB camera coordinates, yielding $\boldsymbol{X}_{RGB} = \boldsymbol{R}\,\boldsymbol{X}_{ToF} + \boldsymbol{t}$. Finally, $\boldsymbol{X}_{RGB}$ is projected onto the RGB plane, using the intrinsic camera matrix $\boldsymbol{K}_{\mathrm{RGB}}$ of the RGB camera, $\boldsymbol{x}_{RGB} = \boldsymbol{K}_{RGB}\,\boldsymbol{X}_{RGB}$ and eventually distorted. As this usually results in a subpixel coordinate in the sensor domain, the color texture of the ToF pixel is calculated by bilinear interpolating between the surrounding pixels in the RGB image.

## 2.3   ToF/RGB Endoscope Prototype

All experiments in this paper are based on a 3-D endoscope prototype manufactured by Richard Wolf GmbH, Knittlingen. The prototype utilizes photonic mixer device technology for acquisition of 3-D surface data. In comparison to the endoscope used in [3], our device acquires surface and color data through one single endoscope optics by using a beam splitter. The RGB sensor acquires data with a resolution of $640{\times}480\,\mathrm{px}$ the low resolution ToF sensor with $64{\times}50\,\mathrm{px}$.

### 2.4 Experiments

For evaluation of the marker identification, 50 ToF amplitude images of the calibration pattern were acquired considering different angels and distances. The detection rate in Sect. 3.1 is calculated on the total amount of automatically detected markers compared to the total amount of markers detected by a human observer in all 50 images. The identification rate is based on the number of all detected markers compared to the number of correct identified markers.

The experiments for evaluating the camera calibration is only considered for the low resolution ToF camera. The reprojection error is calculated using 50 ToF amplitude images. For evaluation of the robustness of our calibration technique, the intrinsic parameters were estimated based on different number of views of the calibration pattern. For each experiment, the image set was randomly chosen out of 110 views of the calibration pattern. The mean and standard deviation were calculated after 20 repetitions for each number of views.

Finally, for qualitative evaluation of the sensor fusion a red pepper was measured.

## 3 Results

In the following sections the results of our previously described experiments are presented in detail. Three different outcomes are distinguished. First, the identification of the markers. Second, the calibration results of the ToF camera and third, the camera fusion.

### 3.1 Marker Identification

In terms of the self-encoded marker identification we achieved a detection rate of 93.1 % and an identification rate of 92.0 %. Note that all erroneously identified markers are eliminated due to the sanity check. Furthermore, let us consider the aspect of time as well. As shown in Fig. 2, in practice, at least 50 images need to be acquired in order to achieve a robust estimation of the intrinsics. The expert has to detect and identify all corners of the checkerboard in the ToF amplitude images. Using our approach the automatic identification of these corners was performed within 50 seconds.

### 3.2 Camera Calibration

For the intrinsic calibration, we investigated the reprojection error and achieved a mean error of 0.23 px using 50 images. Besides, Fig. 2 illustrates the mean and the standard deviation of the focal lengths ($f_x$, $f_y$) and the principal point ($c_x$, $c_y$), given an increasing number $N$ of images considered for calibration. Note that in practice a robust estimation of the camera parameters highly depends on the number of corners that can be detected in each view.

### 3.3   Sensor Fusion

For qualitative evaluation, Fig. 3 shows a checkerboard view of a red pepper, where the texture of the amplitude image of the ToF camera and the merged RGB image are shown in alternating checkerboard patches. Note that at the edges adjacent structures coincide. In order to confirm the qualitative impression, we also computed the normalized mutual information (NMI) [10] as a similarity measure using the RGB image and the amplitude image for both views. Here, we achieved an improvement from 0.84 (0.88) without the alignment to 0.90 (0.93) after mapping the color information using the calculated relative transformation. The value of NMI is located between 0 and 2, where 2 indicates perfect similarity.
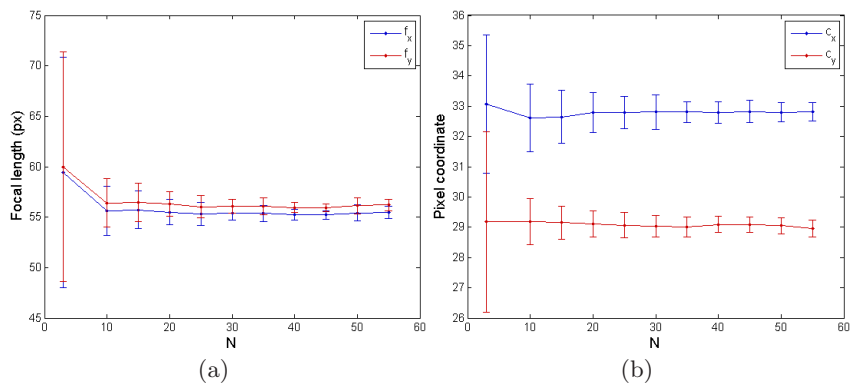


(a)                                    (b)

**Fig. 2.** 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 views $N$ of the checkerboard.



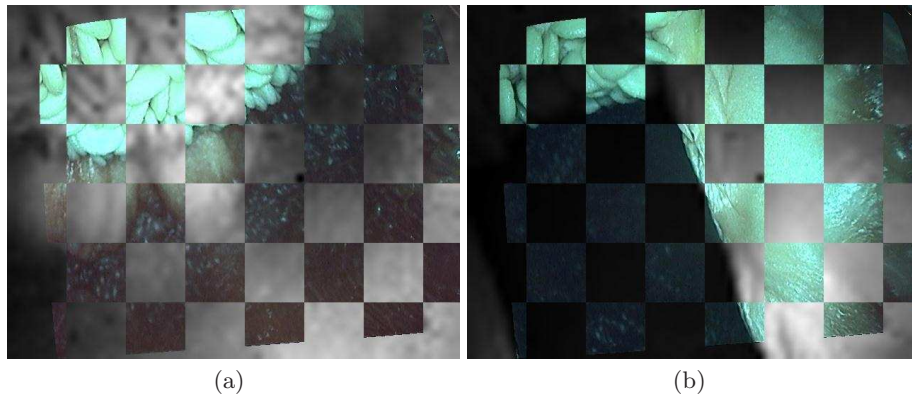(a)                                    (b)

**Fig. 3.** Two checkerboard views of the ToF/RGB fusion result and the corresponding ToF amplitude image. Please note that at the edges adjacent structure coincide.

## 4    Discussion

In this paper we have presented a powerful framework for ToF/RGB fusion for 3-D endoscopy using a fully automatic calibration technique. Thereby, we reduced the calibration effort significantly. Considering a typical grid size of $6{\times}5$ corners and a few seconds for manual selection and identification of the corners, this results in about an hour of annotation work. Using our automatic approach we reduced this to less than a minute. Furthermore, we achieved a detection rate and identification rate for the markers of more than $90\,\%$. For the calibration a mean reprojection of $0.23\,\mathrm{px}$ was calculated.

As the range value has a prominent influence on the camera fusion, ongoing research will concern preprocessing the ToF data for a more accurate projection on the RGB plane. Future work will also address occlusion handling.

## References

1. Röhl S, Bodenstedt S, Suwelack S, et al. Real-time surface reconstruction from stereo endoscopic images for intraoperative registration. In: Proc SPIE. vol. 7964; 2011. p. 796414.
2. Mersmann S, Müller M, Seitel A, et al. Time-of-flight camera technique for augmented reality in computer-assisted interventions. In: Proc SPIE. vol. 7964; 2011. p. 79642C.
3. Penne J, Schaller C, Engelbrecht R, et al. Laparoscopic Quantitative 3D Endoscopy for Image Guided Surgery. Proc BVM. 2010; p. 16–20.
4. Gudmundsson SÁ, Sveinsson JR. TOF-CCD image fusion using complex wavelets. In: Proc ICASSP; 2011. p. 1557–1560.
5. Lindner M, Kolb A. Data-Fusion of PMD-Based Distance-Information and High-Resolution RGB-Images. In: Proc ISSCS. vol. 1; 2007. p. 121–124.
6. Forman C, Aksoy M, Hornegger J, et al. Self-encoded marker for optical prospective head motion correction in MRI. Med Image Anal. 2011;15(5):708–719.
7. Malin DF. Unsharp Masking. AAS Photo-Bulletin. 1977;(16):10–13.
8. Zhang Z. A Flexible New Technique for Camera Calibration. IEEE Trans Pattern Anal Mach Intell. 1998;22:1330–1334.
9. Horn BKP. Closed-form solution of absolute orientation using unit quaternions. J Opt Soc Am A. 1987;4(4):629–642.
10. Studholme C, Hill DLG, Hawkes DJ. An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognit. 1999;32(1):71–86.