# Instrument segmentation in hybrid 3-D endoscopy using multi-sensor super-resolution

S. Haase[1], T. Köhler[1,2], T. Kilgus[3], L. Maier-Hein[3], J. Hornegger[1,2], H. Feußner[4]

[1] Pattern Recognition Lab, Dept. of Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
[2] Erlangen Graduate School in Advanced Optical Technologies (SAOT), Germany
[3] Div. Medical and Biological Informatics Junior Group: Computer-assisted Interventions, German Cancer Research Center (DKFZ) Heidelberg, Germany
[4] Research Group Minimally-invasive interdisciplinary therapeutical intervention, Klinikum rechts der Isar of the Technical University Munich, Germany

Contact: sven.haase@fau.de

*Abstract:*

*In hybrid 3-D endoscopy, photometric information is augmented by range data for guidance in minimally invasive procedures. In this paper, we propose a method for instrument segmentation exploiting sensor data fusion between range data and complementary photometric information. For improved robustness and to overcome the limited spatial resolution of range sensors, we make use of multi-sensor super-resolution to obtain high-quality range images. Data of both modalities is then segmented separately using thresholding techniques. The results are then consolidated into a common segmentation mask. Our approach was evaluated on real image data acquired from a liver phantom and manually labeled ground truth data. Compared to purely color driven segmentation we improved the F-score from 0.61 to 0.73.*

*Keywords: Time-of-Flight, 3-D Endoscopy, Super-Resolution, Segmentation*

## 1    Problem Statement

3-D endoscopy gained high attention as it enables new applications to minimally invasive surgery [1]. Besides structured light [2] and stereo vision [3], Time-of-Flight (ToF) technology was manufactured into a first hybrid 3-D endoscope prototype, recently. In comparison to stereo vision, ToF is independent of texture information. Hence, the endoscope acquires range images with a constant resolution of 64×48 px. As the ToF sensor is manufactured into a conventional endoscope we additionally acquire high-resolution color images with 640×480 px through a common optical system using a beam splitter. Both range and complementary color information can be used to develop robust algorithms for image guided surgery. Haase et al. [4] proposed a tool localization framework that exploits range and color information for increased robustness. Nevertheless, as ToF technology exhibits a low signal-to-noise ratio, preprocessing is a required first step. Different preprocessing techniques have been proposed for ToF range images [5, 6], recently. For instrument segmentation approaches based on geometric information [7] or color information [8] have been investigated. This segmentation result can thereupon be used for further applications, e.g. the avoidance of risk situations as proposed in [9]. However, in comparison to purely 2-D driven approaches we are able to incorporate 3-D surface data as well as 2-D photometric data to improve robustness. Our preliminary framework describes a first approach towards entire instrument segmentation on 3-D surface information using a ToF/RGB endoscope.
We propose a multi-sensor instrument segmentation framework using super-resolution to denoise ToF data and increase spatial resolution [6]. Our framework exploits a data fusion for range and color images [10]. After upsampling ToF data, segmentation is performed on both modalities and then the results are consolidated into a common segmentation mask.

## 2    Material and Methods

The proposed segmentation framework is illustrated in Fig. 1 and is subdivided into (a) super-resolution and (b) multi-sensor segmentation. The preprocessing is applied according to our previous publication [6]. Our approach requires data fusion of range and color images. As the prototype uses a beam splitter to deliver signal to the RGB and the ToF sensor,

we map color information to the surface data using a homography that is estimated beforehand using a modified checkerboard [10]. This mapping results in a high-quality RGB image that is aligned to the range image up to a scale-factor.
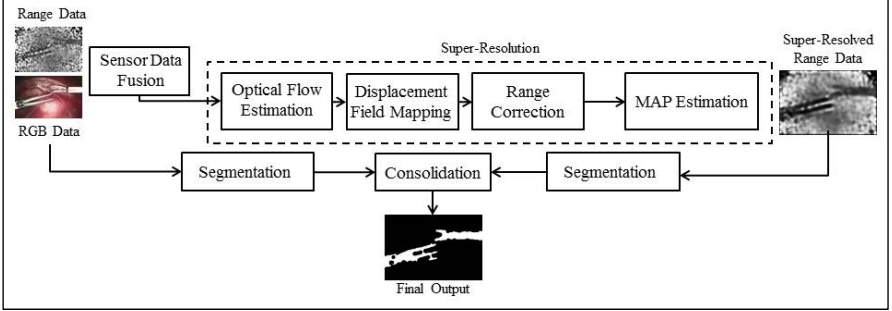
**Figure 1:** Flowchart for our instrument segmentation framework. First, sensor fusion and super-resolution are performed. Second, RGB data and range data are segmented separately. Third, both results are consolidated.

## 2.1 Super-Resolution for Range Image Preprocessing

We cope with the low signal-to-noise ratio of the ToF images by applying super-resolution as described in [6]. The super-resolution approach is subdivided into motion estimation, range correction and numerical optimization. Multi-frame super-resolution employs subpixel displacements between consecutive frames as a cue to obtain a super-resolved image from multiple low-resolution frames. These displacements are induced by navigating the endoscope. Our objective function to obtain a maximum a-posteriori (MAP) estimate $\hat{x}$ for a high-resolution image $x$ is described by:

$$\hat{x} = \arg\min_{x} \left( \sum_{k=1}^{K} \left\| y^{(k)} - \gamma_m^{(k)} W^{(k)} x - \gamma_a^{(k)} \mathbf{1} \right\|_2^2 + \lambda \sum_{n=1}^{N} h_\tau((Dx)_n) \right).$$

The first sum denotes the data term and the second sum is a regularizer based on a pseudo Huber loss function $h_\tau$ of a high-pass filtered version of the input image $x$. $\lambda$ weights the regularizer and $K$ denotes the number of low-resolution input frames and $N$ denotes the number of pixels in the super-resolved output image. The data term describes the distance of the $k^{\text{th}}$ low-resolution input frame $y^{(k)}$ and a mathematical model of our image acquisition.

The system matrix $W^{(k)}$ incorporates blur induced by the point spread function, downsampling and the displacement field of a high-resolution image $x$. As the low signal-to-noise ratio of ToF data limits the accuracy of displacement field estimation, we exploit data fusion to estimate a high-quality displacement field in the color domain using optical flow [11] and transfer it into the range domain. As we acquire images from different angles and distances we have to correct the range data, to have all low-resolution range images in the same plane. This correction is modeled by $\gamma_m^{(k)}$ and $\gamma_a^{(k)}$. For more details considering the multi-sensor super-resolution see Köhler et al. [6].

## 2.2 Multi-Sensor Segmentation

Based on the output of the preprocessing we apply instrument segmentation on data of both modalities. We distinguish between instruments and background by different thresholding techniques [8]. For our segmentation we exploit the fact that instruments are usually closer to the sensor and that instruments are usually grayish. Due to the data fusion in our hybrid 3-D endoscope, we can not only exploit the range data but also incorporate the color information into the segmentation process similar to [9]. Range values $y$ are considered as instruments pixels if $y \leq \alpha \cdot \max(y)$. In the color domain we exploit the value and the saturation channel of the HSV color space to segment the instrument. Here, pixels $z$ are considered as instrument pixels if $z_S \leq \beta \cdot \max(z_S)$ and $z_V \geq \delta \cdot \max(z_V)$, where $z_S$ and $z_V$ denote the saturation channel and the value channel of the color image, respectively. Both binary results are then consolidated into a common segmentation mask by multiplication. For outlier removal caused by noisy data we apply morphological operators to close small holes and remove separated areas with less than 1000 instrument pixels as false instrument candidates.

## 2.3 Experimental Setup

Our algorithm is evaluated on real data with a realistic liver phantom. Data was acquired with a ToF/RGB endoscope manufactured by Richard Wolf GmbH, Knittlingen, Germany. We assembled realistic scenarios including two different endoscopic instruments. For evaluation we investigated the results in two different scenarios, for 6 frames each. The upsampled images had a resolution of 240×160 px. Our instrument segmentation is compared to segmentation for both modalities separately. For ground truth data, the endoscopic instruments were manually segmented by an expert in the

color domain. The threshold parameters were set empirically to          ,          and          by analyzing the first frame. This frame was excluded for further evaluation to separate between training and evaluation data.

## 3    Results

For quantitative evaluation, we analyzed the sensitivity, the specificity and the F-score of our approach in Table 1. Here, we compare our segmentation results to the results of our framework for a purely range driven approach based on super-resolution and for a purely color driven approach.

For qualitative evaluation we illustrate the results of all three approaches in Fig. 2. The benefit of super-resolution for our noisy range data is shown in Fig. 3 with the color overlay encoding the segmentation result.

|  | Super-Resolved (1) | Color (2) | Super-Resolved & Color (3) |
|---|---|---|---|
| Sensitivity | 0.93 | 0.91 | 0.85 |
| Specificity | 0.67 | 0.80 | 0.91 |
| F-Score | 0.50 | 0.61 | 0.73 |

**Table 1:** Comparison of our approach (3) to segmentation on super-resolved data only (2) or on color data only (3).
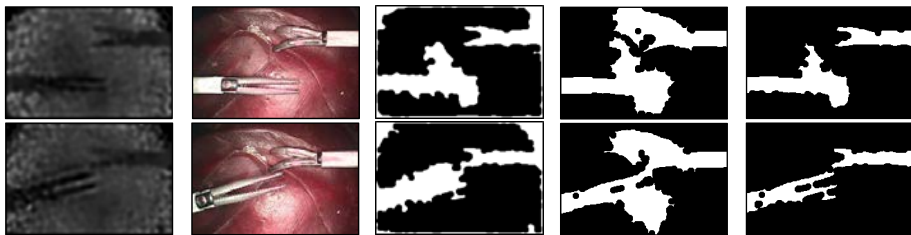


**Figure 2:** Input data (first and second column) and comparison of three segmentations: our approach on super-resolved data only (third column), on color data only (fourth column), on color and range data (last column).
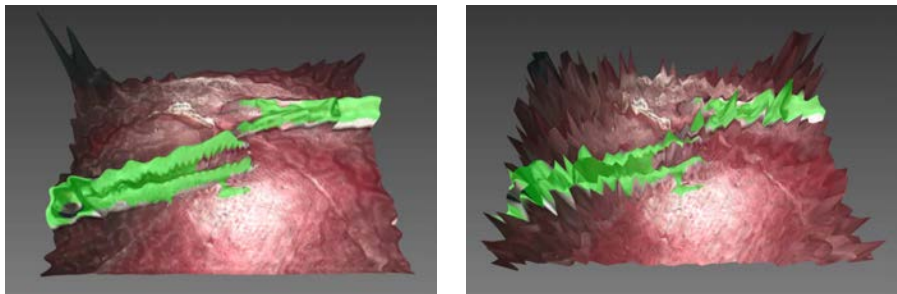


**Figure 3:** 3-D meshes with (left) and without (right) the use of super-resolution. The greenish overlay in both images is the segmentation mask of the proposed approach. [See the electronic publication for a color version of this figure. ]

## 4    Discussion

Table 1 illustrates that our approach results in the best specificity, i.e. only few background pixels are considered as instrument pixels. However, both single-sensor approaches result in satisfying sensitivities, i.e. only few instrument pixels are missing. Nevertheless, the F-score as a measurement of the accuracy indicates a more reliable performance of our approach. Furthermore, as our approach consolidates both modalities, we have a higher robustness considering different threshold parameters. Oversegmentation in one modality can be compensated by the other modality. The qualitative results confirm the comparison in Tab. 1 and highlight that both modalities oversegment the image in areas close to the sensor with surface normals pointing directly to the camera. In those areas the instruments are too close to the tissue to

be distinguished in the range image, but specular highlights exclude the use of the color image, likewise. Our approach achieves a reasonable compromise, where only few instrument pixels are missed and oversegmentation is reduced. The 3-D reconstructions show that most parts of the instruments are segmented correctly in our approach and that preprocessing is required to provide any intuitive visualization.

## 5    Summary

In this paper we proposed an instrument segmentation framework for 3-D ToF/RGB endoscopy. Our method applies robust multi-sensor super-resolution based on motion estimation on high-resolution RGB images to upsample and to denoise low-resolution range images. Due to improved signal-to-noise ratio of the range images we apply instrument segmentation using thresholding techniques and consolidate the results of both modalities. Compared to purely color driven segmentation we improved the F-score from 0.61 to 0.73.

Future work will consider different segmentation techniques and refinement of our super-resolution for further denoising. For the consolidation of both sensor results additional weighting factors will be taken into account as proposed in [4]. In experiments on real organs, we will investigate the robustness of our segmentation in real medical scenarios.

## 6    Acknowledgments

## 7    References

[1]    Röhl, S., Bodenstedt, S., Suwelack, S., Kenngott, S., Mueller-Stich, P., Dillmann, R., Speidel, S., *Real-time surface reconstruction from stereo endoscopic images for intraoperative registration*, Proc SPIE, Volume 7964, 796414-796414 (2011)

[2]    Schmalz, C., Forster, F., Schick, A., Angelopoulou, E., *An Endoscopic 3D Scanner based on Structured Light*, Med Image Anal 16(5), 1063-1072 (2012)

[3]    Field, M., Clarke, D., Strup, S., Seales, W., *Stereo Endoscopy as a 3-D Measurement Tool*, EMBC 2009, 5748-5751 (2009)

[4]    Haase, S., Wasza, J., Kilgus, T., Hornegger, J., *Laparoscopic Instrument Localization using a 3-D Time-of-Flight/RGB Endoscope*, WACV 2013, 449-454 (2013)

[5]    Lenzen, F., In Kim, K., Nair, R., Meister, S., Schäfer, H., Becker, F., Garbe, C., Theobalt, C., *Denoising Strategies for Time-of-Flight Data*, Time-of-Flight Imaging: Algorithms, Sensors and Applications (2012)

[6]    Köhler, T., Haase, S., Bauer, S., Wasza, J., Kilgus, T., Maier-Hein, L., Feußner, H., Hornegger, J., *ToF Meets RGB: Novel Multi-Sensor Super-Resolution for Hybrid 3-D Endoscopy*, MICCAI, LNCS 8149, To Appear (2013)

[7]    Climent, J., Mares, P., *Automatic Instrument Localization in Laparoscopic Surgery*, Electronic Letters on Computer Vision and Image Analysis, 4(1):21-31 (2004)

[8]    Doignon, C.; Nageotte, F.; De Mathelin, M., *Detection of grey regions in color images : application to the segmentation of a surgical instrument in robotized laparoscopy*, Proc of IROS, Volume 4, 3394-3399 (2004)

[9]    Speidel, S., Sudra, G., Senemaud, J., Drentschew, M., Müller-Stich, B. P., Gutt, C., Dillmann, R., *Recognition of Risk Situations Based on Endoscopic Instrument Tracking and Knowledge Based Situation Modeling*, Proc SPIE, Volume 6918, 69180X-69180X-8 (2008)

[10]    Haase, S., Forman, C., Kilgus, T., Bammer, R., Maier-Hein, L., Hornegger, J., *ToF/RGB Sensor Fusion for 3-D Endoscopy*, Current Medical Imaging Reviews 9, 113-119 (2013)

[11]    Liu, C., *Beyond Pixels: Exploring New Representations and Applications for Motion Analysis*, PhD thesis, Massachusetts Institute of Technology (2009)