Laparoscopic Instrument Localization using a 3-D Time-of-Flight/RGB Endoscope

Sven Haase[1], Jakob Wasza[1], Thomas Kilgus[2], Joachim Hornegger[1,3]
[1]Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
[2]Division of Medical and Biological Informatics, DKFZ Heidelberg
[3]Erlangen Graduate School in Advanced Optical Technologies (SAOT)
sven.haase@cs.fau.de

Abstract

Minimally invasive procedures are of importance in modern surgery due to reduced operative trauma and recovery time. To enable robot assisted interventions, automatic tracking of endoscopic tools is an essential task. State-of-the-art techniques rely on 2-D color information only which is error prone for varying illumination and unpredictable color distribution within the human body. In this paper, we use a novel 3-D Time-of-Flight/RGB endoscope that allows to use both color and range information to locate laparoscopic instruments in 3-D. Regarding color and range information the proposed technique calculates a score to indicate which information is more reliable and adopts the next steps of the localization procedure based on this reliability. In experiments on real data the tool tip is located with an average 3-D distance error of less than 4 mm compared to manually labeled ground truth data with a frame-rate of 10 fps.

1. Introduction

Minimally invasive procedures have become popular in the community of abdominal surgery. Compared to conventional open surgery, endoscopic interventions aim at reducing pain, scars and recovery time and thereby hospital stays. Therefore, minimally invasive procedures hold benefits for surgeons as well as patients. To observe the operative scene, an endoscope is usually navigated by a medical assistant to follow the intervention. Thus, the position of the endoscope is error prone to the jitter of the assistant. To eliminate this issue, robot assisted surgery gained popularity with the deployment of the Da Vinci or Zeus system in medical environments. For guiding robotic systems, research on non-invasive imaging techniques for medical applications has evolved within the last decade [1, 7]. Utilizing either structured light, stereo vision or Time-of-Flight (ToF) technology, topological information can be acquired in a markerless and non-invasive manner. With the development of stereo endoscopy, topology measuring devices were introduced for minimally invasive procedures. These devices enable several new applications like metric measurements, the recognition of risk situations [9] or identifying and tracking endoscopic tools [3, 10, 12]. The latter is of importance for robot assisted minimally invasive procedures.

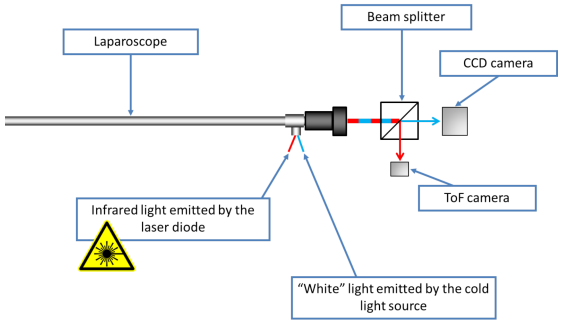


Figure 1. ToF/RGB endoscope system using a beam splitter to acquire data through one optical system.

In this paper we present an approach for locating endoscopic tools using the first ToF/RGB 3-D endoscope that was recently developed. Similar to stereo endoscopes this prototype is capable to acquire conventional 2-D color information enhanced by 3-D metric range information. Compared to a previously proposed ToF endoscope [8] the complementary information is acquired through a single optical system using a beam splitter as depicted in Fig. 1. This enables a mapping of the color information onto the range data without the issues of conventional stereo vision and thereby allows to locate instruments within one common coordinate system.

Our algorithm is based on prior knowledge about

978-1-4673-5052-5/12/$31.00 ©2012 IEEE

the color as well as the geometry of instruments used in minimally invasive procedures. The novelty of our approach is to exploit both color and range information of the new ToF/RGB endoscope to increase robustness and to enable 3-D localization of endoscopic tools. To increase reliability, a scoring system of intermediate results is introduced. This allows to assess the results between both modalities and enables the automatic adoption for subsequent steps to be performed on the best results of the previous calculations.

Related Work Approaches to localizing endoscopic tools can be split into two groups: Tool segmentation and tracking based on color features [3] or by using prior knowledge about the tool geometry [2, 9, 12].

Climent and Mares [2] proposed a technique that relies on the Hough transformation to find straight lines indicating the presence of endoscopic instruments. This approach combined with an heuristic filter achieves robust results being capable to detect the tool tip in 99% of all evaluated cases. However, to reach interactive frame-rates and high robustness, several restrictions have been made, e.g. only radial lines are considered as tool candidates. Furthermore, this technique processes color images only and is thereby error-prone to inhomogeneous illumination. Doignon et al. [3] proposed another technique for recognizing endoscopic tools based on a joint hue saturation color feature. Their approach takes prior knowledge about the color of the endoscopic tool into account to perform an adaptive region growing with the seed point automatically detected by the algorithm. For this purpose, the fact that the endoscopic tool enters the scene from the boundary is utilized to search the border for minima in the joint color feature space as seed candidates. However, due to possible occlusions this assumption does not always hold true for color images. Recently, a technique relying on statistical and geometric modeling was proposed by Wolf et al. [12]. They utilize the insertion point in 3-D and the conventional 2-D color information for more robust tool tracking and enabled tracking in 3-D compared to Climent [2]. Another method for 3-D localization was proposed in [9] where the tool tip was located in 2-D using a stereo endoscope and in a final step its 3-D location was estimated using stereo vision. Our approach extends all those concepts by combining range and color information and utilizing a scoring system for higher robustness.

2. Materials and Methods

In this section we describe our approach to locate endoscopic tools in a multi-modal manner using color and range data. In the following sections $I(\boldsymbol{x})$ always denotes the value at position $\boldsymbol{x}$ in an image identified by its subscript. The scoring value is defined within $[0;1]$ and depicted by S with the subscript denoting the current step. Finally, $\boldsymbol{x}^c$ marks a candidate point of the current step and $\boldsymbol{x}^r$ denotes a point indicating a reliable result for the next step. In this paper we denote the part of the instrument attached to the shaft as the *tool tip*. Regions of interest are abbreviated as *ROI*s.

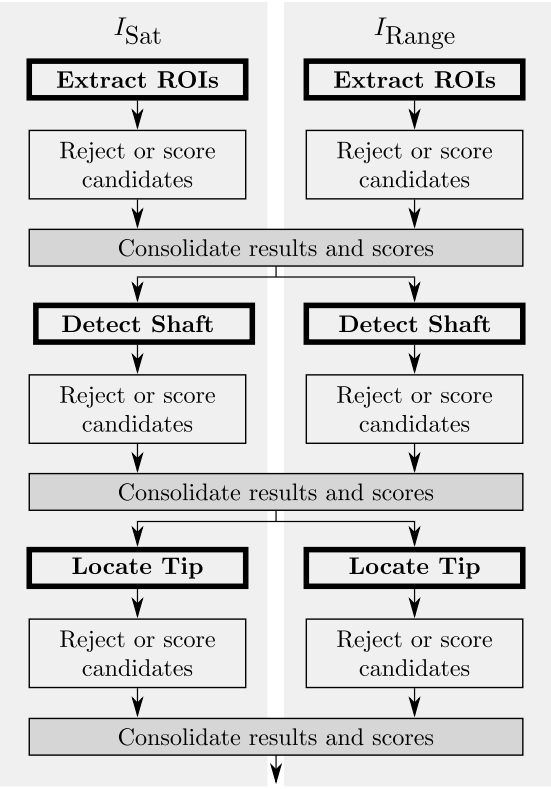


Figure 2. Each of the localization is computed for both modalities denoted by I_{Sat} for the saturation image and by I_{Range} for the range image. After rejecting false candidates and scoring, the results are fused for the next step.

2.1. Preprocessing Pipeline

ToF devices exhibit a low signal-to-noise ratio due to multiple error sources, such as temporal noise or systematic offsets [6]. Therefore, preprocessing the range images is essential for robust tool localization. Due to the high frame-rate of our endoscope and time constraints in a medical environment, we use a real-time preprocessing pipeline similar to [11] for the range images. The color information was denoised using edge-preserving guided filtering [5]. These color images were thereupon used for guided upsampling the ToF images in a joint manner. This allows to work in the same domain for both modalities. For correct texture mapping

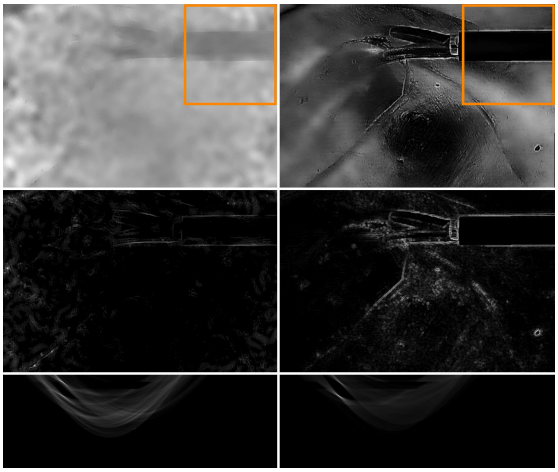


Figure 3. Range image outputs (first column), RGB image outputs (second column). Range and modified saturation input (first row), corresponding edge images (second row), Hough image for the calculated regions of interest denoted by the orange boxes (third row).

an intrinsic and extrinsic calibration is performed [4].

As endoscopic tools are in general grayish, the saturation space is a suitable representation of the color image to distinguish between instruments and body tissue. The saturation image is calculated as:

$$I_{\text{Sat}}(\boldsymbol{x}) = 1 - \frac{3}{I_R(\boldsymbol{x}) + I_G(\boldsymbol{x}) + I_B(\boldsymbol{x})} I^{\min}_{\text{RGB}}, \quad (1)$$

with $I^{\min}_{\text{RGB}}$ denoting the minimal intensity of the three color channels.

2.2. Generic Localization Algorithm

To increase robustness we developed a generic algorithm that can be applied to both color and range information likewise. The algorithm is divided into three steps, each followed by a rejection phase, a scoring phase and a consolidation of all intermediate results as depicted in Fig. 2. First, ROIs are extracted that indicate potential locations of laparoscopic tools. Second, the two lines defining the shaft of the endoscopic tools are detected. Third, the tool tip is located along the centerline of the shaft.

ROI Extraction To reduce computation time and increase robustness all ROIs are to be found where endoscopic tools are expected. We exploit the fact that tools enter the scene from the boundary of the image. This allows us to reduce the search space for the ROIs by analyzing pixels along the border of the images only. As values in range images represent distances to the sensor, small values indicate close points. In saturation images low intensity values indicate uncolored pixels being a typical property of laparoscopic instruments. Thus, detecting local minima along the border for both modalities results in a hypothesis generation $\boldsymbol{x}^{\boldsymbol{c}}$ of an endoscopic tool.

After finding all $\boldsymbol{x}^{\boldsymbol{c}}$ a twofold rejection phase is performed. First, the neighborhood of all $\boldsymbol{x}^{\boldsymbol{c}}$ is expected to have a similar value and therefore is analyzed by its variance. Second, candidates that have an intensity value above the mean μ_{Input} of the corresponding input images I_{Sat} and I_{Range} are determined as unreliable and therefore rejected. Then a simple clustering is performed to fuse candidates of close mutual proximity that refer to the same tool. The representative of each cluster is denoted by $\boldsymbol{x}^r$.

The scoring value for each $\boldsymbol{x}^r$ is calculated as:

$$S_{\text{ROI}}(\boldsymbol{x}^r) = 1 - \frac{I(\boldsymbol{x}^r)}{\mu_{\text{Input}}}. \quad (2)$$

This scoring value converges to 1 the lower the located minimum is compared to the mean of the whole image. The size of the ROIs is defined by δ denoting a fraction of the input image size. The ROIs are determined with the previously found initial points $\boldsymbol{x}^{\boldsymbol{c}}$ being the center.

Shaft Detection On the saturation as well as the range images we apply the Sobel operator to find edges as depicted in Fig. 3. For each of the previously detected ROIs the gradient image is then transformed into Hough space to detect noticeable lines in polar coordinates. This step requires the instrument to be rigid with a cylindrical shaft, which is a valid assumption for laparoscopic tools. Calculating the Hough transformation of an ROI results in two high peaks in Hough space that point to the location of the two lines describing the boundary of the shaft. The crucial step of this part is finding these maxima. To find two separated peaks, first the global maximum in Hough space is found and then the second maximum outside the neighborhood of the first maximum is located.

To reject false candidates $\boldsymbol{x}^{\boldsymbol{c}}$ in Hough space the angle between the corresponding lines is analyzed. Within a tolerable range due to perspective distortion shaft lines are expected to be parallel. Thus, if the angle exceeds a threshold $\Delta\varphi$ the line referred by the second maximum in Hough space is rejected. Otherwise both candidates $\boldsymbol{x}^{\boldsymbol{c}}$ are accepted as reliable $\boldsymbol{x}^r$.

For scoring shaft lines the intensity of the peaks $I_{\text{Hough}}(\boldsymbol{x}^r)$ are evaluated, as a higher intensity value in Hough space indicates more points being assigned to that line. Therefore, the score is computed as:

$$S_{\text{Shaft}}(\boldsymbol{x}^r) = 1 - \frac{\mu_{\text{Hough}}}{I_{\text{Hough}}(\boldsymbol{x}^r)}, \quad (3)$$

with μ_{Hough} denoting the mean value of votings in Hough space. A higher peak compared to the mean results in an increased score.

Tool Tip Localization Finally, we assume that the tip of the endoscopic tool needs to be located along the centerline between both detected shaft lines denoted by a steep gradient on this line. The point with the highest gradient indicates a step from the tool to the background.

As we assume smooth movements between successive frames, the criterion for rejection of the tip location is its distance to the location of the previous frame. Therefore, a located point $\boldsymbol{x}^c$ is considered as a reliable point $\boldsymbol{x}^r$ if the 3-D distance to its location in the previous frame is below a threshold ϵ. The reliability of $\boldsymbol{x}^r$ at the tip of the endoscopic tool is computed as:

$$S_{\text{Tip}}(\boldsymbol{x}^r) = 1 - \frac{\mu_{\text{Sobel}}}{I_{\text{Sobel}}(\boldsymbol{x}^r)}, \tag{4}$$

with μ_{Sobel} denoting the mean of all edge pixels of the input image. A strong gradient compared to all other edges results in a score converging to 1.

2.3. Combining Range and Color Localization

To increase robustness the result of each step is combined for both imaging modalities. For consolidation the scoring is analyzed and a weighting α denoting the reliability of the range sensor is used for each step. α depends on the hardware and scenario. The combined results serve as an input for the next step for both modalities. If their results $\boldsymbol{x}^r$ are similar, the final score $\hat{S}$ for this step is calculated as:

$$\hat{S} = \alpha S_{\text{Range}} + (1-\alpha) S_{\text{Sat}}, \tag{5}$$

and the consolidated point $\hat{x}^r$ is given as:

$$\hat{x}^r = \alpha x^r_{\text{Range}} + (1-\alpha) x^r_{\text{Sat}}. \tag{6}$$

Otherwise, if S of a single modality weighted by α still exceeds a threshold γ, $\hat{S}$ is calculated as:

$$\hat{S} = \begin{cases} \alpha S_{\text{Range}} & \text{, if } \alpha S_{\text{Range}} \geq (1-\alpha) S_{\text{Sat}} \\ (1-\alpha) S_{\text{Sat}} & \text{, if } (1-\alpha) S_{\text{Sat}} > \alpha S_{\text{Range}} \end{cases} \tag{7}$$

The final result $\hat{x}^r$ is then equal to $\boldsymbol{x}^r$ of this modality. If in any step neither similar values nor dominant values for a single modality are found the possible candidate for an endoscopic tool is rejected and the localization procedure for this point is aborted.

2.4. Experiments

All experiments were performed using a 3-D endoscope prototype manufactured by Richard Wolf GmbH, Knittlingen, Germany. The prototype acquires ToF (64×50 px) and RGB (640×480 px) data simultaneously through one optical system with a frame-rate of 30 fps. As this endoscope is still in an early prototype stage, all experiments are performed ex-vivo using a liver phantom and real endoscopic tools.

For all scenarios we set the angular threshold $\Delta\varphi$ to 25°, the scoring threshold γ to 0.2 and the size of each ROI denoted by δ to 25% of the input image size. The threshold ϵ defining smooth movements was set to 5 mm. As in our experiments usually the RGB data showed more reliable results for the initial point the weighting α_{ROI} for the first step was set to 45%. α_{Shaft} was set to 50%. Due to material properties of the instruments, tool tips in the saturation image are usually located at the beginning of the tool tip, whereas the step in range data is expected at the very end of the tool. Therefore, both final tool tip locations differ even though they may describe the same tool tip. As the gradient information of the color image showed more reliable edges α_{Tip} was set to 0.

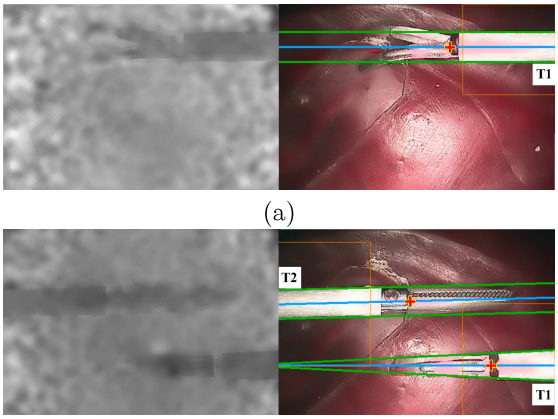
Sequences We evaluate the accuracy of our approach on 15 frames for two different scenarios. These scenarios include a scene with a single endoscopic tool in Fig. 4 (a) and a scene with two endoscopic instruments inserted from different directions in Fig. 4 (b). For quantitative results, the 2-D and 3-D euclidean distance was calculated between the located tool tip and the ground truth data manually labeled on the fused sensor data.

Challenging Scenarios The robustness of our algorithm was evaluated on single frames showing challenging scenarios. These scenarios include intersection of two tools, blood splatter on a tool, occlusion by the surrounding tissue and the absence of any endoscopic tool. These experiments are evaluated by their scoring to show the influence and reliability of each modality.

3. Results

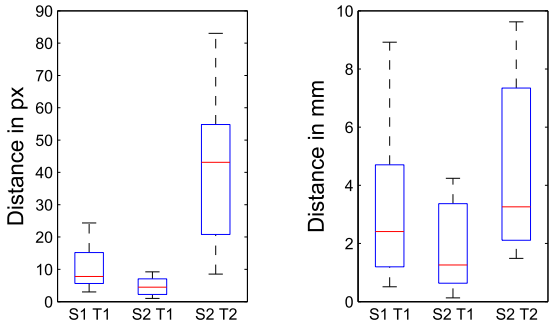
Locating a single endoscopic tool in real data as shown in Fig. 4 was performed with a frame-rate of ∼10 fps using an NVIDIA Quadro FX 1800M.

Sequences Fig. 4 (a) and (b) demonstrate an exemplary output of the first experiments described in Sect. 2.4. Both figures show the mapped color information and the range data acquired by the ToF/RGB



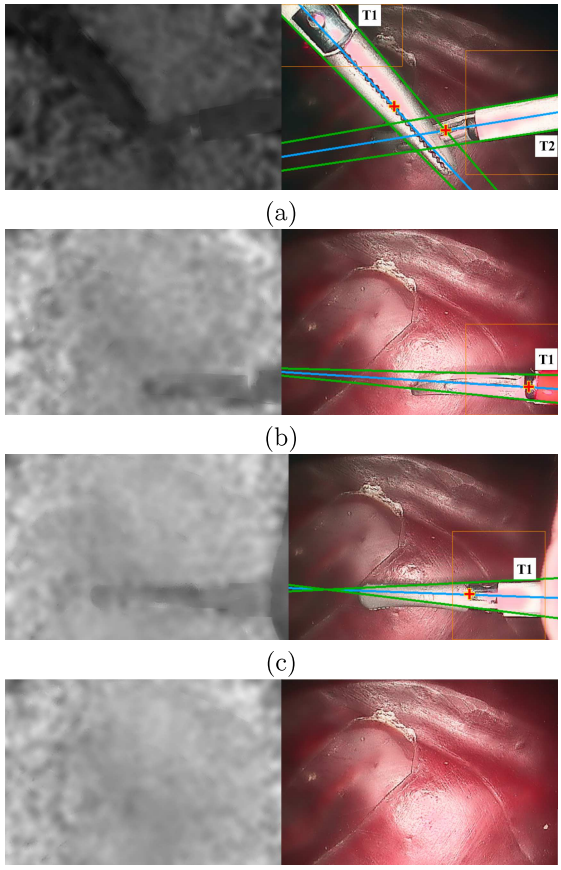
(a) (b)

Figure 4. Range and color images of sequences S1 (a) and S2 (b) acquired for evaluation of the distance error. The red cross marks the detected tool tip, green the shaft boundaries, blue the centerline, orange the regions of interest.



(a) (b)

Figure 5. Distance errors in 2-D and 3-D comparing manually labeled and automatic located tool tips for both sequences (S1, S2) in box-and-whisker diagrams. The median value of all 15 distances of each tool (T1, T2) is colored in red.



(a) (b) (c) (d)

Figure 6. The challenging scenarios described in Sect. 2.4. (a) intersection of tools, (b) blood on the shaft, (c) occlusion at the border, (d) the absence of tools. The red cross marks the detected tool tip, green the shaft boundaries, blue the centerline, orange the regions of interest.

endoscope. The color overlay of the RGB image shows a qualitative outcome of the proposed technique. In Fig. 5 (a) and (b) the euclidean distances between the manually labeled ground truth and the automatically detected tool tips are shown. Note that tool T2 of the second sequence S2 being closer to the light source resulted in varying saturation intensities and thereby led to an increased distance error. Depending on the lighting and reflections the detected locations of the tool tip differ. Furthermore, we want to remark that even points with a high distance to our ground truth are valid points on the tool tip. Due to its closeness to the sensor acceptable 3-D distances may result in noticeable 2-D pixel distances. The mean distance error for both scenes resulted in approximately 23 pixels for the 2-D domain and 3.11 mm for the 3-D domain.

Challenging Scenarios Fig. 6 demonstrates the four challenging scenarios described in Section 2.4. Note that in all scenarios the existing endoscopic instruments were found. Table 1 shows the scoring values of all scenes and all endoscopic tools. As described in Sect. 2.3 the final scoring value is either a weighted average of both modalities if both results coincide, a weighted scoring of a single modality if this scoring exceeds γ. If a field contains no value, all potential candidates are rejected. Note that in the second scene the blood on the tool caused our algorithm to find no valid initial point in the color image due to the increased saturation on the tool. Weighting the initial point detected in the range image by α_{ROI} we determine the correct initial point while the false candidate of the color image is rejected.

Scoring	Intersection			Blood			Occlusion			Absence		
	Sat	Range	Comb	Sat	Range	Comb	Sat	Range	Comb	Sat	Range	Comb
S_{ROI} (T1)	-	0.96	0.43	-	0.49	0.22	0.69	0.71	0.70	0.25	-	-
S_{Shaft} (T1)	0.65	0.86	0.43	0.77	0.69	0.73	0.80	0.74	0.40	-	-	-
S_{Tip} (T1)	0.89	0.90	0.89	0.88	0.95	0.88	0.90	0.99	0.90	-	-	-
S_{ROI} (T2)	0.87	-	0.48	-	-	-	-	-	-	-	0.16	-
S_{Shaft} (T2)	0.88	0.85	0.87	-	-	-	-	-	-	-	-	-
S_{Tip} (T2)	0.89	0.95	0.89	-	-	-	-	-	-	-	-	-

Table 1. The scoring results S of each intermediate step for both saturation and range image separately and combined.

4. Discussion

In this paper we proposed an algorithm to locate endoscopic tools by using a novel 3-D ToF/RGB endoscope. Our method benefits from both color and range information to achieve robust results even in challenging scenarios. By introducing a scoring system and consolidation for each intermediate step we were able to reject false candidates and detect all existing tools in our experiments. Compared to manually labeled ground truth data the proposed algorithm is able to locate the tool tip within a 3-D distance of less than 4 mm with a frame-rate of 10 fps using an off-the-shelf graphics card.

Due to the early stage of the ToF/RGB endoscope prototype, the next generation of hardware is expected to deliver more robust data with higher resolution. For future research the detected tool tip location may serve as an input for a tracking algorithm to improve robot assisted surgery. Furthermore, a fully automatic determination of valid parameter values will be investigated.

Acknowledgments

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.

References

[1] D. Cash, M. Miga, S. Glasgow, B. Dawant, L. Clements, and et al. Concepts and preliminary data toward the realization of image-guided liver surgery. *J Gastrointest Surg*, 11(7):844–859, 2007.

[2] J. Climent and P. Mares. Automatic instrument localization in laparoscopic surgery. *Electronic Letters on Computer Vision and Image Analysis*, 4(1):21–31, 2004.

[3] C. Doignon, P. Graebling, and M. de Mathelin. Real-time segmentation of surgical instruments inside the abdominal cavity using a joint hue saturation color feature. *Real-Time Imaging*, 11(5-6):429–442, 2005.

[4] S. Haase, C. Forman, T. Kilgus, R. Bammer, L. Maier-Hein, and J. Hornegger. ToF/RGB Sensor Fusion for Augmented 3-D Endoscopy using a Fully Automatic Calibration Scheme. In *Bildverarbeitung für die Medizin*, pages 111–116, 2012.

[5] K. He, J. Sun, and X. Tang. Guided Image Filtering. In *Proceedings of European conference on computer vision: Part I*, volume 6311, pages 1–14. Springer, 2010.

[6] A. Kolb, E. Barth, R. Koch, and R. Larsen. Time-of-flight cameras in computer graphics. *Computer Graphics Forum*, 29(1):141–159, 2010.

[7] S. Mersmann, M. Müller, A. Seitel, et al. Time-of-flight camera technique for augmented reality in computer-assisted interventions. In *Proc SPIE*, volume 7964, pages 79642C–79642C–9, 2011.

[8] J. Penne, K. Höller, M. Stürmer, T. Schrauder, A. Schneider, R. Engelbrecht, H. Feußner, B. Schmauss, and J. Hornegger. Time-of-Flight 3-D Endoscopy. In G.-Z. Y. et al., editor, *MICCAI 2009, Part I, LNCS 5761*, volume 5761, pages 467–474, Berlin / Heidelberg, 2009.

[9] 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, pages 69180X–69180X–8, 2008.

[10] S. Voros, J.-A. Long, and P. Cinquin. Automatic detection of instruments in laparoscopic images: A first step towards high-level command of robotic endoscopic holders. *Int. J. Rob. Res.*, 26(11-12):1173–1190, 2007.

[11] J. Wasza, S. Bauer, and J. Hornegger. Real-time preprocessing for dense 3-D range imaging on the GPU: Defect interpolation, bilateral temporal averaging and guided filtering. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 1221 –1227, nov. 2011.

[12] R. Wolf, J. Duchateau, P. Cinquin, and S. Voros. 3d tracking of laparoscopic instruments using statistical and geometric modeling. In *MICCAI 2011, Part I, LNCS 6891*, volume 6891 of *Lecture Notes in Computer Science*, pages 203–210. Springer Berlin Heidelberg, 2011.