

Superpixel-based Structure Classification for Laparoscopic Surgery

Sebastian Bodenstedt^a, Jochen Görtler^a, Martin Wagner^b, Hannes Kenngott^b, Beat Peter Müller-Stich^b, Rüdiger Dillmann^a, and Stefanie Speidel^a

^aInstitute for Anthropomatics and Robotics, Karlsruhe Institute of Technology (KIT), Adenauerring 2, D-76131 Karlsruhe, Germany

^bDepartment of General, Visceral and Transplantation Surgery, University of Heidelberg, Im Neuenheimer Feld 110, D-69120 Heidelberg, Germany

ABSTRACT

Minimally-invasive interventions offers multiple benefits for patients, but also entails drawbacks for the surgeon. The goal of context-aware assistance systems is to alleviate some of these difficulties. Localizing and identifying anatomical structures, maligned tissue and surgical instruments through endoscopic image analysis is paramount for an assistance system, making online measurements and augmented reality visualizations possible. Furthermore, such information can be used to assess the progress of an intervention, hereby allowing for a context-aware assistance. In this work, we present an approach for such an analysis. First, a given laparoscopic image is divided into groups of connected pixels, so-called superpixels, using the SEEDS algorithm. The content of a given superpixel is then described using information regarding its color and texture. Using a Random Forest classifier, we determine the class label of each superpixel. We evaluated our approach on a publicly available dataset for laparoscopic instrument detection and achieved a DICE score of 0.69.

Keywords: Superpixel classification, Instrument detection, Tissue classification, Laparoscopic surgery, Endoscopic image analysis, Endoscopic image segmentation

1. INTRODUCTION

For the patient, minimally-invasive interventions offer multiple benefits in comparison to conventional surgery, e.g. smaller incisions and a shorter stay in hospital. For the surgeon, on the other hand, such an intervention does entail certain drawbacks, such as a loss of depth-perception and a more challenging hand-eye-coordination.

The purposes of a computer-assisted surgery system is to minimize the effect of these drawbacks, for example through quantitative Endoscopy or augmented reality.¹ An assistance system requires knowledge about the surgical scene, such as the locations of surgical instruments, anatomical structures and maligned tissue. Such information can be visualized using augmented reality or it can be used for on-the-fly measurements. Furthermore, such information can also be used to achieve context-awareness, making it possible to track the progress of the intervention and to provide the right assistance at the right instance.

During minimally-invasive interventions, the endoscopic image stream is readily available making, it therefore suggests itself to extract information regarding the progress of the intervention from this feed. Relevant anatomic structures, tumors and surgical instruments can be identified via endoscopic image analysis. Existing methods²³⁴⁵ assign a class label to each pixel and then fuse similar regions together. Information that is potentially contained in the neighborhood of a pixel is hereby disregarded.

In this paper, we present a method that takes neighborhood information into consideration during classification. We first segment the image into regions of connected, similar pixels, so-called superpixels. The content of these superpixels is then described using information regarding its color and texture. The superpixels and manually annotated laparoscopic images are then used to train a Random Forest classifier, which will then be

Further author information: (Send correspondence to S. Bodenstedt.)
S. Bodenstedt: E-mail: bodenstedt@kit.edu

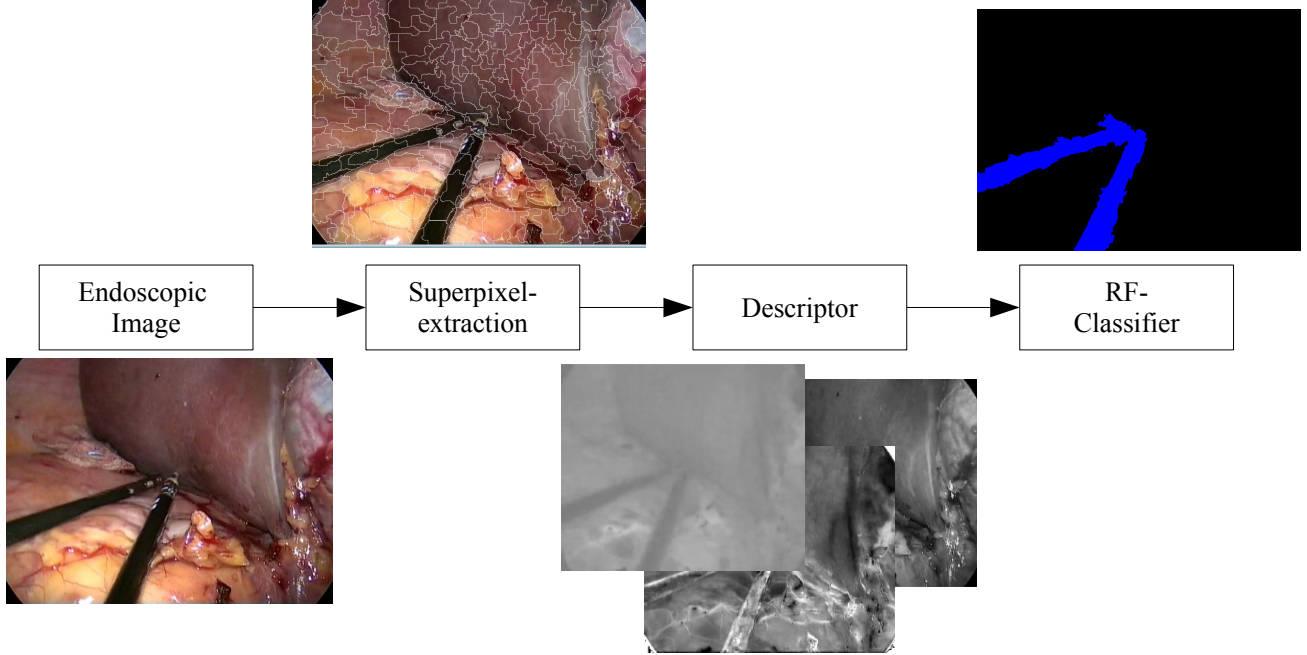


Figure 1. Overview of the classification system

used to assign a label to each superpixel. To our knowledge, this is the first time that a descriptor that describes the texture and color information of a superpixel has been constructed and used to classification.

We determine the optimal parameters for our approach and evaluate using a publicly available dataset for laparoscopic instrument detection.² We furthermore compare the classification results with a pixel-based approach.

The novelty of the approach presented in this work is our method of describing the content of superpixels with a descriptor consisting out of color information (provided by histograms constructed over multiple channels of different color spaces) and texture information using a rotation-invariant version of the local binary pattern and using these descriptors to assign class labels. To the best of our knowledge, this is the first time, a method like this has been used on laparoscopic image data. Furthermore we evaluate our approach on a publicly available dataset, making future comparisons possible.

2. METHODS

Our method is divided into 3 main steps (fig. 1). First a given laparoscopic image is divided into similar regions via superpixel extraction. For each superpixel, a descriptor is then computed using color and texture information. Each descriptor is then labeled using a Random Forest classifier.

2.1 Superpixel Extraction

Given an image with N pixels and let K be the number of wanted superpixels, then the division of one image into superpixels can be described with the following mapping:

$$s : \{1, \dots, N\} \rightarrow \{1, \dots, K\}$$

$s(i)$ is the ID of the superpixel to whom the pixel i belongs. A superpixel A_k can therefore be described in the following manner:

$$A_k = \{i : s(i) = k\}$$

Also let for two superpixels A_k and A_l with $k \neq l$: $A_k \cap A_l = \emptyset$

For computing $s(i)$, we used the SEEDS algorithm outlined in,⁶ due to it being able to compute $s(i)$ in real-time (30Hz with an image resolution of 640x480 pixels), while taking homogeneity and shape of potential superpixel candidate under consideration. An example of such a segmentation can be seen in fig. 2.

2.2 Descriptor

In order to decide what label should be assigned to a given superpixel, we first need a representation of its content. For this, we construct a feature vector (or descriptor) from color and texture information.

To describe the color information contained in a superpixel, we convert the original image into different color spaces (HSV, LAB and Opponent). Since the number of pixels per superpixel is not constant, a histogram with n bins is constructed for each channel of the aforementioned color spaces. n is a variable, whose value we will determine empirically. The combination of channel histograms used in the final descriptor is also determined empirically.

Local Binary Patterns (LBP)⁷ are used to describe the texture contained in a superpixel. A LBP describes for each pixel i the differences between its intensity value compared to those of the pixels in its 3x3 neighborhood via an 8D binary vector. If the intensity of i is large than that of its n -th neighbor, the n -th entry in the binary vector is 1, otherwise 0. The binary vector can be interpreted as an 8-bit integer, which would lead to an rotation-variant representation. Instead, we use the method outlined in⁸ to enter each binary vector into a 10-bin histogram according to the number and positions of 0 to 1 or 1 to 0 flips. We construct one histogram for each superpixel.

Since each histogram can be interpreted as a vector, we stack the previously constructed histograms on top of each other to receive the final superpixel descriptor.

2.3 Classifier

We trained a Random Forest Classifier⁹ to assign each superpixel descriptor a class label. For training the forest, annotated images, in which each pixel i is assigned one of L labels, are required.

$$l : \{1, \dots, N\} \rightarrow \{1, \dots, L\}$$

During training each image is first divided into K superpixels A_k . The label $L(A_k)$ for each superpixel A_k is then determined:

$$L(A_k) = \operatorname{argmax}_{l^* \in \{1, \dots, L\}} |\{p | p \in A_k \wedge l(p) = l^*\}|$$

For each superpixel, we then computed a descriptor, which are then used in combination with its label to train the Random Forest Classifier.

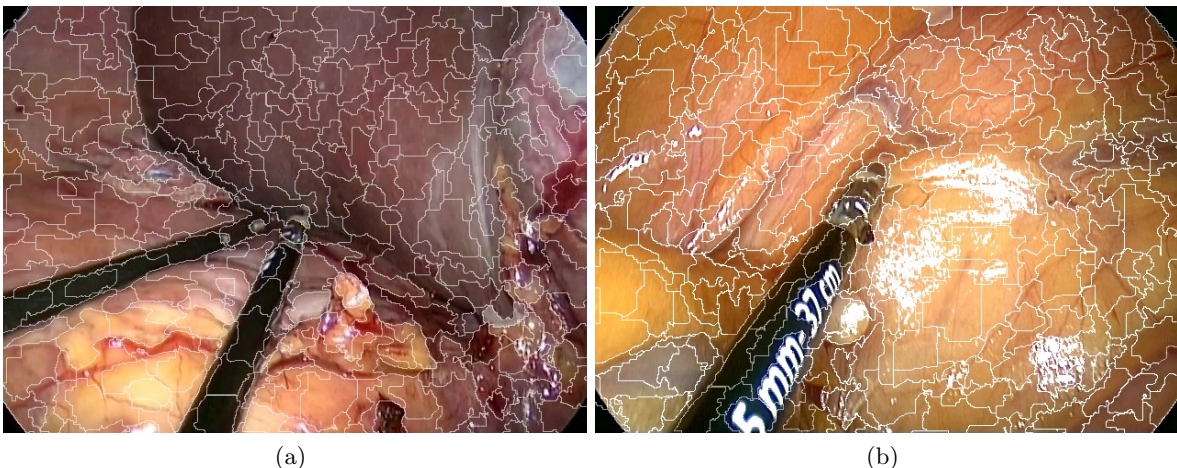


Figure 2. Example SEEDS segmentations of laparoscopic images

	Precision	Recall	DICE
Superspixel-based	0.69	0.73	0.69
Pixel-based ²	0.76	0.58	0.64

Table 1. Comparison of the average precision, recall and DICE coefficient of the superspixel-based and pixel-based classification methods.

3. RESULTS

Our method was developed under Linux in C++ on the basis of the OpenCV library.¹⁰ The evaluation was performed on a computer with 2x Intel Xeon E5-2620 processors and 64GB of RAM.

A publicly available dataset for laparoscopic instrument detection² was used as a basis for evaluation. The dataset contains 120 images from 6 different laparoscopic interventions. All instruments in the images were annotated with crowd-sourcing.

On the dataset, first we empirically determined the optimal parameters for the SEEDS algorithm and also the quality of segmentation. With these parameters, we empirically determined the optimal parameters for the descriptor and the classifier as well as the quality of classification.

3.1 Segmentation Parameter Evaluation

Given a superspixel segmentation $s(i)$ of an annotated image B with N pixels, we determine the quality of segmentation in the following manner:

$$e_s = \frac{|\{i | l(i) = L(A_k) \wedge i \in B \wedge i : s(i) = k\}|}{N}$$

or in other words, the ratio between the number of pixels, whose label correspondences with that of the assigned superspixel, and the number of pixels N . We use this ratio as the criterion for the parameter optimization of the SEEDS algorithm.

The largest value for e_s , 91.4%, was reached with the following parameters: 1000 superspixels, variance of 3, 1 level, 7 bins and 10 iterations. For a detailed description of these parameters please see.⁶

3.2 Classifier Parameter Evaluation

To determine the optimal parameters for the descriptor and the classifier, we performed a leave-one-surgery-out cross-validation for each parameter combination, meaning we trained on 5 interventions and tested on the 6th. As measure of quality, the DICE coefficient was used.

The highest DICE coefficient was reached with the following parameter combination:

- For the descriptor, hue saturation, o1 and o2 of the Opponent color space and LBP were selected. A bin-size of 25 was selected.
- For the Random Forest classifier a maximum depth of 16 and a maximum number of 200 were selected.

The average values for precision, recall and DICE coefficient can be found in table 1.

3.3 Runtime

Currently, our method has an average runtime of 118ms per image with a resolution of 640x480 pixels.

3.4 Discussion

We were able to show that, on a realistic, laparoscopic dataset, an improvement in classification results can be achieved using a superspixel-based method in comparison to a pixel-based method. Table 1 shows that we were able to achieve a significant improvement in recall and could also improve upon the DICE coefficient. Only the precision decreased slightly, which can at least be partially contributed to the error entailed by the superspixel segmentation (fig. 3(a)). In fig. 3 further examples of common error sources can be found. Bleeding (fig. 3(b)) can lead to false positives, which leads us to the conclusion that more training data is required. Also instrument tips with openings can be falsely classified (fig. 3(c)).

A possible cause is that the structure is too small in order to be successfully segmented with a superspixel.

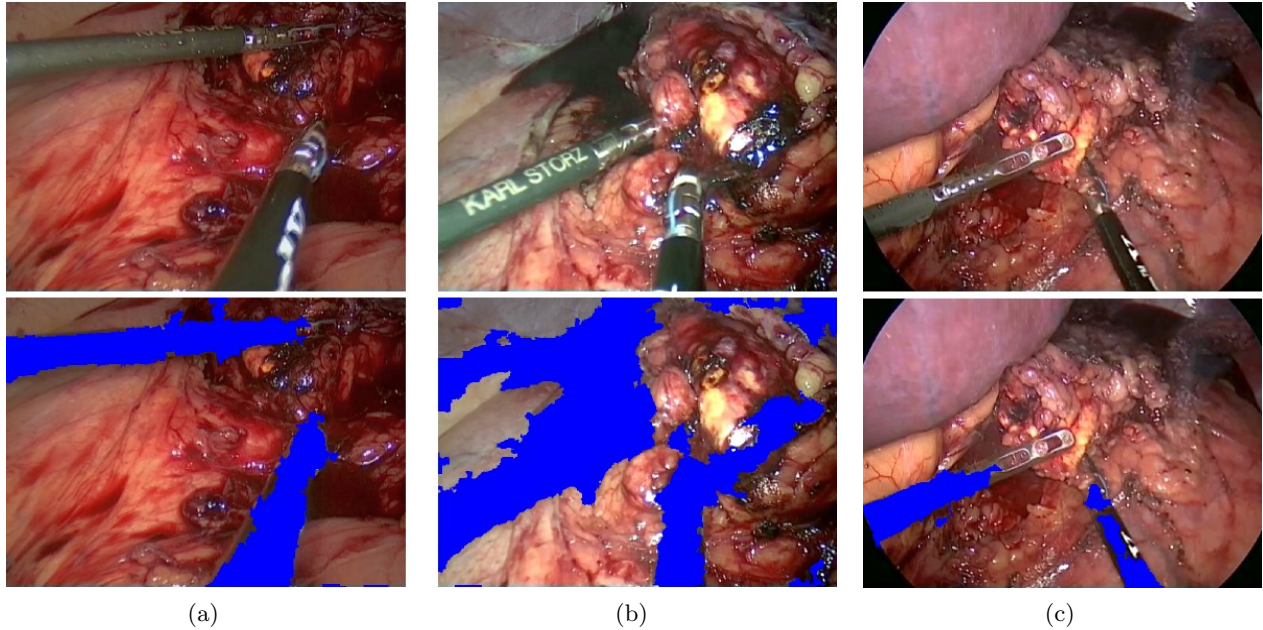


Figure 3. Examples of common error sources: (a) small leakage of the superpixels at the edge of the instrument, (b) Misclassification of a bloody region, and (c) Missing instrument tip.

4. CONCLUSION

In this work, we presented a method for automatically identifying structures in laparoscopic images via superpixels. We evaluated the method on a publicly available dataset and showed an improvement when compared to a pixel-based method.

The goal of future research is to reduce the runtime of the algorithm. Currently, we are achieving 8Hz with a single-core version of the method. Using multiple CPU-cores or the GPU should allow us to achieve a higher frame-rate.

We are planning to evaluate the method on datasets containing labels organs. Furthermore, we are also planning to evaluate new features.

ACKNOWLEDGMENTS

The present research was conducted within the setting of Project A01 of the SFB/Transregio 125 “Cognition-Guided Surgery“ funded by the German Research Foundation. It is furthermore sponsored by the European Social Fund of the State Baden-Wuerttemberg.

REFERENCES

- [1] Katić, D., Wekerle, A.-L., Görtler, J., Spengler, P., Bodenstedt, S., Röhl, S., Suwelack, S., Kenngott, H. G., Wagner, M., Müller-Stich, B. P., et al., “Context-aware augmented reality in laparoscopic surgery,” *Computerized Medical Imaging and Graphics* **37**(2), 174–182 (2013).
- [2] Maier-Hein, L., Mersmann, S., Kondermann, D., Bodenstedt, S., Sanchez, A., Stock, C., Kenngott, H. G., Eisenmann, M., and Speidel, S., “Can masses of non-experts train highly accurate image classifiers?,” in *[Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014]*, 438–445, Springer (2014).
- [3] Allan, M., Ourselin, S., Thompson, S., Hawkes, D. J., Kelly, J., and Stoyanov, D., “Toward detection and localization of instruments in minimally invasive surgery,” *Biomedical Engineering, IEEE Transactions on* **60**(4), 1050–1058 (2013).

- [4] Tjoa, M., Krishnan, S., Kugean, C., Wang, P., and Doraiswami, R., "Segmentation of clinical endoscopic image based on homogeneity and hue," in [*Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*], **3**, 2665–2668, IEEE (2001).
- [5] Chhatkuli, A., Bartoli, A., Malti, A., and Collins, T., "Live image parsing in uterine laparoscopy," in [*Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*], 1263–1266, IEEE (2014).
- [6] Van den Bergh, M., Boix, X., Roig, G., de Capitani, B., and Van Gool, L., "Seeds: Superpixels extracted via energy-driven sampling," in [*Computer Vision–ECCV 2012*], 13–26, Springer (2012).
- [7] Ojala, T., Pietikainen, M., and Harwood, D., "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," in [*Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing, Proceedings of the 12th IAPR International Conference on*], (1), 582–585 (1994).
- [8] Ahonen, T., Hadid, A., and Pietikainen, M., "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **28**(12), 2037–2041 (2006).
- [9] Breiman, L., "Random forests," *Machine learning* **45**(1), 5–32 (2001).
- [10] Bradski, G. et al., "The opencv library," *Doctor Dobbs Journal* **25**(11), 120–126 (2000).