

# Creating Panoramic Images for Bladder Fluorescence Endoscopy

A. Behrens

*The medical diagnostic analysis and therapy of urinary bladder cancer based on endoscopes are state of the art in urological medicine. Due to the limited field of view of endoscopes, the physician can examine only a small part of the whole operating field at once. This constraint makes visual control and navigation difficult, especially in hollow organs. A panoramic image, covering a larger field of view, can overcome this difficulty. Directly motivated by a physician we developed an image mosaicing algorithm for endoscopic bladder fluorescence video sequences. In this paper, we present an approach which is capable of stitching single endoscopic video images to a combined panoramic image. Based on SIFT features we estimate a 2-D homography for each image pair, using an affine model and an iterative model-fitting algorithm. We then apply the stitching process and perform a mutual linear interpolation. Our panoramic image results show a correct stitching and lead to a better overview and understanding of the operation field.*

*Keywords: Image mosaicing, stitching, panoramic, bladder, endoscopy, cystoscopy, fluorescence, photo dynamic diagnosis, PDD.*

## 1 Introduction

Cancer of the urinary bladder is the fourth most common malignancy among males and one of the top eight cancers for women in industrial countries. According to the American Cancer Society, 68,810 new cases and 14,100 deaths are estimated in 2008 in the United States [1]. Bladder cancer tends to occur most commonly in individuals over 60 years of age. The major risk factors are cigarette smoking and exposure to aromatic amines, used in the chemical and dye industry.

Bladder cancer can be diagnosed during a cystoscopy, in which an endoscope is introduced through the urethra into the bladder, which is filled with isotonic saline solution. Malignant tissues of the bladder wall can then be removed with the use of endoscopic tools, e.g. a resectoscope cutting loop.

In white light illumination, small and flat tumors, whose structures do not differ strongly from the surrounding tissue are difficult to recognize and could thus be overlooked during the therapy. To reduce this risk, the visualization of tumor tissue can be improved by a photodynamic diagnosis (PDD) system. This technology uses imaging with fluorescent light, which is activated by the marker substance 5-aminolaevulinic acid (5-ALA), accumulated in malignant tissue. Thus, the contrast between tumor and benign tissue is enhanced and permits easier differentiation, as illustrated in Fig. 1.

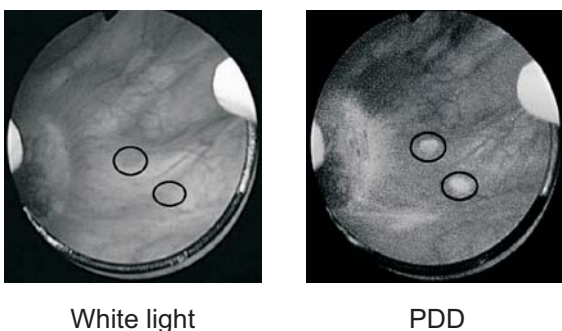


Fig. 1: Papillary tumors in different illuminations

A common disadvantage during an endoscopy is the limited field of view of the endoscope. The physician can examine only a small part of the whole operating field at once. This

causes difficulties in navigation, especially in hollow organs. Instead, a panoramic image provides an overview of the whole region of interest and links images taken from different angles of view. This additional information facilitates visual control and navigation, especially during a cystoscopy, and can be documented in medical evidence protocols.

We have therefore developed an image mosaicing algorithm, which stitches single images of a PDD bladder video sequence and finally provides an expanded panoramic image of the urinary bladder.

This paper is organized as follows: In section 2 we discuss the panorama algorithm in detail. Further optimizations are given in section 3. Section 4 describes the results and perspectives of the algorithm. Finally, section 5 summarizes the proposed approach of our image mosaicing algorithm for endoscopic PDD images.

## 2 Image mosaicing algorithm

The image mosaicing algorithm processes single endoscopic PDD images provided by a video sequence. First, in a preprocessing step we separate the relevant image information of the input images. Then the SIFT features [4] of two images are detected and matched. To refine the feature point correspondences we adapt and apply the RANSAC algorithm [7] to reject outliers. Subsequently we stitch the two images together and interpolate the overlapping region using a linear cross blending method. Then we apply our algorithm iteratively to the next input images. Finally a complete panoramic image of the bladder is built.

### 2.1 Image acquisition

During a cystoscopy the endoscopic images, showing the internal urinary bladder wall, are captured by a PDD video cystoscopy system. In this process the bladder wall is illuminated by a PDD light source. An external camera is attached at the tail end of the rigid cystoscope, as shown in Fig. 2, and captures video images with a resolution of  $720 \times 576$  pixels at a frame rate of 25 frames per second. The video frames are transmitted to the computer video system and are processed.

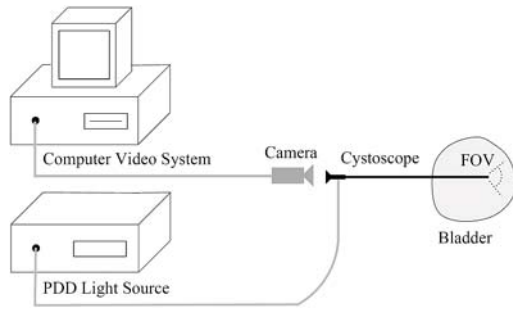


Fig. 2: Principle setup of a PDD video cystoscopy system, showing a cystoscope introduced into the urinary bladder with its limited field of view (FOV). The fluorescence PDD light source provides the illumination and a camera at the tail of the rigid cystoscope transmits the captured image data to a computer video system.

## 2.2. Image preprocessing

In a preprocessing step we subsample the images by a factor of four to reduce computational time and the resolution of the final panoramic image. Then we separate the relevant image information within the elliptical shape from the surrounding dark image region of the input images (see Fig. 3), using Otsu's thresholding method [2].

Thus, we transform the RGB input image to a gray value image and calculate a binary mask, which represents the two classes elliptical and surrounding region. Otsu's algorithm is a thresholding method for separating two classes of pixels so that their between-class variance  $\sigma_b^2$  is maximal. The optimal threshold  $t'$  is then determined by

$$\sigma_b^2(t') = \max_{0 \leq t < T} \sigma_b^2(t) \quad (1)$$

with

$$\sigma_b^2(t) = \omega_1(t)\omega_2(t)(\mu_1(t) - \mu_2(t))^2 \quad (2)$$

and the two class probabilities  $\omega_1, \omega_2$  and their mean levels  $\mu_1, \mu_2$ .

After a further eroding operation we extract the elliptical region using the binary mask, as shown in Fig. 3.

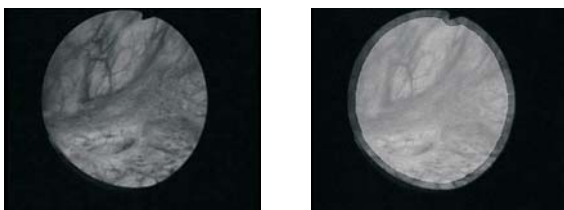


Fig. 3: Left: Original input image, right: Binary mask (transparent overlay)

## 2.3 Feature detection

Common stitching algorithms are based on registrations methods, which can be categorized into pixel-based alignment, feature-based methods, and global registration [3]. In this project we have chosen a feature-based method, since the PDD bladder images generally show a high-contrast vascularization structure. Furthermore it allows a fast stitching

process, since only single feature points have to be matched instead of large pixel blocks.

According to the situation, these image structures can also vary in scale during the video sequence, e.g. caused by a zoom, we use distinctive scale invariant keypoints, calculated by the Scale Invariant Feature Transform (SIFT) [4]. SIFT features are located by detecting the local extrema of a difference-of-Gaussian (DoG) function. This closely approximates the scale-normalized Laplacian of Gaussian (LoG), which is introduced by Lindeberg [5] for effective scale-invariant feature detection. Mikolajczyk [6] showed that this extrema detection generally provides the most stable image features, compared to a range of other possible image functions, such as the gradient, Hessian, or Harris corner function.

The relationship between DoG and LoG can be understood from the heat diffusion equation:

$$\frac{\partial G}{\partial \sigma} = \sigma \nabla^2 G, \quad (3)$$

$$\sigma \nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma} \quad (4)$$

$$\Leftrightarrow \underbrace{G(x, y, k\sigma) - G(x, y, \sigma)}_{DoG} \approx (k-1) \underbrace{\sigma^2 \nabla^2 G}_{LoG}. \quad (5)$$

Eq. (5) shows that the DoG function with a constant scale difference  $k$  approximates the  $\sigma^2$  scale-normalized LoG multiplied by a constant factor  $(k-1)$ , which does not influence the extrema detection. After the features are localized by a maximum operation over space and scale, a further refinement step is applied. Low contrast points and keypoints along edges are rejected, since they are unstable to small amounts of noise. Thus, the ratio of the eigenvalues of the  $2 \times 2$  Hessian matrix

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}, \quad (6)$$

based on the second derivations of the DoG image  $D(x, y, \sigma)$  is calculated and compared to a threshold  $r$ . Keypoints with a large ratio, which means having a large principal curvature across the edge but a small one in the perpendicular direction, are thus rejected. A result of feature detection is shown in Fig. 4.

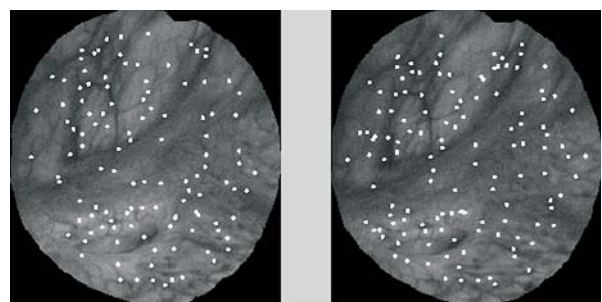


Fig. 4: SIFT features located in image one (left) and image two (right)

## 2.4. Feature matching

Subsequent to the feature localization, described in the preceding section, the feature points of two images are matched. A robust matching algorithm requires distinctive keypoints. So we calculate for each feature point a rotation

invariant SIFT descriptor [4]. The SIFT descriptor takes into account the local neighborhood of the Gaussian smoothed image  $L(x, y)$  by calculating the gradient magnitude  $m(x, y)$  and orientation  $\theta(x, y)$ , accumulated in a histogram, as illustrated in Fig. 5, by

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (7)$$

$$\theta(x, y) = \arctan\left(\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}\right). \quad (8)$$

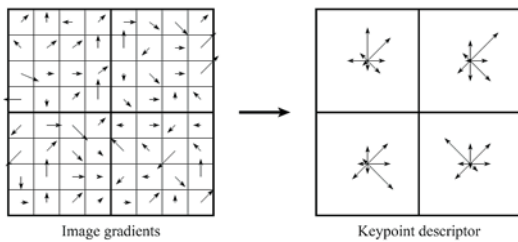


Fig. 5:  $2 \times 2$  Keypoint descriptor array representing the orientation histogram of an  $8 \times 8$  set of samples (principle sketch)

The use of 8 orientations and a  $4 \times 4$  array of histograms leads to a distinctive 128 element feature vector. After each feature point is described by its local descriptor, we apply the matching process using the minimum Euclidean distance measurement

$$d'_i = \min_i \|d_i - d'_i\|_2. \quad (9)$$

In this process one local descriptor  $d'_i$  of the first image is compared to all feature descriptors  $d_i$  of the second image within the 128 dimensional feature space, and vice versa. The minimum squared error then leads to the best corresponding point  $d'_i$ . Afterwards, the next local descriptor  $d'_{i+1}$  is selected and matched. This procedure is repeated until all feature descriptors are processed. Fig. 6 shows the resulting point correspondences of the two sequential images.

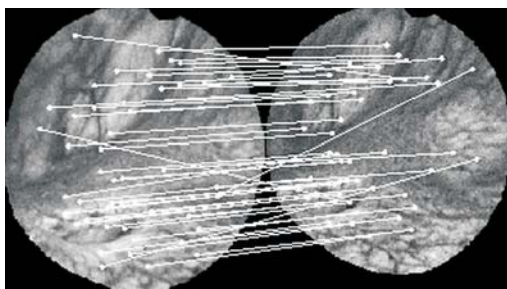


Fig. 6: Matched point correspondences of image one and image two

## 2.5 Homography estimation

In the next step we determine an image-to-image transform, so called 2-D homography, based on the final point correspondences. Since the images of the video sequence describe a non-rigid camera movement, we choose an affine model for the homography. An affine transform provides six degrees of freedom, parametrized by a translation vector  $\vec{t}^T = (t_x, t_y)$ , rotation angle  $\alpha$ , scales  $s_x$  and  $s_y$ , and a shear pa-

rameter  $a$ . In homogeneous coordinates the homography matrix  $\mathbf{M}$  can be written as

$$\mathbf{M} = \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \vec{t} \\ \vec{0}^T & 1 \end{bmatrix} \quad (10)$$

with

$$\mathbf{A} = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix} \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}. \quad (11)$$

Although the matching process shows a good matching result (see Fig. 6), some matching errors can occur, due to the noise and blur of the PDD images. Since matching errors have a high impact on the image transformation, we have to reject these outliers to perform a robust homography estimation. Therefore we employ the RANSAC (RANDOM SAMPLE CONSENSUS) model fitting algorithm [7], which rejects outliers of the matching results during an iterative process. With the adaptation of the fitting model to our 2-D affine model, RANSAC carries out the following steps:

1. Select a set of  $p = 4$  point matches randomly (four point correspondences are required to determine the affine model),
2. Validate the points of being not collinear. If they are collinear, go back to step 1,
3. Calculate the affine 2-D homography between image two and one,
4. Apply the transform to each feature point of image two and the inverse transform to image one, respectively. Count the number of points (so-called inliers), which lie within a spatial error tolerance of  $\epsilon_{\max}$  to the related reference points,
5. If the number of inliers is greater than the best previous score, save the homography matrix.
6. If the maximum number  $N_{\max}$  of trials is reached, terminate the algorithm. Otherwise go back to step 1.

The rejected point correspondences of Fig. 6 performed by the iterative RANSAC algorithm are labeled black in Fig. 7.

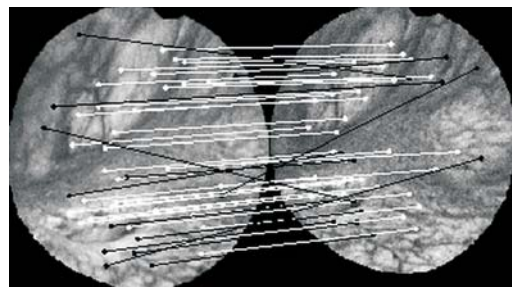


Fig. 7: Rejected outlier points (labeled black) performed by the RANSAC algorithm

Eventually the affine homography matrix between image two and one is determined by the greatest number of inliers.

## 2.6 Stitching and blending

Now we apply the estimated homography to image two and combine image one and two. If a direct composition

method is used, visual artifacts like edges and signal discontinuities occur in the combined image, as shown in Fig. 8.

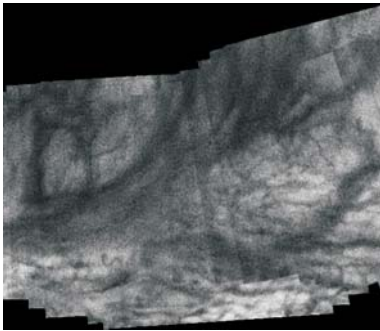


Fig. 8: Direct image compositions without the blending algorithm

This effect results from the inhomogeneous illumination of the bladder wall during the cystoscopy. The illumination intensity decreases from the middle of the image to the borders. To overcome this problem, we apply a cross blending interpolation, suggested by Wald et al. [8], during the composition process. The method performs an interpolation in the overlapping region based on a linear mutual weight distribution, as illustrated in Fig. 9.

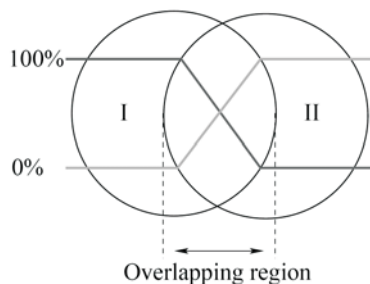


Fig. 9: Cross blending interpolation with a mutual weight distribution in the overlapping region of image I and II

Due to these weight functions, pixels with low illumination at the image borders have less impact on the interpolation than pixels in the image center. This approach reduces the visual artifacts significantly, as can be seen in Fig. 10.

Finally, the two images are stitched and a combined image is built. Then we apply the whole image mosaicing algorithm

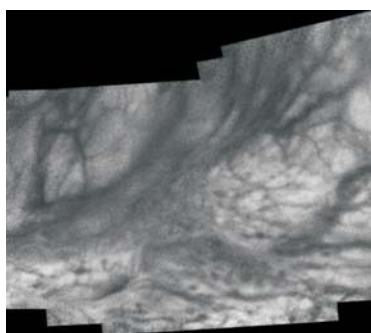


Fig. 10: Image compositions applied with a linear cross blending interpolation

iteratively to the subsequent images, until all frames of the video sequence are processed.

### 3 Optimization

We implement the image mosaicing algorithm firstly in an offline MATLAB program code. To reduce the computational time we apply some further improvements to the algorithm. Since the camera movement along the video sequence is smooth and slow, we dynamically increase the processing frame step size and decrease the overlapping region of the images, respectively. If a sufficient number of reliable matching points is still found, the expansion of the panoramic image will perform faster. Otherwise not enough matching points are found due to low contrast of the image structures or too small image overlap, and the matching process fails. In that case we successively decrease the processing frame step size of the subsequent images, until the matching succeeds.

### 4 Results and perspectives

After all images of the video sequence have been processed by our image mosaicing algorithm, the final panoramic image is built. Fig. 11 represents a panoramic image of an original endoscopic PDD video sequence of 580 frames.

The panorama shows a section of the left part of the urinary bladder, stitched by the single video frames. Fig. 11 show that the papillary tumors are located on the upper left bladder wall related to the left urethral orifice. The spatial relation between adjacent images can now be directly accessed by the panoramic image, since this information is only given implicitly by the camera movement along the video sequence in

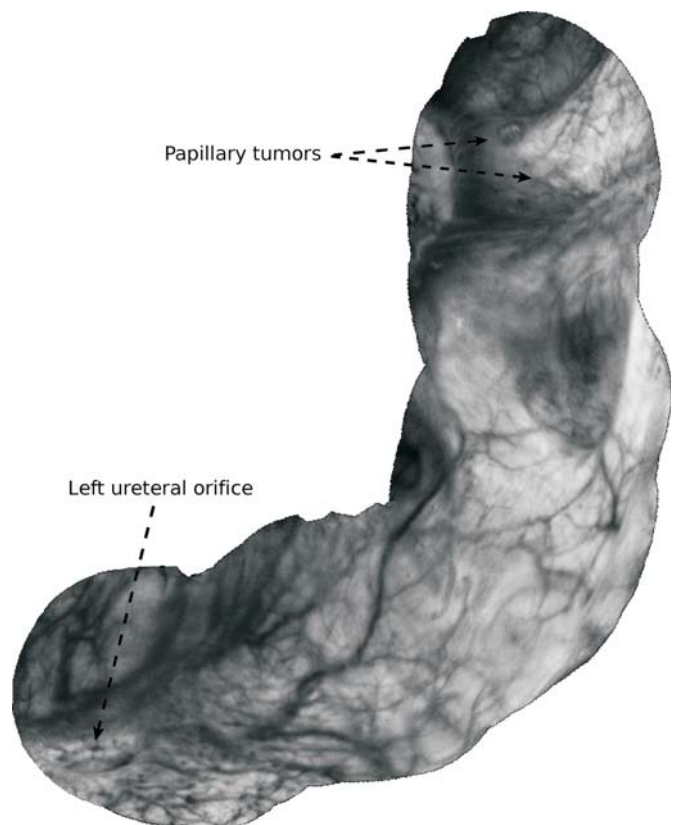


Fig. 11: Panoramic image built from an original endoscopic PDD video sequence of 580 frames

time. In addition, the panoramic image can be documented in medical evidence records to supplement the textual descriptions of the tumor positions in the bladder. This will lead to a better and more intuitive understanding, and can be used for follow-up cystoscopic examinations.

Since the computation time for each image pair takes several seconds, real-time image mosaicing is not supported yet. A software implementation in C++ should improve the performance, and will be applied in future work. Further clinical tests and evaluations also have to be performed. Nevertheless physicians' first comments have indicated that offline results can already provide a high clinical benefit.

## 5 Summary

In this paper we have developed an image mosaicing algorithm for bladder video sequences in fluorescence endoscopy. First, we extracted the relevant image information and applied a feature-based algorithm to get robust and distinctive image feature points for each image pair. Based on our affine parameter model we used an iterative optimization algorithm to estimate the best image-to-image transform according to a mean squared error measurement. Then we described how visual artifacts, caused by inhomogeneous illumination, could be compensated during the stitching process by a mutual linear interpolation function. The results of our iterative image mosaicing algorithm were discussed and illustrated by a panoramic image of an original bladder endoscopic video sequence. Finally we described some optimization steps and perspectives.

## Acknowledgments

I would like to thank Dr. med Andreas Auge, Waldklinikum Gera, Germany and Olympus & Winter Ibe GmbH for technical support, providing the video sequences and giving preliminary feedback for this project.

I would also like to thank Prof. Dr.-Ing. Til Aach, Institute of Imaging and Computer Vision, RWTH Aachen University, Germany, who supervised this project.

## References

- [1] American Cancer Society®, Cancer Facts and Figures 2008 [Online]. Available: <http://www.cancer.org/downloads/STT/2008CAFFfinalsecured.pdf>
- [2] Otsu, N.: A Threshold Selection Method from Gray-Level Histograms, *IEEE Trans. Syst. Man Cybern.*, Vol. **9** (1979), p. 62–66.
- [3] Szeliski, R.: Image Alignment and Stitching: A Tutorial, *Technical Report MSR-TR-2004-92*, Microsoft Research, 2006.
- [4] Lowe, D.: Distinctive Image Features from Scale-Invariant Keypoints, *Int. Journal of Computer Vision*, Vol. **60** (2004), No. 2, p. 91–110.
- [5] Lindeberg, T.: Scale-Space Theory: A Basic Tool for Analysing Structures at Different Scales. *Journal of Applied Statistics*, Vol. **21** (1994), No. 2, p. 224–270.
- [6] Mikołajczyk, K.: Detection of Local Features Invariant to Affines Transformations. *Ph.D. thesis*, Institut National Polytechnique de Grenoble, France, 2002.
- [7] Fischler, M. A., Bolles, R. C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, Vol. **24** (1981), No. 6, p. 381–395.
- [8] Wald, D., Reeff, M., Székely, G., Cattin, P., Paulus, D.: Fliessende Überblendung von Endoskopiebildern für die Erstellung eines Mosaiks. *Bildverarbeitung für die Medizin*, Mar. 2005, p. 287–291.

---

Alexander Behrens  
e-mail: [Alexander.Behrens@lfb.rwth-aachen.de](mailto:Alexander.Behrens@lfb.rwth-aachen.de)

Institute of Imaging & Computer Vision

RWTH Aachen University  
D-52056 Aachen, Germany