Analysis of Feature Point Distributions for Fast Image Mosaicking Algorithms

A. Behrens, H. Röllinger

Abstract

In many algorithms the registration of image pairs is done by feature point matching. After the feature detection is performed, all extracted interest points are usually used for the registration process without further feature point distribution analysis. However, in the case of small and sparse sets of feature points of fixed size, suitable for real-time image mosaicking algorithms, a uniform spatial feature distribution across the image becomes relevant. Thus, in this paper we discuss and analyze algorithms which provide different spatial point distributions from a given set of SURF features. The evaluations show that a more uniform spatial distribution of the point matches results in lower image registration errors, and is thus more beneficial for fast image mosaicking algorithms.

Keywords: feature detection, point distribution, image registration, image mosaicking.

1 Introduction

Alignment and stitching of images into seamless photomosaics is most widely used in computer vision [1]. Besides the approach to minimize pixel-to-pixel dissimilarities directly, a common way to combine image pairs into one image composition is performed using only sparse sets of interest points. Here, distinctive feature points are first extracted and described from the images, and then matched using a similarity measurement. Many different feature-based algorithms, like SIFT [2], SURF [3], GLOH [4], MOPs [5] or [6] have been proposed for extracting distinctive image features. Their robustness, repeatability and invariance to different illumination changes and image transformations have also been widely evaluated by [4, 7, 8, 9].

On the one hand, a high level of distinctiveness of the features, expressed by their descriptors and the filter response values of the detector, leads to robust matching and precise image alignment. On the other hand, these characteristics alone do not always result in a low registration error, since the spatial feature distribution in the images is also relevant. Images which show only a high contrast in local image regions (see Fig. 1) usually create only clusters of strong features and provide an overall non-uniform spatial feature distribution. Since the estimation of a global image transformation between a given image pair is based on matched feature points, image regions with many interest points will create only small registration errors, whereas the error in regions with feature clusters becomes larger. Thus, only the combination of distinctive and also spatially well uniformed distributed feature points leads to good global image transformation with low registration errors. While in many natural image scenes feature points can usually be extracted across the whole image, spatial distribution is often a relevant consideration for medical images, e.g. endoscopic images of the internal urinary bladder wall. These images often show only sparsely located structures with high contrast, e.g. vasculature or lesions (see Fig. 1), and thus impede robust image mosaicking.

Fig. 1: Left: Landscape with a high contrast region in the lower right corner. Right: A noisy and low contrast endoscopic bladder image showing a lesion

During cystoscopy, where a rigid endoscope is introduced into the bladder, an inspection and analysis of bladder cancer, the 4th most common disease with about 52810 new cases among males in the United States estimated in 2009 [10], is performed by a physician based on endoscopic images captured by a video endoscope system. Since sufficient illumination of the bladder wall is only provided if the endoscope is guided close to the bladder wall, the field of view is very small. This results in difficulties in orientation and navigation for the physician. To provide navigation assistance during cystoscopy, a feature-based mosaicking algorithm composing local or global panoramic overview images from single video images has been proposed

by [11, 12, 13, 14]. For this application, a fixed number of feature points is required for real-time processing [14]. It is therefore necessary to generate an adequate sample set of the extracted features list for robust image mosaicking.

While some registration algorithms have been developed that create feature distributions based on local non-maximum suppression [5] or space-partitioning [15], these different techniques have not been evaluated. Thus, in this paper we analyze these different feature distribution algorithms and evaluate their average image registration errors.

The paper is organized as follows: In section 2 the mosaicking algorithm is described. The distribution algorithms for selecting a feature sample set are described in section 3. Section 4 presents the evaluated data, and the conclusions and future work are given in section 5.

2 Mosaicking algorithm

In the following sections only a brief overview is given of the mosaicking algorithm. A more detailed description can be found in [12, 14]. During the mosaicking process, image pairs of a captured video sequence are sequentially stitched and blended to compose a successively growing panoramic overview image. When endoscopic images are used, the lens distortions of the fish-eye optics are compensated in a preprocessing step. Then, distinctive feature points are extracted and described by their local neighborhood region. Based on these sparse feature sets, point correspondences are matched between two subsequent images. Thus, a global image-to-image transformation, a so-called homography, is estimated. The two images are registered by applying the transformation. Finally, the overlap regions are blended by an interpolation algorithm. Iteratively, all images of the video sequence are then sequentially processed to build the final panoramic overview image.

2.1 Feature detection

Feature extraction is performed by the Speeded Up Robust Features (SURF) detector [3]. Based on a very basic approximation of the Hessian matrix, the feature strength is calculated by

$$\det(\mathcal{H}) = \left\| \begin{bmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{bmatrix} \right\| \approx D_{xx}D_{yy} - (0.9 \cdot D_{xy})^2, \tag{1}$$

where L_{xx} represents the convolution of the Gaussian second order derivative $\frac{\partial^2}{\partial x^2}G$ with the input image. Instead of iteratively reducing the image size, integral images and scalable box filters D_{xx}, D_{yy}, D_{xy} are used to speed up the blob filter response. The distinctive feature points are localized in the space domain and over scales by non-maximum suppression. Feature descriptors are then extracted in a square region centered around the point of interest. Split up into smaller 4×4 square sub-regions, the Haar wavelet responses in horizontal h_x and vertical direction h_y are calculated and composed to a four-dimensional descriptor vector $d_{4D} = \left(\sum h_x, \sum h_y, \sum |h_x|, \sum |h_y|\right)$. The concatenation of the sixteen sub-regions then results in a descriptor vector d of length 64.

2.2 Matching and registration

The point correspondences between two images are determined by matching the SURF feature descriptors. Based on the least similarity measurement

$$\Delta d_{ij} = \min_j \left|\left| d_i - d_j \right|\right|_2, \tag{2}$$

a feature descriptor d_i of the first image is compared to all feature descriptors d_j of the second image within the 64 dimensional feature space, and vice versa. The minimum squared error Δd_{ij} then leads to the best point match $d_i \leftrightarrow d_j$. To increase the distinctiveness and robustness of the point correspondences, only matches with the ratio

$$\frac{\Delta d_{ij}}{\Delta d_{ij}^{\text{2nd}}} < \tau \tag{3}$$

between the best and the second best match are considered.

After single points of an image pair are matched, the homography $\mathbf{H}$ is estimated. The applied affine transformation model provides six degrees of freedom, parameterized by a translation vector $t^T = (t_x, t_y)$, rotation angle α, scales s_x and s_y, and skew a. In homogeneous coordinates, the homography matrix $\mathbf{H}$ can be written as

$$\mathbf{H} = \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A} & t \\ 0^T & 1 \end{bmatrix} \tag{4}$$

with

$$\mathbf{A} = \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix} \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix}. \tag{5}$$

To ensure robust homography estimation, unavoidable false point correspondences are rejected by the RANSAC (RANdom SAmple Consensus) model fitting algorithm [16]. Based on iteratively and randomly selected point correspondences p, a homography $\hat{\mathbf{H}}$ is estimated and the number of inliers is determined by the threshold operation

$$\left|\left| p_i - \hat{\mathbf{H}} \cdot p_j \right|\right|_2 < d. \tag{6}$$

If point correspondences $p_i \leftrightarrow p_j$ satisfy eq. 6 with estimation $\hat{\mathbf{H}}$ and the given pixel distance d, they are marked as inliers. Finally, the estimated homography $\hat{\mathbf{H}}$ with the highest number of inliers is selected as the final image transform and is used for registration by warping the second image into the reference system of the first image.

3 Feature selection

Since the computational complexity of the SURF detector increases linearly with number of feature points, the limitation to a fixed number is often desired for real-time mosaicking algorithms [14]. Also calculating of the feature descriptors is computationally more intensive than locating the feature and the filter response value itself (eq. 1). Thus, feature extraction is separated into two steps. First, the SURF feature detector is applied to the whole image to detect the locations and the strengths of potential interest points based on the filter response values. Then a sample set with a fixed number of the points is chosen from all the point candidates. To assure constant computational complexity, only these features are then characterized by the 64-dimensional SURF descriptor and passed to the matching process. Since the usage of fewer feature points leads to less robust homography estimations and registration results, the selection of an adequate and representative sample set is very important.

3.1 Top N selection

A straightforward approach is to select the first N-th strongest feature points from the whole set of interest points. For each feature point, the location and the response value of the blob filter is calculated according to eq. 1, and is used for comparison. After the extracted interest points have been sorted in descending order of their filter response values, the first N-th feature points are selected. Since this algorithm generates only a sample set based on the filter response, no additional information about the spatial distribution of the feature points is exploited.

3.2 Adaptive non-maximal suppression

To select a fixed number of interest points from images which are local maxima and whose response values are also significantly greater than those of all of their neighbors within radius r, Brown et al. [5] developed an adaptive non-maximal suppression (ANMS) strategy. Conceptually, this can be performed by initializing the suppression radius $r = 0$ and then increasing it until the desired number of N feature points is obtained. In practice, this operation can be done by first sorting all local maxima by their response values, and then creating a second list sorted by decreasing suppression radius. The first entry in the list represents the global maximum, which is not suppressed at any radius. As the radius decreases from infinity, feature points are added to the list. To increase the robustness, a second constraint is defined, which requires that a neighbor feature has sufficiently greater strength. Thus, the minimum suppression radius r_i is determined by

$$\begin{aligned} r_i &= \min_j |x_i - x_j|, \\ &\text{with} \quad f(x_i) < c \cdot f(x_j), \quad x_j \in \mathcal{I}, \end{aligned} \tag{7}$$

where x_i describes the position (x, y) of the feature point, $f(x_i)$ its strength, and $\mathcal{I}$ the set of all feature point positions. The parameter $c = 0.9$ represents the robustness factor, which adjusts how significantly greater the strength of a neighbor feature must be for suppression to take place. For each feature point its radius is then determined by eq. 7, and the feature list is sorted by the radius in descending order. Then, the first N entries of the list are selected. This sample set now provides features which are spatially distributed across the whole image (cf. Fig. 4 (right column)).

3.3 k-d tree partitioning

Another selection method that considers the spatial data distribution is based on space-partitioning of high-dimensional data sets. Cheng et al. [15] developed an algorithm using a 2-dimensional k-d tree. Here, the feature points are separated into M rectangular image regions and cells, respectively. In a recursive manner, each partition cuts the region with the current highest variance in two, using the median data value as the dividing line. The position of the dividing line is stored as the node's data. For each non-terminal node of the binary tree, the feature points of the related region are then handed over to the left and right child node, respectively, until the given number of M cells is obtained. Finally, each leaf node contains a list of the features which are located within the related image partition. An example is given in Fig. 2. From each cell $\lfloor N/M \rfloor$, the strongest features are selected as the output sample set, cf. Fig. 4 (middle column). While Cheng [15] uses a balanced tree, which limits the number of leaf nodes to be a power of two, we have extended the algorithm to handle any integer number of points. Since only one feature is taken from each cell, the partitions are divided until the desired number of features N is obtained. Thus, a direct comparison between k-d tree partitioning, ANMS, and top N selection becomes more feasible.

4 Evaluation and results

The three feature distribution algorithms (top N, ANMS, and k-d tree) are evaluated using three data

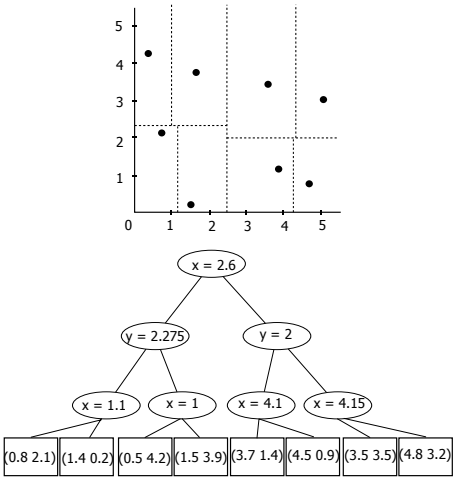
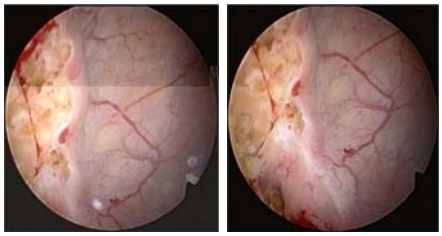


Fig. 2: Above: 2-dimensional k-d tree partitions with feature points. Below: k-d tree

tree-clinic

houses



bladder

Fig. 3: Data sets used for evaluation

sets (Fig. 3) which show varying contrast across the whole image. Each image pair represents one scene from two slightly different point of views, but still providing a large overlap region. Since the images are not synthesized, no ground truth data of the correct image composition is given. Thus, for each image pair control points $p'_i \leftrightarrow p'_j$ are manually set and matched. Based on these points, a homography is also estimated according to eq. 4, and the registration error

$$e = \frac{1}{2}\left(||p'_i - \hat{\mathbf{H}} \cdot p'_j||_2 + ||p'_j - \hat{\mathbf{H}}^{-\mathbf{1}} \cdot p'_i||_2\right) \quad (8)$$

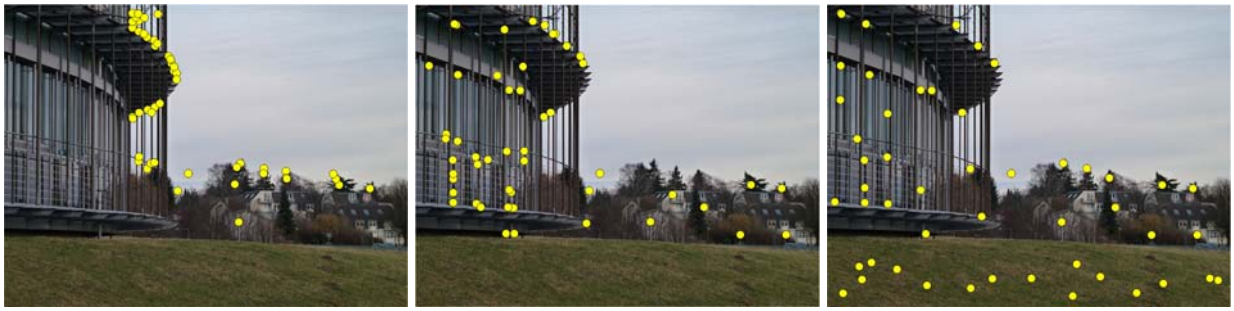
is calculated. These error values are used as references in later evaluations.

To evaluate different feature distributions, SURF features are first extracted and listed for each data set. Each algorithm of section 3 is then applied to the point lists to generate a sample set of only N feature points, resulting in different point distributions. Representative examples for each data set are shown in Fig. 4. The figure shows, that the selection of the top N strongest features leads to single point clusters in all three data sets. In example, many SURF features show a high filter response value in the tree region of the *tree-clinic* sequence, since the dark branches have a high contrast against the bright background. By contrast, the k-d tree and also the ANMS algorithm provide a more spatially uniform distributed set of feature points across the whole image and the relevant overlap regions, respectively. As shown in Fig. 4(c), more feature points are also located in the vasculature in the image center in the case of k-d tree partitioning and ANMS. Since this region is also present in the overlap region of the bladder sequence (cf. Fig. 3), consideration of these features during the homography estimation should provide a small registration error.

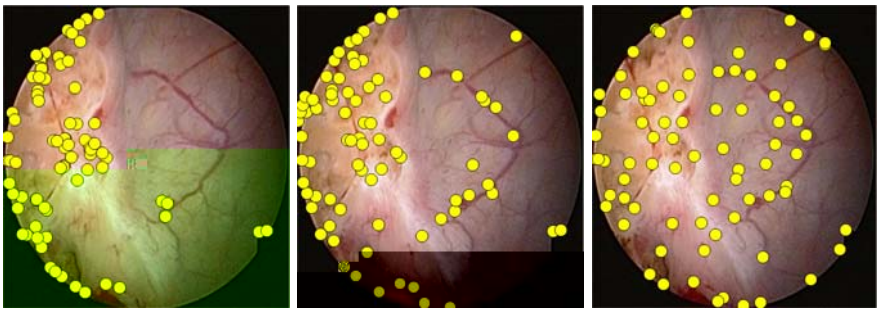
After the different feature lists have been generated, they are passed to the matching process, and a global image-to-image transformation for each sample set is estimated, as described in section 2.2. For a quantitative evaluation, the homographies are then applied to the ground truth control points $p'_i \leftrightarrow p'_j$ and the final registration errors are calculated using eq. 8. Fig. 5 shows the reproduction error characteristics for each distribution method over the selected number of feature points. Since the RANSAC algorithm provides robust point matches and rejects outliers on the basis of a random process, mean error values are calculated. Thus, Fig. 5 shows the averaged registration errors and the number of inliers over the number of feature points for each data set. The results of the *tree-clinic* and *houses* sequence are averaged by 15 measurements, and the graphs of the *bladder* sequence are based on 20 measurements.

As can be seen from the characteristics, the mean registration errors calculated from feature points distributed by k-d tree partitioning and ANMS are usually smaller than with the top N method. Especially for a small number of points, the error difference increases. At higher numbers, the different algorithms provide almost the same registration errors. This is obvious, since a larger point set usually leads to a spatial distribution with a higher variance. The results of the *tree-clinic* sequence also show that robust matching of the first N strongest features is not feasible until at least a minimum number of $N = 210$ feature points

tree-clinic sequence with 200 features points



houses sequence with 50 features points



bladder sequence with 80 features points

Fig. 4: For each data set *top N* selection (left), *k-d tree* partitioning (middle), and *ANMS* (right) is performed

is obtained, but still resulting in high error. By contrast, the variance of the mean errors of the k-d tree and ANMS sample sets is much smaller. Although the registration errors of k-d tree partitioning and ANMS are almost equal in all image sequences, the number of inliers used for robust homography estimation is higher in the case of ANMS. Against this, the complexity of ANMS with up to $O((N-1)!)$ is much higher than $O(\log N))$ of the k-d tree. Thus, the integration of the k-d tree partitioning in fast and real-time image mosaicking algorithms like [14] should be preferred.

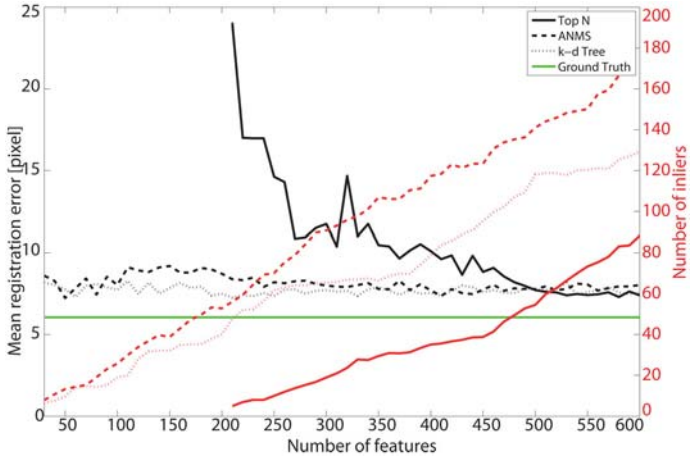
5 Conclusions

An evaluation of the ANMS and k-d tree algorithms, which provide a spatially well uniformed feature distribution, has shown that the mean error of an image pair registration can be greatly reduced, compared to a sample set based on strongest feature selection. This difference becomes significant in images showing regions of different varying contrast. For applications that require a fixed or limited number of interest points, e.g. real-time image mosaicking algorithms for medical computer assistance systems, the usage of feature distribution algorithms is also beneficial. In future work, further evaluations will be made on more medical image data sets and time measurements of the distribution algorithms integrated in a real-time image mosaicking algorithm for fluorescence endoscopy. The generation of feature point sets based on an adaptive algorithm considering the feature strengths and their distribution according to the current image information will also be analyzed.

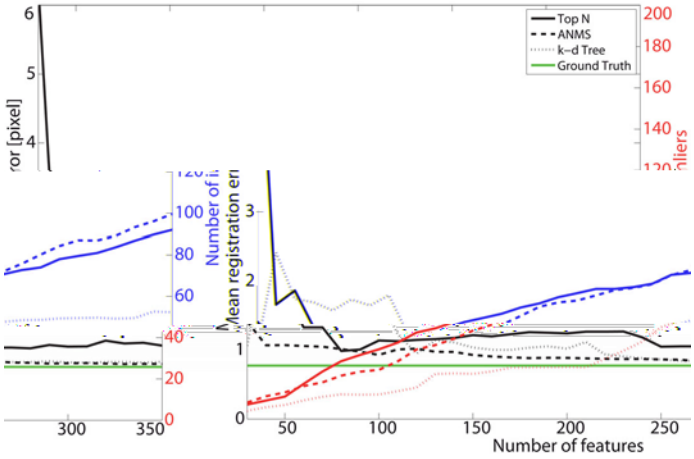
Acknowledgement

We would like to thank Prof. Dr.-Ing. Til Aach, Institute of Imaging and Computer Vision, RWTH Aachen University, Germany, for supervising this project.

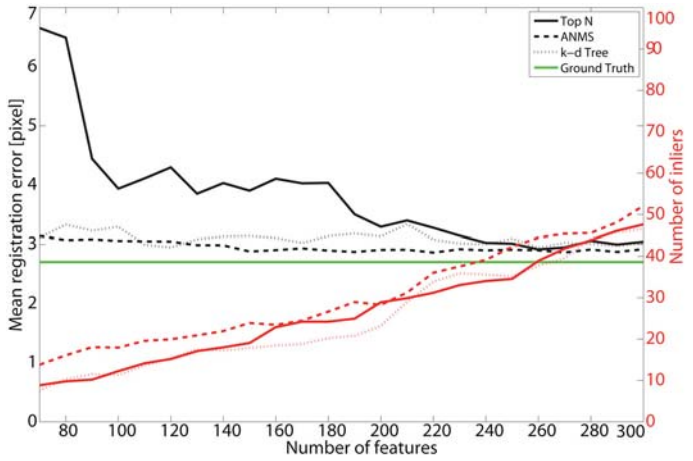
Additionally we would like to thank Olympus & Winter Ibe GmbH for providing the endoscopic image sequences.



tree-clinic



houses



bladder

Fig. 5: Mean registration errors (black) and numbers of inliers (red) of each distribution method (Top N, ANMS, k-d Tree) over the number of feature points. The registration errors using the ground truth homography are highlighted in green

References

[1] Szeliski, R.: Image Alignment and Stitching: A Tutorial, *Microsoft Research*, 2006, Tech. Rep. MSR-TR-2004-92.

[2] Lowe, D.: Distinctive image features from scale-invariant keypoints, In *International Journal of Computer Vision*, 2004, vol. **60**, no. 2, pp. 91–110.

[3] Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: SURF: Speeded Up Robust Features, In *Computer Vision and Image Understanding (CVIU)*, 2008, vol. **110**, no. 3, pp. 346–359.

[4] Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors, In *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2005, vol. **27**, no. 10, pp. 1 615–1 630.

[5] Brown, M., Szeliski, R., Winder, S.: Multi-Image Matching using Multi-Scale Oriented Patches, In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, vol. **1**, pp. 510–517.

[6] Shi, J., Tomasi, C.: Good Features to Track, In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*, 1994, pp. 93–600.

[7] Dorko, G.: *Selection of Discriminative Regions and Local Descriptors for Generic Object Class Recognition*, PhD thesis, 2006.

[8] Schmid, C., Mohr, R., Bauckhage, C.: Evaluation of Interest Point Detectors, In *International Journal of Computer Vision*, 2000, vol. **37**, no. 2, pp. 151–172.

[9] Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L.: A Comparison of Affine Region Detectors, In *Int. J. Comput. Vision*, 2005, vol. **65**, no. 1–2, pp. 43–72.

[10] American Cancer Society: *Cancer Facts and Figures 2009*, 2009.

[11] Miranda-Luna, R., Daul, C., Blondel, W., Hernandez-Mier, Y., Wolf, D., Guillemin, F.: Mosaicking of Bladder Endoscopic Image Sequences: Distortion Calibration and Registration Algorithm, In *Biomedical Engineering, IEEE Transactions on*, 2008, vol. **55**, no. 2, pp. 541–553.

[12] Behrens, A.: Creating Panoramic Images for Bladder Fluorescence Endoscopy, In *Acta Polytechnica Journal of Advanced Engineering*, 2008, vol. **48**, no. 3, pp. 50–54.

[13] Behrens, A., Stehle, T., Gross, S., Aach, T.: Local and Global Panoramic Imaging for Fluorescence Bladder Endoscopy, In *Engineering in Medicine and Biology Society, EMBC 2009. 31th Annu. Int. Conf. of the IEEE*, 2009, pp. 6 990–6 993.

[14] Behrens, A., Bommes, M., Stehle, T., Gross, S., Leonhardt, S., Aach, T.: A Multi-Threaded Mosaicking Algorithm for Fast Image Composition of Fluorescence Bladder Images, In *Proceedings SPIE Medical Imaging 2010: Visualization, Image-Guided Procedures, and Modeling*, 2010, vol. **7 625**.

[15] Cheng, Z., Devarajan, D., Radke, R. J.: Determining vision graphs for distributed camera networks using feature digests, In *EURASIP Journal on Adcances in Signal Processing*, 2007, Article ID 57034.

[16] 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*, 1981, vol. **24**, no. 6, pp. 381–395.

About the authors

Alexander BEHRENS was born in Bückeburg, Germany in 1980. He received his Dipl.-Ing. degree in electrical engineering from the Leibniz University of Hannover, Hannover, Germany, in 2006. After receiving the degree, he worked as a Research Scientist at the university's Institut für Informationsverarbeitung. Since 2007, he has been a Ph.D. candidate at the Institute of Imaging and Computer Vision, RWTH Aachen University, Aachen, Germany. His research interests are in medical image processing, signal processing, pattern recognition, and computer vision.

Hendrik RÖLLINGER was born in 1983 in Aachen, Germany. He has been studying Computer Engineering at RWTH Aachen University since 2003, and will graduate with a Dipl.-Ing. degree in 2010.

Alexander Behrens
Hendrik Röllinger
E-mail: alexander.behrens@lfb.rwth-aachen.de,
hendrik.roellinger@lfb.rwth-aachen.de
Institute of Imaging & Computer Vision
RWTH Aachen University
52056, Aachen, Germany