Endoscopic image mosaics for real-time color video sequences

W. Konen[a], M. Naderi[a], M. Scholz[b]

[a]*Institute for Informatics, University of Applied Science Cologne, Germany*
[b] *Department of Neurosurgery, Ruhr-University Bochum, Germany*

...

Abstract. We present an algorithm capable of making in real time image mosaics with enlarged field-of-view from the endoscopic video data stream. The method is applied for the first time to neuroendoscopic color video. It is shown that radial lens distorsion correction leads to improved registration.

Keywords: image mosaicing, endoscopy, neurosurgery, visual navigation, optical flow.

---

1. Introduction

Endoscopy is an important tool for surgical and diagnostic applications, e.g in orthopedic surgery, neuroendoscopic interventions, urology or gynecology. It permits minimally invasive procedures, but the surgeon has often to deal with a rather limited field of view which can cause navigational difficulties. It is therefore desirable to have a tool which combines automatically many endoscopic video frames to a larger, metrically accurate field-of-view (image mosaicing or 'panoramic overview'). We present in this contribution an algorithm capable of making in real time image mosaics with enlarged field-of-view from the endoscopic color video data stream. The algorithm is based on the method of Kourogi et al. [1] which we extend to the case of endoscopic masks. The algorithm finds *fully automated* the optimal affine transform between video frames and builds the enlarged field-of-view as an intervention-free side task. We apply our algorithm to endoscopic video sequences and compare it to the well-known image-mosaicing algorithm of Szeliski [2].

Many algorithms on image mosaicing are known, however only relatively few of them can work fully automatic, e.g. [1,2,3,4] and under real-time conditions [1,3,4]. Only in one case [3] these methods have been applied to endoscopic video. In [3] special care is taken to account for tubular environements as they occur for example in bronchoscopy. Another method for stitching together endoscopic views is the use of external optical tracking [5]. This however requires additional tracking equipment and additional registration procedures in the operating theatre.

All of the above methods (with the exception of [5]) do not account for radial distorsions as they typically occur in endoscopic lens systems.

2. Methods

The goal of our algorithm based on [1] is to estimate the motion field between successive frames *I(t-1)* and *I(t)* of a video sequence. This is done with an improved optical flow algorithm which calculates at each pixel *(x,y)* the so-called pseudo motion

$$(1)\quad \begin{pmatrix} u_p \\ v_p \end{pmatrix} = \begin{pmatrix} -I_t^{(c)} / I_x \\ -I_t^{(c)} / I_y \end{pmatrix} + \begin{pmatrix} u_c \\ v_c \end{pmatrix} \qquad \text{with} \qquad I_t^{(c)} = I(x+u_c, y+v_c, t) - I(x,y,t-1)$$

where I_x and I_y denote the spatial gradient and (u_c,v_c) is the so-called compensated motion at this pixel location. *I(x,y,t)* is the luminance signal at pixel *(x,y)* in frame *t*.

Our algorithm proceeds as follows: Initially we start with $(u_c,v_c)=0$ or with an estimate from the previous frame. Then the following steps are carried out in a loop: **(A)** Calculate the pseudo motion (u_p,v_p) for each pixel inside the endoscopic mask. **(B)** Accept only those pixel which fulfil the following criteria: (a) I_x and I_y are not 0, (b) $(x+u_p,y+v_p)$ is inside the endoscopic mask and (c) $|I(x+u_p,y+v_p,t) - I(x,y,t\text{-}1)| < T$. Here, T is a suitable grey level threshold, e.g. $T=5$. **(C)** Find the affine parameters $\boldsymbol{a} = \{a_1,\ldots,a_6\}$ for a global motion field best-fitting the pseudo motion at all *accepted* pixel locations i, i.e. solve the overdetermined system of equations

$$(2) \quad \begin{aligned} a_1 x_i + a_2 y_i + a_3 &= u_{p,i} \\ a_5 x_i + a_5 y_i + a_6 &= v_{p,i} \end{aligned}$$

in a least-square sense. Use the motion field given by $\boldsymbol{a}$ (the LHS of Eq. (2)) as a new estimate for (u_c,v_c) and continue with step **(A)**.

The loop is terminated either after a fixed number of iterations or when the change in the global motion field drops below a certain threshold.

Some care has to be taken when setting up the masks: In order to avoid large errors at the mask boundary, the gradient calculation with a [-1 0 1] filter is allowed only at those pixels which come from a smaller region, namely the morphological erosion of the mask with a 3x3 cross. Likewise the bilinear interpolation can only be done at pixels from a region being an erosion with a 2x2 square of the original mask.

After a frame is registered, the 'new' portion of it is added to the image mosaic using bilinear interpolation. This is done on the original color image and it can be done rather fast since each frame adds only a small new region to the existing mosaic.

Endoscopic lens systems are characterized by large radial distorsions. Our algorithm therefore has the option to 'undistort' every frame prior to the above mosaicing process. The radial distorsion parameters (κ_1, piercing point) are known from previous camera calibration [6] and can be used to reconstruct a distorsion-free image by bilinear interpolation. We call this process 'undistorting'.

3. Results

In a first experiment we created a short endoscopic video sequence (30 frames) of grey level images where each frame is connected to the next by a known affine transform. We tested two algorithms applied to the same task: The first one is our method described above, based on Kourogi's algorithm [1] with acceptance threshold $T=5$. The second one is based on the well known Szeliski image mosaicing algorithm [2]. Both algorithms work *fully automated* on video sequences, i.e. they had no other information than the sequence itself (no start parameters).

The resulting image mosaic (panoramic view) gives the surgeon a much better overview than the single frames. It is free of mosaicing artefacts and close to the original base image in Kourogis case (shown in Fig. 1 on the right). In Fig. 2, left, we compare the frame-to-frame accuracy, measured as the mean motion error Δu (in pixel) between the true and the estimated motion field. Kourogi's method has much lower error and we do not have any outliers in the frame sequence. Szeliski's method has outliers in frames #1-4. This is important because a single outlier will make all subsequent frames in the image mosaic wrong which is seen clearly in Fig. 2, right: it has clear errors in translation estimation and errors in aspect ratio estimation.

In a second experiment (Fig. 3) we obtained results on a real neuroendoscopic color sequence where the true motion is not known. The matching used solely the luminance signal. Although the lighting conditions change from frame to frame (observe how the border region appears darker than the center), our method is able to find very good registrations for the image mosaic. Small border artefacts (green rings with the size of the endoscop mask) are hardly visible. All images were 'undistorted' prior to image mosaicing. A parallel run of the same experiment on the original distorted images resulted in an inferior mosaic, indicated by a decline of 4% in the rate of accepted pixels (number of accepted pixels divided by the number of pixels inside the endoscopic mask) in step **(B)**.

Finally we note that our method based on Kourogi's algorithm [1] is by a factor of 3 faster than the algorithm of Szeliski [2]. An implementation of our algorithm as Java-based ImageJ-plugin [8] runs at a speed of 10-20 fps (frames per second) on a standard 1.6 GHz Pentium M PC, depending on image size and algorithmic settings.

4. Conclusion and Outlook

We have shown how to build image mosaics from endoscopic video sequences and applied it successfully to color images of neuroendoscopic interventions. Of course our work is only a first step towards an integrated system for real-time endoscopic image mosaicing. Nevertheless this first step is promising, since the algorithm turns out to be robust, does not need any manual intervention or starting values, is faster and at the same time more accurate than comparable algorithms. Our method has shown to be fast enough to run as a side task in the operating room.

There are many directions we plan to investigate in the near future to make the algorithm more robust for a wider set of endoscopic images, even if they are lower in quality: The class of transforms should be extended from affine to projective to account for more general camera movements. Changes in lighting should be accounted for: Global contrast and brightness will vary slowly from frame to frame. Special care should be taken to account for varying illumination (the outer rim of the field of view is usually darker than the center). Different strategies for combining videos to image mosaics will be explored. To increase the signal-to-noise ratio we plan to use color features also in the matching process. With those features incorporated we will finally integrate the image mosaicing module into our VN system [6,7]. The whole system will be tested with respect to ergonomic requirements by surgeons working in daily routine with endoscopic images.

References

[1] Kourogi M, Kurata T, Hoshino J et al.: Real-time image mosaicing from a video sequence. Procs ICIP99, vol. 4, 133-137, 1999.

[2] Szeliski R: Image Mosaicing for Tele-Reality Applications. TR 94/2, Digital Equipment Corporation, Cambridge Research Lab, June 1994.

[3] Seshamani S, Lau WW, Hager GD: Real-Time Endoscopic Mosaicking. MICCAI'2006, Lecture Notes in Computer Science, vol. 4190, Springer, 355-363 2006.

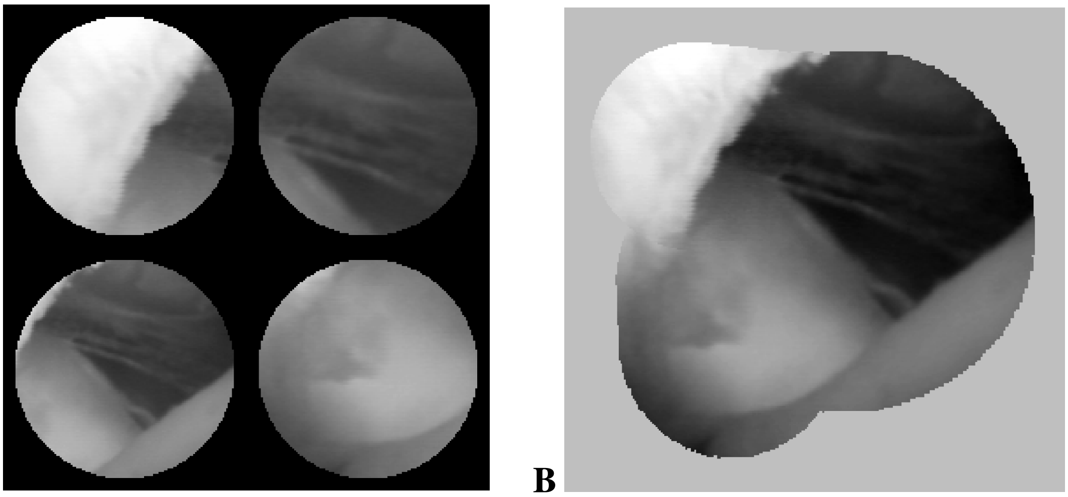
[4] Robinson JA: A Simplex-Based Projective Transform Estimator. Procs Visual Information Engineering (VIE), Guildford, 290-293, July, 2003.

[5] Dey, D, Gobbi DG, Slomka PJ et al.: Automatic Fusion of Freehand Endoscopic Brain Images to Three-Dimensional Surfaces. IEEE Trans. Med. Imaging 21(1): 23-30 (2002)

[6] Konen W, Scholz M, Tombrock S: The VN-project: Endoscopic image processing for neurosurgery. Computer Aided Surgery 3, No. 3, pp. 144-148, 1998.

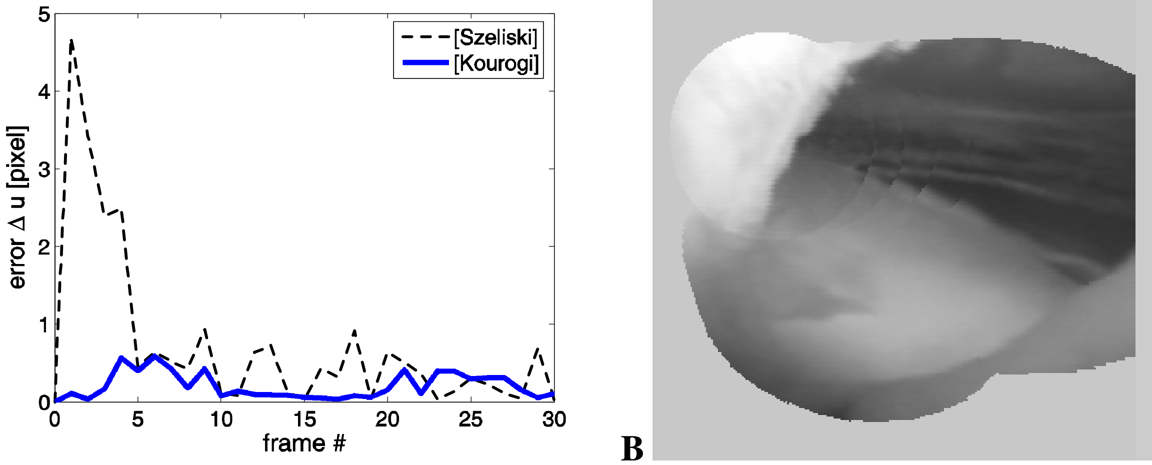
[7] Scholz M, Dick S, Fricke B et al.: Consideration of ergonomic aspects in the development of a new endoscopic navigation system, Br. J. of Neurosurgery, 19(5), 402-408, 2005.

[8] Wayne Rusband: ImageJ. http://rsb.info.nih.gov/ij. NIH, 2004-2007.



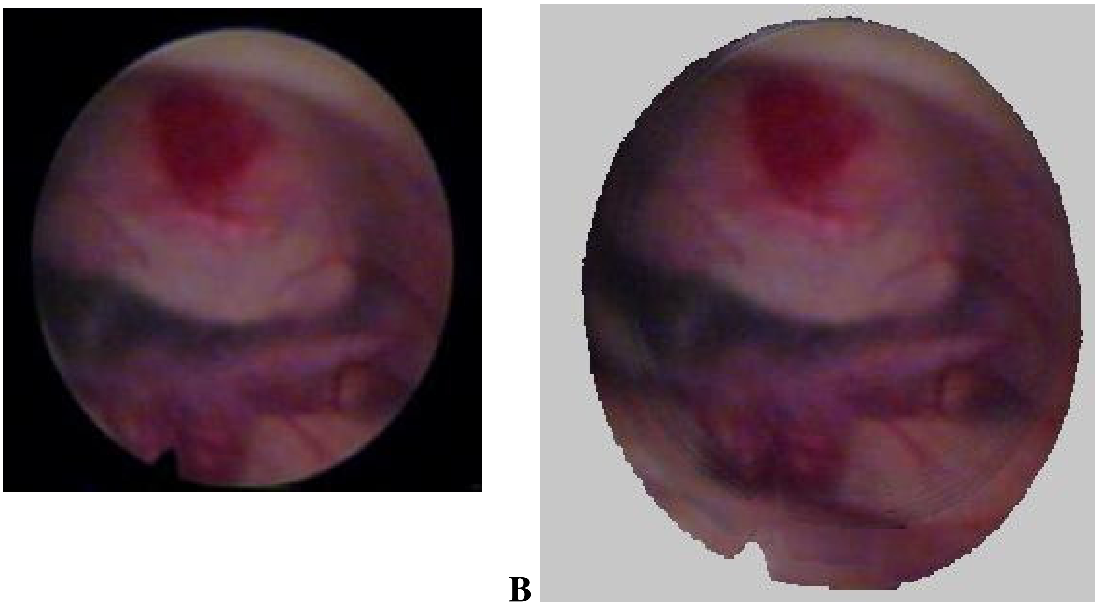
A B

Fig. 1: **A**: 4 out of 30 frames from a facial video sequence. **B**: Image mosaic resulting with Kourogi's algorithm. The mosaic is nearly undistinguishable to the original overall image.



A B

Fig. 2: **A**: Accuracy comparision between Szeliski's and Kourogi's algorithm for the video sequence of Fig. 1. **B**: Image mosaic resulting with Szeliski's algorithm. Clear artefacts from the registration error in frame #1-4 are visible.



A B

Fig. 3: **A**: First frame from the neuroendoscopic video. **B**: Color image mosaic created from seven consecutive frames