Real-Time Image Mosaic for Endoscopic Video Sequences

Wolfgang Konen[1], Beate Breiderhoff[1], Martin Scholz[2]

[1]Institut für Informatik, Fachhochschule Köln, 51643 Gummersbach
[2]Neurochirurgische Universitätsklinik, Ruhr-Universität Bochum, 44892 Bochum
Email: wolfgang.konen@gm.fh-koeln.de

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 algorithm is based on the method of Kourogi et al. (1999) which we extend to the case of endoscopic masks. The algorithm automatically finds 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 (1994). Our method turns out to be more robust, more than 3 times faster, having at the same time a 4 times smaller average motion estimation error: 0.19 pixel instead of 0.72 pixel between successive frames.

1   Introduction

In endoscopic interventions the surgeon has often to deal with a rather limited field of view which can cause navigational difficulties. It would be therefore desirable to have a tool which combines automatically many endoscopic video frames to a larger, metrically accurate field-of-view ('panoramic overview'). To our knowledge such an endeavour has not been undertaken so far, with the exception of a very recent (Oct'06) paper [1] which was brought to our attention after submitting this paper. They show interesting results obtained independently on endoscopic retinal images. Our approach is similar, using and evaluating a new method based on optical flow.

Many algorithms on image mosaicing are known, however only relatively few of them can work fully automatic, e.g. [2, 3, 1, 4] and under real-time conditions [2, 1, 4]. Except for [1] they have not yet been applied to endoscopic video. On the other hand, interesting work on combining and improving endoscopic images exist: Wald et al. [5] combine two endoscopic frames using manual control points and show how the image quality can be improved by using a smoothing cross dissolve technique. Vogt et al. [6] show how to reduce colour errors and mark specular lights in endoscopic images. This is an important prerequisite for image mosaics.

Fig. 1. The main idea of Kourogi's algorithm shown for the 1D-case
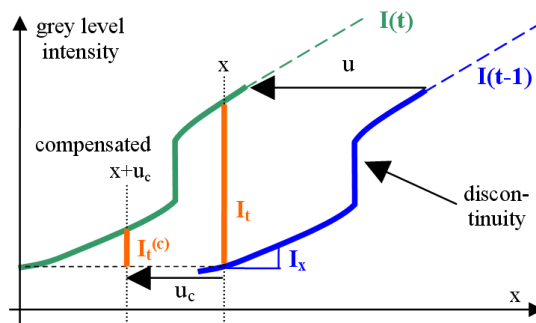
2   Methods

The goal of Kourogi's algorithm [2] 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

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

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. If we set $(u_c, v_c) = 0$ then $I_t^{(c)}$ becomes the time derivative I_t and we have the usual pseudo motion equation, which is however known to be non-robust and bound to fail at discontinuities or non-linearities in grey level distribution (Fig. 1). Due to the shown discontinuity the estimate $-I_t/I_x$ will be larger than the true motion u. If, on the other hand, u_c is an estimate for the true motion, it is likely that we avoid the discontinuity and get with $-I_t^{(c)}/I_x + u_c$ a good estimate for u. Note that $I_t^{(c)}$ has to be calculated with subpixel accuracy.

Our algorithm proceeds now as follows: First the compensated motion is initialized either with zero or with an estimate from the previous frame. Then the following steps are carried out in a loop:

1. Calculate the pseudo motion for each pixel of the endoscopic mask.
2. 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) $|I(x + u_p, y + v_p, t - I(x, y, t - 1)| < T$. Here, T is a suitable grey level threshold, e.g. $T = 5$.

3. Find the affine parameters $\mathbf{a} = \{a_1, \ldots, a_6\}$ for a global motion field best-fitting the pseudo motion at all accepted pixel locations i, i.e. solve the overdetermined system of equations

$$a_1 x_i + a_2 y_i + a_3 = u_{p,i} \quad \text{and} \quad a_4 x_i + a_5 y_i + a_6 = v_{p,i} \tag{1}$$

in a least-square sense. Use the motion field given by $\mathbf{a}$ as a new estimate for (u_c, v_c) and continue with step (1.).

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 can be done rather fast since each frame adds only a small new region to the existing mosaic. Currently no special blending occurs but a blending strategy as reported in [5] can be easily incorporated.

3 Results

We tested our algorithm on short endoscopic video sequences. In this first step the goal was to measure its accuracy and to compare its performance with another well-known image mosaicing algorithm [3] applied to the same task. We created a short endoscopic video sequence (30 frames) where each frame is connected to the next by a known affine transform, for example translations up to 10 pixel, size changes up to 8% and combinations thereof. These transforms correspond to simple camera movements. The motion field differs from frame to frame. Figs. 2a and 2c show 4 frames out of these sequences.

We tested two algorithms: The first one is our method described in Sec. 2, based on Kourogi's algorithm [2] with acceptance threshold $T = 5$. The second one is based on the well known Szeliski image mosaicing algorithm [3]. 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 Kourogi's case (Fig. 2b and 2d). In the case of Szeliski's algorithm (not shown here) it has clear artefacts and errors in aspect ratio estimation.

In Fig. 3, 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 (e.g. frames #1-4). This is

Fig. 2. Results: (a) 4 out of 30 frames from a facial video sequence, (b) image mosaic resulting with Kourogi's algorithm, (c) and (d) the same for the neuroendoscopic video sequence
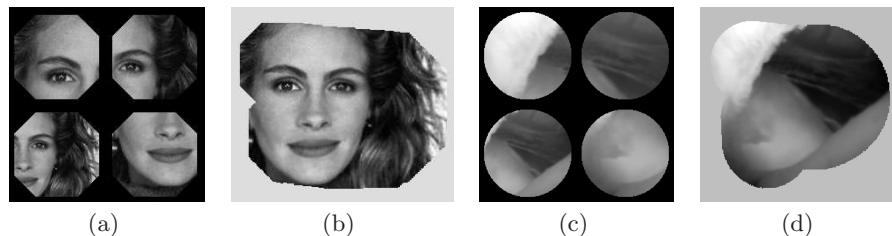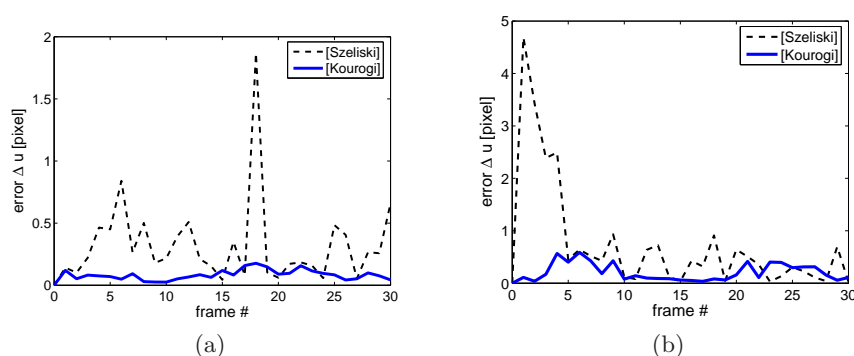

(a)   (b)   (c)   (d)

Fig. 3. Accuracy of Szeliski's and Kourogi's algorithm: (a) facial video, (b) neuroendoscopic video

(a)   (b)

important because a single outlier will make all subsequent frames in the image mosaic wrong w.r.t. the frames before.

While having a 4 times better accuracy, the method of Kourogi is at the same time faster by a factor of 3 (Tab. 1). The computation time of 4.3 sec/frame is obviously not yet real time. But this is only due to the fact, that our current implementation is a first development step with slow Matlab code. We plan to rebuild the system in C or C++ and have no doubt that we can reach with today's standard PCs a performance of at least 10 fps as Kourogi [2] reported it 8 years ago.

4 Conclusion and Outlook

We have shown how to build image mosaics from endoscopic video sequences. Of course our work is only a very 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.

Table 1. Comparison of main features of both image mosaicing algorithms

	computation time	avg. frame-to-frame accuracy $< \Delta u >$	outlier frames ($< \Delta u >$ > 1 pixel)
Szeliski	15.5 sec/frame	0.72 pixel	4
Kourogi	4.3 sec/frame	0.19 pixel	0
improvement	factor 3	factor 4	

We believe that the crucial feature of Kourogi's method is the pixel test in step (2.), which allows to base any estimate only on those pixels where the information seems reliable. This flexibility of the proposed algorithm has a further advantage when dealing with specular lights: If a system along the lines of [6] detects specular lights, these can be easily accounted for by marking them as 'non-acceptable' pixels during the registration process.

There are many directions we plan to investigate in the near future: The class of transforms should be extended from affine to projective to account for more general camera movements. The distorsion of an endoscope lens system has to be taken into account. We know from our previous work [7] how to calibrate an endoscope camera system and therefore believe that this step is quite straightforward. Changes in lighting should be accounted for: Global contrast and brightness will vary slowly from frame to frame, but specular lights can vary quite rapidly. We plan to port the system to a real-time environment and test it with real endoscope sequences, leading finally into the integration in our VN system [8]. Different strategies for combining videos to image mosaics will be explored and will be tested with respect to their ergonomic requirements by surgeons working in daily routine with endoscopic images.

References

1. Seshamani S, Lau WW, Hager GD. Real-time endoscopic mosaicking. Procs MICCAI 2006;4190:355–363.
2. Kourogi M, Kurata T, Hoshino J, et al. Real-time image mosaicing from a video sequence. In: International Conference on Image Processing. vol. 4; 1999. 133–137.
3. Szeliski R. Image mosaicing for tele-reality applications. Cambridge Research Lab; 1994.
4. Robinson JA. A Simplex-based projective transform estimator. In: Visual Information Engineering; 2003. 290–293.
5. Wald D, Reeff M, Székely G, et al. Fliessende Überblendung von Endoskopiebildern für die Erstellung eines Mosaiks. Procs BVM 2005; 287–291.
6. Vogt F, Klimowicz C, Paulus D, et al. Bildverarbeitung in der Endoskopie des Bauchraums. Procs BVM 2001; 320–324.
7. Konen W, Scholz M, Tombrock S. The VN-project: endoscopic image processing for neurosurgery. Computer Aided Surgery 1998;3(3):144–148.
8. Scholz M, Dick S, Fricke B, et al. Consideration of ergonomic aspects in the development of a new endoscopic navigation system. Br J Neurosurg 2005;19(5):402–408.