METHOD OF MOTION ESTIMATION FOR IMAGE STABILIZATION

Marius Tico, Sakari Alenius, Markku Vehvilainen

Nokia Research Center, POBox 100, FIN-33721 Tampere, Finland

ABSTRACT

In this paper we introduce a novel approach to global motion estimation for image stabilization application. The method is robust to image degradations characteristic to image stabilization, e.g. image blur caused by motion or out of focus. In addition, due to its low computational complexity, the proposed method could be included in a real-time digital image stabilization system. The ability of the proposed registration approach to capture the global motion of the camera in the presence of image degradations and outliers, have been evaluated through a large number of experiments. The results reveal that, in spite of its lower computational complexity, the proposed method achieves sub-pixel motion estimation accuracy, close to the performance achieved by the state of the art approaches to image registration.

1. INTRODUCTION

Image stabilization objective is to remove the effect of unwanted motion fluctuations from video data. An image stabilizer comprises two parts: camera motion estimation, and unwanted motion compensation. If inertial motion sensors are not available, the camera motion must be estimated from the video data using a global motion estimation algorithm. In the second part of the stabilization algorithm, the unwanted component of the camera motion is canceled by warping the video frames accordingly [1, 2]. The final result of the stabilizer could be either a stabilized video stream, or a still image compounded (restored) based on several overlapping frames.

The camera motion estimation is essentially an image registration problem, where the images to be registered are successive frames in the video stream. The complexity and the magnitude of the motion between successive frames depends of factors like frame rate, focus distance, physical properties of the device hosting the camera, etc. Using hand held devices, at normal video frame rates (i.e. 25-30Hz), and normal focus, the inter-frame motion could be modeled as a rigid transformation (translation and rotation). However, in order to reduce the computational complexity, the motion model is often simplified further to a simple translation [1].

The existent approaches to image registration could be classified in two categories: feature based, and featureless methods [3]. The feature based methods rely on determining the correct correspondences between different types of visual features extracted from the images. In some applications, the feature based methods are the most effective ones, as long as the images are always containing specific silent features (e.g. minutiae in fingerprint images [4]). On the other hand, when dealing with natural images of various nature, it is more difficult to define a certain type of silent features that are detectable in sufficient number in all images. A more robust alternative for such applications could be a featureless image registration approach, that utilizes directly the intensity information in the image pixels, without searching for specific visual features.

In general a parametric model for the two-dimensional mapping function that overlaps an "input" image over a "reference" image is assumed. Let us denote such mapping function by $\mathbf{f}(\mathbf{x};\mathbf{p}) = [f_x(\mathbf{x};\mathbf{p})\ f_y(\mathbf{x};\mathbf{p})]^t$, where $\mathbf{x} = [x\ y]^t$ stands for the coordinates of an image pixel, and $\mathbf{p}$ denotes the parameter vector of the transformation. Denoting the "input" and "reference" images by I and R respectively, the objective of a featureless image registration approach is to estimate the parameter vector $\mathbf{p}$ that minimizes a cost function (e.g. the sum of square differences) between the transformed input image $I(\mathbf{f}(\mathbf{x};\mathbf{p}))$ and the reference image $R(\mathbf{x})$.

The minimization of the cost function, can be achieved in various ways. A trivial approach would be to adopt an exhaustive search among all feasible solutions by calculating the cost function at all possible values of the parameter vector. Although this method ensures the discovery of the global optimum, it is usually avoided due to its tremendous complexity. To improve the efficiency several alternatives to the exhaustive search technique have been developed by reducing the searching space at the risk of losing the global optimum, e.g. logarithmic search, three-step search, etc. Another category of featureless image registration approaches, known as gradient-based methods, assume that an approximation to image derivatives can be consistently estimated, such that the minimization of the cost function could be achieved by applying a gradient-descent technique [5, 6]. An important efficiency improvement, for gradient-based algorithms, has been proposed in [5], under the name of "Inverse Compositional Algorithm" (ICA). The improvement results from the fact that the Hessian matrix of the cost function, needed in the optimization process, is not calculated in each iteration, but only once in a pre-computation phase.

In this paper we introduce a novel approach to camera motion estimation based on a featureless image registration technique. The proposed approach satisfies well the specific requirements of an image stabilization application, achieving a sub-pixel accuracy with rather low computational cost in comparison to other approaches.

2. THE PROPOSED METHOD

The main challenges faced by an image registration approach used to estimate the camera motion in the context of image stabilization application are as follows.

- The visual quality of the video frames could be often degraded by noise and various types of blur. Thus, the very reason why stabilization is needed is the presence of unwanted high frequency motions of the camera during video capturing. These motions are not only displacing the video frames one with respect to another, but they occur also during the exposure time of each frame causing motion blur degradations of the images.
- The presence of outliers represented by independently moving objects in the scene.

- The need to achieve a low computational complexity. A real-time implementation of a stabilization solution is desirable in most cases, and often the target device has rather limited computational power. Consequently, the low complexity requirement of the image stabilization solution is essential for practical acceptability.

In the following we describe our approach by addressing the requirements formulated above in the context of image stabilization application.

Image degradations are mainly affecting the high frequency components of an image, destroying the fine details present in the original images. On one hand, blurring, caused by various factors like camera motion during exposure, or out of focus optical system, is a form of bandwidth reduction of the original image. On the other hand, additive noise present in the image affects all frequency bands, but due to high spatial correlation of natural images, it dominates only at high frequencies. The degradation of high spatial frequencies has a negative impact on the accuracy achieved by a registration method.

To reduce this effect we apply a low-pass filtering operation on the two images before registration, in order to attenuate the high frequency components that are likely to be disturbed. Thus, in the presence of image degradations, the two smoothed images are more similar than the original ones, the differences between them being mainly of a geometrical nature that should be actually resolved by the registration algorithm. For low-pass filtering we employed a method inspired from the fast dyadic wavelet decomposition algorithm [7]. The method consists of iteratively smoothing the original image I such that to obtain smoother and smoother versions of it. Let $\tilde{I}_\ell$ denotes the smoothed image resulted after ℓ-th low-pass filtering iterations ($\tilde{I}_0 = I$). The smoothed image at next iteration is calculated by applying one-dimensional filtering along the image rows and columns as follows:

$$\begin{array}{ll} \text{Tmp}(x,y) & = \sum_c h_c \tilde{I}_\ell \left(x - 2^\ell c, y\right), \\ \tilde{I}_{\ell+1}(x,y) & = \sum_r h_r \text{Tmp}\left(x, y - 2^\ell r\right), \end{array} \quad (1)$$

where h_k are the taps of the low-pass filter. In our work we use a symmetric filter of size 3, whose taps are respectively $h_1 = 1/4$, $h_0 = 1/2$, and $h_1 = 1/4$.

Due to low-pass filtering, the resulted image after L smoothed iteration ($\tilde{I}_L$) is over-sampled, and hence it could be reconstructed by a subset of its pixels. This property allows to enhance the efficiency of the registration process by using only a subset of the smoothed image pixels in the registration algorithm. The advantage offered by the availability of the smoothed image, is that the set of pixels that can be used in the registration is not unique. A broad range of geometrical transformations could be thereby approximated by simply choosing a different set of pixels to describe the smoothed image. In this way, the smoothed image is regarded only as a "reservoir of pixels" for different warped low-resolution versions of the image, which may be needed at different stages in the registration algorithm.

Let $\mathbf{x}_{n,k} = [x_{n,k}\ y_{n,k}]^t$, for n, k integers, denote the coordinates of the selected pixels into the smoothed image ($\tilde{I}_L$). A low-resolution version of the image ($\hat{I}$) can be obtained by collecting the values of the selected pixels: $\hat{I}(n,k) = \tilde{I}_L(\mathbf{x}_{n,k})$. Moreover, given an invertible geometrical transformation function $\mathbf{f}(\mathbf{x};\mathbf{p})$, the warping version of the low resolution image can be obtained more efficiently by simply selecting another set of pixels from the area of the smoothed image, rather than warping and interpolating the low-resolution image $\hat{I}$. This is: $\hat{I}'(n,k) = \tilde{I}_L(\mathbf{x}'_{n,k})$, where $\mathbf{x}'_{n,k} = \text{round}\left(\mathbf{f}^{-1}(\mathbf{x}_{n,k};\mathbf{p})\right)$.

The process described above is illustrated in Fig.1, where the images shown on the bottom row represent two low-resolutions warped versions of the original image (shown in the top-left corner). The two low-resolution images are obtained by sampling different pixels from the smoothed image (top-right corner) without interpolation.

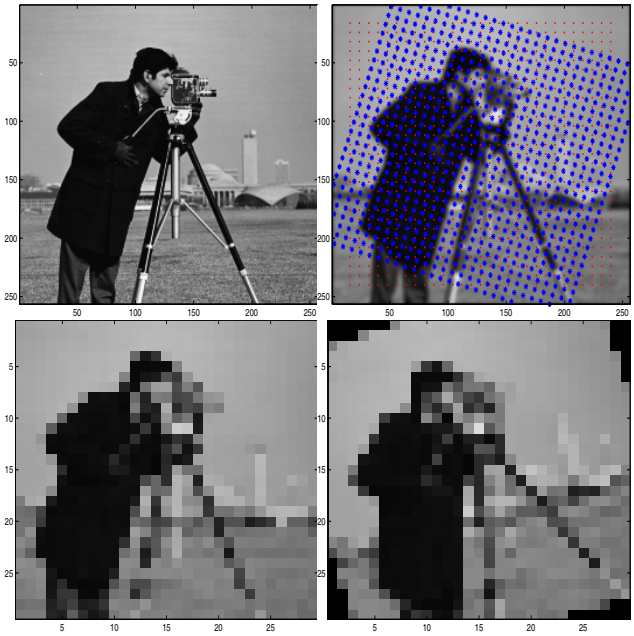

Fig. 1. Low-resolution image warping by re-sampling the smoothed image.

At this point we can formulate the registration algorithm used in our approach. The algorithm follows the ICA framework [5], by taking advantage of the efficiency improvements proposed there. In addition, the efficiency is further improved by simplifying the image warping operations needed in each iteration of the optimization procedure. Adopting a three parameter rigid motion model, the proposed registration algorithm is briefly described as follows:

Input: the input and reference images plus an initial guess of the parameter vector $\mathbf{p} = [p_1\ p_2\ p_3]^t = [t_x\ t_y\ \theta]^t$, where (t_x, t_y) and θ denotes respectively the translation and rotation between the images.

Output: the parameter vector that overlaps the input image (I) over the reference image (R)

Pre-computation:

1. Calculate the smoothed images $\tilde{I}_L$, $\tilde{R}_L$
2. Set the initial position of the sampling points $\mathbf{x}_{n,k}$ in the vertex of a rectangular lattice of period $D = 2^L$, over the area of the two smoothed images.
3. Construct the reference image: $\hat{R}(n,k) = \tilde{R}_L(\mathbf{x}_{n,k})$.
4. Approximate the gradient $\hat{R}_x$, $\hat{R}_y$ of the reference image by applying on $\hat{R}$ the operators: $\begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}$, and $\begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}$.
5. For each parameter p_i of the warping function calculate the image

$$J_i(n,k) = \hat{R}_x(n,k)\frac{\partial f_x(\mathbf{x};\mathbf{0})}{\partial p_i} + \hat{R}_y(n,k)\frac{\partial f_y(\mathbf{x};\mathbf{0})}{\partial p_i}$$

6. Calculate the first order approximation of the 3×3 Hessian matrix, whose element (i, j) is given by:

$$\mathbf{H}(i,j) = \sum_{n,k} J_i(n,k) J_j(n,k)$$

Iteration:

1. Construct the warped low-resolution input image in accordance to the warping parameters estimated so far: $\hat{I}(n,k) = \tilde{I}_L(\text{round}\left(\mathbf{f}^{-1}(\mathbf{x}_{n,k}; \mathbf{p})\right))$.
2. Calculate the error image $E(n,k) = \hat{I}(n,k) - \hat{R}(n,k)$, and smooth it by applying a 2×2 constant mask.
3. Calculate the 3×1 vector of elements:

$$\mathbf{g}(i) = \sum_{n,k} E(n,k) J_i(n,k), \; i = 1, 2, 3.$$

4. Update the parameter vector

$$\mathbf{p} = \mathbf{p} + \text{diag}\,(D, D, 1)\,\mathbf{H}^{-1}\mathbf{g}.$$

There are several possible choices for the iteration stoping criterion. In our work we use a crietarion based on the value of the mean absolute error MAE $= \sum_{n,k} |E(n,k)|$. Thus, in each iteration where MAE becomes smaller than its minimum value achieved so far, the algorithm stores the parameter vector and updates the minimum MAE. The iteration process ends when the MAE do not go bellow the last minimum for a specific number of iterations. A maximum number of allowed iterations is also specified in order to ensure stoping the process in any conditions.

The outlier rejection is achieved by running the above algorithm a few times and analyzing the error image $E(n,k)$ at the end of each such running. Based on error image values, the pixels are classified either as *inliers* or *outliers* after each running, following to use only the inliers in the next running of the algorithm. Denoting by O a binary image used as outlier segmentation mask (i.e. $O(n,k) = 0$ if (n,k) outlier), we have after each running of the algorithm:

$$O(n,k) = \begin{cases} 1 & \text{if } |E(n,k)| < \tau, \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

where the threshold τ is calculated based on the estimated standard deviation of the error image. In our experiments we observed that the above process converges quite fast, such that after a very few iterations of the algorithm (typically 2, 3), the outlier segmentation mask do not changed anymore.

3. EXPERIMENTAL RESULTS

We tested the proposed registration algorithm on natural images representing various scenes, as shown in Fig.2. The images have the same resolution of 512x512 pixels. In order to evaluate the accuracy of the registration method we conducted a set of experiments in which a given transformation function, applied on each image, should be recovered by the algorithm. The transformation parameter vector that we used in this experiments was $\mathbf{p} = [10\ 10\ 10^\circ]^t$, and the initial guess for all tests was the identity transformation (i.e. null parameter vector). A pair of "input" and "reference" images have been created from each test image by applying the above transformation. The transformation estimated by a certain algorithm was then evaluated based on its ability to overlap the corresponding pixels of the two images. In our work, we used as error criterion the distance between the centers of two most distanced corresponding pixels after registration.

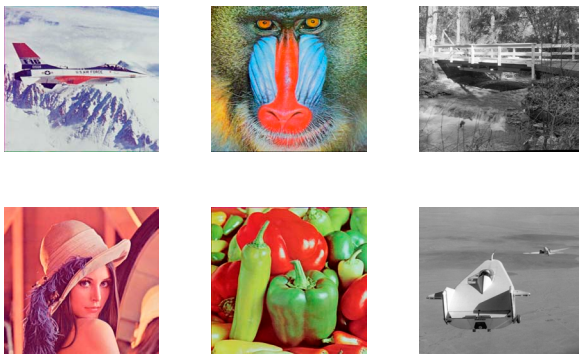

Fig. 2. The set of various natural images used for evaluation tests.

The following registration algorithms have been used for comparison:

a The proposed method.

b The image registration algorithm proposed in [6]. The algorithm uses cubic spline interpolation for image warping, and employs a coarse-to-fine strategy (pyramid approach).

c A course-to-fine Gaussian pyramid approach that employs the ICA algorithm [5] at each level of the pyramid. Cubic interpolation was used in order to accomplish the various image warping operations needed at each level during the iterative optimization procedure.

d A reduced variant of (b) in which only the coarsest level of the pyramid is used. This approach is the most similar with our approach that processes the image only at one level. The main difference is that in our approach the coarsest level image is not sub-sampled like in a typical pyramid decomposition.

For all experiments the coarsest decomposition level of the images was set to 5, with the first level corresponding to the original image resolution.

The experiments have been divided in three sets, according to the degradations applied to each pair of "input" and "reference" images before submitting them to the registration algorithm. These are:

- **Clean image tests:** No image degradation have been added to the images before registration. For each test image, a registration experiment was performed with each one of the registration methods considered here.
- **Motion blur tests:** In these experiments the images have been artificially degraded by a linear motion blur of length 15 pixels, plus a zero mean Gaussian noise of normalized variance 0.001 (i.e. using Matlab `imnoise()` method). We performed a number of 100 experiments per image. In each such experiment a new realization of the Gaussian noise and new motion blur orientations have been randomly generated in each one of the two images submitted for registration.
- **Out of focus tests:** In these experiments the images have been artificially degraded by a uniform out of focus blur modeled by a circular PSF of radius 11 and 1 in "input" and "reference" images respectively. In addition a zero mean Gaussian noise of normalized variance 0.001 was also added to the

Fig. 3. Three video frames containing changing (moving) foreground.

two images before registration. As in the motion blur test described above, we performed 100 experiments per image by generating different realization of the additive noise in each of the two images before registration.

Image degradation	Clean image	Motion blur	Out of focus
Method	Error (pixels)		
	avg.(std.) above 1	avg.(std.) above 1	avg.(std.) above 1
(a)	0.13 (0.06) 0%	0.23 (0.12) 0%	0.25 (0.11) 0%
(b)	0.13 (0.00) 0%	0.22 (0.10) 0%	0.25 (0.11) 0%
(c)	0.07 (0.05) 0%	1.22 (0.43) 67%	1.97 (0.75) 100%
(d)	0.76 (0.82) 17%	0.84 (0.64) 26%	0.73 (0.30) 17%

Table 1. Image registration accuracy achieved by the four methods described in the text. The entries of the table show the average and standard deviation of the errors found in all experiments, as well as the percent of experiments in which the error was above 1 pixel.

The results in Table 1 show that the proposed approach and the approach (b) clearly outperforms the other two approaches in the presence of image degradations. However, the method proposed here achieves a much lower computational complexity than (b). This is because our approach uses for registration operations only a small number of pixels selected from the smoothed images, and in addition no interpolation is used for image warping during the optimization process. It is of importance to emphasize also that the method (d) which carries out operations only at one level of the image pyramid has inferior performance to the method proposed here. The same we can tell about method (c) in the presence of image degradations. However, we note also that in the absence of any image degradations (i.e. clean image tests), the method (c), outperforms the other three methods considered here.

In order to demonstrate the ability of the proposed method to eliminate outliers we use the video frames shown in Fig. 3. The task of the algorithm is to estimate the camera motion with respect to the background making abstraction of the car which is passing in front of the camera. Fig. 4 shows the overlapped frames as well as the outlier rejection masks calculated by the algorithm. We note that the outliers represented by the area of the moving car are segmented quite well by the algorithm, resulting in a robust camera motion estimation with respect to the background.

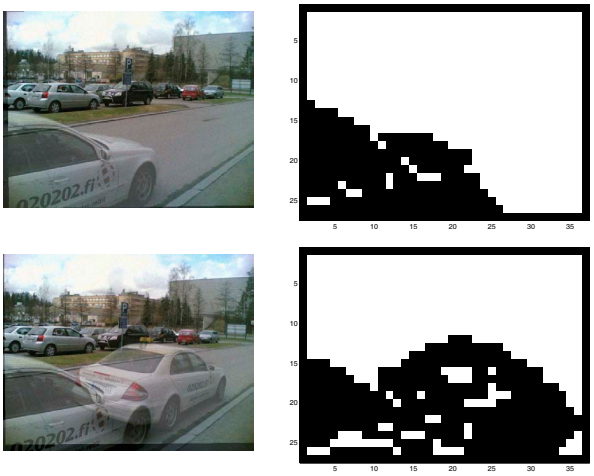

Fig. 4. The overlapped frames 1,2 (up) and 2,3 (down), along with the corresponding outlier segmentation masks calculated by the algorithm.

4. CONCLUSIONS

We introduced a novel approach to camera motion estimation for image stabilization. The specific requirements imposed by this application have been formulated, and the proposed method has been designed accordingly. Following a gradient-based procedure, the method achieves a significant reduction in complexity by replacing the image warping operations with simple pixel selections from a smoothed image version. The robustness to various image degradations is achieved by using only the low-frequency components of the images. The proposed approach has been demonstrated through several experiments and comparisons.

5. REFERENCES

[1] S. Erturk, "Digital image stabilization with sub-image phase correlation based global motion estimation," *IEEE Transaction on Consumer Electronics*, vol. 49, no. 4, pp. 1320–1325, 2003.

[2] Marius Tico and Markku Vehvilainen, "Constraint motion filtering for video stabilization," in *Proc. of the IEEE International Conference of Image Processing (ICIP)*, Genova, Italy, Sep. 2005, vol. 3, pp. 569–572.

[3] Barbara Zitova and Jan Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, pp. 977–1000, 2003.

[4] Marius Tico and Pauli Kuosmanen, "Fingerprint matching using an orientation-based minutia descriptor," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 1009–1014, 2003.

[5] Simon Baker and Iain Matthews, "Lucas-kanade 20 years on: A unifying framework," *International Journal of Computer Vision*, 2004.

[6] P. Thevenaz and M. Unser, "A Pyramid Approach to Subpixel Registration Based on Intensity," *IEEE Transactions on Image Processing*, vol. 7, no. 1, pp. 27–41, 1998.

[7] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, USA, 1990.