ESTIMATION OF CAMERA PLANAR MOTION FROM BLURRED IMAGES

Barbara Zitová and Jan Flusser

Institute of Information Theory and Automation
Pod vodárenskou věží 4
Academy of Sciences of the Czech Republic
182 08 Praha 8, Czech Republic

ABSTRACT

The topic of the paper is to estimate parameters of the camera motion using images taken at initial and current positions of the camera. An original method handling defocused images or images blurred in other ways is proposed. The method is based on registering the images by means of detected control points (CPs) and so-called *combined invariants*. These features are invariant both to rigid-body transform and to blurring caused by convolution with an arbitrary centrosymetric point-spread function (PSF). Thanks to this, the camera position can be estimated directly from defocused images without any deblurring.

1. INTRODUCTION

To estimate a current position of an uncalibrated moving camera with respect to its previous position is a frequent task in computer vision. Information which is available consists of images taken at the initial and current positions of the camera, sometimes additional prior knowledge can be available (for example the kind of motion). Provided the images have at least partial overlap, an accurate estimation of the parameters of the spatial transformation between the images is the key to solving this problem.

There are basically two main approaches to the estimation of transform parameters: correspondence-based and correspondenceless ones. In the first one, any method of image-to-image registration can be employed to find corresponding CPs and to estimate the parameters of the transform via interpolation or least-square fit. This approach was chosen for instance in the well-known experiment by Zheng and Chellappa [1]. They were estimating the wind velocity by means of images taken from a floating balloon.

Correspondenceless methods look at the images globally and try to find the transform parameters directly. Typical examples are phase correlation in the Fourier domain [2] and image normalization. The latter approach uses global image moments to transform each image into a canonical form [3]. Between-image transform is obtained as a composition of canonical transformations. The correspondenceless methods are simpler but their common drawback is their globality. They require a complete overlap of the images that is not realistic in practice.

In this paper, our attention is paid to the estimation of the moving camera position in the situation when one of the images is blurred in an unknown manner. The blur can be introduced by various factors such as wrong focus, atmospheric turbulence (important particularly in remote sensing and astronomy), diffraction and camera vibration, among others. The motion blur is not addressed explicitly because short-exposure imaging is assumed.

Since general-purpose registration methods do not employ any blur invariant property, they are not supposed to handle blurred images correctly. Only few papers considering blurred images have been published. Myles and Lobo [4] proposed an iterative method originating from the direct motion estimation approach [5] and modified it particularly for blurred images. Although it was proven to converge in most cases and to be robust to noise, it has severe limitations: the blurring function is assumed to be a Gaussian or pillbox-like, the scene must be flat (or, at least, it must consist of large planar patches) and a significant image overlap is required. Moreover, good initial estimate of the translation is required. Zhang et al. [6] proposed an affine normalization method that does not depend on the blurring function. To bring the images into canonical form, blur invariant moments (introduced by Flusser and Suk [7]) were used to define the normalization constraints. Thanks to this, neither the type nor the level of the blur influences the motion parameter estimation. Pei and Wu [8] used the same idea to register images deformed by a rigid-body transform. These two methods are robust to noise and computationally efficient but they require the complete image overlap to yield reliable results. Kubota et al. [9] proposed a two-stage registration method based on a hierarchical matching. In the first stage, they perform a full search in the space of transform parameters, without taking the blur into account at all, to find a coarse match. Then the refined match is found

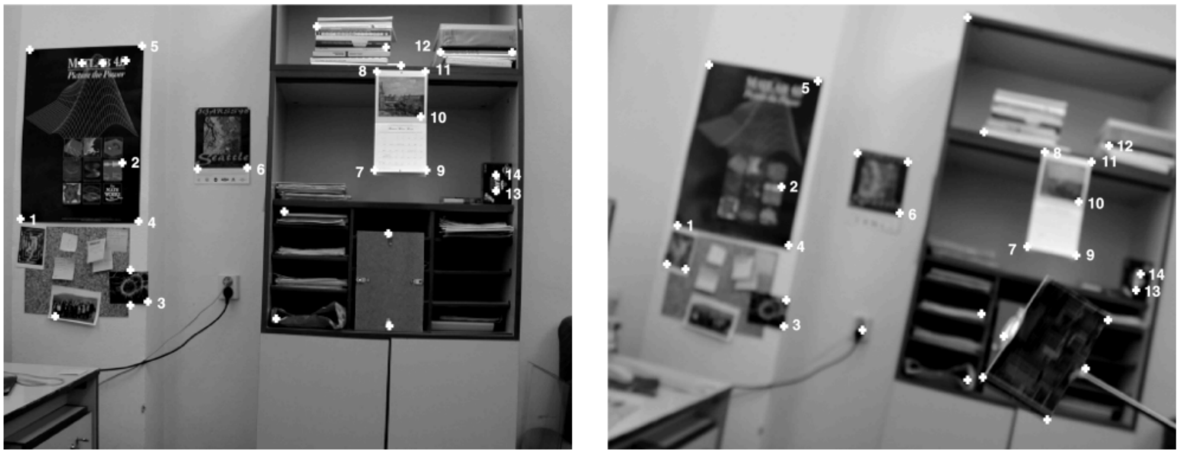

Fig. 1. Images of the indoor scene, sized 384×512 pixels, (a) - left image: camera at the initial position, (b) - right image: after the camera moved. The image taken by the moved camera has an out-of-focus blur, caused by the inserted foreign object. 30 detected CPCs are marked by crosses. The CPCs which form the corresponding CP pairs are numbered.

block-wise, considering the blur radius as another parameter of the search space. All possible Gaussian blurs are generated and examined to find an optimal match.

We present a new correspondence-based method of the camera motion estimation from blurred images. We assume, that the relationship between the images taken at initial camera position ($f(x,y)$) and at the current camera position ($g(x,y)$) can be described as

$$g(\tau(x,y)) = a(f * h)(x,y) + n(x,y). \qquad (1)$$

In this model, $n(x,y)$ is additive random noise, a is a constant describing the global change of contrast, $*$ denotes 2-D convolution and $h(x,y)$ is the PSF of the camera at the time of the acquisition of $g(x,y)$. Ideally, $h(x,y)$ equals to Dirac delta-function, in practice the PSF is a composition of Airy function. The PSF might have also a component depending on changes in imaging geometry. Since we do not attempt to recover the blur and all the PSF components are considered unknown, we can group them under one symbol $h(x,y)$ without loss of generality. In the following text the PSF is assumed to be centrosymmetric (i.e. $h(x,y) = h(-x,-y)$) and energy preserving ($\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)dxdy = 1$). This is not a significant limitation from practical point of view, because these assumptions hold for most real sensors and blurring functions. τ on the left hand side of the equation (1) stands for a transform of spatial coordinates due to the changes in imaging geometry. The type of τ depends on the number of degrees of freedom of the camera motion and on the character of the scene. Methodology of the motion estimation is influenced by the actual character of τ. We solve the task for τ consisting of translation and rotation. Although not addressed explicitly, small scale changes are allowed too.

2. IMAGE REGISTRATION APPROACH

The proposed method of the camera motion estimation is based on image registration — the process of overlaying two or more images of the same scene acquired from different viewpoints and/or at different times so that the pixels of the same coordinates in the images correspond to the same part of the scene. Here, we do not need actually to overlay the images, we only look for the coordinate transform that describes the spatial relationship between them.

Image registration starts with the selection of control point candidates (CPCs) both in $f(x,y)$ and $g(x,y)$ frames. Significant corners and other corner-like dominant points are considered as the candidates. To detect them, a method developed particularly for blurred images [10] is employed.

The most difficult step is a matching of the CP candidates. The correspondence between the CPCs from the initial frame and those from the current frame must be established and the candidates having no counterparts should be rejected. This step is realized by means of a new class of features, called *combined invariants*, introduced in [11]. The combined invariants I are invariant simultaneously to convolution with an arbitrary centrosymmetric PSF and rotation around the origin (additional invariance to translation and scaling can be obtained) and are defined as follows:

$$I = \prod_{j=1}^{n} K(p_j, q_j)^{k_j}, \qquad \sum_{j=1}^{n} k_j(p_j - q_j) = 0, \qquad (2)$$

where $n \geq 1$ and k_j, p_j and $q_j; j = 1, \cdots, n,$ are non-negative integers such that $(p_j + q_j)$ is odd for each j. The

Image	R angle (^{o})	R diff	V shift	V diff	H shift	H diff
m, *s*	6.38	0.10	-25.3	0.6	-35.3	-0.9
m, *b*	10.65	-0.02	-40.3	0.1	-56.0	0.0
A, *s*	6.20	0.07	-20.4	-0.6	-14.8	0.2
A, *b*	11.00	0.08	-37.7	0.0	-41.0	0.2
B, *s*	6.28	0.11	-19.7	0.0	-15.7	-0.9
B, *b*	10.82	0.03	-38.1	0.2	-41.1	-0.3
mean		0.06		0.1		-0.3
STD		0.05		0.4		0.5

Table 1. Parameters of the camera motion. Estimates of the rotation (R) and of the translations in vertical (V) and horizontal (H) directions (in pixels) and the difference between the computed values and the ground truth. Images were blurred (manual defocus, inserted objects A and B), rotated (*big* angle, approximately 10.8^o; *small* angle, approximately 6.2^o) and translated.

$K(p_j, q_j) : \mathcal{Z} \times \mathcal{Z} \to \mathcal{C}$ ($\mathcal{Z}$ is the set of non-negative integers and $\mathcal{C}$ is the set of complex numbers) are convolution invariants ([11]). If $(p_j + q_j)$ is even then $K(p_j, q_j)^{(f)} = 0$, if $(p_j + q_j)$ is odd then

$$K(p_j, q_j)^{(f)} = c^{(f)}_{p_j q_j} - \frac{1}{c^{(f)}_{00}} \sum_{n=0}^{p_j} \sum_{\substack{m=0 \\ 0<n+m<p_j+q_j}}^{q_j} \binom{p_j}{n} \binom{q_j}{m} \times$$

$$\times K(p_j - n, q_j - m)^{(f)} \cdot c^{(f)}_{nm}.$$

The $c^{(f)}_{nm}$ in the definition stands for the *complex moments*. The complex moment $c^{(f)}_{nm}$ of order $(n+m)$, where $n \geq 0$ and $m \geq 0$, of image $f(x, y)$ is defined as

$$c^{(f)}_{nm} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + iy)^n (x - iy)^m f(x, y) dx dy, \quad (3)$$

where i denotes the imaginary unit.

A vector of such invariants is computed for each CPC over its circular neighborhood. Two CPC pairs with the minimum distance of their invariant vectors are found as the most-likely corresponding CPCs (more robust *matching likelihood coefficients* algorithm [12] can be used for correspondence estimation). The other CPCs from the current frame are transformed using a rigid-body transform the coefficients of which are calculated by means of the two above mentioned CPC pairs. The correspondence between the transformed CPCs from the current frame and CPCs in the initial frame is found via the thresholded nearest neighbor rule in the spatial domain. Finally, for each CP in the current frame, its improved position is found in its local neighborhood by means of combined invariants.

As soon as the control point correspondence is established, we can find an ”optimal” rigid-body mapping function whose parameters are calculated via least-square technique. Knowing these parameters (rotation angle, scaling factor and translation vector) and the initial camera position, the current position can be easily estimated.

Since our approach is based on local properties of images, a partial overlap of the frames is sufficient. The method is time effective. In the correlation-like and direct methods, each registration parameter increases the dimension of the search space. For instance, in case of rigid-body transform and blurring we have 5-D optimization problem which is time expensive even if multiscale approach is employed. In the invariant-based approaches, as our method is, the use of the invariants reduces the search space significantly. Further reduction is achieved by the preliminary detection of control points.

3. INDOOR EXPERIMENT

The proposed method was tested on the estimation of the motion parameters and the current position of the camera in the real indoor scene. The camera (Nikon Coolpix 950, 250 cm distant from the wall of the room) was variously rotated and translated from its initial position (common for all cases under consideration). Using images (sized 384×512) taken at the start and at the end of moves of the camera, the motion parameters were estimated. Fig. 1 shows images taken at the initial position (a) and after the camera was moved (b). Images were blurred by out-of-focus blur caused by a manual defocus or by the insertion of foreign objects between the camera and the scene (see Fig. 1 (b)). In the latter, the blur is generated by the automatic focusing of the camera on the foreign object in front. The inserted objects produces also partial occlusion of the scene. The parameters of camera motion were estimated by means of the proposed method (sets of 30 CPCs were used; in Fig. 1, detected CPCs are marked by crosses, the matched CPs are numbered).

The accuracy of the method was evaluated by the comparison of the computed values with the ground truth. Illustrative examples of the parameter comparison are summarized in the Table 1. Six situations are introduced here, images were blurred by the out-of-focus blur (the manual defocus and the insertion of two different foreign objects A and B), translated, and differently rotated (*small* and *big* rotation, approximately 6.2^o and 10.8^o), respectively. The change of scale in the images is negligible and the same in all cases. The translation parameters are given in pixels (1 pixel ≅ 0.4 cm). In the last two rows there are means and standard deviations of the errors.

In all cases the estimates corresponds well to the reference values and the standard deviations are small (the differences are mostly below the level of the discretization error). Even when parts of the scene are occluded by the foreign objects, the variations are negligible.

The experiments which were carried out proved the applicability of the proposed method for the estimation of the camera motion. In contrast to other existing methods even in situations when the scene is partially occluded and the images have only partial overlap due to camera rotation and translation the achieved estimation accuracy is satisfactory. The proposed method was able to handle situations when images are blurred without strong limitation on the type of the blur.

4. CONCLUSION

In this paper, a new method of estimation of camera motion parameters from blurred images is introduced. The method is based on registering the images by means of control points and their invariant descriptors. Combined invariants, invariant to translation, rotation, and to blurring caused by a convolution with an arbitrary centrosymmetric PSF are employed. Thus, the camera position can be estimated from defocused images directly without any deblurring.

Our approach has numerous advantages. Firstly, it is able to handle blurred frames. Since in our method the registration is established by means of local features, a partial overlap is sufficient. Thus, full overlap of the frames, as in the method designed particularly for blurred images [6], [8], using global image normalization, is not required. Comparing to iterative registration methods [4], proposed algorithm does not require any initial motion estimation and is faster. Moreover, our method works well for any type of blurring, it is not limited to certain type of blur (Gaussian, pillbox-like). The only assumption is the centrosymmetry of the PSF. The blur can even vary within the frame (but not within the neighborhood of one CP). Thanks to this, our method can handle images of 3-D scene where the local defocusation depends on the depth of the scene.

The major limitation of our method is that it cannot register images with significant unknown scaling differences. The invariants can be easily made invariant also to scaling but the CP neighborhoods (if taken of the same size in the both frames) comprise different areas of the scene. Thus, the invariant features of the actually corresponding CPs would be different. Theoretically, this can be avoided by using multiscale approach. However, using k resolution levels increases k-times the dimension of the feature space which makes the computation time-expensive. In practice, the scaling difference should be roughly estimated (preferably using prior knowledge about the scene or about the camera setting) before the registration algorithm is applied.

Acknowledgment

This work was supported by the grants 102/00/1711 and 102/01/P065 of the Grant Agency of the Czech Republic.

5. REFERENCES

[1] Q. Zheng and R. Chellapa, "A computational vision approach to image registration," *IEEE Transactions on Image Processing*, vol. 2(3), pp. 311–325, July 1993.

[2] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation and scale-invariant image registration," *IEEE Transaction on Image Processing*, vol. 5(8), pp. 1266–1271, August 1996.

[3] S. C. Pei and C. N. Lin, "Image normalization for pattern recognition," *Image and Vision Computing*, vol. 13(10), pp. 711–723, December 1995.

[4] Z. Myles and N. V. Lobo, "Recovering affine motion and defocus blur simultaneously," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20(6), pp. 652–658, June 1998.

[5] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," in *Proc. European Conf. Computer Vision ECCV'92*, pp. 237–252, 1992.

[6] Y. Zhang, C. Wen, and Y. Zhang, "Estimation of motion parameters from blurred images," *Pattern Recognition Letters*, vol. 21, pp. 425–433, 2000.

[7] J. Flusser and T. Suk, "Degraded image analysis: An invariant approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20(6), pp. 590–603, 1998.

[8] S. C. Pei and C. C. Wu, "Recover affine defocused images using blur invariants," *submitted to IEEE Transaction on Image Processing*, 1999.

[9] A. Kubota, K. Kodama, and K. Aizawa, "Registration and blur estimation methods for multiple differently focused images," in *Proc. 1999 Int. Conf. on Image Proc.*, 1999, vol. II, pp. 447–451.

[10] B. Zitová, J. Kautsky, G. Peters, and J. Flusser, "Robust detection of significant points in multiframe images," *Pattern Recognition Letters*, vol. 20(2), pp. 199–206, 1999.

[11] J. Flusser and B. Zitová, "Combined invariants to linear filtering and rotation," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 13(8), pp. 1123–1136, 1999.

[12] J. Flusser, "Object matching by means of matching likelihood coefficients," *Pattern Recognition Letters*, vol. 16, pp. 893–900, 1995.