Camera Motion Estimation from Defocused Images

Jan Flusser, *Member, IEEE,* **and Barbara Zitová**

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4, 182 08 Prague 8, Czech Republic
E-mail: {flusser,zitova}@utia.cas.cz

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 so-called combined invariants. These features are invariant both to rigid-body transform and to blurring caused by convolution with an arbitrary centrosymetric PSF. Thanks to this, the camera position can be estimated directly from defocused images without any deblurring.

Index Terms: Motion estimation, blurred image, registration, convolution and rotation invariants.

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, such as the kind of motion, the number of degrees of freedom of the camera, coordinates of the starting position, etc. can be available. Provided the images have at least partial overlap (the task is not solvable otherwise), an accurate estimation of the parameters of the spatial transformation between the images is the key to solving this problem.

There are basically two different 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 control points and to estimate the parameters of the transform model via interpolation or least-square fit. This approach was chosen in the well-known experiment by Zheng and Chellappa [1], for instance. They registered two images taken from a floating balloon in order to measure the wind velocity.

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], [3] and image normalization. The latter approach uses global image moments to transform each image into a canonical form [4], [5]. Between-image transform is then 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 (caused by 3-D nature of the scene, camera motion along optical axis or zooming), atmospheric turbulence (important particularly in remote sensing and astronomy), diffraction and camera vibration, among others. Although the motion blur could be also considered in our method, we do not address it explicitly because short-exposure imaging is assumed. We assume the blur, regardless of its particular type, can be modeled as a convolution of the original image with an unknown point-spread function (PSF) of the imaging system.

Only few papers dealing with motion estimation from blurred images have been published. Myles and Lobo [6], [7] proposed an iterative method for affine motion estimation. 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. Zhang et al. [8] proposed an affine normalization method that does not depend on the blurring function. To bring the images into canonical form, blur invariant moments originally introduced by Flusser and Suk [9] as a tool for object recognition 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 [10] used the same idea but they restricted themselves to a rigid-body transform only. Last two methods are robust to noise and computationally efficient but they require the complete image overlap to yield reliable results.

In this paper, we propose a new motion estimation method, which is based on the image registration by means of control points (CP's). Since at least one of the images is supposed to be blurred, a new registration technique, which is insensitive both to the type and amount of blurring and does not require any deblurring/restoration of the images, is presented. First, CP candidates are detected in both initial and current frames. Matching of the CP candidates is performed in the space of invariant signatures called *combined invariants*. They are calculated over a neighborhood of each CP candidate. As soon as the matched CP pairs are found, the parameters of the image-to-image transform are calculated and, finally, the new camera position is estimated. Since the registration is based on local properties of images, a partial overlap of the frames is sufficient.

The paper is organized as follows. Basic assumptions and necessary mathematical background are presented in Section 2. In Section 3, the theory of blur-rotation invariants is introduced briefly. The registration algorithm is described in Section 4. Section 5 demonstrates practical application of the proposed method to the camera position estimation in a real indoor scene.

2. BASIC ASSUMPTIONS AND MATHEMATICAL BACKGROUND

In this Section, we introduce basic terms and propositions that are used later in the paper.

Let us denote the image taken at the initial position of the camera as $f(x,y)$ and the image taken at the current camera position (i.e. at the position we want to estimate) as $g(x,y)$. Both $f(x,y)$ and $g(x,y)$ are assumed to have bounded supports and nonzero integrals. We assume ideal acquisition at the initial position and, furthermore, we assume the relationship between $f(x,y)$ and $g(x,y)$ can be described as

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

In this model, $h(x,y)$ is the PSF of the camera at the time of the acquisition of $g(x,y)$, $n(x,y)$ is additive random noise, a is a constant describing the global change of contrast (if the illumination of the scene does not change then $a = 1$), τ stands for a transform of spatial coordinates due to the changes in imaging geometry and $*$ denotes 2-D convolution.

In case of ideal imaging, $h(x,y)$ equals to Dirac delta-function. In practice, the PSF is a composition of Airy function that arises from diffraction and of an unknown blurring function introduced by defocusation, atmospheric turbulence and similar factors. In the following text the PSF is assumed to be centrosymmetric (i.e. $h(x,y) = h(-x,-y)$) and energy preserving, that means

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)dxdy = 1.$$

This is not a significant limitation from practical point of view, because these assumptions hold for most real sensors and blurring functions.

The type of the spatial transform τ depends on the number of degrees of freedom of the camera motion and on the character of the scene. If the scene is flat and the camera moves in parallel without any rotation, τ becomes pure translation. If the camera has six degrees of freedom that allows unlimited motion in a 3-D environment and the scene is assumed to be flat again, τ becomes a perspective projection. In a general case when the scene is not planar and the camera can move arbitrarily, τ might become very complex nonlinear transform. Methodology of the camera motion estimation is of course influenced by the actual character of τ. In this paper we solve the task for τ consisting of translation, rotation and small scale change.

The invariants introduced later in this paper are based on *complex moments*. Complex moment $c_{pq}^{(f)}$ of order $(p+q)$, where $p \geq 0$ and $q \geq 0$, of image $f(x,y)$ is defined as

$$c_{pq}^{(f)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+iy)^p (x-iy)^q f(x,y)dxdy, \tag{2}$$

where i denotes the imaginary unit. Complex moments have useful *rotation property*, which is employed later in the construction of invariants. Let f' be a rotated version (around the origin) of f, i.e. $f'(r,\theta) = f(r,\theta+\alpha)$ where α is the angle of rotation. Let us denote the complex moments of f' as c'_{pq}. Then

$$c'_{pq} = e^{-i(p-q)\alpha} \cdot c_{pq}. \tag{3}$$

To prove this assertion we only need to express (2) in polar coordinates

$$c_{pq}^{(f)} = \int_{0}^{\infty}\int_{0}^{2\pi} r^{p+q+1} e^{i(p-q)\theta} f(r,\theta)d\theta dr \tag{4}$$

and substitute f' for f.

3. INVARIANTS TO CONVOLUTION AND ROTATION

In this Section, we introduce a new class of features, called *combined invariants*, invariant simultaneously to convolution with an arbitrary centrosymmetric PSF and rotation. Theorem 1 introduces invariants to convolution that are further used in Theorem 2 to define combined invariants.

Theorem 1: Let $f(x,y)$ be an image function. Let us define the following function $K^{(f)} : \mathcal{Z} \times \mathcal{Z} \rightarrow \mathcal{C}$, where $\mathcal{Z}$ is the set of non-negative integers and $\mathcal{C}$ is the set of complex numbers. If $(p+q)$ is even then

$$K(p,q)^{(f)} = 0.$$

If $(p+q)$ is odd then

$$K(p,q)^{(f)} = c_{pq}^{(f)} - \frac{1}{c_{00}^{(f)}} \sum_{n=0}^{p} \sum_{\substack{m=0 \\ 0<n+m<p+q}}^{q} \binom{p}{n}\binom{q}{m} K(p-n,q-m)^{(f)} \cdot c_{nm}^{(f)}. \tag{5}$$

Then $K(p,q)^{(f*h)} = K(p,q)^{(f)}$ for any p and q and for any centrosymmetric $h(x,y)$.

For the proof of Theorem 1, we refer to our earlier paper [9], where a similar theorem is formulated and proven for standard moments. The proof in case of complex moments is analogous.

As can be easily proven by induction, the invariants $K(p,q)$ have the same rotation property as the complex moments themselves. Thus,

$$K'(p,q) = e^{-i(p-q)\alpha} \cdot K(p,q). \tag{6}$$

Theorem 2: Let $n \geq 1$ and let k_j, p_j and $q_j; j = 1, \cdots, n$, be non-negative integers such that $(p_j + q_j)$ is odd for each j and that

$$\sum_{j=1}^{n} k_j(p_j - q_j) = 0.$$

Then

$$I = \prod_{j=1}^{n} K(p_j, q_j)^{k_j} \tag{7}$$

is invariant to rotation around the origin and to convolution with a centrosymmetric PSF.

Proof: The invariance to convolution follows immediately from Theorem 1. Eq (6) implies the invariance to rotation:

$$I' = \prod_{j=1}^{n} K'(p_j, q_j)^{k_j} = \prod_{j=1}^{n} e^{-i(p_j-q_j)\alpha k_j} \cdot K(p_j, q_j)^{k_j} = e^{-i\alpha \sum_{j=1}^{n} k_j(p_j-q_j)} \cdot \prod_{j=1}^{n} K(p_j, q_j)^{k_j} = I$$

□

According to Theorem 2, simple examples of the combined invariants are $K(1,0)K(0,1)$, $K(1,0)K(1,2)$, $K(2,1)K(1,2)$, $K(2,1)^3K(0,3)$, etc. Most invariants (7) are complex. If one prefers to have real-valued features, one can consider their real and imaginary parts separately.

Theorem 2 allows us to construct an infinite number of the combined invariants, but some of them are algebraically dependent on the others. They must be discarded before practical usage because they do not contribute to discrimination power at all. A general theory how to select a minimal complete and independent basis of the combined invariants can be found in our recent paper [11], where its robustness and other numerical properties are investigated too.

For practical purposes it is often useful to make the combined invariants (7) invariant also to translation and scaling. Translation invariance is achieved easily by using central coordinates in the definition of complex moments. However, $K(1,0) = K(0,1) = 0$ in that case and, consequently, all invariants containing $K(1,0)$ and/or $K(0,1)$ are also zero for any image function $f(x,y)$. If normalized complex moments

$$\nu_{pq} = \frac{c_{pq}}{c_{00}^{(p+q+2)/2}}$$

are used for the construction of the convolution invariants, the scaling invariance can be obtained.

4. IMAGE REGISTRATION ALGORITHM

The proposed method of the camera motion estimation is based on image registration. Generally, image registration is 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. In our case, we do not need actually to overlay the images, we only look for the coordinate transform that describes the spatial relationship between them.

First, control point candidates (CPC's) are found both in the initial frame $f(x, y)$ as well as in the current frame $g(x, y)$, taken at the camera position we want to estimate. Significant corners and other corner-like dominant points are considered as the candidates. To detect them, a method developed particularly for blurred images [12] is employed.

The most difficult step is a matching of the CP candidates. The correspondence between the CPC's from the initial frame and those from the current frame must be established and the candidates having no counterparts should be rejected. The candidate matching by means of the combined invariants is realized by the following algorithm.

Algorithm MATCH:

1. *Invariant vector computation.* A vector of invariants is computed for each CPC over its circular neighborhood of a certain radius. The number of the invariants involved and the radius of the neighborhood are user-defined parameters.

2. *Matching of the candidates.* Two CPC pairs with the minimum distance of their invariant vectors are found as the most-likely corresponding CPC's. The other CPC's 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 CPC's from the current frame and CPC's in the initial frame is found via the thresholded nearest neighbor rule in the spatial domain. Knowing the correspondence, the set of matched control point pairs is established.

3. *Improvement of the CP localization in the current frame.* For each CP in the current frame, its improved position is found in its local neighborhood. For every pixel from the neighborhood its invariant vector is computed according to Step 1. The point having the minimum distance between its invariant vector and the invariant vector of the CP counterpart is found and set as the improved position of the CP.

Algorithm MATCH has several user-defined parameters, namely the number of the invariants involved, the radius of the neighborhood the invariants are calculated from, the distance threshold used in Step 2 and the radius of the neighborhood used for localization improvement. The choice of the parameters is influenced by the type of the scene and an extent of the camera motion and the blurring degradation, among others. However, there is no explicit relationship between parameter values and the type of image deformation, just heuristic conclusions can be made. The experiments we have performed indicate the following. From 3 to 9 invariants usually provide sufficient discriminative power. The more blurred the image is, the larger neighborhood for calculating the invariants should be. Common values range from 30 to 90 pixels. The distance threshold in Step 2 should be low (typically less than 7 pixels), false matches could be found otherwise. The neighborhoods for position improvement should be small (2 or 3 pixel radius) because of high computational cost of this step and because the corner detection algorithm should not produce higher deviations of the candidate positions.

To make Step 2 more robust, the first two matching pairs can be found by more sophisticated but computationally more expensive techniques than by minimum-distance rule. Using so-called *matching likelihood coefficients* [13] is particularly suitable for that purpose.

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.

Figure 1: Images of the indoor scene, sized 384×512 pixels, (a) - top image: camera at the initial position, (b) - bottom image: after the camera moved (*big* rotation, approximately 10.8^o and slight translation). The image taken by the moved camera has an out-of-focus blur, caused by the inserted foreign object. The inserted object produces the occlusion of a part of the scene too.

5. 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. Images were blurred by out-of-focus blur caused either 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.

30 CPCs were detected in all images and matched to the CPCs in the initial frame using the combined invariants ($K(2,1)K(1,2)$, $K(3,0)K(1,2)^3$, $K(3,2)K(1,2)$). In Fig. 2, CPCs

Figure 2: The images from Fig. 1 with 30 detected CPCs. Detected CPCs are marked by crosses. The CPCs which form the corresponding CP pairs are numbered.

detected in images from Fig. 1 are marked by crosses, the matched CPs are numbered. The CPC correspondence was found using the MATCH algorithm together with the matching likelihood coefficients. Finally, the coefficients of a rigid-body mapping function (s – uniform scale change, α – the angle of rotation and (t_x, t_y) – the vertical and horizontal translations)

$$x^{'} = s(\cos(\alpha)x - \sin(\alpha)y) + t_x$$

$$y^{'} = s(\sin(\alpha)x + \cos(\alpha)y) + t_y$$

were calculated via least-square approximation.

To visually evaluate the efficiency of the method, backward camera motion was simulated using the computed parameters. An image which would be acquired at the pseudo-initial position, estimated in this way, was generated. Theoretically, if the true motion parameters would be available, the difference of the generated image and the initial frame should be zero on smooth uniform areas and non-zero on edges. High frequency details would be emphasized because the generated image is a blurred version of the initial image and because image details are the most affected parts of an image by blur. An example of the absolute difference of the initial image and the backward-transformed one is shown in Fig. 3. The emphasize of the edges is apparent. The

Figure 3: The absolute difference of the initial image and the image generated from the Fig. 1(b) data using the computed estimates of the camera motion parameters. Theoretically, the difference image should be zero with the emphasized details (like edges, signs, etc.) because of the blur in the generated image.

bright thick line, corresponding to the the table edge, at the left bottom part of the difference image, growing in the toward direction, is a consequence of the 3D characteristic of the scene and the chosen model of the image-to-image transform. The present nonlinear geometric deformation between images due to the varying depth of the scene cannot be well modeled by the rigid-body transform in this case. However, if the distance between the camera and the object is comparatively larger than the size of the object (which holds for objects in the middle part of the scene), the rigid-body transform model can be used without considerable limitations.

Besides the visualization, the accuracy of the parameter estimation was evaluated by the comparison of the computed values with the reference parameter values. Sets of corresponding CP pairs, evenly distributed over the images, were manually detected and the reference parameters were computed by means of these sets. Illustrative examples of the parameter comparison are summarized in the Table 1. Six situations are introduced here, when images were blurred by the out-of-focus blur (the manual defocus and the insertion of two different foreign objects A and B) and differently rotated (*small* and *big* rotation, approximately $6,2^o$ and $10,8^o$), respectively. The translations were present in the camera motion too.

For each situation, the estimates of the rotation angle and of the translations in vertical and horizontal directions are given together with the reference values (in the parenthesis) and errors between the computed and the reference values. The change of scale in the images is negligible and the same in all cases. The translation parameters are given in pixels. One pixel corresponds approximately to 0.4 cm (using the known size of objects in the scene). In the last two rows there are means and standard deviations (STDs) of the errors. The row in the Tab. 1, emphasized by double line, refers to the experiment with the initial and current frames shown in Fig. 1. Its detected CPCs and matched CPs are marked in Fig. 2 and the difference image is in Fig. 3.

In all cases the estimates corresponds well to the reference values and the STDs are small. Moreover the differences between the computed and reference values are mostly below the level of the discretization error. Even in the situations, when parts of the scene are occluded by the foreign objects, the variations are negligible.

Image (blur, rotation)	Rotation angle (^{o})	Rotation difference	V-shift	V-shift difference	H-shift	H-shift difference
manual, *small*	6.38 (6.28)	0.10	-25.3 (-25.9)	0.6	-35.3 (-34.4)	-0.9
manual, *big*	10.65 (10.67)	-0.02	-40.3 (-40.4)	0.1	-56.0 (-56.0)	0.0
object A, *small*	6.20 (6.13)	0.07	-20.4 (-19.8)	-0.6	-14.8 (-15.0)	0.2
object A, *big*	11.00 (10.92)	0.08	-37.7 (-37.7)	0.0	-41.0 (-41.2)	0.2
object B, *small*	6.28 (6.17)	0.11	-19.7 (-19.7)	0.0	-15.7 (-14.8)	-0.9
object B, *big*	10.82 (10.79)	0.03	-38.1 (-38.3)	0.2	-41.1 (-40.8)	-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 and of the translations in vertical (V) and horizontal (H) directions and the difference between the computed values and the reference ones. The reference parameter values are in the parenthesis. Images were blurred (manual defocus, inserted objects - A, B), rotated (*big* angle, approximately $10,8^o$, *small* angle, approximately $6,2^o$) and slightly translated.

The experiments, which have been 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 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.

6. 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. The CP matching algorithm employs so-called combined invariants that are calculated from a neighborhood of each CP candidate. In addition to translation and rotation invariance, these signatures are invariant also to blurring caused by a convolution with an arbitrary centrosymmetric PSF. Thanks to this, the camera position can be estimated from defocused images directly without any deblurring. Experiments demonstrated an applicability of the proposed method.

In comparison with the other methods, our approach has numerous advantages. Severe limitation of the traditional correspondence-based methods is their inability to handle blurred frames. On the other hand, the correspondenceless method designed particularly for blurred images [8], [10] use global image normalization, i.e. they require 100% overlap of the frames. Since in our method the registration is established by means of local features, a partial overlap is sufficient. Comparing to the method of Myles [6], [7], the proposed algorithm is faster because it is not iterative. Moreover, the Myles method works for Gaussian and pillbox-like blurs only and the blur must be uniform over large regions of the image. Our method works well for any type of blurring, 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. In practice, the scaling difference must be roughly estimated (preferably using prior knowledge about the scene or about the camera setting) before the registration algorithm is applied.

Acknowledgment

This work has been supported by the grant No. 102/00/1711 of the Grant Agency of the Czech Republic.

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] E. de Castro and C. Morandi, "Registration of translated and rotated images using finite Fourier transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9(5), pp. 700–703, September 1987.

[3] 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.

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

[5] I. Rothe, H. Susse, and K. Voss, "The method of normalization to determine invariants," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, pp. 366–375, April 1996.

[6] 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.

[7] Z. Myles and N. V. Lobo, "Recovering affine motion and defocus blur simultaneously," *Proc. CVPR '96*, pp. 756–763, San Francisco, California, 1996.

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

[9] 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.

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

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