Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 41 (2008) 2263–2282
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

A homotopy-based approach for computing defocus blur and affine transform simultaneously

F. Deschênes[a,*], D. Ziou[b], P. Fuchs[c]

[a] Université du Québec en Outaouais, 283, boul. Alexandre-Taché, C.P. 1250 succursale Hull, Gatineau, Québec Canada J8X 3X7
[b] Département d'informatique, Université de Sherbrooke, 2500, boul. de l'Université, Québec Canada J1K 2R1
[c] Centre de robotique, École des Mines de Paris, 60, boul. Saint-Michel, 75272 Paris Cedex 06, France

Received 4 July 2007; received in revised form 8 December 2007; accepted 14 December 2007

Abstract

This paper presents a homotopy-based algorithm for a simultaneous recovery of defocus blur and the affine parameters of apparent shifts between planar patches of two pictures. These parameters are recovered from two images of the same scene acquired by a camera evolving in time and/or space and for which the intrinsic parameters are known. Using limited Taylor's expansion one of the images (and its partial derivatives) is expressed as a function of the partial derivatives of the two images, the blur difference, the affine parameters and a continuous parameter derived from homotopy methods. All of these unknowns can thus be directly computed by resolving a system of equations at a single scale. The proposed algorithm is tested using synthetic and real images. The results confirm that dense and accurate estimation of the previously mentioned parameters can be obtained.

Keywords: Computer vision; Unified model; Defocus blur; Affine matching; Homotopy method; Generalized moment expansion

1. Introduction

Image matching is a key issue of many applications in computer vision. Among those we may cite optical flow estimation, stereovision, depth from zooming, and feature tracking. Traditional approaches that address this problem in the spatial domain often assume that apparent shifts of brightness patterns correspond to constant translations within a local window [1–4]. Many researchers in the late 1980s and early 1990s suggest to overcome this basic limitation by replacing the translational model with the affine model [5–8]. Nevertheless, both types of methods still rely on simplistic assumptions. In fact, most of those techniques implicitly assume that radiometric transformations (due to blur, camera geometry, changes in illumination, etc.) are negligible. This common hypothesis, called *brightness constancy assumption*, implies a perfect control of both the environment and the acquisition system, which is difficult and even insufficient in many practical situations [9–11]. Such a constraint may hence lead to poor performance of the algorithms. Quite recent researches have thus been done to integrate radiometric and geometric transformations [11–16]. In this line of thoughts, we are interested in modeling simultaneously the affine deformations (rotation, scaling, translation, etc.) and the radiometric transformations due to defocus blur between two images.

Specifically, in this paper we propose a homotopy-based algorithm for a simultaneous and cooperative estimation of defocus blur and affine transformations of planar patches between a pair of images of a static scene acquired by a camera evolving in time and/or space. Assuming perspective projection, passive image formation system, Gaussian point spread function (PSF) and locally constant blur, we show that one of the images (respectively, its partial derivatives) may be locally expressed as a function of its partial derivatives, the partial derivatives of the other image, the blur difference, the affine parameters of the geometric transformation, and a continuous parameter derived from homotopy methods. Hence, all of these parameters

* Corresponding author. Tel.: +1 819 595 3980; fax: +1 819 595 3825.
E-mail addresses: Francois.Deschenes@usherbrooke.ca (F. Deschênes), Djemel.Ziou@usherbrooke.ca (D. Ziou), Philippe.Fuchs@ensmp.fr (P. Fuchs).

0031-3203/$30.00 © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2007.12.005

can be computed by resolving a system of equations based on limited Taylor's expansion. All computations required by our algorithm are local and are carried out in the spatial domain. Furthermore, no image model has been used to derive the algorithm. The resulting system of equations thus yields a dense and accurate estimation of the previously mentioned depth cues (defocus blur and affine transform). In the next section, we will first summarize related work. In Section 3, we will derive a family of systems of equations for a simultaneous estimation of defocus blur and affine deformations which is based on the use of homotopy methods. The resulting algorithm is described in Section 4. Section 5 presents the experimental results.

2. Related work

As mentioned the idea of locally approximating apparent motion by an affine model instead of a translational model was introduced in the late 1980s and early 1990s [5–7]. That is researchers propose to replace the well-known constraint

$$I(\vec{r}, t + \delta t) = I(\vec{r} + \vec{\delta} r, t), \tag{1}$$

where $\vec{r}=[x, y]^{\mathrm{T}}$, $\vec{\delta}r=[\delta_x, \delta_y]^{\mathrm{T}}$ represents constant translations within a local neighborhood, by

$$I(\vec{r}, t + \delta t) = I(\mathbf{A}\vec{r} + \vec{T}, t), \tag{2}$$

where

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \text{and} \quad \vec{T} = \begin{bmatrix} T_x \\ T_y \end{bmatrix} \tag{3}$$

are the affine parameters. Based on this more realistic constraint, several algorithms have then been developed to estimate apparent motion in spatial domain [7,8,17–19]. However, radiometric transformations between a pair of images (due to blur, illumination changes, etc.) are often neglected, which may lead to erroneous results in many practical situations [11]. In order to circumvent this problem, only few matching techniques have been proposed up to date in order to simultaneously consider affine transforms and illumination changes. The first work in that sense concerns simultaneous estimation of affine transforms and unidentified illumination changes. In extension to the work of Negahdaripour [12] who experimentally showed that the brightness constancy assumption leads to erroneous results in image regions with significant (ir)radiance variations, Jin et al. [16] proposed an algorithm for real-time feature tracking which simultaneously considers image patches undergoing affine deformations and radiomatric transformations. The resulting constraint is

$$I(\vec{r}, t + \delta t) = I(\mathbf{A}\vec{r} + \vec{T}, t) + \delta I(\vec{r}, t), \tag{4}$$

$\delta I(\vec{r}, t) = \lambda(\vec{r}, t) I(\vec{r}, t) + \delta c(\vec{r}, t)$ is the resulting radiometric transformation associated to a given pixel (x, y, t). δI is thus explicitly defined in terms of a multiplier ($\lambda(\vec{r}, t)$) and an offset ($\delta c(\vec{r}, t)$). Thus the brightness constancy assumption holds when $\lambda(\vec{r}, t) = \delta c(\vec{r}, t) = 0$. The authors linearize the resulting relation using first-order Taylor expansion then solve it using Newton–Raphson's minimization algorithm. Black et al. [14] propose to model image brightness changes using probabilistic mixtures associated to different causes. Specifically, they define four causes of appearance changes that can be used to construct or explain an image. Deformation parameters are thus estimated using a mixture of causes which corresponds to the probability of observing $I(\mathbf{A}\vec{r} + \vec{T}, t)$ from $I(\vec{r}, t - \delta t)$. In addition to the two previously mentioned causes, that is object/camera motion and illumination phenomena (Eq. (4)), they also take into account specular reflection and iconic changes (complex occlusions and changes in the material properties of the objects). A third approach based on explicit physical/optical modeling of radiometric and geometric transformations was introduced by Myles and Lobo [13]. They propose an iterative process for a simultaneous computation of defocus blur and affine motion parameters. To this end, they extended the algorithm of Manmatha [8] for the recovery of affine transforms and they include defocus blur computation. More especially, in the case of a thin lens, the image is formed by the convolution of the ideal projected image and the camera PSF $g_\beta(\vec{r})$, where β is the blur parameter at pixel (x, y). The more blurred image may thus be locally expressed as

$$I(\vec{r}, t + \delta t) = I(\mathbf{A}\vec{r} + \vec{T}, t) * g_\beta(\mathbf{A}\vec{r} + \vec{T}), \tag{5}$$

where $*$ is the convolution operator. From Eq. (5) they derived a system of equations based on first-order Taylor's series expansion with respect to β and the affine parameters. They, however, mention that the resulting system might not be stable if solved at a single scale. They thus form a system of equations using several scales to estimate the unknowns. The overall computation is embedded in a two-level pyramid scheme. The parameters are first estimated using the higher level (lower resolution), and are then propagated as initial estimates for the detailed level (higher resolution).

Kubota et al. [20] propose to first register the images without taking defocus blur into account. For this purpose they decrease resolution level in order to reduce the effect of blur. Blur radius can then be computed from registered images. Similarly, Zhang et al. [21,22], inspired the moment invariants introduced by Flusser and Suk [23], suggest to first normalize the images using blur invariant moments. Normalized images are used to compute translational or affine parameters. The blur differences are then recovered using both the original images and the estimated motion parameters. Deschênes et al. [24] introduces the use of homotopy to simultaneously estimate translations and defocus blur between pictures. They propose to solve a family of systems of equations which is based on limited Taylor's expansion. Mudenagudi and Chaudhuri [25] introduce a MAP-MRF approach to estimate depth using defocused stereo image pairs. They first compute defocus blur using two defocused images (i.e., the two left images or the two right images). Then, they determine the disparity using the blur estimates, the known camera parameters, and the relation between defocus blur and stereo disparity in this specific framework. Notice that four of the last five approaches compute defocus blur and geometric