Simultaneous Computation of Defocus Blur and Apparent Shifts in Spatial Domain

F. Deschênes[1,2] D. Ziou[1] P. Fuchs[2]
[1]Dépt. de mathématiques et d'informatique
Université de Sherbrooke, Canada, J1K 2R1
{deschene, ziou}@dmi.usherb.ca
[2]Centre de robotique
École des Mines de Paris, France
fuchs@paris.ensmp.fr

Abstract

This paper presents an algorithm for a cooperative and simultaneous estimation of depth cues: defocus blur and spatial shifts (stereo disparities, 2D motion, and/or zooming disparities). These cues are estimated 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. This algorithm is based on generalized moment expansion. We show that the more blurred image may be expressed as a function of the partial derivatives of the two images, the blur difference and the horizontal and vertical shifts. Hence, these depth cues can be computed by resolving a system of equations. The proposed algorithm is tested using synthetic and real images. The results are dense and accurate. They confirm that defocus blur and spatial shifts can be simultaneously computed at a single scale.

1 Introduction

In computer vision, the three-dimensional perception of a real scene allows us to understand the 3D relationship of objects in world space. 3D perception is generally related to the computation of depth information which involves the extraction of relevant image features (*depth cues*) such as shadows, motion, blur, disparity, etc. Most existing techniques compute depth cues independently and usually rely on simplistic assumptions. For example, one of the most common hypothesis for spatial shift estimation is the brightness constancy assumption [1, 2, 3]. Concerning defocus blur estimation, most existing approaches assume that spatial shifts between a pair of images of the same scene are negligible [4, 5, 6]. However, all of these implicit assumptions imply a perfect control of both the environment and the acquisition system, which is usually difficult and even insufficient in many practical cases [7, 8, 9].

Inspired by the 3D perception system of human being, many searchers suggest that the use of only one depth cue is insufficient and that it is essential to consider complementary sources of information [10, 11, 12]. They affirm that the limitations associated to the use of a single depth cue can be overcome by taking into account complementary information obtained from additional cues. In this line of thoughts, we are interested in simultaneous and cooperative estimation of blur differences (depth from defocus) and spatial shifts (stereo disparities, 2D motion, and/or zooming disparities). Let us consider $I(x(t_i,p_i),y(t_i,p_i),t_i,p_i,\gamma_i)$, $i=1,2$, a pair of images of a real scene obtained by a camera evolving in time (t) and space and for which the values of the extrinsic (p: position and orientation) and intrinsic parameters (γ: aperture, focal length, lens radius, etc.) are known. In order to simplify notation let us replace $x(t_i,p_i)$ with x_i and $y(t_i,p_i)$ with y_i in what follows. We propose to use general and flexible constraint given by:

$$I(x_2,y_2,t_2,p_2,\gamma_2) = I(x_1,y_1,t_1,p_1,\gamma_1) * g_\beta(x,y), \quad (1)$$

where $*$ is the convolution operator, g_β the PSF and β the blur parameter. Based on this relation, we derive an unified approach for the estimation of spatial shifts (stereo disparities, 2D motion, and/or zooming disparities) and defocus blur which considers their mutual interdependence. Using generalized moment expansion and assuming perspective projection, passive image formation system, PSF Gaussian and blur locally constant, we will show that the more blurred image may be expressed as a function of its partial derivatives, the partial derivatives of the other image, the blur difference (β) and the horizontal and vertical shifts δ_x and δ_y. Hence, β, δ_x, and δ_y can be computed by resolving a system of equations. All computations required by our algorithm are local and are carried out in the spatial domain. Our algorithm yields a dense estimation of the previously mentioned depth cues. A major interest of such a model is thus to simultaneously take different depth cues into account in order to potentially reduce estimation errors related to simplistic models.

In the next section, we will first summarize related work. In section 3, we will derive a system of equations for simultaneous estimation of defocus blur and spatial shifts. The resulting algorithm is described in section 4. Section 5 presents the experimental results.

2 Related Work

Approaches that compute depth from several cues can be classified in two distinct categories: 1) Independent estimation of depth cues followed by a merging process; 2) Simultaneous and cooperative estimation of depth cues. The first class includes approaches that merge information retrieved independently from stereopsis and apparent motion [13] or defocus blur and binocular disparity in active vision systems [10, 11]. The second category, which is quite recent, includes algorithms that consider the interactions between depth cues during the extraction process, as the one we are proposing. In this line of thoughts, Myles and Lobo [12] propose an iterative process for a simultaneous computation of defocus blur and affine motion parameters. From a relation similar to (1), they derived a system of equations based on first order Taylor's series expansions. They however mention that the resulting relation might however not be stable if solved at a single scale. They thus form a system of equations using several scales. According to the authors, the main disadvantage of this method is the need of relatively large planar patches within the images (usually more than 20×20 pixels). Moreover they assume that a rough estimate of the image translation is known and that both the scaling and the rotation are global. That is they depend on a center point which is known and unique for both images. Zhang *et al.* [14] suggest to first normalize the images using blur invariant moments. Normalized images are used to compute motion parameters. The blur differences are then recovered using both the original images and the estimated affine motion parameters.

Based on this overview, our work thus addresses a problem similar to the one addressed by Myles and Lobo. However, there are four major differences. First, we consider higher-order polynomial expansion. Second, the proposed scheme allows simultaneous computation of depth cues from the images themselves but also from the image derivatives. This leads to denser and more accurate estimates. Third, a complete system of equations is directly derived at a single scale. Finally, we do not consider the entire affine transforms. More especially, we focus on simultaneous computation of spatial shifts (stereo disparities, 2D motion, and/or zooming disparities) and defocus blur. This means that scaling and rotation between two images are estimated as apparent local shifts and thus the transformation centers do not have to be known *a priori*.

3 Proposed Framework

In this section we present the mathematical development behind the proposed approach for a simultaneous estimation of defocus blur and spatial shifts (stereo disparities, 2D motion and/or zooming disparities). This development corresponds to an application of the generalized moment expansion, that is a generic convolution formula for the images and their derivatives for the case of arbitrary order polynomials which takes spatial shifts into account [9]. In order to simplify notation we will replace $I(x_i, y_i, t_i, p_i, \gamma_i)$ by $I_i(x_i, y_i)$ in what follows. $I_2(x_2, y_2)$ and $I_1(x_1, y_1)$ can thus be considered as a pair of images of the same scene obtained from a passive image formation system by varying one or more intrinsic and/or extrinsic camera parameters. Let us assume that $I_2(x_2, y_2)$ is more blurred than $I_1(x_1, y_1)$. This assumption has no consequence for this work, since it is possible to determine which image is more blurred. Obviously, the image which has the smaller grey-level variance is the more blurred one. Under a perspective projection, when the PSF is Gaussian, $I_2(x_2, y_2) = (I * g_{\sigma_2})(x_2, y_2)$ and $I_1(x_1, y_1) = (I * g_{\sigma_1})(x_1, y_1)$, where $I(x, y)$ is the focused image, $*$ the convolution operator, and $\sigma_2 > \sigma_1$. The relation between $I_2^{(n)(m)}(x_2, y_2)$ and $I_1^{(n)(m)}(x_1, y_1)$, that is between the original images or their partial derivatives ($I_i^{(n)(m)}(x, y) = \frac{\partial^{n+m} I_i(x,y)}{\partial x^n \partial y^m}$, $n, m = 0, 1, 2, \cdots$), can thus be expressed by:

$$I_2^{(n)(m)}(x_2, y_2) = (I_1^{(n)(m)} * g_\beta)(x_1, y_1), \qquad (2)$$

where $x_1 = x_2 + \delta_x$ and $y_1 = y_2 + \delta_y$ are the relations between both coordinate systems and $\beta = \sqrt{\sigma_2^2 - \sigma_1^2}$ is the blur difference between the two images. It must be mentioned here that considering the derivatives of the images for depth cue estimation may lead to a denser and more accurate solution, as shown in [6] and [9].

Estimating defocus blur and spatial shifts consists in computing β, δ_x and δ_y. To this end, eq. (2) can be rewritten as:

$$I_2^{(n)(m)}(x_2 - \tfrac{\delta_x}{2}, y_2 - \tfrac{\delta_y}{2}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_1^{(n)(m)}(x_2 + \tfrac{\delta_x}{2} - x, y_2 + \tfrac{\delta_y}{2} - y) g_\beta(x, y) dx dy. \qquad (3)$$

Let us consider the Taylor expansions of both images at (x_2, y_2):

$$I_1(x_2 + \tfrac{\delta_x}{2} - x, y_2 + \tfrac{\delta_y}{2} - y) = \sum_{p=0}^{\infty}\sum_{q=0}^{\infty} \frac{(\frac{\delta_x}{2} - x)^p (\frac{\delta_y}{2} - y)^q I_1^{(p)(q)}(x_2, y_2)}{p!q!}, \qquad (4)$$

and

$$I_2(x_2 - \tfrac{\delta_x}{2}, y_2 - \tfrac{\delta_y}{2}) = \sum_{p=0}^{\infty}\sum_{q=0}^{\infty} \frac{(-\frac{\delta_x}{2})^p (-\frac{\delta_y}{2})^q I_2^{(p)(q)}(x_2, y_2)}{p!q!}. \qquad (5)$$

Substituting (4) and (5) in (3), the relation between the two images can be expressed as:

$$I_2^{(n)(m)}(x_2, y_2) = I_1^{(n)(m)}(x_2, y_2) + \sum_{p=0}^{\infty}\sum_{q=0, p+q\geq 1}^{\infty} \left(s_{p,q}(\tfrac{\delta_x}{2}, \tfrac{\delta_y}{2}, \beta) I_1^{(n+p)(m+q)}(x_2, y_2) - s_{p,q}(-\tfrac{\delta_x}{2}, -\tfrac{\delta_y}{2}, 0) I_2^{(n+p)(m+q)}(x_2, y_2) \right), \qquad (6)$$

where

$$s_{p,q}(\delta_x,\delta_y,\beta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{(\delta_x-x)^p(\delta_y-y)^q}{p!q!}g_\beta(x,y)dxdy. \tag{7}$$

$g_\beta(x,y)$ is separable (i.e., $g_\beta(x,y)=g_\beta(x)g_\beta(y)$), thus

$$s_{p,q}(\delta_x,\delta_y,\beta)=s_p(\delta_x,\beta)s_q(\delta_y,\beta), \tag{8}$$

where

$$s_p(\delta_x,\beta)=\int_{-\infty}^{\infty}\frac{(\delta_x-x)^p}{p!}g_\beta(x)dx. \tag{9}$$

In this relation, a coefficient $s_p(\delta_x,\beta)$ is proportional to the p^{th} moments of $g_\beta(x)$ centered on δ_x, and is a function of the blur β and the spatial shifts δ_x. In other words, given $I_1^{(n)(m)}(x,y)$, $I_2^{(n)(m)}(x,y)$ and $s_p(\delta_x,\beta)$ we are able to estimate β , δ_x and δ_y. In order to reduce computation cost of moments, we propose an efficient computation procedure for $s_p(\delta_x,\beta)$ [9]:

$$s_p(\delta_x,\beta)=\begin{cases}\frac{1}{p!}\sum_{i=0}^{r}C_{2i}^{p}\delta_x^{2i}\mu_{2r-2i}(\beta) & \text{If } p=2r,\\ \frac{1}{p!}\sum_{i=0}^{r}C_{2i+1}^{p}\delta_x^{2i+1}\mu_{2r-2i}(\beta) & \text{If } p=2r+1,\\ 1 & \text{If } p=0,\\ 0 & \text{Otherwise.}\end{cases} \tag{10}$$

where

$$\mu_p(\beta)=\begin{cases}(p-1)(p-3)\cdots 3\beta^p & \text{If } p>0 \text{ and } p \text{ is even},\\ 1 & \text{If } p=0,\\ 0 & \text{Otherwise.}\end{cases} \tag{11}$$

For example, when $p\leq 0$, the value of $s_p(\delta_x,\beta)$ is directly given and thus integration over the entire domain can be avoided. For the sake of clarity, let us examine eq. (6). In practice, it has been shown that the series on the right-hand side of this equation converges rapidly if the two images are first smoothed and if the size s of the smoothing filter is properly selected [9]. The result of the convolution operation can thus be accurately approximated using the first few terms of the series. Based on this assumption, let us consider a finite summation (i.e., high-order terms are neglected):

$$\begin{aligned}&I_2^{(n)(m)}(x_2,y_2)=I_1^{(n)(m)}(x_2,y_2)+\\&\sum_{p=0}^{P}\sum_{q=0,p+q\geq 1}^{Q}\Big(s_p(\tfrac{\delta_x}{2},\beta)s_q(\tfrac{\delta_y}{2},\beta)I_1^{(n+p)(m+q)}(x_2,y_2)\\&\quad -s_p(-\tfrac{\delta_x}{2},0)s_q(-\tfrac{\delta_y}{2},0)I_2^{(n+p)(m+q)}(x_2,y_2)\Big).\end{aligned} \tag{12}$$

In other words, let us fit a multi-variate polynomial of degree P and Q to image brightness within small neighborhoods. According to eq. (10), $s_{0,0}(\delta_x,\delta_y,\beta)=1$, $s_{1,0}(\delta_x,\delta_y,\beta)=\delta_x$, $s_{0,1}(\delta_x,\delta_y,\beta)=\delta_y$, $s_{1,1}(\delta_x,\delta_y,\beta)=\delta_x\delta_y$, $s_{2,0}(\delta_x,\delta_y,\beta)=\left(\frac{\beta^2}{2}+\frac{\delta_x^2}{2}\right)$, $s_{0,2}(\delta_x,\delta_y,\beta)=\left(\frac{\beta^2}{2}+\frac{\delta_y^2}{2}\right)$, $s_{2,1}(\delta_x,\delta_y,\beta)=\delta_y\left(\frac{\beta^2}{2}+\frac{\delta_x^2}{2}\right)$, $s_{1,2}(\delta_x,\delta_y,\beta)=\delta_x\left(\frac{\beta^2}{2}+\frac{\delta_y^2}{2}\right)$, $s_{2,2}(\delta_x,\delta_y,\beta)=\left(\frac{\beta^2}{2}+\frac{\delta_x^2}{2}\right)\left(\frac{\beta^2}{2}+\frac{\delta_y^2}{2}\right)$, etc. The first few terms ($P=Q=2$) of eq. (12) thus become:

$$\begin{aligned}&I_2^{(n)(m)}(x_2,y_2)=I_1^{(n)(m)}(x_2,y_2)\\&+\tfrac{\delta_x}{2}\left(I_1^{(n+1)(m)}(x_2,y_2)+I_2^{(n+1)(m)}(x_2,y_2)\right)\\&+\tfrac{\delta_y}{2}\left(I_1^{(n)(m+1)}(x_2,y_2)+I_2^{(n)(m+1)}(x_2,y_2)\right)\\&+\tfrac{\delta_x\delta_y}{4}\left(I_1^{(n+1)(m+1)}(x_2,y_2)-I_2^{(n+1)(m+1)}(x_2,y_2)\right)\\&+(\tfrac{\beta^2}{2}+\tfrac{\delta_x^2}{8})I_1^{(n+2)(m)}(x_2,y_2)-\tfrac{\delta_x^2}{8}I_2^{(n+2)(m)}(x_2,y_2)\\&+(\tfrac{\beta^2}{2}+\tfrac{\delta_y^2}{8})I_1^{(n)(m+2)}(x_2,y_2)-\tfrac{\delta_y^2}{8}I_2^{(n)(m+2)}(x_2,y_2)\\&+\tfrac{\delta_x\delta_y^2}{16}I_2^{(n+1)(m+2)}(x_2,y_2)+\tfrac{\delta_y\delta_x^2}{16}I_2^{(n+2)(m+1)}(x_2,y_2)\\&+\tfrac{\delta_x}{2}(\tfrac{\beta^2}{2}+\tfrac{\delta_y^2}{8})I_1^{(n+1)(m+2)}(x_2,y_2)\\&+\tfrac{\delta_y}{2}(\tfrac{\beta^2}{2}+\tfrac{\delta_x^2}{8})I_1^{(n+2)(m+1)}(x_2,y_2)\\&+(\tfrac{\beta^2}{2}+\tfrac{\delta_x^2}{8})(\tfrac{\beta^2}{2}+\tfrac{\delta_y^2}{8})I_1^{(n+2)(m+2)}(x_2,y_2)\\&-\tfrac{\delta_x^2\delta_y^2}{64}I_2^{(n+2)(m+2)}(x_2,y_2).\end{aligned} \tag{13}$$

The equation in (13) explicitly shows the relationship between $I_1^{(n)(m)}$, $I_2^{(n)(m)}$, β, δ_x and δ_y. Hence, it is clear that eq. (12) may provide a set of equations with three unknowns (β, δ_x and δ_y) that can be used to simultaneously compute the blur difference and the spatial shifts between two images at a single scale.

Recall that the equation implemented by Myles and Lobo is based on first-order Taylor's expansion (i.e. low-order polynomial fitting) and it implicitly assumes that $n=m=0$. Moreover since their relation might not be stable at a single scale, they form a system of equations using several scales. Our estimation model thus suggests the explicit use of higher-order polynomial fitting and the computation of depth cues from the images and their derivatives at a single scale.

4 Summary of the Algorithm

The resulting algorithm for a simultaneous computation of defocus blur and spatial shifts (stereo disparities, 2D motion, and/or zooming disparities) can be derived directly from the mathematical developments presented in the previous section. Assuming the convergence of the series in (12), that is applying a smoothing filter of size s to the two images, the defocus blur β and the spatial shifts δ_x and δ_y can be estimated by resolving a system of equations obtained from at least three values (n,m). However, in practice the imprecision of the numerical differentiation could make it difficult to fulfill equation (12). In order to circumvent this problem, we propose to estimate the blur differences and the spatial

shifts by minimizing:

$$E_{n,m}(x_2,y_2) = \sum_{u=-U}^{U}\sum_{v=-V}^{V}\Bigg[\sum_{p=0}^{P}\sum_{q=0}^{Q}\Big(s_{p,q}(\tfrac{\delta_x}{2},\tfrac{\delta_y}{2},\beta)I_1^{(n+p)(m+q)}(x_2+u,y_2+v) - s_{p,q}(-\tfrac{\delta_x}{2},-\tfrac{\delta_y}{2},0)I_2^{(n+p)(m+q)}(x_2+u,y_2+v)\Big)\Bigg]^2 \quad (14)$$

for different values of n and m, where U and V determine the size of a local neighborhood of (x_2,y_2). Minimizing $E_{n,m}$ implies the computation of roots of polynomials of degrees greater than 4. Since there is no direct computation rule for such cases, it is necessary to use a numerical scheme. To this end, many algorithms have been proposed for the minimization of nonlinear equation of the form $\sum_i(y(x_i;\mathbf{a}) - b_i)^2$, where $\mathbf{a} \in \mathbb{R}^n$ is the vector of parameters. Among those, we suggest to use the Levenberg-Marquardt algorithm. This method has become a standard of nonlinear least-squares routines. For a given point (x_2,y_2), the inputs of this algorithm are the default parameter values, which are arbitrary selected in this case (e.g., $\beta = 0, \delta_x = 0, \delta_y = 0$), and the partial derivatives $\frac{\partial I_2^{(n)(m)}}{\partial\beta}$, $\frac{\partial I_2^{(n)(m)}}{\partial\delta_x}$ and $\frac{\partial I_2^{(n)(m)}}{\partial\delta_y}$, which are computed using the right hand-side of equation (12). For every couples (n,m), we thus obtain for each pixel (x,y) zero, one or more triplets $(\beta,\delta_x,\delta_y)$ that correspond to the local minima of $E_{n,m}$ and that constitute an estimate of the real values of the defocus blur and the horizontal and vertical shifts. From all those possible solutions we choose the one which minimizes the following quadratic error:

$$E_f(x_2,y_2) = \prod_{n=0}^{N}\prod_{m=0}^{M}\sum_{u=-U}^{U}\sum_{v=-V}^{V}\Big((I_1^{(n)(m)} * g_\beta)(x_2+\delta_x+u, y_2+\delta_y+v) - I_2^{(n)(m)}(x_2+u,y_2+v)\Big)^2. \quad (15)$$

Other definitions of quadratic error can also be used.

To summarize, given the two images $I_1(x,y)$ and $I_2(x,y)$ and the parameter values s, N, M, P, Q, U and V, the algorithm is formed by three steps: 1) Smoothing of the images and computation of their partial derivatives. 2) The estimation of $(\beta,\delta_x,\delta_y)$ by minimizing the quadratic difference in (14) for different values of n and m. This can be done by using the Levenberg-Marquardt algorithm. 3) The selection of the best solution $(\beta,\delta_x,\delta_y)$ according to the quadratic error in (15).

It must be mentioned here that the resulting estimates, for which it has experimentally been shown that they are close to the desired solution (ground truth), may nevertheless contain aberrant values. Adding a regularization term to equation (14) could fix the problem but in practice it does not increase global accuracy and it tends to slow down the minimization process. For computer vision applications that need a higher level of accuracy, median filtering could be applied to the resulting estimates.

Finally, notice that the proposed algorithm is independent of the geometry of the acquisition system. It is thus not subject to the drawbacks originating from the use of the epipolar constraint [15].

5 Performance Evaluation

5.1 Experimental Results

We have tested our algorithm using synthetic and real grey-level images. For all of those experiments, the smoothing filter is Gaussian and its scale σ_s is set to 2 (i.e., the size of the filter is $s = k\sigma_s$, where $k \in [3\sqrt{2}, 5\sqrt{2}]$). Notice that the choice of this parameter is a compromise among several factors. Large values cause image details to be flattened out while small ones preserve sharp discontinuities that are not well approximated by relatively low-order polynomial fitting. The image derivatives are computed by convolving the images with the appropriate partial derivatives of the Gaussian function. The parameter values for equation (14) are $N = 1$, $M = 1$, $P = 2$, $Q = 2$, $U = 5$, $V = 5$ and the maximum number of iterations for the minimization process is set to 10. Aberrant values are removed by using median filtering (mask 5×5). In order to validate our algorithm, we generated three kinds of results from image pairs that are supposed to contain: 1) Defocus blur only; 2) Spatial shifts only (stereo disparities, 2D motion, and zooming disparities); and 3) Defocus blur and spatial shifts.

5.1.1 Defocus Blur Estimation

The performance of the unified approach has been evaluated on several real images. Those images are generally used to test existing depth from defocus algorithms under the assumption that there is no spatial displacements. Figures 1.a and 1.b are blurred images of a scene made of two planar objects slanted inward to meet at a vertically oriented straight line. They have 384×288 pixels. They were obtained from an optical camera. The focal length and the f-number of the camera used to grab those pictures are $F = 2.5$ cm and $f = 4$, respectively. Focusing ranges are 120 cm (Fig. 1.a) and 90 cm (Fig. 1.b). The intersection of the planar objects is 120 cm away from the camera while the nearest point is at a distance of 105 cm. The estimates of defocus blur and spatial shifts for points lying on textured regions are plotted in Fig. 1.c, 1.e and 1.f. From the plot of blur estimates, we observe for a given value of y a parabolic distribution. Such a distribution seems to be accurate since it fulfills the relation between the depth z and the amount of blur σ_b [6]. Figure 1.d presents the resulting depth estimates computed from blur values. As shown, the distance of the nearest point is about 105 cm and the planes meet

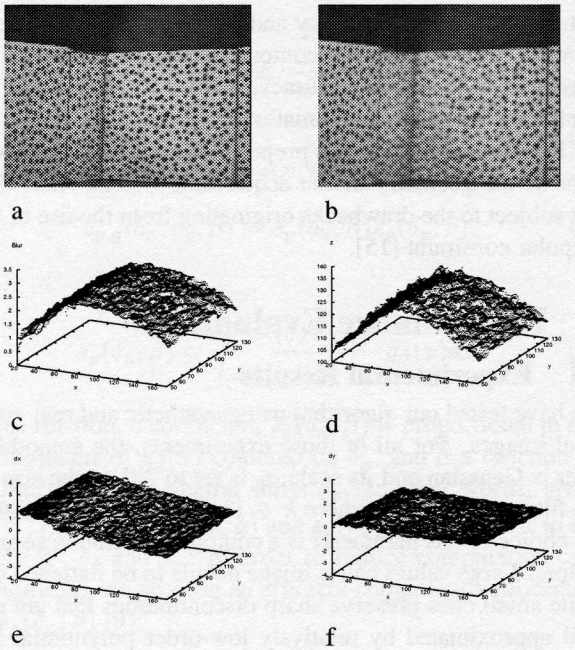

Figure 1: *a) and b) Real images of two planes slanted inward. Courtesy of S. Chaudhuri of the Department of Electrical Engineering at Indian Institute of Technology-Bombay. Estimated depth cues for textured regions: c) Defocus blur, d) and e) Horizontal (δ_x) and vertical (δ_y) shifts.*

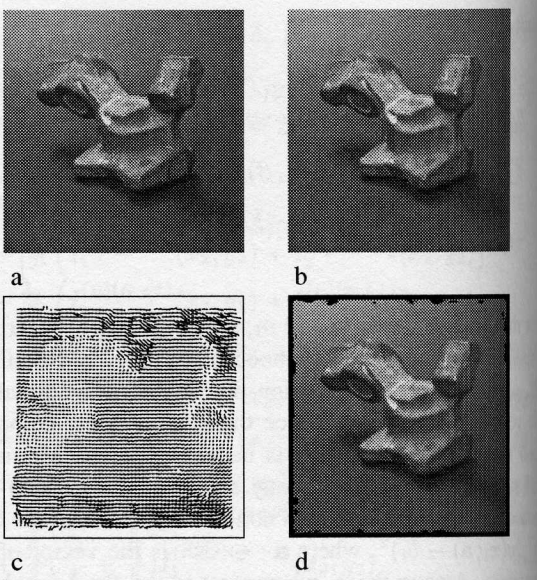

Figure 2: *a) and b) Stereo images of a car part. Courtesy of G. Medioni of the USC Institute for Robotics and Intelligent Systems at Carnegie Mellon University. c) Estimated disparities as a needle diagram. d) Right image reconstructed using estimated depth cues.*

at about 120 cm, that is the ground truth. It thus confirms that the estimated blur differences are accurate. Concerning spatial displacements (δ_x and δ_y), their estimates are effectively close to zero. However, they do not equal zero for every pixels as it is usually assumed by depth from defocus approaches (*cf.* Section 1). In fact, $\delta_x \in [-0.5, 1.5]$ and $\delta_y \in [-0.5, 0.25]$ and their variations are linear as a function of x and y. The extrinsic or intrinsic camera parameters may thus have been modified during the acquisition process. For example, the camera might have been slightly moved or its optical center might have been shifted when modifying focus (*cf.* [8]). Recall that it is difficult to ensure that $\delta_x = \delta_y = 0$ in practical computer vision. Hence, most existing depth from defocus techniques, which assume that spatial shifts between a pair of images of the same static scene are negligible, may lead to a lack of accuracy. Consequently, an algorithm for depth from defocus estimation which is tolerant to spatial shifts seems to be essential.

5.1.2 Spatial Shift Estimation

The performance of the proposed approach has also been evaluated on several pairs of images including spatial shifts only. In this section we will present specific results about the estimation of: 1) stereo disparity, 2) relative 2D motion, and then 3) spatial shifts resulting from zooming.

Figures 2.a and 2.b present a real pair of stereo images of a car part (256 × 256 pixels). Figure 2.c exhibits the spatial shifts (δ_x, δ_y) evaluated for all points as a field of vectors. We observe that the magnitude of the relative shift of pixels belonging to the car part is larger on the right side of the object, which is correct since it is closer to the camera. Moreover, the estimated displacements clearly exhibit the translational and rotational components. The accuracy of both the orientation and the magnitude of the spatial shifts are confirmed by Fig. 2.d. This figure corresponds to the reconstruction of the right image using the left image (Fig. 2.a), and the spatial shifts. The mean reconstruction error in grey-level is equal to 2.9, that is 1.14%, which is quite satisfying.

Figures 3.a to 3.c present a real sequence of images (256 × 240 pixels) of a laboratory scene captured by a camera that was moving to the right. The estimates of the horizontal and vertical shifts between frames $t = 91$ and $t = 93$ are displayed as a needle diagram in Fig. 3.d. As expected, the vertical shift is approximately equal to zero for most of the pixels. In order to facilitate visualization, the main component (horizontal shift) is shown as a grey-level image in Fig. 3.e. For each pixel, the grey level is proportional to the estimated shift between the two images. It is thus obvious that the estimated values are quite satisfying since they are proportional to the inverse of the distance between the camera and the objects. However, notice that computed shifts are less accurate within the neighborhood of occlusions. This type of error was anticipated since some pixels of the first image cannot be matched with pixels of the second image (occlusion problem). Fig. 3.f exhibits a dense and accurate reconstruction of frame $t = 93$ using frame $t = 91$ and the

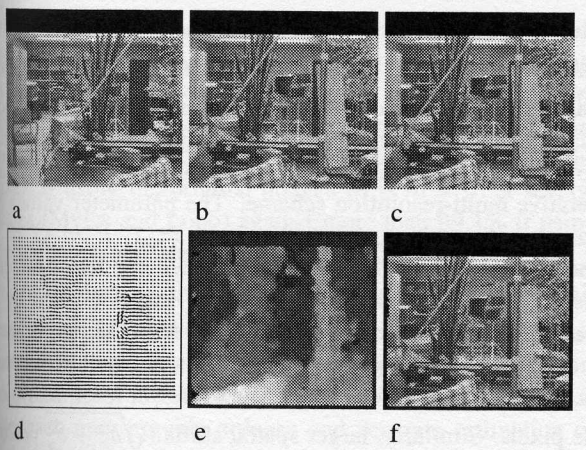

Figure 3: *a), b) and c) Real sequence of images acquired by moving the camera rightwards ($t = 25$, 70 and 93, respectively). d) Estimated 2D motion using frames $t_1 = 91$ and $t_2 = 93$. e) Horizontal shift as a grey-level image. f) Reconstruction of frame $t = 93$ using frame $t = 91$ and the estimated depth cues.*

estimated depth cues (δ_x and δ_y). The mean reconstruction error in grey-level is 4.6 (1.8%), which is once more quite low.

Figures 4.a and 4.b are two images of a real scene acquired by zooming out. The original images have 576×384 pixels. The estimated horizontal and vertical shifts are displayed as a vector field in Fig. 4.c (textured regions only). This figure clearly shows an accurate zooming effect. Every vectors that do not belong to the borders are oriented towards an unique central point. Moreover, the vector magnitude gradually increases as a function of the distance between the current position and the central point. Figure 4.d shows the reconstruction of the image in Fig. 4.b using the other image and the computed values δ_x and δ_y. The mean reconstruction error of this image in grey-level is 2.3, that is 0.92%. It thus confirms that we get dense and accurate estimates of the spatial shifts introduced by zooming.

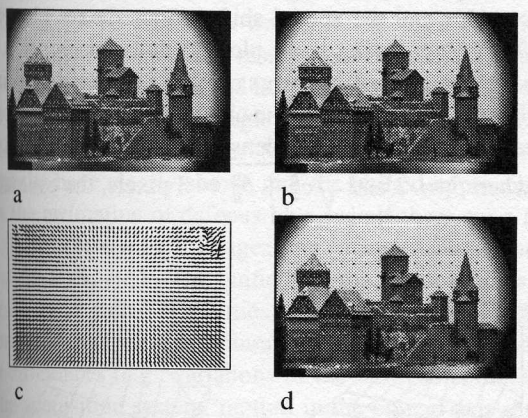

Figure 4: *a) and b) Pair of real images acquired using different zoom values. Courtesy of the Calibrated Imaging Laboratory at Carnegie Mellon University. c) Spatial shifts related to zooming effect as a vector field (textured regions). d) Image in b) reconstructed using retrieved depth cues.*

All of these results clearly show that the proposed algorithm is an effective and accurate approach for computing spatial shifts. This approach provides dense and accurate estimates of the horizontal and vertical shifts originating from binocular disparities, 2D motion, and/or zooming effect.

5.1.3 Simultaneous Estimation

The performance of the proposed approach has finally been evaluated on several grey-level images including both blur differences and spatial shifts (horizontal and vertical). The pair of images in Figs. 5.a and 5.b is composed of the original pictures in Fig. 1. In this case, we moved the image in Fig. 1.b two pixels rightward and two pixels downward. The estimated values of the blur differences for pixels lying on textured regions are shown in Fig. 5.c. By comparing this plot to the one in Fig. 1.c, we realize that the estimated values are quite similar, which is totally accurate. Moreover, by comparing the recovered horizontal and vertical displacements (Figs. 5.d and 5.e) to those in Figs. 1.d and 1.e, it is clear that both distributions are globally moved down by two pixels. That is the additional shifts we introduced. More especially, $\delta_x \in [-2.5, -0.5]$ and $\delta_y \in [-2.5, -1.75]$. It

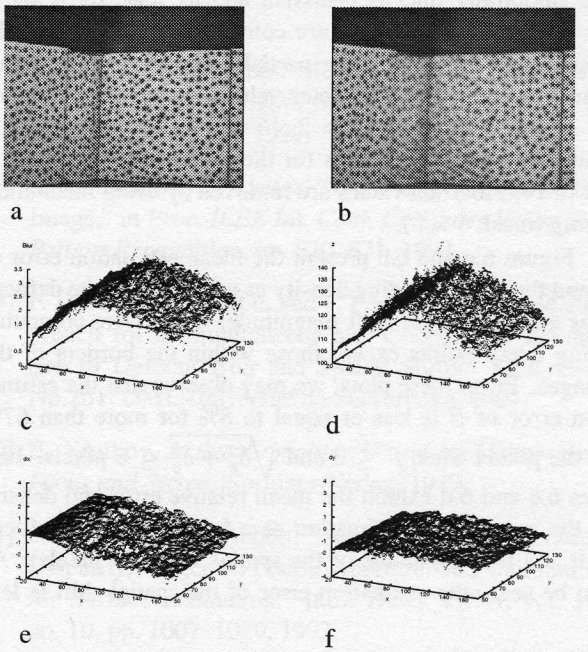

Figure 5: *a) and b) Two vertical planes slanted inwards where the more blurred image is spatially moved ($\delta_x = -2, \delta_y = -2$). c) and d) Defocus blur and corresponding depth estimates for points lying on the planes (textured regions). e) and f) Recovered horizontal and vertical displacements.*

thus confirms that the unified approach provides an accurate estimation of defocus blur even if there are spatial shifts between both images.

All of these results lead to the conclusion that the proposed approach allows a simultaneous and cooperative estimation of the blur differences (β) and the spatial shifts (δ_x, δ_y) between two images at a single scale.

5.2 Mean Estimation Errors and Densities

Let us now examine the influence of the defocus blur and the spatial shifts on their respective estimation error. To this end, these depth cues are estimated using a set of 840 pairs of images that was created from 12 real images that have been successively blurred ($\beta = 1, 2, \cdots, 6$) and shifted ($\sqrt{\delta_x^2 + \delta_y^2} = 1, 2, \cdots, 10$). For all of those experiments, the smoothing filter is Gaussian and its scale σ_s is set to 6. The image derivatives are computed by convolving the image with the appropriate partial derivatives of the Gaussian function. The parameter values for equation (14) are $N = M = 1$, $P = Q = 2$, $U = V = 5$, and the maximum number of iterations for the minimization process is set to 10. Aberrant values are removed by using median filtering (mask 5×5).

Figure 6.a and 6.b present the mean estimation error of β and the corresponding density as a function of the defocus blur and the spatial shift magnitude. They were computed using every points except those within the borders of the images. From these plots, we may observe that the estimation error of β is less or equal to 8% for more than 67% of the pixels when $\beta \leq 3$ and $\sqrt{\delta_x^2 + \delta_y^2} \leq 5$ pixels. Figures 6.c and 6.d exhibit the mean relative error and density of the spatial shift estimation as a function of the defocus blur and the magnitude of the spatial shifts (in pixels). As can be seen, the estimation error of the spatial shift is less than 0.5 pixel for more than 69% of the pixels when $\beta \leq 2$ and $\sqrt{\delta_x^2 + \delta_y^2} \leq 5$ pixels. Notice that for estimating spatial shifts greater or equal to 6 pixels many techniques such as multi-resolution can be used in conjunction with our algorithm. As an example, we tested our algorithm with an iterative multi-resolution scheme. The parameter values of our algorithm are $\sigma_s = 2$, $N = M = 1$, $P = Q = 2$ and $U = V = 5$. For the first test, we use two different resolutions (the entire image and the image reduced by a factor of 2). In this case, spatial shifts up to 12 pixels have been estimated with a relative error less or equal to 4% and an absolute error inferior to 0.5 pixel for at least 92.4% of the pixels. Similarly, larger spatial shifts ($\sqrt{\delta_x^2 + \delta_y^2} = 20$) were computed using a four stage resolution pyramid. The resulting relative error and density are equal to 1.1% and 96.1%, respectively. It thus confirms that using our algorithm in conjunction with a multi-resolution scheme makes it possible to estimate large spatial shifts.

Since we deal with a particular case of the entire affine transforms, we also compute the mean estimation errors resulting from the use of the equation proposed by Myles and Lobo [12] for comparison purpose. In this case, as mentioned earlier, rotation and scaling are neglected (they are respectively set to zero and one) and the initial translations are initialized to zero. As suggested by Myles and Lobo, we use five different scale values: 1.75, 2.5, 3.0, 3.5 and 4.5. The corresponding errors and densities as a function of the defocus blur and the spatial shift magnitude are plotted in Fig. 7. By comparing these plots to those in Fig. 6, it appears that the estimation errors resulting from our algorithm are globally inferior to the one obtained with the equation of Myles and Lobo, while the densities are greater. For example, when $\beta = 2$ and $\sqrt{\delta_x^2 + \delta_y^2} = 3$ pixels, the estimation

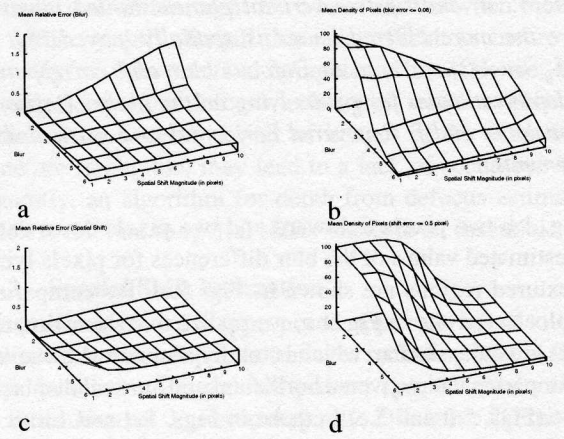

Figure 6: *a) to d) Mean estimation errors and densities as a function of the defocus blur and the spatial shift magnitude.*

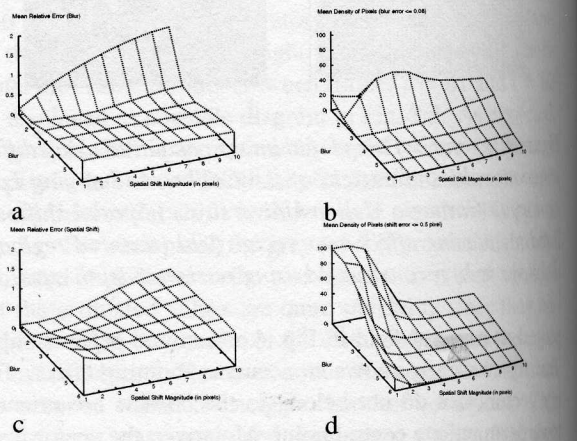

Figure 7: *a) to d) Mean estimation errors and densities as a function of the defocus blur and the spatial shift magnitude obtained from the equation of Myles and Lobo.*

error of β is less or equal to 10% for 51.6% of the pixels if using their system of equations, while it is less or equal to 5.1% for 92.1% of the pixels with our approach. Under the same conditions, the estimation error of the spatial shift is less or equal to 0.5 pixel for more than 54% of the pixels if using the original images ($n = m = 0$) at several scales (Myles and Lobo) against more than 99.1% if using the proposed approach. Moreover notice that when $\beta > 2$ or $\sqrt{\delta_x^2 + \delta_y^2} > 3$, the accuracy and the density of the estimates provided by the equation of Myles and Lobo decrease faster than those originating from our approach.

From all of these results, it is obvious that the proposed approach may provide a dense and accurate estimation of defocus blur and spatial shifts at a single scale.

6 Conclusions

We proposed an unified approach for a simultaneous estimation of defocus blur and spatial shifts (stereo disparities, 2D motion, and/or zooming disparities) from a pair of images of the same scene. Assuming perspective projection, passive image formation system, PSF Gaussian, and locally constant blur, we have shown that the more blurred image, respectively its derivatives, can be expressed as a function of its partial derivatives, the partial derivatives of the other image, the blur differences and the horizontal and vertical shifts. Hence these depth cues can be computed by resolving a system of equations at a single scale.

The proposed algorithm was tested using several synthetic and real images. The results confirmed that it simultaneously provides dense and accurate estimates of defocus blur and spatial shift at a single scale from the original images and/or their derivatives. More especially, experimental results have clearly shown that the unified approach allows the estimation of defocus blur even if there are a spatial shifts between both images (e.g., magnification due to focus changes, etc.). The unified approach also allows the estimation of stereo disparities, 2D motion, and zooming disparities when there are illumination changes induced by blur differences (e.g., variations of the intrinsic parameters of the acquisition system, motion in the optical axis direction, etc.). It must also be mentioned that all computations are done without considering the particular geometry of the acquisition system, that is without using the epipolar constraint. Consequently, another main advantage of this algorithm is that it is not subject to the drawbacks originating from the reduction of the 2D matching problem to a 1D problem [9].

References

[1] B. Horn and B. Schunck, "Determining Optical Flow," *Artificial Intelligence*, vol. 17, pp. 185–204, 1981.

[2] B. Super and W. Klarquist, "Patch-based Stereo in a General Binocular Viewing Geometry," IEEE *Trans. PAMI*, vol. 18, pp. 247–253, March 1997.

[3] J. Lavest, C. Delherm, B. Peuchot, and N. Daucher, "Implicit Reconstruction by Zooming," *CVIU*, vol. 66, no. 3, pp. 301–315, 1997.

[4] M. Subbarao and G. Surya, "Depth from Defocus: A Spatial Domain Approach," *Int. Jour. of Computer Vision*, vol. 13, no. 3, pp. 271–294, 1994.

[5] A. Rajagopalan and S. Chaudhuri, "Space-Variant Approaches to Recovery of Depth from Defocused Images," *CVIU*, vol. 68, pp. 309–329, 1997.

[6] D. Ziou and F. Deschenes, "Depth from Defocus Estimation in Spatial Domain," *CVIU*, vol. 81, no. 2, pp. 143–165, 2001.

[7] S. Olsen, "Image Point Motion When Zooming and Focusing," in *Proc. 10th Scandinavian Conf. Image Analysis*, pp. 65–70, 1997.

[8] R. Willson and S. Shafer, "What is the Center of the Image," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp. 670–671, 1993.

[9] F. Deschênes, D. Ziou, and P. Fuchs, "An Unified Approach for a Simultaneous and Cooperative Estimation of Defocus Blur and Spatial Shifts." Tech. Rep. No.261, DMI, Université de Sherbrooke, 2001.

[10] E. Krotkov, *Active Computer Vision by Cooperative Focus and Stereo*. Springer-Verlag, 1989.

[11] N. Ahuja and A. Abbott, "Active Stereo: Integrating Disparity, Vergence, Focus, Aperture, and Calibration for Surface Estimation," IEEE *Trans. PAMI*, vol. 15, no. 10, pp. 1007–1029, 1993.

[12] Z. Myles and N. V. Lobo, "Recovering Affine Motion and Defocus Blur Simultaneously," IEEE *Trans. PAMI*, vol. 20, no. 6, pp. 652–658, 1998.

[13] G. Sudhir, S. Banerjee, K. K. Biswas, and R. Bahl, "Cooperative Integration of Stereopsis and Optic Flow Computation," *J. Opt. Soc. Am. A*, vol. 12, no. 12, pp. 2564–2572, 1995.

[14] Y. Zhang, C. Wen, and Y. Zhang, "Simultaneously Recovering Affine Motion and Defocus Blur Using Moments," in *Proc. 15th Int. Conf. Pattern Recognition*, pp. 881–884, 2000.

[15] S. Roy, "Stereo Without Epipolar Lines: A Maximum-Flow Formulation," *Int. J. Computer Vision*, vol. 34, no. 2/3, pp. 147–161, 1998.