

# A Contrast Invariant Approach to Motion Estimation

V. Caselles, L. Garrido, and L. Igual\*

Universitat Pompeu Fabra, Barcelona (Spain)  
{vicent.caselles, luis.garrido, laura.igual}@upf.edu

**Abstract.** Motion estimation is one of the key tools in many video processing applications. Most of the existing motion estimation approaches use the brightness constancy assumption in order to model the movements of the objects present in the scene. In this paper the motion of objects is modeled from a geometrical-based point of view, leading thus to a contrast invariant formulation. The present approach is region-based and assumes affine motion model for each region.

## 1 Introduction

Computing the apparent motion of objects in a sequence of images is one of the key problems in computer vision known as the optical flow computation. Its numerous applications make it the object of current research (see [23] for an account of it).

Most known motion estimation methods, in one form or another, employ the *optical flow constraint* which states that the image intensity remains unchanged from frame to frame along the true motion path. The *optical flow equation* is derived from the optical flow constraint:

$$\partial_x I u + \partial_y I v + \partial_t I = 0 \tag{1}$$

where  $I(t, x, y)$  denotes the image sequence and  $(u, v)$  the motion vector field. The movement of the objects present in the scene may be recovered by minimizing an error measure based on the optical flow equation [23]. Furthermore, it is known that motion estimation is an "ill-posed" problem, indeed, the solution may not be unique, and/or solutions may not depend continuously on the data [4]. Current motion estimation approaches try to solve the latter issue by imposing additional assumptions about the structure of the 2D motion field. The latter constraints are introduced into the error measure either by adding a smoothness term to it, or by restricting it to a particular motion model. The

---

\* The first and second authors acknowledge partial support by the Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya and by PNPGC project, reference BFM2003-02125. L. Igual acknowledges support by the French Space Agency (CNES) and the company THALES (France).

former strategies are called *dense motion field estimation* approaches, whereas the latter ones are usually called *parametric motion estimation* approaches.

The Horn-Shunck's method is a classical method for dense motion field estimation. It seeks for a motion field that satisfies the optical flow equation (1) with a minimum pixel-to-pixel variation between the flow vectors:

$$\min \int_{\Omega} (\partial_x I u + \partial_y I v + \partial_t I)^2 + \alpha^2 ((\partial_x u)^2 + (\partial_y u)^2 + (\partial_x v)^2 + (\partial_y v)^2)$$

where  $\Omega$  is the image domain and  $\alpha$  may be used to control the influence of the constraint. Larger values of  $\alpha^2$  increase the influence of the constraint.

The Lucas-Kanade method can be considered a *parametric motion estimation*, since it estimates the motion by assuming that the motion vector associated to the optical flow equation remains unchanged over a particular block of pixels. The method thus allows to estimate a translational motion vector for that block. A very interesting combination of both previous methods with an efficient implementation has been proposed in [7]. For the interested reader, a good review of current motion estimation techniques can be found in [3, 22].

The optical flow constraint assumption is generally violated in image sequences taken from the real world. Global or local changes in illumination due to, for instance, a moving camera or a change in the shade of an object may prevent the correct motion to be estimated. Alternatives to the classical brightness constancy assumption have been already proposed in the literature. A common approach to handle non constant intensity is through explicit modelling of the illumination change in the optical flow equation [18]. The approach requires complex minimization since, in addition to the motion field, illumination fields must also be estimated.

In [4] a constraint based on spatial gradient's constancy is proposed. It relaxes the classical assumption, but requires that the amount of dilation and rotation in the image be negligible, a limitation often satisfied in practice according to [22]. The technique has been demonstrated to be very robust in the presence of time-varying illumination. More recently, it has been shown that the direction of the intensity gradient is invariant to global illumination changes [10]. The work presented in [8] is based on this property.

In this paper we propose to substitute the optical flow equation, derived from the brightness constancy assumption, by the assumption that the shapes of the image move along the sequence. We identify the shapes of the image with the family of its level lines [9] and we assume that they move along the image sequence (with possible deformation). This assumption permits us to design contrast invariant estimate of the optical flow. The approach is in fact based on the invariance of the gradient direction to contrast changes. However, no restriction to the amount of dilation and rotation is imposed.

The paper is organized as follows. Section 2 describes the contrast-invariant model that has been developed, whereas Sect. 3 introduces the region-based strategy that has been implemented. Sect. 4 gives some details about the implementation. Finally, Sect. 5 presents the results obtained with the proposed method and Sect. 6 ends up with the conclusions and future research work.

## 2 A Contrast-Invariant Functional

Let  $\Omega$  be the image domain, which we may assume to be normalized to  $[0, 1]^2$ . Let  $I : \Omega \rightarrow R$  be a given image. Mathematical morphology offers an image description in terms of its level sets, be upper  $X^\lambda I = \{\mathbf{p} \in \Omega : I(\mathbf{p}) \geq \lambda\}$ , or lower  $X_\lambda I = \{\mathbf{p} \in \Omega : I(\mathbf{p}) \leq \lambda\}$ . Level sets provide a complete image description, in particular, the image  $I$  can be reconstructed from its (upper) level sets by the formula  $I(\mathbf{p}) = \sup\{\lambda : \mathbf{p} \in X^\lambda I\}$  (a similar formula exists for the lower level sets) where  $\sup$  denotes the supremum operator, and  $\mathbf{p} = (x, y)^T$  denotes a point in  $\Omega$ . Level sets give a contrast invariant representation of the image [21].

We call level lines the boundaries of the level sets. In the discrete framework, any level set can be described in terms of its boundary. Indeed, the connected component of each level set can be described in terms of its external and the family of its internal boundaries [9, 21]. Thus, we may use the family of level lines as basic contrast invariant geometric description of the image  $I$ . As an analytical tool, we shall use the unit normals to the level lines to describe them.

Let  $I(t, \mathbf{p})$  be a given image sequence,  $t$  being in the time interval  $[T_0, T_1]$  and  $\mathbf{p} \in \Omega$ . We assume that the image sequence has been sampled at points multiple of  $\Delta t$ , the sampling points being  $t_j = T_0 + j\Delta t$ ,  $j = 0, \dots, N$  ( $t_N = T_1$ ). Let us denote by  $\phi^j(\mathbf{p})$  the coordinates at time  $t_j + \Delta t$  of the point whose coordinates at time  $t_j$  are  $\mathbf{p}$ ,  $j = 0, \dots, N - 1$ . The map  $\phi^j : \Omega \rightarrow \Omega$  is nothing else than the motion path starting from time  $t_j$  and we may think about it as a deformation. We do not assume in this section any particular motion model for  $\phi^j$ . That is, the image objects may suffer any deformation over time. For simplicity, when no confusion arises, the arguments of the previous function will be dropped out.

Assume for a while that  $j$  is fixed and let  $\phi = (\phi_1, \phi_2)$  be any of the maps  $\phi^j$ , where  $\phi_1$  and  $\phi_2$  are the components of  $\phi^j$ . Let  $X = (x(s), y(s))^T$  be the arclength parameterization of a given level line  $\mathcal{C}$  of the image  $I(t_j, \mathbf{p})$ ,  $s$  being the arc length parameter. The curve  $\mathcal{C}$  may be described by its normal vectors  $Z = (-y'(s), x'(s))^T$ , where  $(\cdot)'$  denotes the first derivative with respect to  $s$ . Note that  $Z$  has unit norm.

Let us describe the normal vectors to the curve  $\phi(\mathcal{C})$  in terms of  $\phi$  and the normal vectors to  $\mathcal{C}$ . Since the curve  $\phi(\mathcal{C})$  is described by

$$\bar{X} = (\bar{x}(s), \bar{y}(s))^T = \phi(x(s), y(s)),$$

the tangent vector to the deformed curve  $\phi(\mathcal{C})$  is given by

$$\bar{X}' = \begin{pmatrix} \bar{x}'(s) \\ \bar{y}'(s) \end{pmatrix} = \begin{pmatrix} \partial_x \phi_1 & \partial_y \phi_1 \\ \partial_x \phi_2 & \partial_y \phi_2 \end{pmatrix} \begin{pmatrix} x'(s) \\ y'(s) \end{pmatrix} = D\phi X'$$

where  $\partial_x$  and  $\partial_y$  denote the partial derivative with respect to  $x$  and  $y$  respectively. Thus, the normal vector  $Z$  of the deformed curve is

$$Z = \begin{pmatrix} -\bar{y}'(s) \\ \bar{x}'(s) \end{pmatrix} = \begin{pmatrix} \partial_y \phi_2 & -\partial_x \phi_2 \\ -\partial_y \phi_1 & \partial_x \phi_1 \end{pmatrix} \begin{pmatrix} -y'(s) \\ x'(s) \end{pmatrix}. \tag{2}$$

Observe that the matrix in the right hand side of (2) is the *cofactor matrix* associated to  $D\phi$  which we shall denote by  $(D\phi)^\dagger$ . Thus, the normal vectors of the deformed curve are related to the original normal vectors by means of the cofactor matrix. Observe that  $s$  is not necessarily the arclength parameter of  $\phi(\mathcal{C})$ , hence  $Z$  is not, in general, a unit vector. We normalize it to be of unit norm by redefining

$$\bar{Z}_\phi = \frac{(D\phi)^\dagger Z}{\|(D\phi)^\dagger Z\|} \quad \text{if } (D\phi)^\dagger Z \neq 0; \quad 0 \text{ otherwise,} \quad (3)$$

where  $\|\cdot\|$  denotes the modulus of a vector in  $R^2$ .

In the context of this work,  $Z^t(\mathbf{p})$  will be the vector field of unit normals to the level lines of  $I(t, \mathbf{p})$ . Usually, the energy functional whose minimum gives the optical flow tries to impose the brightness constancy equation (1). Instead, our main assumption will be that shapes move with possible deformation along the sequence. We interpret it in the following way: (\*) we may find the boundary of a connected component of the level set  $[I(t_j, \cdot) \geq \lambda]$ ,  $\lambda \in R$ , eventually deformed by  $\phi^j(\cdot)$ , as a level curve of  $I(t_{j+1}, \cdot)$  at some other level  $\lambda'$ . Observe that if two consecutive frames are related by the motion model and a global illumination change, i.e., if (\*\*)  $I(t_{j+1}, \phi^j(\mathbf{p})) = g_j(I(t_j, \mathbf{p}))$  for some contrast change  $g_j$ , then (\*) holds (in this case  $\lambda' = g_j(\lambda)$ ). Our assumption (\*) is more general than (\*\*) since the former is local: the level at which we may find the boundary of the connected component of  $[I(t_j, \cdot) \geq \lambda]$  may depend on the connected component itself, besides of depending on  $\lambda$ . Thus, our purpose will be to align the level lines of two consecutive frames at times  $t_j$  and  $t_{j+1}$  by a map  $\phi^j$ . Using the description of level lines in terms of unit normals, we propose to compute the optical flow  $\phi^j$  by aligning the unit normal vector field  $Z^{t_{j+1}}(\mathbf{p})$  with the transformed vector field of  $Z^{t_j}(\mathbf{p})$  by the map  $\phi^j$  (i.e., the vector field obtained by (3)). Thus, we propose to compute the motion estimation by minimizing the energy functional

$$E(\phi) = \sum_{j=0}^{N-1} \int_{\Omega} \|Z^{t_{j+1}}(\phi^j(\mathbf{p})) - \bar{Z}_{\phi^j}(\mathbf{p})\|^2 \mu_j(\mathbf{p}) \, dx \, dy. \quad (4)$$

where  $\mu_j(\mathbf{p})$  represent weight functions that will be later discussed. The vector field  $Z^{t_j}(\mathbf{p})$  is computed by

$$Z^{t_j}(\mathbf{p}) = \frac{\nabla I(t_j, \mathbf{p})}{\|\nabla I(t_j, \mathbf{p})\|} \quad \text{if } \nabla I(t_j, \mathbf{p}) \neq 0; \quad 0 \text{ otherwise,} \quad (5)$$

where  $\nabla := (\partial_x, \partial_y)^T$  denotes the 2D gradient. Note that (5) computes the normal vector of the level line that passes through point  $\mathbf{p}$ .

Since for any smooth strictly increasing function we have  $\nabla g(I) = g'(I)\nabla I$ , it is easy to check that if  $\mu_j(\mathbf{p}) = 1$ , then the energy (4) is contrast invariant. In case that we decide to give more weight to edges, we may take  $\mu_j(\mathbf{p}) = \|\nabla I(t_j, \mathbf{p})\|$ ,  $j = 0, \dots, N - 1$ , in this case, if  $I(t_j + 1, \phi^j(\mathbf{p})) = g_j(I(t_j, \mathbf{p}))$  for some contrast change  $g_j$ , then the estimate of  $\phi^j$  obtained by minimizing the corresponding term in (4) does not depend on the contrast change  $g_j$ .

If our assumption (\*) does not hold, then the minimum of (4) (plus some regularization terms for  $\phi^j$ ) can be considered only as an approximation to the optical flow in terms of that criterion, and further validation is required.

The same functional was used by Droske and Rumpf, together with suitable regularizations, for morphological image registration [13]. Other authors ([12]) have also used alignment of unit normals and other geometric features like curvature for registration. Another contrast invariant functional, based on Bayesian inference, was proposed in [11]. The main part of their functional is the integral of  $(I_t + uI_x + vI_y)^2$  divided by the norm of  $(1, u, v)$  times the norm of  $(I_t, I_x, I_y)$ . As we shall also do, the authors assume a parametric piecewise affine motion model. Let us finally mention the work [6] where authors minimize a robust functional which incorporates deviations from the brightness constancy assumption and the gradient constancy assumption, and compute a dense optical flow. Thus, this functional incorporates gradients, hence normal directions and geometry. Finally, let us mention that other contrast invariant functionals can be constructed based on mutual information [19].

### 3 Region-Based Motion Estimation

The energy functional together with a regularization term for  $\phi^j$ ,  $j = 0, \dots, N - 1$ , could be used to compute a dense motion field. In this work, we shall assume that the motion fields can be expressed locally by an affine model and we shall follow a region-based strategy to minimize (4).

Our approach will be similar to the one presented in [14]. In this paper, two images at two different time instants, generally consecutive, of an image sequence, are taken. The first of them is partitioned into connected regions with disjoint interior. These regions are assumed to be extracted from the image using a particular partitioning strategy, such as a luminance homogeneity criterion. Matching of regions is carried out by minimizing a cost functional based on the brightness constancy assumption. Moreover, the technique is embedded in a multiresolution scheme in order to improve the robustness of the method.

For the rest of the paper, the motion is estimated between two consecutive frames of a sequence, denoted by  $I(t)$  and  $I(t+1)$ . The vector fields of the normals to the level lines of  $I(t)$  and  $I(t+1)$  are denoted by  $Z^t$  and  $Z^{t+1}$  respectively. Suppose that we are given a partition  $\mathcal{R}$  into disjoint connected regions of the image  $I(t)$ . The partition may be computed for instance with a segmentation algorithm like the Mumford-Shah functional [17] which may be subordinated to the topographic map [2]. We denote by  $\phi$  the displacement field between  $I(t)$  and  $I(t+1)$ .

In the present context, we can write functional (4) for discrete images as

$$E_{\mathcal{R}}(\phi) = \sum_{R \in \mathcal{R}} \sum_{\mathbf{p} \in R} \|Z^{t+1}(\phi(\mathbf{p})) - \bar{Z}_{\phi^t}(\mathbf{p})\|^2 \mu(\mathbf{p}) \Delta_R \quad (6)$$

for a weighting function  $\mu(\mathbf{p})$  and where  $\Delta_R = \Delta x \Delta y$ ,  $\Delta x$ ,  $\Delta y$  being the discretization steps which coincide with the interpixel distance in the  $x$  and  $y$  axis. For later convenience, let us denote by  $E_R$  the term

$$E_R(\phi) = \sum_{\mathbf{p} \in R} \|Z^{t+1}(\phi(\mathbf{p})) - \bar{Z}_{\phi^t}(\mathbf{p})\|^2 \mu(\mathbf{p}) \Delta_R \quad (7)$$

Recall that rigid motions of planar objects in 3-D space induce quadratic motion models in 2-D images [23]. This quadratic motion model is a good approximation when the depth of the objects is small compared to their distance to the camera. The affine motion model is a good approximation under orthographic projection, i.e., when  $f \rightarrow \infty$ , being  $f$  the distance from the center of the lens to the surface of the film. As a first approximation, we shall assume the affine motion model on each region. Such motion can be described by a six parameter affine model [23]:

$$\phi(\mathbf{p}) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix} \quad (8)$$

where  $e, f$  are the translation parameters and  $a, b, c, d$  are the parameters that model the linear transformation (thus, including scaling, rotation and shearing) [23]. From now on,  $e$  and  $f$  are called *zero-order parameters* whereas the remaining ones are called *first-order parameters*.

In this case the cofactor matrix is

$$(D\phi)^\dagger = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}$$

Observe that we have no information in a region  $R$  when we have no level lines in it. In this case, we would have  $Z = 0$  at time  $t$  inside  $R$ , and  $Z \neq 0$  is the unit normal on its boundary; the proposed functional is looking for a region at time  $t + 1$  which is free of level lines in its interior and matches the unit normals of boundary of  $R$  by  $\phi^t$ . In this case, it could be useful to consider the brightness constancy assumption for this region.

As it is presented, this model does not take into account the fact that new objects may appear or disappear due to motion of objects or to geometric variations produced by local contrast changes. In a further extension of this work we consider statistical validation of the estimated motion and we believe that the appearance/disappearance of an object will lead to incorrect estimations.

## 4 Implementation

From a practical point of view, it is necessary to define a strategy in order to find a minimum of (6) in an efficient and robust manner. We describe in this section some details of our implementation.

**Functional Minimization.** We assume that each region that composes the partition moves independently, thus (6) may be minimized by minimizing (7) for each region. The parameters that minimize (7) are those that satisfy  $\nabla_{\mathbf{m}} E_R(\phi) = 0$ , where  $\mathbf{m}$  is the vector made up of the motion parameters,  $\mathbf{m} = (a, b, c, d, e, f)^T$  and  $\nabla_{\mathbf{m}}$  is the gradient operator with respect to the unknown motion parameters,  $\nabla_{\mathbf{m}} = (\partial_a, \partial_b, \partial_c, \partial_d, \partial_e, \partial_f)$ . The strategy adopted in this work to find the motion

parameters that minimize (6) is the *conjugate-gradient* method. In this paper the conjugate update directions are computed using Polak-Ribiere method. At each iteration, Brent's line minimization is used to find the minimum along the selected update direction [20].

**Coordinate System Selection.** Let us discuss how the selection of the coordinate system may affect the convergence of the descent method to minimize  $E_R$ . For simplicity, assume that  $Z^{t+1}$  and  $\bar{Z}_{\phi^t}^t$  are unit vectors. The differentiation of each term in (7) with respect to any motion parameter  $m \in \mathbf{m}$  gives the contribution

$$\langle \partial_m \bar{Z}_{\phi^t}^t, Z^{t+1} \rangle + \langle \bar{Z}_{\phi^t}^t, \partial_m Z^{t+1} \rangle.$$

Let us compute  $\partial_m Z^{t+1}$ . For that, assume that the origin of the coordinate system is located at the upper left corner of the image, whereas samples are placed on a squared grid at a distance of one. Let us denote by  $(\hat{x}, \hat{y})^T$  the coordinates of  $\mathbf{p}$  in this coordinate system. Let  $Z_i^{t+1}$ ,  $i = 1, 2$ , be the  $x$  and  $y$  components, respectively, of  $Z^{t+1}$ . Observe that the partial derivative of  $Z_1^{t+1}$  with respect to the motion parameter  $a$  is

$$\partial_a Z_1^{t+1}(\phi(\mathbf{p})) = \partial_{\phi_1} Z_1^{t+1} \partial_a \phi_1 + \partial_{\phi_2} Z_1^{t+1} \partial_a \phi_2 = \partial_{\phi_1} Z_1^{t+1} \hat{x} \quad (9)$$

since  $\partial_a \phi_1 = \hat{x}$  and  $\partial_a \phi_2 = 0$ . The derivative of  $Z_1^{t+1}$  with respect to the motion parameter  $e$  is

$$\partial_e Z_1^{t+1}(\phi(\mathbf{p})) = \partial_{\phi_1} Z_1^{t+1} \partial_e \phi_1 + \partial_{\phi_2} Z_1^{t+1} \partial_e \phi_2 = \partial_{\phi_1} Z_1^{t+1} \quad (10)$$

since  $\partial_e \phi_1 = 1$  and  $\partial_e \phi_2 = 0$ . With similar computations we see that the partial derivatives of  $Z_1^{t+1}$  and  $Z_2^{t+1}$  with respect to  $a$ ,  $b$ ,  $c$  or  $d$  depend proportionally on the pixel coordinates (either  $\hat{x}$  or  $\hat{y}$ ), while its partial derivatives with respect to  $e$  or  $f$  do not exhibit such dependence. Thus the derivatives with respect to  $a, b, c, d$  have a stronger contribution in the descent than the ones with respect to  $e, f$  and affect its convergence.

In [14], the authors propose to normalize the pixel coordinates simply by the image dimensions. We shall use a different normalization for each region. The origin of the coordinate system,  $(\tilde{x}^c, \tilde{y}^c)$ , will be the centroid of the region, and the axis will be re-scaled by  $\sigma$  where  $\sigma^2 = \sum_{\mathbf{p} \in R} ((\hat{x} - x^c)^2 + (\hat{y} - y^c)^2)$  is the variance of the distance of the pixel coordinates to the centroid.

If  $(x, y)^T$  denote the coordinates of  $\mathbf{p}$  in this new coordinate system, then its relation with  $(\hat{x}, \hat{y})^T$  is given by

$$x = \frac{\hat{x} - x^c}{\sigma} \quad y = \frac{\hat{y} - y^c}{\sigma}. \quad (11)$$

The partial derivatives of  $Z_1^{t+1}$  with respect to  $a$  and  $e$  have the same form as in (9), (10), with a different interpretation of the coordinates:

$$\partial_a Z_1^{t+1}(\phi(\mathbf{p})) = \partial_{\phi_1} Z_1^{t+1} \frac{(\hat{x} - x^c)}{\sigma} \quad \text{and} \quad \partial_e Z_1^{t+1}(\phi(\mathbf{p})) = \partial_{\phi_1} Z_1^{t+1}. \quad (12)$$

Comparing these expressions with (9) and (10), we notice that the normalized coordinate system has the effect that no particular derivative value is predominant with regard to the others. Our experiments have shown that in this case a large range of motions may be recovered since no particular type of motion is prioritized.

On the other hand, if the  $(\hat{x}, \hat{y})^T$  coordinate system is used the conjugate gradient algorithm selects update directions in which the first order motion parameters are predominant. As a result, the algorithm will “try to explain” the motion present in the image using only first order parameters (zoom, rotations and skew) even if only translational motion is present in the image. Motions such as translations may not be recovered in this case. Thus, the selection of the proper coordinate system affects directly the gradient values and thus the convergence of our estimator.

**Multiresolution Analysis.** Both to avoid local minima and for computational efficiency, motion estimation is usually embedded in a multiresolution scheme [1, 14, 22, 23]. The basic idea is to obtain a set of coarse to fine images which are obtained by means of a low-pass filter. Starting the parameter estimation at the coarsest resolution level, the motion is estimated on each level successively using the resulting motion parameters of the current resolution level as input to initialize the gradient descent on the next level. Lower resolution levels allow to obtain an approximation of the motion parameters, whereas finer resolution levels are used to improve and fine-tune the motion parameter estimation. Multiresolution representations allow to deal with large zero and first order parameters.

In our experiments a set of three (including the original image) levels are used. At each level the image is lowpass filtered with respect to the previous level [24]. As proposed in [23], the downsampling step is skipped. Thus the pyramid contains images that are all the same size by successively more blurred as we go to the coarser resolution levels. This permits us to maintain the geometry of the region.

However,  $\Delta x$  and  $\Delta y$  are divided by two between successive levels, hence the area of the region is scaled by 4. Thus,  $\Delta x$  and  $\Delta y$  act as a scale parameter. If such area is small at a fixed level, only the zero-order parameters are estimated. This is due to that the texture information of small regions present at the coarser levels is poor, and thus the minimum may not be well defined for the first-order parameters.

**Differentiation.** Differentiation is an ill-posed problem [4], and regularization may be used to obtain good numerical derivatives. Such regularization may be accomplished with a low-pass filter such as the Gaussian, and is essential for motion estimation [3, 22]. More recently, [15] proposes to use a matched pair of low pass and differentiation filters as a gradient operator.

Notice that, for motion estimation applications, it may be necessary to compute the gradient at non integer points, since non integer displacements are allowed. In such cases, a simple way to proceed is a two step process: in a first step, the image is interpolated at the required points using an interpolation kernel such as [16], and in the second step the derivative is computed. Since both

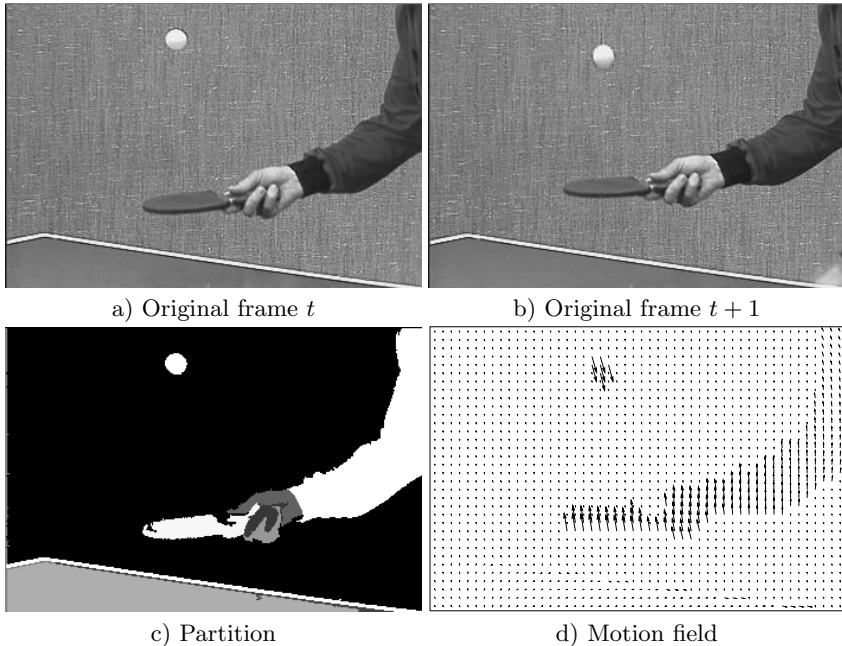
are linear operators they may be performed in one step: the derivative filter is interpolated at the required non-integer positions, then the derivative can be computed at integer pixel positions using the interpolated filter taps.

## 5 Results

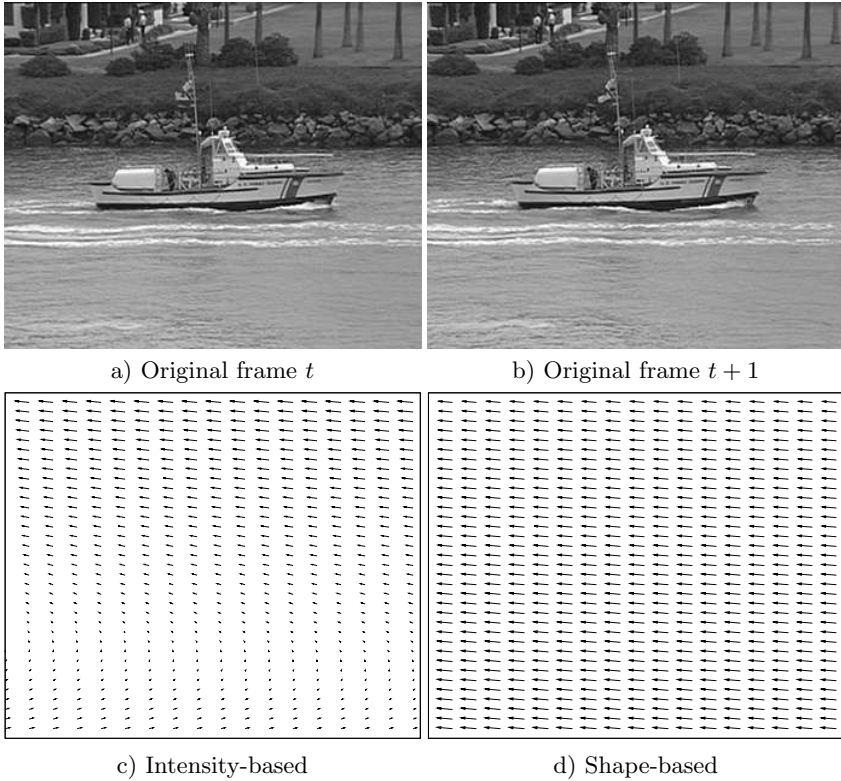
In all experiments below, we assume that  $\mu_j(\mathbf{p}) = 1$  (see (6)).

The original images in our first example correspond to the table tennis sequence (frames #4 and #1), and are displayed in Fig. 1a and Fig. 1b respectively. In these images, the ball moves downwards, the arm moves upwards and the background is static. The associated partition has been computed using the Mumford-Shah segmentation functional subordinated to the topographic map [2], and is shown in Fig. 1c. The motion field recovered by our estimator is shown in Fig. 1d and corresponds to our above description.

An interesting point is to compare our results with those obtained with the classical motion estimation approach based on minimization of the squared prediction error, defined as:  $E_R^{\text{int}}(\phi) = \sum_{\mathbf{p} \in R} (I(t+1, \phi(\mathbf{p})) - I(t, \mathbf{p}))^2$ . The previous error measure is in fact based on the brightness constancy assumption. The latter approach has been implemented using the techniques described in Sect. 4. In order to simplify nomenclature, the latter approach is called *intensity*-based



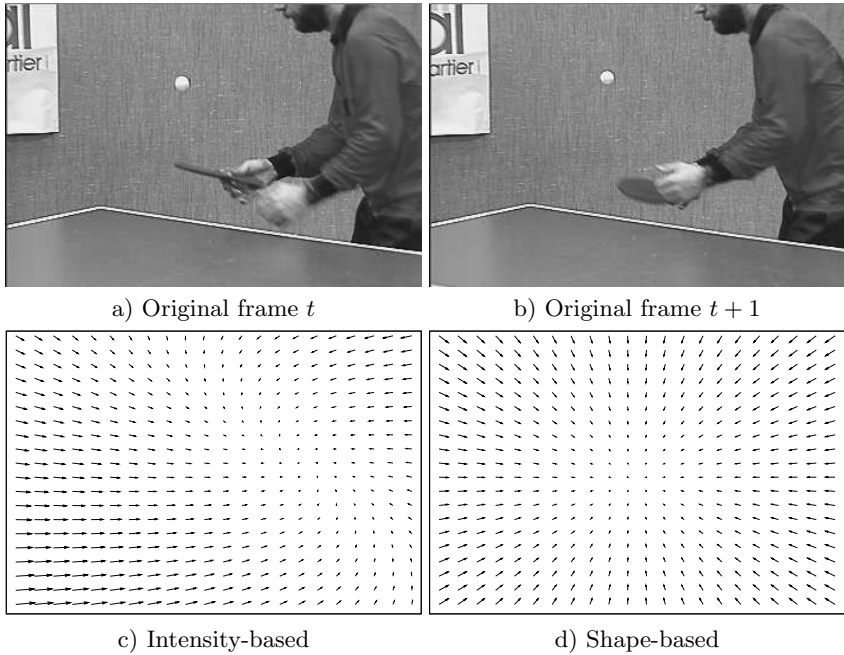
**Fig. 1.** Region-based motion example. a) Original frame  $t$ , b) Original frame  $t + 1$ , c) Partition of original frame  $t$ , d) Recovered motion field



**Fig. 2.** Global motion estimation example. The purpose is to extract camera's movement between images a) and b). Results for intensity and shape-based approach are shown in images c) and d) respectively

motion estimation, whereas the technique presented in this paper is called *shape-based* motion estimation.

The next experiments deal with global motion estimation, that is, the extraction of the camera motion. This is very useful in many scene analysis approaches, where first the camera motion is detected and then the moving objects in the scene are detected and tracked. Fig. 2 shows two frames from the *coastguard* sequence (frames #170 and #176). The frames show a moving boat and a static background. In these frames the camera follows the displacement of the boat, thus the apparent motion of the boat is zero (i.e. no motion) whereas the background has an apparent motion which corresponds to the camera's movement. We choose a partition made up of *one region* which includes the whole image support. Thus, the global motion between the two frames is estimated. The resulting motion vector fields are shown in Fig. 2c and Fig. 2d. Note that our approach has been able to properly extract the camera's motion. We believe that the intensity-based motion estimation has failed due to the strong influence of the high gradient of the boat. Since the apparent motion of the boat is



**Fig. 3.** Global motion estimation example. The purpose is to extract camera’s movement between images a) and b). They correspond to frames #45 and #47 of the table tennis sequence. Recovered motion fields for intensity and shape-based approach are shown in images c) and d) respectively

zero, the motion estimation algorithm tries to set to zero the motion at the boat pixel locations. This is not the case of the shape-based approach, where gradient modulus has no effect. Motion in the latter case is recovered by interpreting the image as a set of moving level lines. Thus, the boat is treated as an outlier. Note that the correct motion parameters may be recovered using an intensity based energy if robust estimation techniques are used [5, 22].

Fig. 3 shows another example of global motion estimation. The camera performs a zoom out of the scene. Even though the tennis player and the ping-pong ball is moving, our approach has been able to properly recover the zoom.

## 6 Conclusions and Future Work

We have presented a contrast invariant model for the computation of the optical flow. We interpret the image sequence as a set of moving level lines and we propose to compute the deformation between the level lines of two consecutive frames. Several topics have to be further developed in the future: a) the selection of regions bounded by level lines where motion is estimated by an affine model, b) joint motion segmentation techniques, c) the computation of a dense motion field from the image sequence without imposing a particular motion model.

## References

1. L. Alvarez, J. Weickert, and J. Sánchez. A scale-space approach to nonlocal optical flow calculations. In *Scale-Space Theories in Computer Vision*, pages 235–246. Springer Verlag, 1999.
2. C. Ballester, V. Caselles, and L. Igual. Minimal morphological shape selection for segmentation and encoding. *Preprint*, 2004.
3. J. Barron, D. Fleet, and S. Beauchemin. Performance of optic flow techniques. *IJCV*, 12:43–77, 1994.
4. M. Bertero, T.A. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76(8):869–889, August 1988.
5. M.J. Black and P. Anandan. The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. *CVIU*, 63(1):75–104, January 1996.
6. T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *European Conf. on Computer Vision*, pages 25–36. Springer Verlag, 2004.
7. A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger, and C. Schnoerr. Real-time optic flow computation with variational methods. In *Int. Conf. on Computer Analysis of Images and Patterns*, pages 222–229. Springer Verlag, August 2003.
8. P.Y. Burgi. Motion estimation based on the direction of intensity gradient. *Image and Vision Computing*, 22:637–653, 2004.
9. V. Caselles, B. Coll, and J-M. Morel. Topographic maps and local contrast changes in natural images. *IJCV*, 33(1):5–27, 1999.
10. H. Chen, P. Bellhumeur, and D. Jacobs. In search of illumination invariants. In *Int. Conf. on Computer Vision and Pattern Recognition*, pages 254–261, 2000.
11. D. Cremers and S. Soatto. Motion competition: a variational approach to piecewise parametric motion segmentation. *IJCV*, 2004.
12. C. Davatzikos. Spatial transformation and registration of brain images using elastically deformable models. *CVIU*, 1997.
13. M. Droske and M. Rumpf. A variational approach to non-rigid morphological image registration. *SIAM Journal Applied Mathematics*, 2004.
14. J.L. Dugelay and H. Sanson. Differential methods for the identification of 2D and 3D motion models in image sequences. *Image Communication*, 7:105–127, 1995.
15. H. Farid and E.P. Simoncelli. Differentiation of discrete multidimensional signals. *IEEE Trans. on IP*, 13(4), April 2004.
16. R.G. Keys. Cubic convolution interpolation for digital image processing. *IEEE Trans. on ASSP*, 29(6):1153–1160, December 1981.
17. J.M. Morel and S. Solimini. *Variational methods in image segmentation*. Birkhäuser Verlag: Basel, 1995.
18. J.M. Obobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models applied to complex scenes. *JVCIR*, 6(4):348–365, December 1995.
19. J.P.W. Pluim, J.B.A. Maintz, and M.A. Viergever. Mutual-information-based registration of medical images. *IEEE Trans. on MI*, 22(8):986–998, August 2003.
20. W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C: the art of scientific computing*. Cambridge Univ. Press, 1992.
21. J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, 1982.
22. C. Stiller and J. Konrad. Estimating motion in image sequences. *IEEE Signal Processing Magazine*, pages 70–91, July 1999.
23. A.M. Tekalp. *Digital Video Processing*. Prentice-Hall, 1995.
24. M. Unser, A. Aldroubi, and M. Eden. Enlargement or reduction of digital images with minimum loss of information. *IEEE Trans. on IP*, 4(3):247–258, March 1995.