Available online at www.sciencedirect.com
SCIENCE DIRECT
Computer Vision and Image Understanding 103 (2006) 89–100
Computer Vision and Image Understanding
www.elsevier.com/locate/cviu

Joint optical flow estimation, segmentation, and 3D interpretation with level sets

H. Sekkati, A. Mitiche *

Institut National de la recherche Scientifique, INRS-EMT, Place Bonaventure, 800, de la Gauchetière Ouest, Suite 6900, Montréal, Que., Canada H5A1K6

Received 8 July 2005; accepted 16 November 2005
Available online 5 January 2006

Abstract

This paper describes a variational method with active curve evolution and level sets for the estimation, segmentation, and 3D interpretation of optical flow generated by independently moving rigid objects in space. Estimation, segmentation, and 3D interpretation are performed jointly. Segmentation is based on an estimate of optical flow consistent with a single rigid motion in each segmentation region. The method, which allows both viewing system and viewed objects to move, results in three steps iterated until convergence: (a) evolution of closed curves via level sets and, in each region of the segmentation, (b) linear least squares computation of the essential parameters of rigid motion, (c) estimation of optical flow consistent with a single rigid motion. The translational and rotational components of rigid motion and regularized relative depth are recovered analytically for each region of the segmentation from the estimated essential parameters and optical flow. Several examples with real image sequences are provided which verify the validity of the method.

Keywords: 3D structure; 3D motion; Segmentation; Optical flow; Variational method; Level sets

> A major branch of picture processing deals with *image analysis* or *scene analysis*; ... the desired output is a *description* of the given picture or scene.
> Segmentation is basically a process of *pixel classification*; the picture is segmented into subsets by assigning the individual pixels to classes ... The approach discussed ... minimizes the expected classification error, ..., more generally, (an) *expected cost* ... It should be pointed out that if range or *velocity information* is available for each pixel, obtained by special sensors or derived from stereopairs or image sequences, we can use this information as a basis for segmenting the picture.
> (At) an edge, the gray level is relatively consistent in each of two adjacent, extensive regions, and changes abruptly as the border between the regions is crossed ... this is a special case of pixel classification.
>
> *Azriel Rosenfeld* in A. Rosenfeld and A. C. Kak *Digital picture processing*, Academic Press, Second edition, 1982.

Foreword

Azriel Rosenfeld founded the field of image processing four decades ago. He identified and investigated fundamental subjects which are still the focus of intense research. The book *Digital picture processing* by A. Rosenfeld and A. C. Kak, published in 1976, was structured based on these subjects and described research in a textbook style. This made its contents informative, accessible, and attractive so as to bestir extensive and sustained research. This paper deals with image *segmentation/edge-detection* and *description*, subjects that the quotes from the book highlight. The basis for segmentation/edge-detection is the *velocity field* derived from an image sequence, and the saught description is the three-dimensional (3D) structure and motion of the objects in the observed scene. The velocity field and its 3D interpretation are derived from *cost function*

* Corresponding author.
E-mail addresses: sekkati@emt.inrs.ca (H. Sekkati), mitiche@emt.inrs.ca (A. Mitiche).

1077-3142/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2005.11.002

minimization, using some of the latest analysis tools: the variational formalism and active curves via level sets.

1. Introduction

Optical flow analysis plays fundamental roles in computer vision. For instance, it serves movement detection, tracking, segmentation, and 3D interpretation [1,2]. Optical flow 3D interpretation is of considerable interest in applications such as robot autonomous navigation, virtual and augmented reality, medical diagnosis, and surveillance.

The primary focus of this study is *dense* 3D interpretation of optical flow of rigid objects moving independently relative to a viewing system. Within the general context of 3D interpretation of image sequences, *sparse* interpretation, which seeks depth and 3D motion from views of a sparse set of points in space, has been the subject of a substantial number of well documented studies [3–5,2,6]. Dense optical flow interpretation has been investigated relatively little because it is considerably more complex, particularly when the viewing system and the viewed objects are allowed to move, [7–25].

Most current methods of dense optical flow interpretation consider the problem of a viewing system moving in a stationary environment [7–19]. In this case, the camera motion is the single 3D motion to recover and moving object segmentation is no longer an issue. In spite of these simplifications, interpretation remains complicated by image noise and depth discontinuities. Such complications occur in optical flow and disparity estimation as well and have been resolved by modeling and anisotropic diffusion within a variational framework [26]. The variational formalism allows the coding of models, assumptions, and constraints in a single functional to minimize. This leads to efficient, tractable algorithms [27]. The variational methods of optical flow 3D interpretation in [19,8,12] have not fully exploited the formalism.

A few studies have addressed the problem when both the viewing system and the viewed objects are allowed to move independently [20–22,24,23,25]. In this case, segmentation of the image domain into regions corresponding to differently moving objects in space is essential. This complicates the problem significantly. Non variational methods have used traditional grouping processes for segmentation, such as region growing [22] and clustering [20]. Variational formulations have been investigated in [24,23,25]. In [24,23], the methods are not based on active curves/level sets and do not address segmentation explicitly. The active curves/level sets method in [25] states the problem as 3D-motion segmentation with simultaneous 3D-motion and dense depth estimation within each region of the segmentation.

This study investigates a new level set method for joint estimation, segmentation, and 3D interpretation of the optical flow generated by multiple rigid objects in space moving independently relative to the viewing system. Optical flow is estimated so as to correspond to a single rigid 3D motion in each segmentation region via the essential parameters representation of rigid motion [28,29,2]. The formulation results in an algorithm where three steps are interated until convergence: (a) evolution of closed curves via level sets and, in each segmentation region, (b) least squares computation of the essential parameters of rigid motion and (c) estimation of optical flow consistent with a single rigid 3D motion. The translational and rotational components of rigid motion and regularized relative depth are recovered analytically for each segmentation region from the estimated essential parameters and optical flow. Several examples with real image sequences are provided to verify the validity of the method.

2. Basic models

In this section, we describe a constraint which relates optical flow to the representation of rigid 3D motion by essential parameters. This constraint will be used in the formulation to compute an estimate of optical flow conform to a single rigid 3D motion in each segmentation region. We also describe a mapping between the interiors $\{\mathbf{R}_{\gamma_k}\}_{k=1}^{N-1}$ of a family of $N-1$ closed plane curves $\{\gamma_k\}_{k=1}^{N-1}$ and regions $\{\mathbf{R}_k\}_{k=1}^{N}$ which always forms a partition of the image domain. This mapping will be used in the formulation to guarantee that the computed segmentation corresponds to a partition of the image domain.

2.1. Optical flow and essential parameters of rigid 3D motion

Let $\mathbf{T}=(t_1,t_2,t_3)$, $\boldsymbol{\omega}=(\omega_1,\omega_2,\omega_3)$ be the translational and rotational components of the motion of a rigid body $\mathscr{B}$ in space. Let $\mathbf{p}$ be the image of a point $\mathbf{P}\in\mathscr{B}$ (refer to

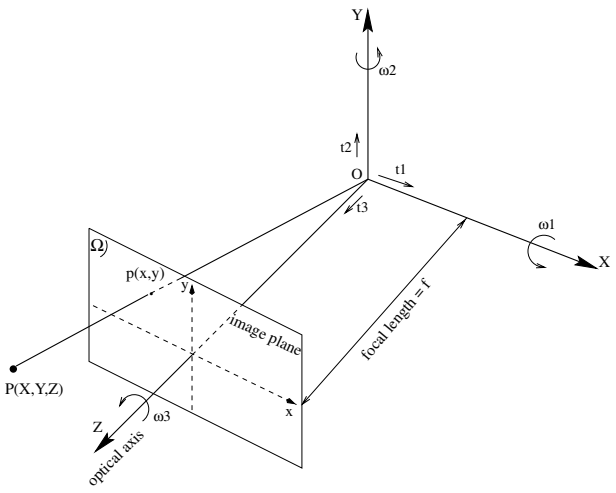

Fig. 1. The viewing system is modeled by an orthonormal coordinate system and central projection through the origin.

Fig. 1). Finally, let (x,y) be the coordinates of $\mathbf{p}$ and $\mathbf{w}=(u,v)$ the optical flow at (x,y) (the optical velocity of $\mathbf{p}$) at some instant of observation. Optical flow and the components $\mathbf{T}$, $\boldsymbol{\omega}$ of the rigid motion of $\mathscr{B}$ are related by the following homogeneous linear equation [28,29,2]:

$$\mathbf{d}\cdot\mathbf{e}=0, \tag{1}$$

where $\mathbf{d}$ is a vector of dimension nine whose components depend on image coordinates and optical flow:

$$\mathbf{d}=(x^2,y^2,f^2,2xy,2xf,2yf,-fv,fu,-uy+vx) \tag{2}$$

and $\mathbf{e}$ is the vector of *essential parameters* related to $\mathbf{T}$ and $\boldsymbol{\omega}$ by

$$\begin{aligned}
&t_1=e_7,\quad t_2=e_8,\quad t_3=e_8,\\
&2\omega_1t_1=e_1-e_2-e_3,\\
&2\omega_2t_2=e_2-e_3-e_1,\\
&2\omega_3t_3=e_3-e_2-e_1,\\
&\omega_2t_1+\omega_1t_2=2e_4,\\
&\omega_1t_3+\omega_3t_1=2e_5,\\
&\omega_2t_3+\omega_3t_2=2e_6.
\end{aligned} \tag{3}$$

2.2. A partition of the image domain from closed curves

In segmentation by active curve evolution, the standard representation of a region maps the region to the interior of a closed simple curve. A problem arises when two or more curves intersect because the membership of the points in the intersection becomes ambiguous. It is essential that a segmentation algorithm yields a partition of the image domain, albeit at convergence [30,31–33]. Here, we define regions from curves so as to guarantee a partition [33]. This definition is as follows. Refer to Fig. 2 for an illustration.

Let $\{\gamma_k\}_{k=1}^{N-1}$ be a family of closed plane curves, and let $\mathbf{R}_{\gamma_k}$ designate the interior of $\gamma_k, k=1,\ldots,N-1$. We define region $\mathbf{R}_1$ as $\mathbf{R}_{\gamma_1}$, i.e., the interior of γ_1. Region $\mathbf{R}_2$ of the segmentation, however, is identified with $\mathbf{R}^c_{\gamma_1}\cap\mathbf{R}_{\gamma_2}$, rather than $\mathbf{R}_{\gamma_2}$ as in the standard correspondence. Similarly, region $\mathbf{R}_3$ of the segmentation is defined as $\mathbf{R}^c_{\gamma_1}\cap\mathbf{R}^c_{\gamma_2}\cap\mathbf{R}_{\gamma_3}$ and, more generally, $\mathbf{R}_k=\mathbf{R}^c_{\gamma_1}\cap\mathbf{R}^c_{\gamma_2}\cap\cdots\cap\mathbf{R}^c_{\gamma_{k-1}}\cap\mathbf{R}_{\gamma_k}, k\leqslant N-1$. The last region, $\mathbf{R}_N$, is defined as in the standard correspondence by $\mathbf{R}_N=(\cup_{k=1}^{N-1}\mathbf{R}_k)^c$, i.e., the complement of the union of the other regions. Therefore, the resulting family $\{\mathbf{R}_k\}_{k=1}^{N}$ is a partition of the image domain for any given family of closed curves $\{\gamma_k\}_{k=1}^{N-1}$.

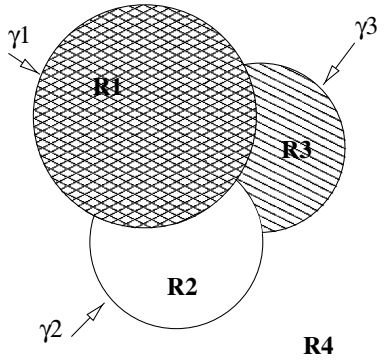

Fig. 2. Representation of a partition of the image domain by explicit correspondence between regions of segmentation and regions defined by closed plane curves (illustration for four regions).

3. Formulation

The problem of segmenting optical flow into N regions, each corresponding to a single rigid 3D object, can be stated as follows: minimize

$$\begin{aligned}
\mathscr{E}(\{\gamma_k\}_{k=1}^{N-1})=&\sum_{k=1}^{N}\int_{\mathbf{R}_k}\left(\left(u^*-\bar{u}_k^*\right)^2+\left(v^*-\bar{v}_k^*\right)^2\right)\mathrm{d}\mathbf{x}\\
&+\lambda\sum_{k=1}^{N-1}\int_{\gamma_k}\mathrm{d}s,
\end{aligned} \tag{4}$$

where the family of regions $\{\mathbf{R}_k\}_{k=1}^{N}$ is defined from closed curves $\{\gamma_k\}_{k=1}^{N-1}$ as in Section 2.2; (u^*, v^*) is a motion field consistent, in *each* region, with a *single* rigid motion described by essential parameters and computed jointly with the essential parameters by

$$\begin{aligned}
(u^*,v^*,\mathbf{e}_k)=\arg\min_{u,v,\mathbf{e}_k}\int_{\mathbf{R}_k}&\Big[(\mathbf{d}\cdot\mathbf{e}_k)^2+\mu(\nabla I\cdot\mathbf{w}+I_t)^2\\
&+v(g(\|\nabla u\|)+g(\|\nabla v\|))\Big]\mathrm{d}\mathbf{x},\quad k=1,\ldots,N
\end{aligned} \tag{5}$$

and, $\bar{u}_k^*(\bar{v}_k^*)$ is the mean of u^* (v^*) in region $\mathbf{R}_k$, $k=1,\ldots,N$,

$$\bar{u}_k^*=\frac{1}{\int_{\mathbf{R}_k}\mathrm{d}\mathbf{x}}\int_{\mathbf{R}_k}u^*\mathrm{d}\mathbf{x}$$

$$\bar{v}_k^*=\frac{1}{\int_{\mathbf{R}_k}\mathrm{d}\mathbf{x}}\int_{\mathbf{R}_k}v^*\mathrm{d}\mathbf{x}.$$

Function g in (5) is of class C^2. With the quadratic function, we have the Horn-and-Shunck regularization term. This function can also be chosen to preserve motion boundaries within each region, as with the Aubert–Deriche–Kornprobst function $g(s)=2\sqrt{1+s^2}-2$ [26].

This formulation can be explained as follows. The purpose is 3D interpretation of the optical flow induced by 3D objects simultaneously and independently moving. First, this requires the segmentation of the objects. Second, it requires the optical flow. Third, the optical flow must be related to the variables of the 3D interpretation within each region of the segmentation. These requirements are all embedded in Functionals (4) (segmentation) and (5) (estimation of optical flow and 3D interpretation) of the formulation. The two functionals are tightly linked because the formulation seeks simultaneously a segmentation and an estimate of optical flow consistent with a *single* rigid body motion in *each* region of the segmentation. There are two important observations to make:

(A) In image and optical flow segmentation [34,35], a functional such as (4) is sometimes referred to as a piecewise constant segmentation functional . It is the simplest form of the general linear parametric functional [36]. The focus of such formulations is not on computing an accurate parametric representation of an image or a flow field, but on partitioning the image domain into regions which (are assumed to) differ by the parameters of the representation. This is the case in this formulation which seeks a segmentation under the assumption that the *desired* regions, i.e., where each region corresponds to a single rigid object in space, differ by their average optical velocity. This assumption is generally valid and, if necessary, an affine model [35] can be used without affecting the segmentation paradigm. However, 3D interpretation requires both an accurate optical flow estimate [2] and a segmentation where each region corresponds to a single rigid object in space. These requirements justify Functional (5). For *each* region of the segmentation, its minimization yields an estimate of optical flow which is smooth (the third term of the integrand), conforms to the spatio-temporal data (second term), and is consistent with a *single* rigid body motion described by its essential parameters (the first term). Optical flow and the essential parameters are estimated jointly. The contribution of the 3D interpretation term (the first term) to both the segmentation and the estimation of optical flow will be illustrated in the experiments (Section 6). Note that smoothness defined by total variation as in (5) cannot be a basis for segmentation when regions are explicitly referenced in the objective functional as in (4).

(B) In regard to joint optical flow estimation and segmentation, this formulation is most related to the one in [35], and can be seen as a generalization because it also jointly estimates 3D structure and motion. Other notable differences are: with this formulation (a) the estimated optical flow is consistent with rigid 3D motion, (b) each region of the segmentation is consistent with a single 3D rigid body motion, (c) segmentation into N regions uses $N-1$ curves as described in Section 2.2, whereas in [35] it uses $\log N$ curves following the construction in [34]. Our construction has the advantage of being straightforward to implement for an arbitrary number of regions [33]. Finally, (d) the method in [35] uses a single functional, whereas this method affords more flexibility with two different functionals, one for segmentation, the other for optical flow estimation (and 3D interpretation).

In regard to 3D interpretation, this formulation is most related to the active curves/level sets formulation in [25]. The formulation in [25] uses a term of conformity to data based on the Horn-and-Schunck optical flow constraint in which it substitutes for optical velocity its expression in terms of depth and the rigid 3D motion components of translation and rotation. Therefore, one significant difference is that optical flow estimation in this formulation replaces the estimation of depth in [25]. This has a major implication computationally because optical flow involves sparse systems of linear equations which, although large, can be solved quite efficiently by Gauss–Seidel iterations or the semi-quadratic algorithm, whereas the estimation of depth involves large systems of nonlinear equations solved by gradient descent. Another difference is that segmentation in [25] is based directly on 3D motion, rather than on optical flow constrained by 3D motion. On all the examples we ran, both methods produce similar results but this method executes significantly faster.

4. Algorithm

The formulation can be implemented by an algorithm which iterates three consecutive steps: computation of the essential parameters for each region, estimation of the optical flow in each region, and curve evolution.

4.1. Initialization

The segmentation curves and the optical flow are initialized. The optical flow is initialized to zero. Alternatively, it can be initialized with the result of a few iterations of the Horn-and-Schunck algorithm. An initial partition is defined (Section 2.2) using closed curves arbitrarily placed over the image domain.

4.2. Step 1: essential parameters

With the optical flow and the segmentation curves fixed, Functionals (5) are each minimized with respect to the essential parameters. Because the essential parameters are constant in each region, this reduces their estimation, up to a scale factor, to a linear least squares fit within each region. This can be done efficiently by the singular value decomposition method [37].

4.3. Step 2: optical flow estimation

With the curves of the segmentation and essential parameters fixed, Functionals in (5) are minimized with respect to optical flow. The corresponding Euler–Lagrange equations for region $\mathbf{R}_k$, $k=1,\ldots,N$, are

$$\begin{cases} (fe_{k,8}-ye_{k,9})\mathbf{d}\cdot\mathbf{e}_k+\mu I_x(\nabla I\cdot\mathbf{w}+I_t)-\nu\mathrm{div}\left(\frac{g'(\|\nabla u\|)}{\|\nabla u\|}\nabla u\right)=0,\\ (-fe_{k,7}+xe_{k,9})\mathbf{d}\cdot\mathbf{e}_k+\mu I_y(\nabla I\cdot\mathbf{w}+I_t)-\nu\mathrm{div}\left(\frac{g'(\|\nabla v\|)}{\|\nabla v\|}\nabla v\right)=0, \end{cases} \tag{6}$$

where $e_{k,j}$ designates the jth component of $\mathbf{e}_k$, the vector of essential parameters corresponding to region $\mathbf{R}_k$, and $\mathbf{w}=(u,v)$. When g is the quadratic function, we have a large sparse linear system equations which can be solved efficiently by Jacobi or Gauss–Seidel iterations, as with the Horn-and-Schunck algorithm. When g is the Aubert–Deriche–Kornprobst function, we have a large sparse system of nonlinear equations which can be solved efficiently by the half-quadratic algorithm [26]. In both cases, computations are confined to the regions interior; this has the effect of

preserving 3D-motion boundaries because regions differ by their 3D-motion via the essential parameters.

4.4. Step 3: curves evolution

The curves evolution is governed by the Euler–Lagrange descent equations corresponding to the minimization of Functional (4) with respect to the curves:

$$\frac{\mathrm{d}\gamma_k}{\mathrm{d}\tau} = -\frac{\delta \mathscr{E}}{\partial \gamma_k}, \quad k = 1, \ldots, N-1, \tag{7}$$

where each curve γ_k is embedded in a one-parameter family of curves indexed by the (algorithmic) time parameter τ.

With our representation of a partition from closed curves (Section 2.2), and with optical flow and the essential parameters fixed, the expression of the descent Eq. (7) are (see [33] for generic details):

$$\begin{cases} \frac{\mathrm{d}\gamma_1}{\mathrm{d}\tau} &= -(\xi_1 - \psi_1 + \lambda\kappa_1)\mathbf{n}_1, \\ \frac{\mathrm{d}\gamma_2}{\mathrm{d}\tau} &= -\left(\chi_{\mathbf{R}_1^c}(\xi_2 - \psi_2) + \lambda\kappa_2\right)\mathbf{n}_2, \\ & \vdots \\ \frac{\mathrm{d}\gamma_k}{\mathrm{d}\tau} &= -\left(\chi_{\mathbf{R}_1^c}\cdots\chi_{\mathbf{R}_{k-1}^c}(\xi_k - \psi_k) + \lambda\kappa_k\right)\mathbf{n}_k, \\ & \vdots \\ \frac{\mathrm{d}\gamma_{N-1}}{\mathrm{d}\tau} &= -\left(\chi_{\mathbf{R}_1^c}\cdots\chi_{\mathbf{R}_{N-2}^c}(\xi_{N-1} - \psi_{N-1}) + \lambda\kappa_{N-1}\right)\mathbf{n}_{N-1}, \end{cases} \tag{8}$$

where $\chi_{\mathbf{R}}$ designates the characteristic function of region $\mathbf{R}$, i.e., $\chi_{\mathbf{R}}(\mathbf{x}) = 1$ if $\mathbf{x} \in \mathbf{R}$ and $\chi_{\mathbf{R}}(\mathbf{x}) = 0$ otherwise; $\mathbf{n}_k$ is the outward unit normal function to γ_k and κ_k its mean curvature function. For $k = 1,\ldots,N$, functions ξ_k and ψ_k are defined by

$$\xi_k = (u^* - \bar{u}_k^*)^2 + (v^* - \bar{v}_k^*)^2, \quad k = 1, \ldots, N,$$

$$\psi_k = \xi_{k+1}\chi_{\mathbf{R}_{\gamma_{k+1}}} + \xi_{k+2}\chi_{\mathbf{R}^c_{\gamma_{k+1}}}\chi_{\mathbf{R}_{\gamma_{k+2}}} + \cdots + \xi_{N-1}\chi_{\mathbf{R}^c_{\gamma_{k+1}}}\cdots\chi_{\mathbf{R}^c_{\gamma_{N-2}}}\chi_{\mathbf{R}_{\gamma_{N-1}}} + \xi_N\chi_{\mathbf{R}^c_{\gamma_{k+1}}}\cdots\chi_{\mathbf{R}^c_{\gamma_{N-2}}}\chi_{\mathbf{R}_{\gamma^c_{N-1}}}.$$

The corresponding level set evolution equations [38,33] are

$$\begin{cases} \frac{\partial\Phi_1}{\partial\tau} &= -(\xi_1 - \psi_1 + \lambda\kappa_{\Phi_1}) \mid \nabla\Phi_1 \mid, \\ \frac{\partial\Phi_2}{\partial\tau} &= -\left(\chi_{\mathbf{R}^c_{\Phi_1}}(\xi_2 - \psi_2) + \lambda\kappa_{\Phi_2}\right) \mid \nabla\Phi_2 \mid, \\ & \vdots \\ \frac{\partial\Phi_k}{\partial\tau} &= -\left(\chi_{\mathbf{R}^c_{\Phi_1}}\cdots\chi_{\mathbf{R}^c_{\Phi_{k-1}}}(\xi_k - \psi_k) + \lambda\kappa_{\Phi_k}\right) \mid \nabla\Phi_k \mid, \\ & \vdots \\ \frac{\partial\Phi_{N-1}}{\partial\tau} &= -\left(\chi_{\mathbf{R}^c_{\Phi_1}}\cdots\chi_{\mathbf{R}^c_{\Phi_{N-2}}}(\xi_{N-1} - \psi_{N-1}) + \lambda\kappa_{\Phi_{N-1}}\right) \mid \nabla\Phi_{N-1} \mid, \end{cases} \tag{9}$$

where Φ_k is the level set function corresponding to γ_k taken positive inside γ_k; $\mathbf{R}_{\Phi_k} = \{\mathbf{x} \in \Omega \mid \Phi_k(\mathbf{x}, t) > 0\}$; $\kappa_{\Phi_k} = \nabla \cdot (\nabla\Phi_k / \mid \nabla\Phi_k \mid)$. Functions ψ_k are given by

$$\psi_k = \xi_{k+1}\chi_{\mathbf{R}_{\Phi_{k+1}}} + \xi_{k+2}\chi_{\mathbf{R}^c_{\Phi_{k+1}}}\chi_{\mathbf{R}_{\Phi_{k+2}}} + \cdots + \xi_{N-1}\chi_{\mathbf{R}^c_{\Phi_{k+1}}}\cdots\chi_{\mathbf{R}^c_{\Phi_{N-2}}}\chi_{\mathbf{R}_{\Phi_{N-1}}} + \xi_N\chi_{\mathbf{R}^c_{\Phi_{k+1}}}\cdots\chi_{\mathbf{R}^c_{\Phi_{N-2}}}\chi_{\mathbf{R}_{\Phi^c_{N-1}}}.$$

We summarize the algorithm as follows:

Initialize the curves and optical flow.
Iterate
1. Compute the essential parameters in each region by linear least-squares.
2. Estimate optical flow using (6).
3. Evolve curves via level sets using (9).

Until convergence

In Functional (4), the term for region $\mathbf{R}_k$ in the sum corresponding to the $u(v)$ component of the motion field is proportional to the variance of the component in the region. Therefore, assuming that the smoothing of motion which occurs at step 2 of the algorithm decreases this variance, Functional (4) will decrease at each iteration and the algorithm will converge.

5. Recovery of structure

The components $\mathbf{T}$, $\boldsymbol{\omega}$ of rigid motion for each segmentation region can be recovered analytically from the region essential parameters. From (3), the translational component $\mathbf{T} = (t_1, t_2, t_3)$, up to a sign and a positive scale factor, is given by $t_1 = e_7$, $t_2 = e_8$, $t_3 = e_9$. The rotational component is obtained, also analytically, from the other equations in (3), i.e., from the linear system of equations

$$\begin{pmatrix} 0 & -e_8 & -e_9 \\ -e_7 & 0 & -e_9 \\ -e_7 & -e_8 & 0 \\ \frac{e_8}{2} & \frac{e_7}{2} & 0 \\ \frac{e_9}{2} & 0 & \frac{e_7}{2} \\ 0 & \frac{e_9}{2} & \frac{e_8}{2} \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{pmatrix}.$$

Depth cannot be recovered if $\mathbf{T} = \mathbf{0}$. When $\mathbf{T} \neq \mathbf{0}$, the relations [29]

$$\begin{cases} u &= \frac{1}{Z}(ft_1 - xt_3) - \frac{xy}{f}\omega_1 + \frac{f^2+x^2}{f}\omega_2 - y\omega_3, \\ v &= \frac{1}{Z}(ft_2 - yt_3) - \frac{f^2+y^2}{f}\omega_1 + \frac{xy}{f}\omega_2 + x\omega_3 \end{cases} \tag{10}$$

give (relative) depth in terms of (u, v) and $(\mathbf{T}, \boldsymbol{\omega})$:

$$Z = \left(\frac{(ft_1 - xt_3)^2 + (ft_2 - yt_3)^2}{(u + \frac{xy}{f}\omega_1 - \frac{f^2+x^2}{f}\omega_2 + y\omega_3)^2 + (v + \frac{f^2+y^2}{f}\omega_1 - \frac{xy}{f}\omega_2 - x\omega_3)^2}\right)^{\frac{1}{2}}. \tag{11}$$

Because the components of **T** appear in a ratio with depth in (10), translation and depth are recovered up to a common scale factor, which is fixed when the essential parameters are computed under a fixed norm constraint Eq. (1) is homogeneous). Once depth is computed, the sign of **T** can be adjusted, if necessary, to correspond to positive depth [28,2].

Note that the algorithm recovers a regularized estimate of depth because the estimated optical flow and essential parameters from which it is computed (Eq. 11) are regularized solutions of functional minimization.

6. Experimental results

To evaluate the proposed method, we ran several experiments with real image sequences. Reconstructed objects are displayed using triangulation-based surface rendering with back-projection of the texture of the original input images. We also provide anaglyphs of stereoscopic images constructed from the estimated depth [25,39]. The results of the segmentation and optical flow estimation are shown and compared to those of other methods. For each example, we experimented with different curve initializations by placing the curves so that the image of a moving object is (a) completely inside a curve, (b) completely outside and, (c) partly inside/outside. In all the examples, these initializations led to the same results. However, the algorithm convergence time varied with the initialization.

This first example uses two consecutive frames of the *Berber* sequence of real images. The figurine rotates and moves forward to the right. The scene surfaces are textured. The parameters are: $\nu = 1$, $\mu = 10^3$, and $\lambda = 10^{-2}$. The initial segmentation curve is shown in Fig. 3A, superimposed on the first of the two frames used. The final segmentation is shown in Fig. 3B. Views of the reconstructed figurine, back-shaded using the original texture, are shown in Figs. 4A–C. The estimated optical flow, and the optical flow computed by the Horn-and-Schunck and the Aubert–Deriche–Kornprobst algorithms are displayed in Figs. 5A–C respectively. The methods of Horn-and-Schunck and Aubert–Deriche–Kornprobst are benchmark variational methods which use,

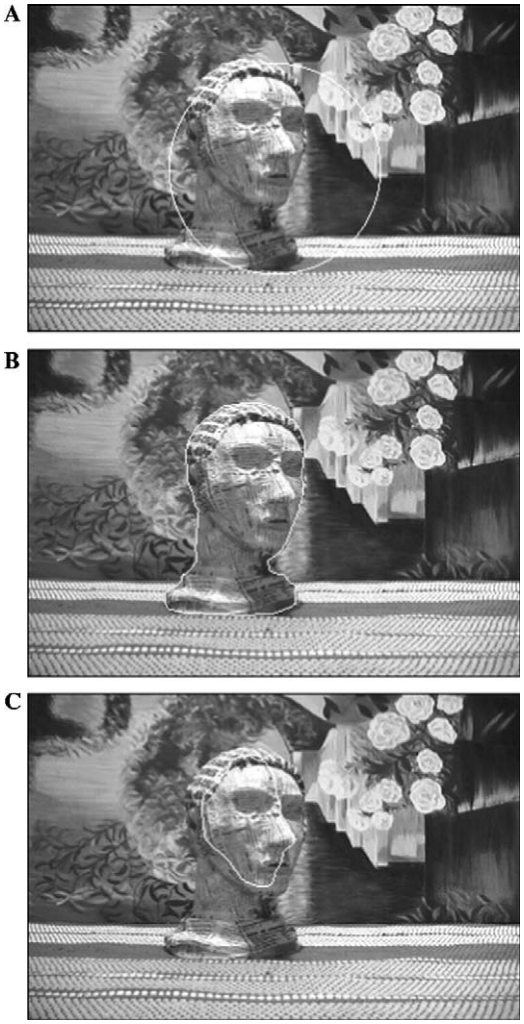

Fig. 3. (A) The first frame of the *Berber* sequence and the initial level set; (B) the computed segmentation; (C) the segmentation without the 3D interpretation term.

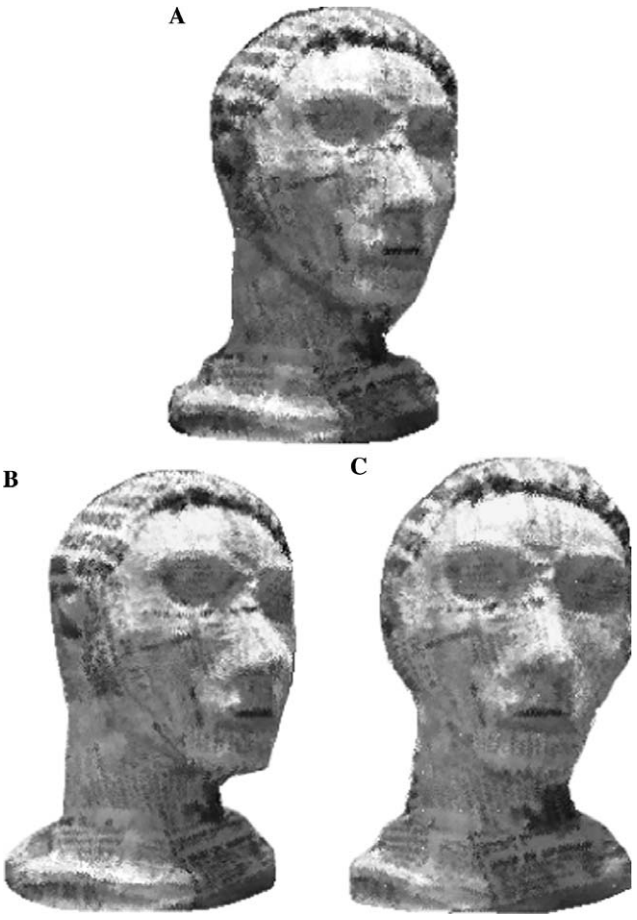

Fig. 4. The reconstructed 3D structure of the moving figurine in the *Berber* sequence from different viewpoints.

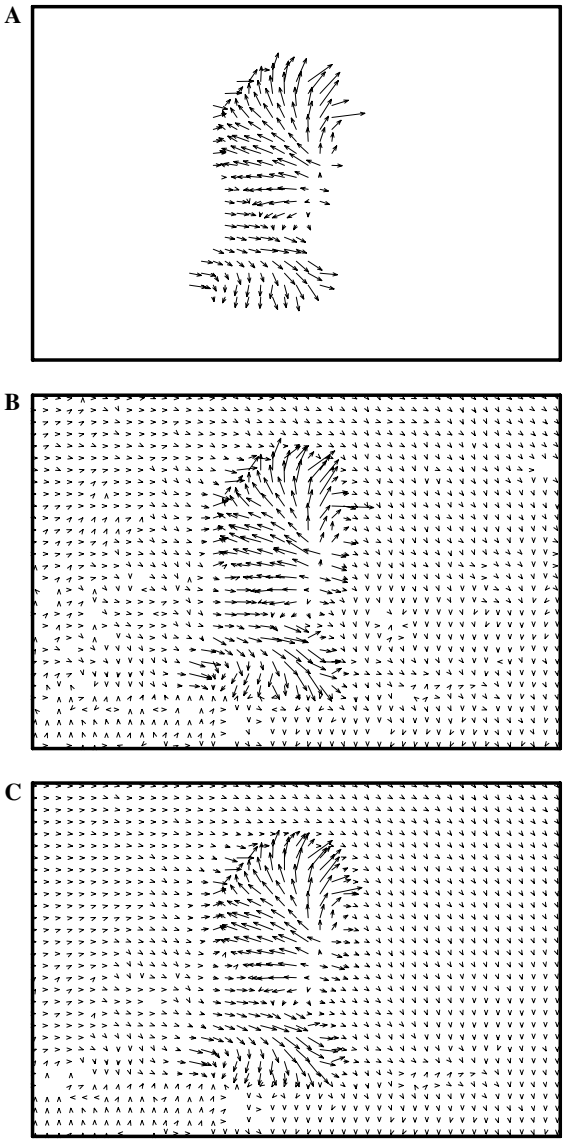

Fig. 5. Optical flow estimated by (A) this method; (B) the Horn-and-Schunck method; (C) the Aubert–Deriche–Kornprobst method.

respectively, isotropic and anisotropic regularization. Fig. 13A displays an anaglyph of a stereoscopic image constructed from the first image and the estimated depth.

To evaluate the contribution of the 3D interpretation term to the estimation of optical flow, we computed the displayed frame difference (DFD) for this method and the methods of Horn and Schunck and Aubert, Deriche, Kornprobst. As can be seen in Table 1, the DFD for this method is the lowest. We also show the contribution of

Table 1
The DFDs for the *Berber* sequence

Method	DFD
Horn-and-Schunck	0.0345
Aubert–Deriche–Kornprobst	0.0278
This method	0.0123

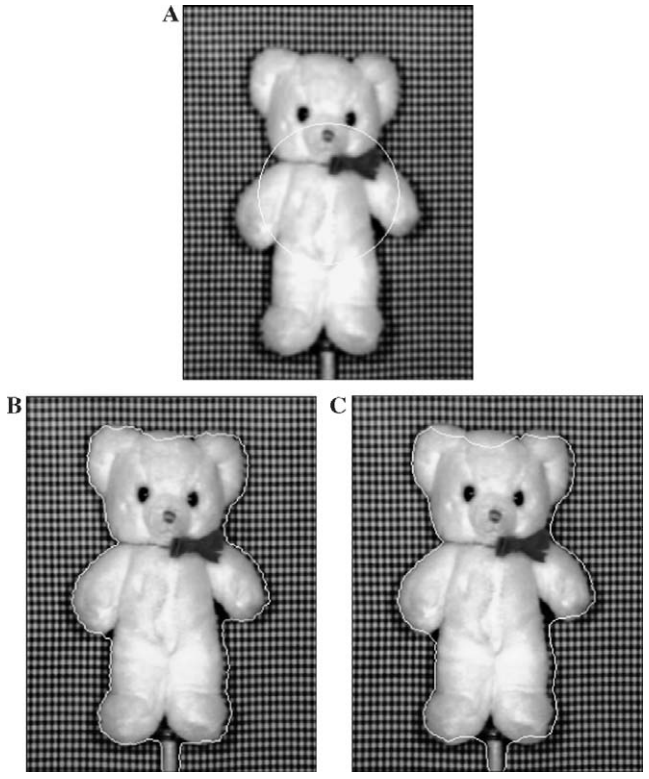

Fig. 6. (A) The first frame of *Teddy1* sequence and the initial level set; (B) the computed segmentation; (C) the segmentation without the 3D interpretation term.

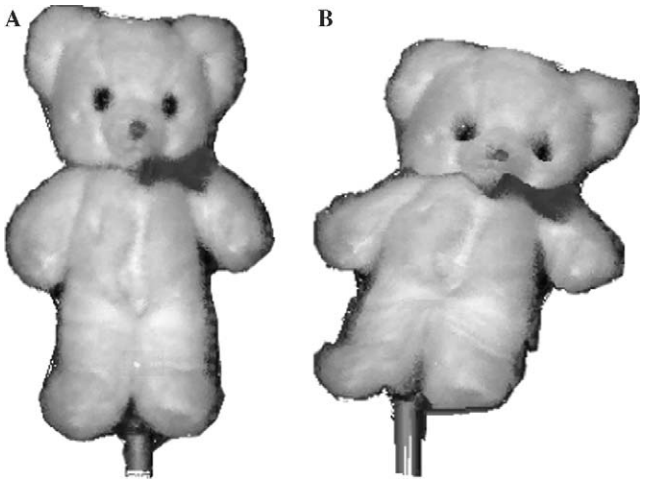

Fig. 7. The reconstructed 3D structure of the moving object from different viewpoints.

Table 2
The DFDs for the *Teddy1* sequence

Method	DFD
Horn-and-Schunck	0.0633
Aubert–Deriche–Kornprobst	0.0428
Method in [25]	0.0195
This method	0.0205

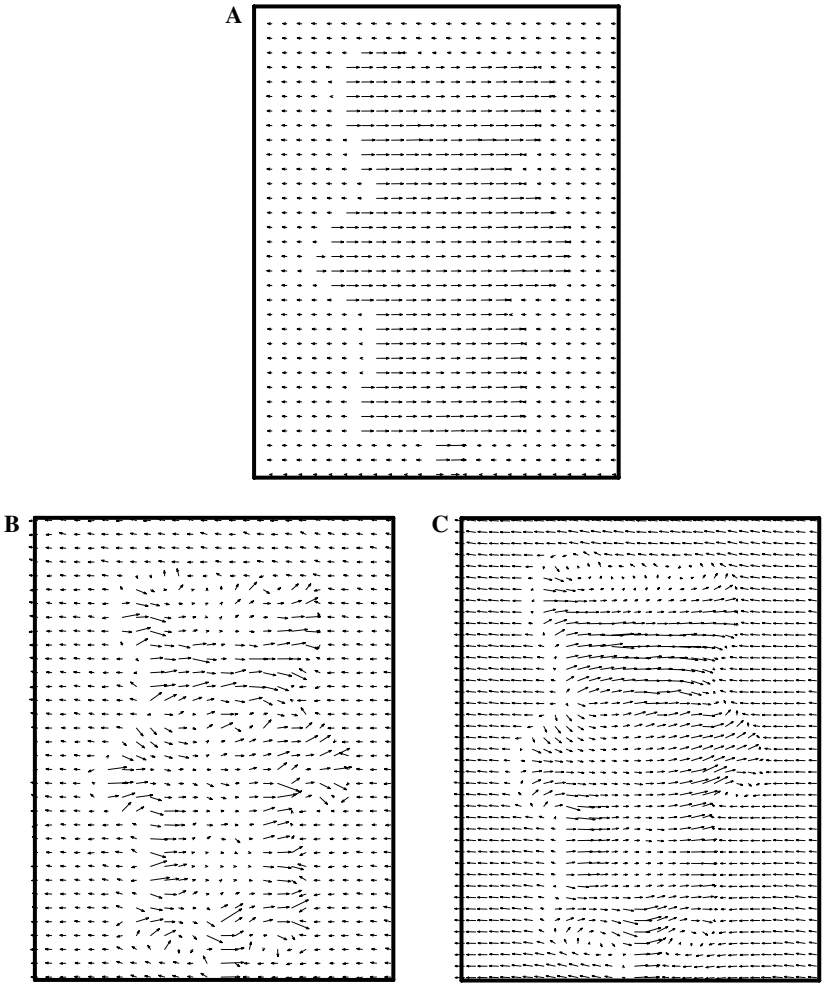

Fig. 8. Optical flow estimated by (A) this method; (B) the Horn-and- Schunck method; (C) the Aubert–Deriche–Kornprobst method.

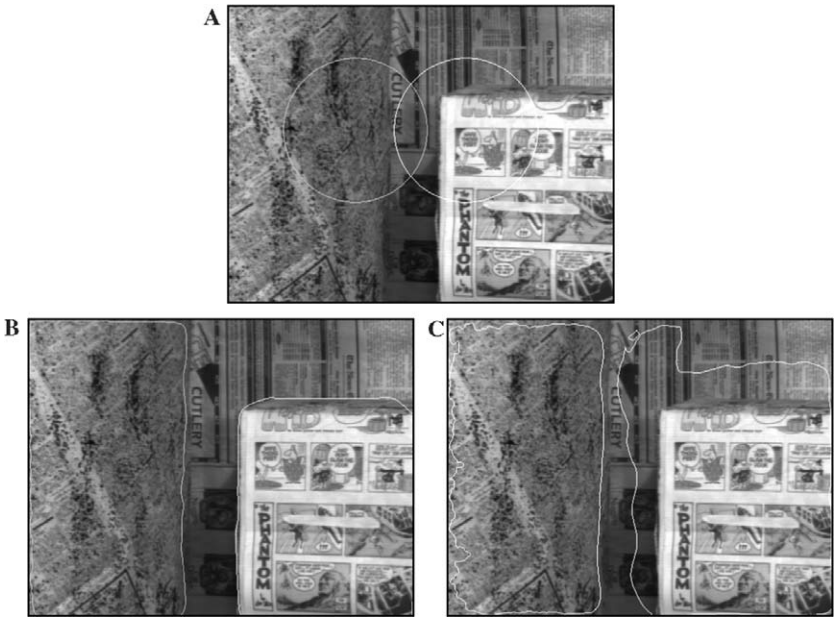

Fig. 9. (A) The first of the two frames used of the *Cylinder* sequence with the initial curves; (B) the computed segmentation; (C) the segmentation without the 3D interpretation term.

the 3D interpretation term to the segmentation. Fig. 3C gives the segmentation obtained without the 3D interpretation term. Although in one region the optical flow is consistent with a single rigid 3D motion, it is due to two different motions in the other (background and part of the object).

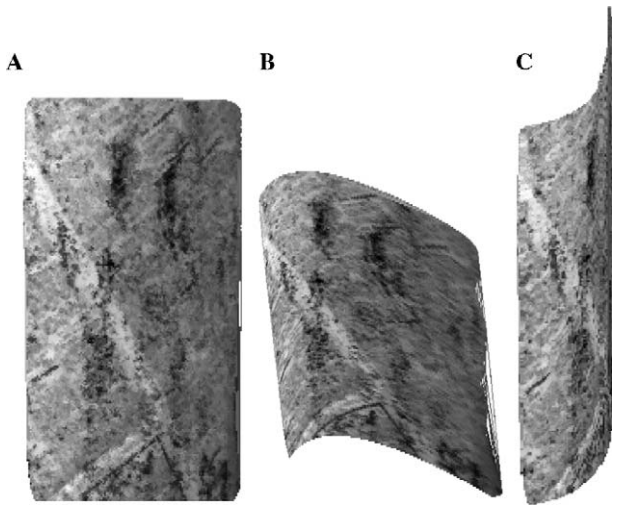

Fig. 10. The reconstructed 3D structure of the moving cylinder from different viewpoints.

By contrast, the 3D term in (5) constrained the optical flow to be consistent with a single rigid 3D motion in each segmentation region Fig. 3B.

In this second example (*Teddy1* sequence, Carnegie Mellon image database), the stuffed animal moves approximately laterally against a background which moves in the opposite direction. In contrast with the sequence of the previous example, this sequence has large areas (on the object) with weak spatio-temporal variations, as well as small areas without texture (eyes and nose tip of the object). The parameters in this experiment are set to $v = 1$, $\mu = 0.5 \times 10^3$ and $\lambda = 10^{-2}$. The initial segmentation curve is shown in Fig. 6A superimposed on the first of the two images used in this example. The final segmentation is shown in Fig. 6B. Views of the reconstructed object surface, back-shaded using the original texture, are shown in Figs. 7A and B. Estimated optical flow, optical flow by the Horn-and-Schunck algorithm and the Aubert–Deriche–Kornprobst algorithm are displayed in Figs. 8A–C, respectively. Fig. 13B displays an anaglyph of a stereoscopic image constructed from the first image and the estimated depth.

The DFDs for this method, the methods of Horn-and-Schunck and Aubert, Deriche, Kornprobst, and the method in [25], are given in Table 2. The DFDs

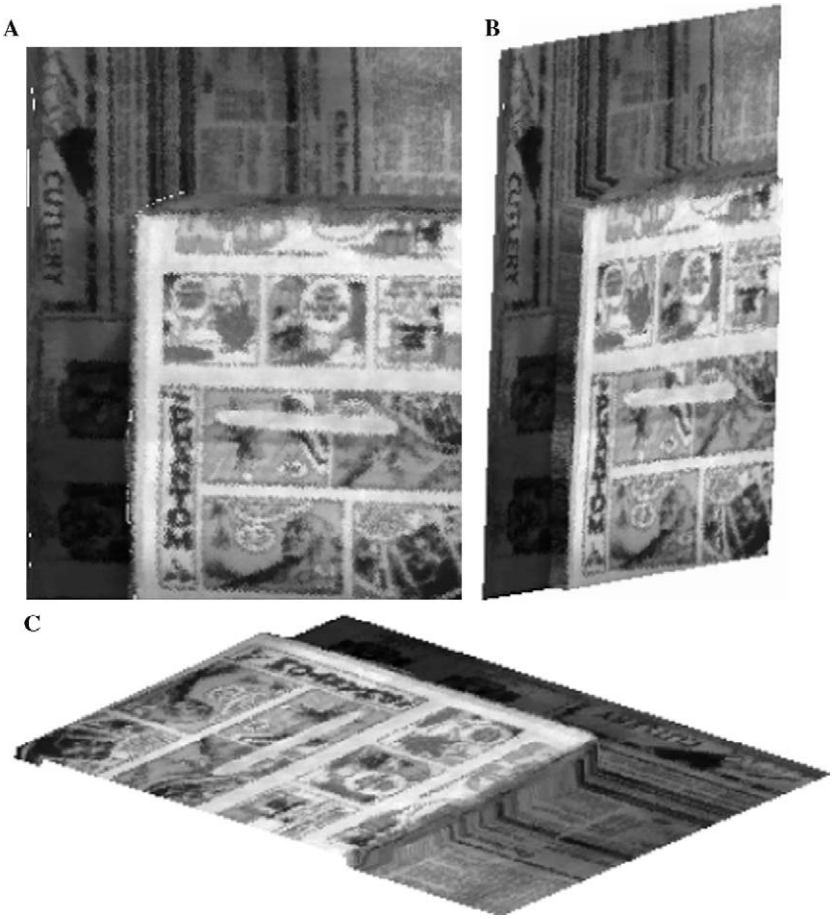

Fig. 11. The reconstructed 3D structure of the moving box shown with the background from different viewpoints.

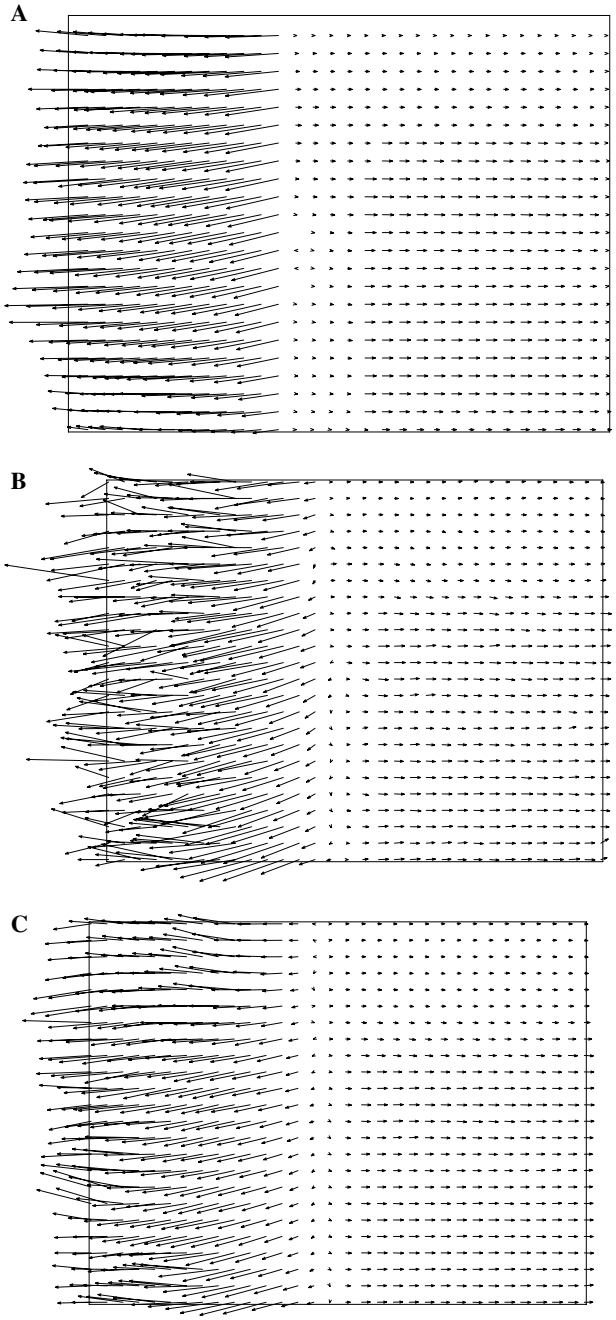

Fig. 12. Optical flow estimated by (A) this method; (B) the Horn-and-Schunck method; (C) the Aubert–Deriche–Kornprobst method.

for this method and the method in [25] are about the same and are lower than the other two. The segmentation obtained without the 3D interpretation term is shown in Fig. 6C. Here, the segmentation is similar to the one obtained with the inclusion of the 3D interpretation term.

This last example uses the *Cylinder* sequence (courtesy of S. Debrunner who constructed it and used it for sparse 3D reconstruction [40]). This is also a sequence of real images where a curved, cylinder-like object rotates about one degree per frame around a nearly vertical axis while translating to the right with an image velocity of about 0.14 pixel per frame. There is also a box moving right at an image rate of about 0.31 pixel per frame in front of a flat backgroung also moving right at 0.14 pixel per frame. The parameter are $\nu = 1$, $\mu = 10^4$ and $\lambda = 10^{-1}$. The two initial segmentation curves are shown in Fig. 9A, superimposed on the first of the two frames used. The challenge, here, is to segment the scene into tree regions because the 3D motion of box and background are very close (they were considered as one motion in [40]). The final segmentation is shown in Fig. 9B. The reconstructed depth of the objects, triangulated and textured using the texture in the original images, is shown from different viewpoints in Figs. 10A–C and 11A–C.

Note that these surfaces are smooth (regularized depth), and that the placement of boundaries is accurate. The estimated image motion by this method, by the Horn-and-Schunck algorithm, and the Aubert–Deriche–Kornprobst method are displayed in Figs. 12A–C, respectively. Fig. 13C displays an anaglyph of a stereoscopic image constructed from the first image and the estimated depth.

The DFDs are shown in Table 3. As in the preceding example, the 3D methods have similar DFDs, lower than the others. The segmentation without the 3D interpretation term is displayed in Fig. 9C. As in the first example, the exclusion of the 3D term from the formulation yields a segmentation where the optical flow estimate is not consistent with a single 3D motion in every region.

7. Conclusion

This paper investigated a level set method for the segmentation and 3D interpretation of the optical flow of multiple rigid objects moving independently in space. Estimation, segmentation, and 3D interpretation were performed jointly. This variational formulation saught simultaneously a segmentation into regions differing by their mean estimated optical flow, and a regularized estimate of optical flow consistent with both the Horn-and-Schunck optical flow constraint and a single rigid body motion in each region of the segmentation. The method, which allows both the viewing system and the viewed objects to move, resulted in three steps interated until convergence: evolution of closed curves via level sets and, in region of segmentation, computation of essential parameters of rigid motion by linear least squares, and estimation of optical flow consistent with a single rigid motion. Following the segmentation, the translational and rotational components of rigid motion and regularized relative depth are recovered analytically for each region of the segmentation from the estimated essential parameters and optical flow. The validity of the method has been verified in several experiments with real image sequences.

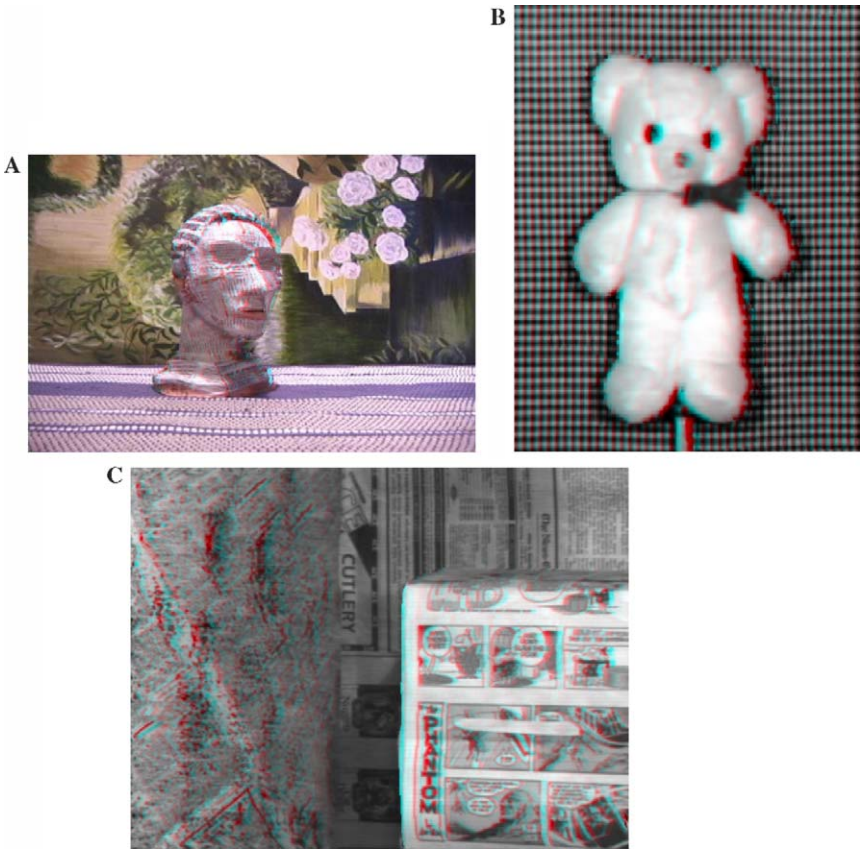

Fig. 13. Anaglyphs (to be viewed with red-blue glasses) of the sequences, (A) *Berber*; (B) *Teddy1*; and (C) *Cylinder*.

Table 3
The DFDs for the *Cylinder* sequence

Method	DFD
Horn-and-Schunck	0.0412
Aubert–Deriche–Kornprobst	0.0325
Method in [25]	0.0225
This method	0.0275

Acknowledgment

This research was supported in part under NSERC Grant OGP0004234.

References

[1] K. Nakayama, J. Loomis, Optical velocity patterns, velocity-sensitive neurons and space perception, Perception (3) (1974) 63–80.
[2] A. Mitiche, Computational Analysis of Visual Motion, Plenum Press, 1994.
[3] J. Aggarwal, N. Nandhakumar, On the computation of motion from a sequence of images: A review, Proc. IEEE 76 (1988) 917–935.
[4] T. Huang, A. Netravali, Motion and structure from feature correspondences: A review, Proc. IEEE 82 (1994) 252–268.
[5] O. Faugeras, Three Dimensional Computer Vision: A Geometric Viewpoint, MIT Press, 1993.
[6] A. Zisserman, R. Hartley, Multiple View Geometry in Computer Vision, Cambridge University Press, Cambridge, 2000.
[7] G. Adiv, Determining three-dimensional motion and structure from optical flow generated by several moving objects, IEEE Trans. Pattern Anal. Mach. Intell. 7 (4) (1985) 384–401.
[8] B. Horn, E. Weldon, Direct methods for recovering motion, Int. J. Comput. Vis. 2 (2) (1988) 51–76.
[9] S. Srinivasan, Extracting structure from optical flow using the fast error search technique, Int. J. Comput. Vis. 37 (3) (2000) 203–230.
[10] Y. Hung, H. Ho, A kalman filter approach to direct depth estimation incorporating surface structure, IEEE Trans. Pattern Anal. Mach. Intell. 21 (6) (1999) 570–575.
[11] B. Shahraray, M. Brown, Robust depth estimation from optical flow, Int. Conf. Comput. Vis. (1988) 641–650.
[12] M. Taalebinezhaad, Direct recovery of motion and shape in the general case by fixation, IEEE Trans. Pattern Anal. Mach. Intell. 14 (8) (1992) 847–853.
[13] E. De Micheli, F. Giachero, Motion and structure from one dimensional optical flow, in: Proc. Internat. Conf. on Computer Vision and Pattern Recognition, 1994, pp. 962–965.
[14] N. Gupta, N. Kanal, 3D motion estimation from motion field, Artif. Intell. (78) (1995) 45–86.
[15] Y. Xiong S. Shafer, Dense structure from a dense optical flow, in: Proc. Internat. Conf. on Computer Vision and Image Understanding, 1998, pp. 222–245.
[16] T. Brodsky, C. Fermuller, Y. Aloimonos, Structure from motion: Beyond the epipolar constraint, Int. J. Comput. Vis. 37 (3) (2000) 231–258.
[17] H. Liu, R. Chellapa, A. Rosenfeld, A hierarchical approach for obtaining structure from two-frame optical flow, in: Proc. Workshop on Motion and Video Computing, 2002.
[18] D. Heeger, A. Jepson, Subspace methods for recovering rigid motion I: algorithm and implementation, Int. J. Comput. Vis. 7 (2) (1992) 95–117.

[19] A. Bruss, B. Horn, Passive navigation, Comput. Graph. Image Process. (1983) 3–20.
[20] W. MacLean, A. Jepson, R. Frecher, Recovery of egomotion and segmentation of independent object motion using the em algorithm, Proc. Br. Mach. Vis. Conf. BMVC 94 (1994) 13–16.
[21] S. Fejes, L. Davis, What can projections of flow fields tell us about visual motion, Internat. Conf. on Computer Vision (1998) 979–986.
[22] J. Weber, J. Malik, Rigid body segmentation and shape description from dense optical flow under weak perspective, IEEE Trans. Pattern Anal. Mach. Intell. 19 (2) (1997) 139–143.
[23] H. Sekkati, A. Mitiche, Dense 3D interpretation of image sequences: a variational approach using anisotropic diffusion, Proc. Internat. Conf. on Image Analysis and Processing, Mantova, Italy, 2003, pp. 424–429.
[24] A. Mitiche, S. Hadjres, MDL estimation of a dense map of relative depth and 3D motion from a temporal sequence of images, Pattern Anal. Appl. (6) (2003) 78–87.
[25] H. Sekkati, A. Mitiche, Concurrent 3D-motion segmentation and 3D interpretation of temporal sequences of monocular images, IEEE Trans. Image Process. (in press).
[26] R. Aubert, G. Deriche, P. Kornprobst, Computing optical flow via variational techniques, SIAM J. Appl. Math. 60 (1) (1999) 156–182.
[27] G. Aubert, P. Kornprobst, Mathematical Problems in Image Processing, Springer, 2002.
[28] X. Zhuang, R. Haralick, Rigid body motion and the optical flow image, Proc. First Int. Conf. on Artificial Intelligence Applications (1984) 366–375.
[29] H. Longuet-Higgins, K. Prazdny, The interpretation of a moving retinal image, Proc. R. Soc. London, B 208 (1981) 385–397.
[30] M. Rousson, R. Deriche, A variational framework for active and adaptive segmentation, Proc. IEEE Workshop Motion and Video Computing (2002) 56–61.
[31] T. Chan, L. Vese, An active contour model without edges, Proc. Int. Conf. Scale-Space Theor. Comput. Vis. (1999) 141–151.
[32] C. Vazquez, A. Mitiche, I.B. Ayed, Image segmentation as regularized clustring: a fully global curve evolution method, Int. Conf. on Image Processing ICIP04 (2004) 3467–3470.
[33] A. Mansouri, A. Mitiche, C. Vazquez, Image partitioning by level set multiregion competition, Int. Conf. on Image Processing, ICIP04 (2004) 2721–2724.
[34] T. Chan, J. Shen, L. Vese, Variational PDE models in image processing, Notices of the AMS 50 (1) (2033) 14–26.
[35] D. Cremers, S. Soatto, Motion competition: a variational approach to piecewise parametric motion segmentation, Int. J. Comput. Vis. 62 (3) (2005) 249–265.
[36] C. Vazquez, A. Mansouri, A. Mitiche, Approximation of images by basis functions for multiple region segmentation with level sets, Int. Conf. on Image Processing, ICIP04 (2004) 549–552.
[37] G. Forsythe, M. Malcom, C. Moler, Computer Methods for Mathematical Computations, Prentice-Hall, 1977.
[38] J. Sethian, Level Set Methods and Fast Marching Methods, Cambridge University Press, Cambridge, 1999.
[39] E. Dubois, A projection method to generate anaglyph stereo images, Proc. ICASP III (2001) 1661–1664.
[40] C. Debrunner, N. Ahuja, Segmentation and factorization-based motion and structure estimation for long image sequences, IEEE Trans. Pattern Anal. Mach. Intell. 20 (2) (1998) 206–211.