JOINT DENSE 3D INTERPRETATION AND MULTIPLE MOTION SEGMENTATION OF TEMPORAL IMAGE SEQUENCES: A VARIATIONAL FRAMEWORK WITH ACTIVE CURVE EVOLUTION AND LEVEL SETS

H. Sekkati and A. Mitiche

INRS-EMT
Place Bonaventure, 800, de la Gauchetiere W.
Suite 6900, Montreal, Quebec H5A 1K6
sekkati{mitiche}@inrs-emt.uquebec.ca

ABSTRACT

The aim of this study is to introduce a novel method for the simultaneous motion segmentation and dense 3D interpretation of temporal sequences of monocular images. The problem is to recover simultaneously 3D structure, 3D motion, and a motion-based segmentation from the image sequence spatio-temporal variations. Motion in space is considered relative to the viewing system so that both the viewing system and environmental objects are allowed to move. The problem is stated as a 3D motion segmentation problem with simultaneous depth estimation within the regions of segmentation. The Euler-Lagrange equations of minimization of the objective functional lead to curve evolution PDEs implemented via level sets.

1. INTRODUCTION

Dense 3D interpretation of temporal image sequences is an important problem in computer vision with numerous useful applications ranging from object modeling to tracking. Most approaches that have adressed this problem are based on the brightness constraint of Horn and Schunck and can be classified in two categories: those based on an accurate computation of optical flow for subsequent 3D interpretation, and those which proceed directly to 3D interpretation without prior estimation of optical flow. Recently, direct methods seem to have gained interest [1, 2, 3] for the promising results they allowed, and for the possibility of stating the problem so as to account directly for 3D interpretation discontinuities. In this study, however, we take motion of viewed objects relative to the viewing system so that both viewing system and viewed object are allowed to move independently. The main curent issues are the recovery of an accurate 3D interpretation, at least where the image spatio-temporal variations are information bearing, and preserving the interpretation discontinuities at motion boundaries. The purpose of this study is to adress these issues and investigate a novel method for joint motion segmentation and dense 3D interpretation. We state the problem as a 3D motion segmentation problem with simultaneous depth estimation within the region of segmentation. The Euler-Lagrange equations of minimization of the objective functional lead to curve evolution PDEs implemented via level sets. Recovery of dense 3D structure from *stereoscopic images*, albeit under simplifying conditions such as textured object surfaces and non-intervening background, has been stated, recently, as parametric surface evolution by level set PDEs [4], but there are significant underlying differences that make the problem more complex with *temporal image sequences* which we deal with in our study. For instance, in contrast with stereoscopic images, there is no epipolar geometry to exploit with temporal sequences, viewed objects can move as well as the viewing system, and images are acquired continually at regular intervals of time.

The novelty of our formulation has several aspects: 1) motion segmentation is performed simultaneously to 3D interpretation within the regions of segmentation, and this for better interpretation accurracy and to preserve the discontinuities of interpretation at motion boundaries, 2) curve evolution for boundary placement accuracy and level sets for numerical stability, and 3) a major common constraint is relaxed by allowing both movement of the viewing system and viewed objects.

2. BASIC MODELS

Let I be an image sequence function defined over a domain $D = \Omega \times [0, +\infty[$, where Ω is an open subset of $\mathbf{R}$. I is thus a map $I(\mathbf{x}, t) : \quad D \longmapsto R$. Assume that there are $(N-1)$ moving objects on a background. As we allow the camera to move so the background has its own motion. Then the image can be partitioned into N regions $\{\Omega_k\}_{k=1}^{N} \in \Omega$ and each region Ω_k is caracterized by its 3D motion parameters $(\mathbf{T}_k, \omega_k)$ and its depth $\{Z(\mathbf{x}), \mathbf{x} \in \Omega_k\}$. Our aim is to determine each region and its parameters.

2.1. The brightness constraint for rigid objects

Using the Horn and Schunk constraint and the expression of image motion velocity in terms of depth and the parameters of 3D rigid motion [5], we have, for each point in region Ω_k:

$$I_t + \mathbf{s} \cdot \frac{\mathbf{T}_k}{Z} + \mathbf{q} \cdot \omega_k = 0 \qquad \text{, k=1, } \ldots \text{ ,N} \tag{1}$$

where $\mathbf{s}$ and $\mathbf{q}$ are 3D vectors given by:

$$\mathbf{s} = \begin{pmatrix} fI_x \\ fI_y \\ -xI_x - yI_y \end{pmatrix}, \quad \mathbf{q} = \begin{pmatrix} -fI_y - \frac{y}{f}(xI_x + yI_y) \\ fI_x + \frac{x}{f}(xI_x + yI_y) \\ -yI_x + xI_y \end{pmatrix}$$

I_x, I_y and I_t are the spatio-temporal derivatives of I, and f is the focal length. We note here that the camera is modeled using perspective projection. The left-hand side of (1) will serve the purpose of expressing a segmentation conformity to data in the objective functional to be discussed in the next section.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

2.2. The objective functional

We seek to partition Ω into N regions $\{\Omega_k\}_{k=1}^N$ by considering the evolution of $N-1$ closed parametric curves $\{\gamma_k(s)\}_{k=1}^{N-1}$, each delimiting a region, and simultanously estimate the parameters $(\mathbf{T}_k, \omega_k, Z)$ for each region. Ω_k corresponds to the interior of $\gamma_k, k = 1, \ldots, N-1$ and Ω_N is defined to be the complement of the union of regions Ω_k, $k = 1, \cdots, N-1$. The problem can be stated as the minimization of the following energy functional, which can be shown to derive from MAP estimation:

$$E[\{\gamma_k\}_{k=1}^{N-1}, \{\mathbf{T}_k\}_{k=1}^N, \{\omega_k\}_{k=1}^N, Z] =$$

$$\sum_{k=1}^N \left[\int_{\Omega_k} \psi_k^2(\mathbf{x})d\mathbf{x} + \mu \int_{\Omega_k} \|\nabla Z\|^2\right] + \lambda \sum_{k=1}^{N-1} \oint_{\gamma_k} ds \tag{2}$$

where $\psi_k(\mathbf{x}) = I_t + \mathbf{s} \cdot \frac{\mathbf{T}_k}{Z} + \mathbf{q} \cdot \omega_k$, and μ, λ are positive constants to weigh the contribution of each term in (2). The first term in (2) measures conformity to data in Ω_k through the brightness constraint. The other two integrals are regularization terms, one of smoothness of depth in Ω_k, the other of smoothness of the boundary of Ω_k. It is interesting to note that energy (2) takes the form of the generalized Mumford-Shah [6] multiphase piecewise-smooth functional relatively to depth, and of the generalized Mumford-Shah multiphase piecewise-constant segmentation relatively to 3D motion. Minimization of (2) will seek a solution biased toward regions of segmentation that have smooth boundaries, with conformity to data and smooth depth within each region. Note that no smoothing of depth is done accross region boudaries, which is important to preserve motion boudaries. The optimization problem is then to minimize the functional (2) simultanously with respect to γ_k, $k = 1, \cdots, N-1$, to 3D motion parameters, and to depth.

3. ENERGY MINIMIZATION

To minimize (2), we adopt a greedy algorithm which, after initialization, consists of three iterated steps. At each step, we fix two of the sets of variables of $\{\gamma_k\}_{k=1}^{N-1}$, depth, and motion parameters, and solve for the remaining one so as to decrease the objective functional.

3.1. Step 1. 3D motion estimation by least squares minimization

With $\{\gamma_k\}_{k=1}^{N-1}$ and depth fixed, the energy to minimize is:

$$E(\{\mathbf{T}_k\}_{k=1}^N, \{\omega_k\}_{k=1}^N) = \sum_{k=1}^N \int_{\Omega_k} \psi_k^2(\mathbf{x})d\mathbf{x} \tag{3}$$

Minimizing (3) is equivalent to solve constraint (1) for each region Ω_k. In the discrete case of digital images, this amounts to a least squares minimization. Indeed, let N_k be the number of points of the image positional array within region Ω_k, and let $\mathbf{b}$ be the 1×6 vector:

$$\mathbf{b} = (\frac{s_1}{Z}, \frac{s_2}{Z}, \frac{s_3}{Z}, q_1, q_2, q_3)$$

We write Equation (3) for each point $\mathbf{x}_i$ in the region Ω_k to oblain the linear system,

$$\mathbf{B}_k \, \rho_k = \mathbf{c}_k \quad k = 1, \ldots, N \tag{4}$$

where $\rho_k = (\mathbf{T}_k, \omega_k)$ is the 6×1 vector representing the 6D rigid motion components of region Ω_k (3 for translation and 3 for rotation). The $N_k \times 6$ matrix $\mathbf{B}_k$ and $N_k \times 1$ vector $\mathbf{c}_k$ are defined respectively as follow,

$$\mathbf{B}_k = \begin{pmatrix} \mathbf{b}(\mathbf{x}_1) \\ \vdots \\ \mathbf{b}(\mathbf{x}_{N_k}) \end{pmatrix} \quad \mathbf{c}_k = \begin{pmatrix} -I_t(\mathbf{x}_1) \\ \vdots \\ -I_t(\mathbf{x}_{N_k}) \end{pmatrix}$$

The overdetermined linear system (4) can be solved by singular value decomposition of the matrix $\mathbf{B}_k$. The 6D motion vector ρ_k is updated by the least square solution vector of this system. As known, the 3D translational componant of motion is determined only to a scale factor. Therefor, we fix this scale by imposing the constraint

$$\|\mathbf{T}_k\| = constant, \qquad k = 1, \cdots, N$$

3.2. Step 2. Depth estimation by gradient descent

In the second stage, we fix the 3D motion parameters and $\{\gamma_k\}_{k=1}^{N-1}$ and solve for depth. The functional to minimize for recovering depth is:

$$E(Z) = \sum_{k=1}^N \left[\int_{\Omega_k} \psi_k^2(\mathbf{x})d\mathbf{x} + \mu \int_{\Omega_k} \|\nabla Z\|^2\right] \tag{5}$$

The Euler-lagrange equations lead to the following system of partial differential equations (PDEs) that can be solved independently for each region Ω_k:

$$\begin{cases} \frac{\mathbf{s}\cdot\mathbf{T}_k}{Z^2}\psi_k(\mathbf{x}) + \mu\triangle Z = 0 \quad \forall \mathbf{x} \in \Omega_k \\ \\ \frac{\partial Z}{\partial \mathbf{n}_k} = 0 \quad \forall \mathbf{x} \in \gamma_k \end{cases} \tag{6}$$

where $\triangle$ is the Laplacien operator and $\mathbf{n}_k$ is the exterior unit normal vector to the curve γ_k. Equations (6) are resolved by gradient descent:

$$\frac{dZ(\mathbf{x},t)}{dt} = \frac{\mathbf{s}\cdot\mathbf{T}_k}{Z^2}\psi_k(\mathbf{x}) + \mu\triangle Z \tag{7}$$

3.3. Step3. Motion segmentation by multiregion competition: curve evolution via level sets

For multiple regions, the minimization of a functional such as (2) with respect to $\{\gamma_k\}_{k=1}^{N-1}$ can be done in different ways [7, 8, 9, 10]. Here, we adopt the method in [8] where the energy to minimize takes the following form:

$$E(\{\gamma_k\}_{k=1}^{N-1}) = \sum_{k=1}^N \int_{\Omega_k} \xi_k(\mathbf{x})d\mathbf{x} + \lambda \sum_{k=1}^{N-1} \oint_{\gamma_k} ds \tag{8}$$

where $\xi_k(\mathbf{x}) = \psi_k^2(\mathbf{x}) + \mu\|\nabla Z\|^2\chi_k(\mathbf{x})$, with $\chi_k(\mathbf{x})$ is the characteristic of region Ω_k. Let γ_k be parametrised by $\gamma_k(s,t) : [0,1] \times [0, +\infty[\mapsto \mathbf{R}$. The Euler-Lagrange descent equations corresponding to the minimization of (8) are given by:

$$\frac{d\gamma_k(\mathbf{x})}{dt} = -\left(\xi_k(\mathbf{x}) - \varphi_k(\mathbf{x}) + \lambda\kappa_k(\mathbf{x})\right)\mathbf{n}_k(\mathbf{x}) \tag{9}$$

where $\kappa(\mathbf{x})$ is the mean curvature of $\gamma_k(\mathbf{x})$ and the functions $\varphi_k(\mathbf{x})$ are defined by

$$\varphi_k(\mathbf{x}) = \min_{i \neq k} \xi_i(\mathbf{x})$$

For a stable numerical implementation of (9), we use the level set formulation. We represent each curve γ_k implicitly as the zero level set of a function $\Phi_k : \Omega \times \mathbf{R}^+ \mapsto \mathbf{R}$. We adopt the convention that the inside of γ_k will correspond to the set $\{\Phi_k > 0\}$, and obtain the following system of coupled partial differential equations [8, 11]:

$$\frac{d\Phi_k(\mathbf{x})}{dt} = -\left(\xi_k(\mathbf{x}) - \varphi_k(\mathbf{x}) + \lambda H_{\Phi_k}(\mathbf{x})\right) \|\nabla \Phi_k(\mathbf{x})\| \qquad k = 1, \cdots, N-1 \quad (10)$$

where the mean curvature H_{Φ_k} is given by $H_{\Phi_k} = div(\frac{\nabla \Phi_k}{\|\nabla \Phi_k\|})$ and $\varphi_k(\mathbf{x})$ by

$$\varphi_k(\mathbf{x}) = \begin{cases} \min\limits_{\substack{i \neq k \\ \Phi_x(\mathbf{x}) > 0}} \xi_i(\mathbf{x}) & \text{if } \exists k \in \{1, \cdots, N-1\} \\ \xi_N(\mathbf{x}) & \text{else} \end{cases}$$

4. RECOVERING MOTION BOUNDARY PRESERVING OPTICAL FLOW

The relation between optical flow, depth, and the parameter of rigid motion is given by [5]:

$$\begin{cases} u &= \frac{1}{Z}(f\mathbf{T}_{k_1} - x\mathbf{T}_{k_3}) - \frac{xy}{f}\omega_{k_1} + \frac{f^2+x^2}{f}\omega_{k_2} - y\omega_{k_3} \\ v &= \frac{1}{Z}(f\mathbf{T}_{k_2} - y\mathbf{T}_{k_3}) - \frac{f^2+y^2}{f}\omega_{k_1} + \frac{xy}{f}\omega_{k_2} + x\omega_{k_3} \end{cases} \quad (11)$$

Solving equations (11) for each region and knowing its boudaries allows an accurate recovery of optical flow.

5. EXPERIMENTAL RESULTS

In this section, we report on experiments that verify the validity of the proposed formulation and of its implementation. Figure 1(a) shows an image of a two-frame real image sequence with synthetic motion which consists of a moving object (square annulus) on a moving background. The image motion of the square annulus is $(-1, -1)$. The background moves verticaly by $(0, 1)$. We initialised depth to the value of 100 in all the image and the level set as shown in the first frame of the sequence (Figure 1(a)). The final segmentation is shown in the Figure 1(b). The segmentation is very accurate. Figure 1(c) shows the recovered depth map. The estimated 3D motion for the object and background are, respectively $\mathbf{T}_1 = (-7.13, -7.02, -0.0092)$ and $\mathbf{T}_2 = (0.057, 9.99, 0.010)$. The estimated means of depths in both of regions are $Z_1 = 73.5$ and $Z_2 = 99.3$ having a ratio of $\frac{Z_2}{Z_1} = 1.33$. One can show that the theoretical ratio is $\sqrt{2}$. Figure 1(d) shows the recovered optical flow, which is accurate with motion boundaries preserved.

In this second example, we use two consecutive frames from a sequence of real images with synthetic motion consisting of two moving squares on a moving background. The image motions of the upper and lower squares, and of the background are, respectively, equal to $(1, -1)$, $(-1, 2)$ and $(0, 1)$. We initialize the two level sets as shown in Figure 2(a). Recovered depth is shown in Figure 2(c). The estimated translations are $\mathbf{T}_1 = (7.13, -7.02, 0.003)$, $\mathbf{T}_2 = (-4.51, 8.01, -0.01)$, $\mathbf{T}_3 = (0.21, 9.99, 0.014)$ and the respective estimated mean depths are $Z_1 = 69.01$, $Z_2 = 46.01$ and $Z_3 = 99.05$. The ratios between the objects and the background depths are nearly equal to the theoretical ratio ($\frac{Z_3}{Z_1} = \sqrt{2}$ and $\frac{Z_3}{Z_2} = \sqrt{5}$). The recovered optical flow is shown in figure 2(d). Both the magnitude and directions of the flow are faithfuly recovered.

6. CONCLUSION AND FUTURE WORK

We presented a novel method to simultanously segment multiple motions and estimate 3D motion and depth in a temporal sequence of images where the viewing system so that both the viewing system and environmental objects are allowed to move. The problem was stated as a 3D motion segmentation problem with simultaneous depth estimation within the regions of segmentation. The Euler-Lagrange equations of minimization of the objective functional led to curve evolution PDEs implemented via level sets. We are currently validating the formulation on various real image sequences with real motions. We are also investigating other priors for depth, to account for objects that present sharp changes in depth.

7. REFERENCES

[1] A. Mitiche and S. Hadjres, "Mdl estimation of a dense map of depth and 3d motion from a temporal sequence of images," *Pattern Analysis and Applications*, 2003 (to appear).

[2] H. Sekkati and A. Mitiche, "Dense 3d interpretation of image sequences: A variational approach using anisotropic diffusion," International Conference On Image Analysis and Processing, 2003, Mantova, Italy.

[3] S. Stein and M. Shashua, "Model-based brightness constraints: on direct estimation of structure and motion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 9, pp. 993–1005, 2000.

[4] O. Faugeras and R. Keriven, "Variational principles, surface evolution, pde's, level set methods, and the stereo problem," *IEEE Trans. Image Processing*, vol. 7, no. 3, pp. 336–344, 1998.

[5] A. Mitiche, *Computational Analysis of Visual Motion*, Newyrok, Plenum Press, 1994.

[6] D. Mumford and J. Shah, "Optimal approximation by piecewise smooth functions and associated variational problems," *Comm. Pure Applied. Math.*, , no. 42, pp. 577–685, 1989.

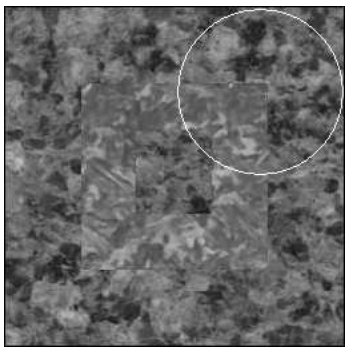
[7] T. Chan and L. Vese, "An active contour model without edges," In Proc. Int. Conf. Scale-Space Theories in Computer Vision, 1999, pp. 141–151, Corfu, Greece.

[8] A.R. Mansouri and J. Konrad, "Multiple motion segmentation with level sets," *IEEE Trans. Image Processing*, vol. 12, no. 2, pp. 201–220, 2003.

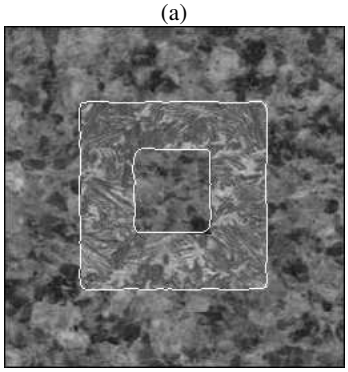
[9] A. Mansouri, A. Mitiche, and C. Vazquaz, "Image segmentation by multiregion competition," Reconnaissance de Formes et Intelligence Artificielle Conference, RFIA-04, 2004, Toulouse, France.

[10] C. Vazquez, A. Mitiche, and I. Ben Ayed, "Segmentation of vectorial images by a global curve evolution method," Reconnaissance de Formes et Intelligence Artificielle Conference, RFIA-04, 2004, Toulouse, France.

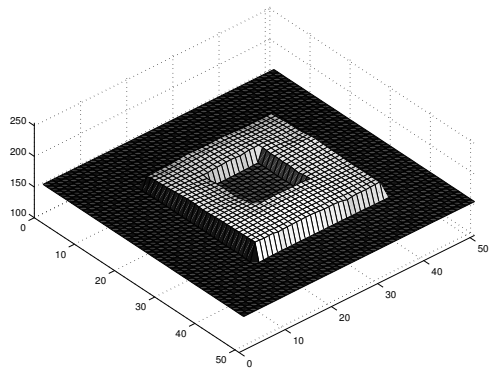
[11] J. Sethian, *Level set methos and fast marching methods*, Cambridge University Press, 1999.



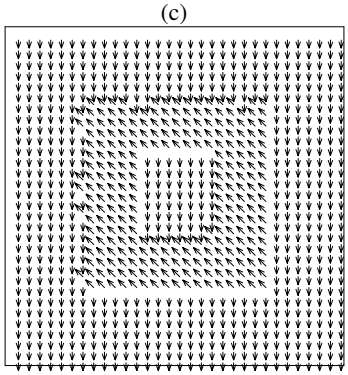
(a)



(b)

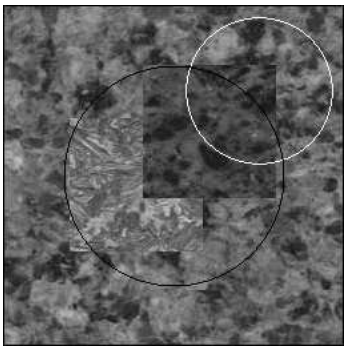


(c)

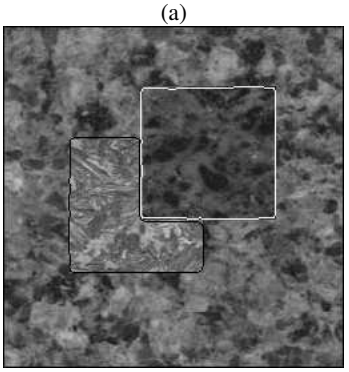


(d)

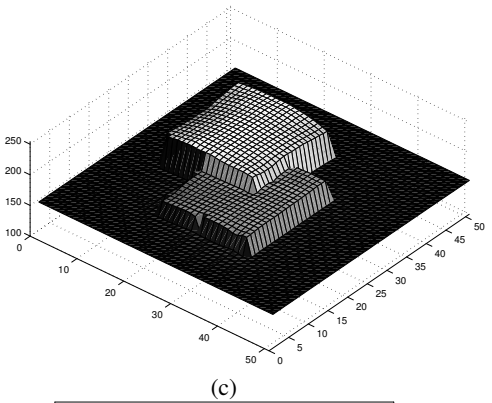
Fig. 1. (a) An image of the square-bande sequence with level set initialisation (b) The resultin motion segmentation (c) The 3D estimated depth mape. (d) Optical flow computed using the estimated depth map and 3D motion.



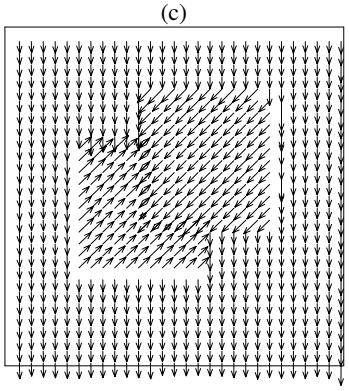
(a)



(b)



(c)



(d)

Fig. 2. (a) An image of two moving squares on moving background with level set initialisation (b) The resulting motion segmentation (c) The 3D estimated depth mape. (d) Optical flow computed using the estimated depth map and 3D motion.