A Level Set Method for Joint 3-D Motion Estimation and Segmentation in Range Image Sequences

Abderrahman Hmimia Amar Mitiche

hmimia@emt.inrs.ca mitiche@emt.inrs.ca

Institut National de la Recherche Scientifique, INRS-EMT
Place Bonaventure, 800, de la Gauchetière ouest, Suite 6900, Montreal, Quebec, H5A 1K6 Canada

Abstract

The purpose of this study is to investigate a variational method for joint 3-D motion segmentation and estimation in a sequence of range images of a scene containing moving rigid objects viewed by a possibly moving camera. The objective functional contains two terms. One term measures the conformity of the interpretation within each region of 3-D motion segmentation to the range images sequence, and the other term is for the regularity of segmentation boundaries. Minimization of the functional result in concurrent 3-D motion segmentation by curve evolution, and 3-D motion by least squares within each region of segmentation. Curve evolution is implemented via level sets for topology independence and numerical stability. Several sequences of range image are provided which verify the method and its implementation.

1 INTRODUCTION

Motion analysis remains a challenging and fundamental problem of computer vision, and plays several important roles such as estimation of motion, image segmentation, movement detection and tracking. Within the general context of motion analysis, 3D motion segmentation and estimation is of considerable interest for its methodological challenges and practical applications. It serves in domains such as medicine, robotics and surveillance.

Intuitively, it seems obvious that 3-D motion segmentation and estimation will be easier if we are given all the three coordinates of the position vectors of image points (as in range image) rather than two coordinates (as in intensity images). However, because of the unreliability associated with range image motion segmentation, the most widely used form of image in motion segmentation so far has been the intensity image.

In the four last decades intensity images have gotten more attention than other kinds of images, particularly due to the good quality of intensity sensors. However, the interpretation of intensity images is very difficult, because they offer no explicit tridimensional (3-D) information. In order to overcome this problem several techniques were developed: stereo-vision, shape from shading, etc....

In parallel, a lot of different sensors have been greatly enhanced, especially range sensors [1], [2], [3], [4], [5], [6], the SR3000 camera and the new camera SR4000 which represent the 4th generation of time-of-flight cameras designed by MESA in September 2008 [7].

Recent advances in range imaging technology make it realistic and necessary to address the problem of 3-D motion segmentation and estimation from a sequence of range images.

The objective of this paper is to focus on the important computer vision task of 3-D motion segmentation and estimation in a range image sequence of a scene containing moving rigid objects viewed by a possibly moving camera. The problem is to segment the range image domain into regions corresponding to differently moving objects in space (3-D segmentation) and determine the motion of each object. For the problem we are addressing, we will show that a variational formulation can yield an efficient algorithm by encoding 3-D motion segmentation and estimation in a single objective functional which embeds segmentation by active curves.

2 BASIC MODELS

We will write the range flow motion constraint equation, to be used to gauge conformity of 3-D motion field to image spatiotemporal data.

2.1 Range flow motion constraint equation

The depth can be viewed as a function of space and time $Z = Z(X,Y,\tau)$. Taking a full time derivative of Z via the chain rule, we obtain the elevation rate constraint equation [8] also called range flow motion constraint equation RFMC [9],[10]:

$$\frac{dZ}{d\tau} = \frac{\partial Z}{\partial X}\frac{dX}{d\tau} + \frac{\partial Z}{\partial Y}\frac{dY}{d\tau} + \frac{\partial Z}{\partial \tau} \tag{1}$$

This can be written in the less intimidating form

$$Z_X\dot{X} + Z_Y\dot{Y} - \dot{Z} + Z_\tau = 0 \tag{2}$$

Where Z_X, Z_Y, Z_τ are the partial derivative of Z with respect to X, Y, τ respectively, and $\dot{X}$, $\dot{Y}$, $\dot{Z}$ are the derivative with respect to the time of X, Y, and Z respectively. let $\dot{X} = U$, $\dot{Y} = V$, and $\dot{Z} = W$, then the 3D velocity vector $\vec{f}$ can be written as $\vec{f} = [U\ V\ W]^T$. The range flow motion constraint equation RFMC become

$$Z_X\ U\ +\ Z_Y\ V\ -\ W\ +\ Z_\tau\ \ =\ \ 0 \tag{3}$$

In order to evaluate the RFMC equation we need to compute partial derivatives of the depth function with respect to the world coordinates. We use the approach that uses derivative kernels presented in [11].

$$Z_X = \frac{Y_y Z_x - Y_x Z_y}{Y_y X_x - X_y Y_x} = \frac{Z_x}{X_x} \tag{4}$$

$$Z_Y = \frac{X_x Z_y - X_y Z_x}{Y_y X_x - X_y Y_x} = \frac{Z_y}{Y_y} \tag{5}$$

Let $\mathbf{T} = (t_1, t_2, t_3)$ be the translational component and $\omega = (\omega_1, \omega_2, \omega_3)$ the rotational component of the motion of a regid body B in space. Let $\mathbf{R} = (X, Y, Z)^T \in B$ be a vector to a point on the surface, the velocity of $\mathbf{R}$ is

$$\frac{\mathbf{dR}}{d\tau} = T + \omega \times \mathbf{R} \tag{6}$$

which can be rewritten as

$$\left\{ \begin{array}{lcl} U & = & t_1 + \omega_2 Z - \omega_3 Y \\ V & = & t_2 + \omega_3 X - \omega_1 Z \\ W & = & t_3 + \omega_1 Y - \omega_2 X \end{array} \right. \tag{7}$$

Substituting these equation into the RFMC itself yields

$$Z_X t_1 + Z_Y t_2 - t_3 + r\omega_1 + s\omega_2 + l\omega_3 + Z_\tau = 0 \tag{8}$$

Where $r = -Z_Y Z - Y$, $s = Z_X Z + X$ and $l = Z_Y X - Z_X Y$ which can be rewritten as

$$Z_\tau + A \cdot T + B \cdot \omega = 0 \tag{9}$$

where A and B are 3-D vectors given by

$$A = \begin{pmatrix} Z_X \\ Z_Y \\ -1 \end{pmatrix}, \quad B = \begin{pmatrix} r \\ s \\ l \end{pmatrix} \tag{10}$$

We will use this equation in the objective functional (Next section) to express conformity of a 3-D motion field to image spatiotemporal data.

2.2 Objective functional

Let $Ir : \Omega \times]0, \tau[\to \mathbb{R}$ be an range image sequence function, acquired by a possibly moving viewing system. Where Ω is an open subset of $\mathbb{R}^2$ and τ is the duration of the sequence. Let γ be a closed planar curve parameterized by arc parameter $s \in [0,1]$, Ω_1 the region enclosed by γ, and $\Omega_2 = \Omega_1^C$. Each region Ω_k is characterized by its 3-D motion parameters (T_k, ω_k). The problem returns to segment each region and determine its 3-D motion parameters. The problem can be stated as the minimization of the following energy functional:

$$\begin{array}{l} E\left(\gamma, \{T_k\}_{k=1}^2, \{\omega_k\}_{k=1}^2\right) \\ = \sum_{k=1}^{2} \int_{\Omega_k} \psi_k^2(X,Y) dX dY + \lambda \oint_\gamma dS \end{array} \tag{11}$$

Where

$$\psi_k(X,Y) = Z_\tau + A.T_k + B.\omega_k \tag{12}$$

And λ is positive real constant to weigh the contribution of the terms in (11). The first integral in (11) measures conformity of the 3-D motion field to the sequence spatiotemporal variations in Ω_k via range flow motion constraint equation. The second term is regularization term of smoothness of γ. Our problem returns to minimize the functional (11) simultaneously with respect to γ, and to 3-D motion parameters.

3 FORMULATION AND ALGORITHM FOR TWO REGIONS

Since the objective functional (11) depends on two groups of variables, namely, the segmentation boundaries γ, and 3-D rigid motion parameters $T_1, T_2, \omega_1, \omega_2$, we will adopt a greedy algorithm which, after initialization, consists of two iterated steps. At each step, we fix one of the two groups of variables and solve for the other group of variables.

3.1 Initialization

An initial set of one curve provide the two-region initial segmentation.

3.2 Step1. 3-D Motion estimation by least-squares

With γ fixed, the energy to minimize is

$$\mathrm{E}\left(\{T_k\}_{k=1}^2, \{\omega_k\}_{k=1}^2\right) = \sum_{k=1}^{2} \int_{\Omega_k} \psi_k^2(X,Y) dX dY \tag{13}$$

Since ψ_k depends linearly on T_k and ω_k, $k \in \{1,2\}$, the minimization of (13) reduces to the linear least-squares estimation of the parameters within each region. In the discrete case of digital images, this estimation is done as follows. Let N_k be the number of points of the image positional array within region Ω_k, $k \in \{1,2\}$, and let m be the 1×6 vector associated to each point of the image. referred to equation (9 and 10) we write m as

$$m(z_i) = (Z_X, Z_Y, -1, r, s, l)_{z_i}. \tag{14}$$

The index z_i under the right bracket means that all vector elements are evaluated at point z_i. We write (9) for each point z_i in the region Ω_k to obtain the linear system

$$M_k \rho_k = n_k \quad k = 1,2 \tag{15}$$

where $\rho_k = (T_k, \omega_k)^T$ is the 6×1 vector representing the six-dimensional rigid motion components of region Ω_k (three for translation and three for rotation). The $N_k \times 6$ matrix M_k and the $N_k \times 1$ vector n_k are defined, respectively, as follows:

$$M_k = \begin{pmatrix} m(z_1) \\ \vdots \\ m(z_{N_k}) \end{pmatrix} \quad n_k = \begin{pmatrix} -Z_\tau(z_1) \\ \vdots \\ -Z_\tau(z_{N_k}) \end{pmatrix} k = 1,2 \tag{16}$$

An effective method to solve the overdetermined linear system (15) is a singular value decomposition of matrix M_k. And ρ_k is updated by the least-squares solution vector of this system.

3.3 Step2. Curve evolution by level sets for 3-D motion segmentation

With the 3-D rigid motion parameters $(T_k, \omega_k)_{k=1}^2$ fixed, the energy to minimize with respect to the curve γ is

$$\mathrm{E}(\gamma) = \sum_{k=1}^{2} \int_{\Omega_k} \xi_k(X,Y) dX dY + \lambda \oint_\gamma ds \tag{17}$$

where $\xi_k(X,Y) = \psi_k^2(X,Y)$. To obtain their evolution equation, curve γ_k is embedded in a family of one-parameter curve. Let γ be parameterized by $\gamma(s,\tau):[0,1]\times[0,+\infty[\mapsto\Re$. Following [12], [13], we obtain the following functional derivatives for the evolution of γ:

$$\frac{\partial E}{\partial \gamma} = \left(\xi_1 - \xi_2 + \lambda\kappa_\gamma\right)\mathbf{n} \tag{18}$$

with corresponding Euler-Lagrange descent equations

$$\begin{aligned}\frac{d\gamma}{d\tau}(s,\tau) &= -\left(\xi_1(\gamma(s)) - \xi_2(\gamma(s)) + \lambda\kappa_\gamma(\gamma(s))\right) \\ &\quad \times \mathbf{n}(\gamma(s))\end{aligned} \tag{19}$$

where the dependence on τ is left implicit for simplicity, κ_γ is the mean curvature function of γ, $\mathbf{n}$ is the outward unit normal function to the curve γ.

3.4 Level-set implementation

For the implementation of equation (19), we use the level set formulation which is known for his numerical stability and topology independence [14]. In this formulation the evolving curve γ is implicitly represented by the zero level set of a function $\Phi:\Omega\times\Re^+\mapsto\Re$. With the convention that $\Phi>0$ inside the zero level set γ. One can show [14] that, if the evolution of a curve γ is described by the equation

$$\frac{d\gamma}{d\tau}(s,\tau) = F(\gamma(s))\,\mathbf{n} \tag{20}$$

Where F is a real-valued function defined on $\Re^2$, then the evolution of the corresponding level set function Φ is given by

$$\frac{\partial\Phi}{\partial\tau}(X,Y,\tau) = F(X,Y)\parallel\nabla\Phi(X,Y)\parallel \tag{21}$$

In our case, the level set function evolution equation corresponding to (19) is given by

$$\begin{aligned}\frac{\partial\Phi}{\partial\tau}(X,Y,\tau) &= -\left(\xi_1(X,Y) - \xi_2(X,Y) + \lambda\kappa_\Phi(X,Y)\right) \\ &\quad \parallel\nabla\Phi(X,Y)\parallel\end{aligned} \tag{22}$$

where the curvature function κ_Φ is given by $\kappa_\Phi = div\left(\nabla\Phi/\parallel\nabla\Phi\parallel\right)$.

4 GENERALIZATION TO MULTIPLE REGIONS

Segmentation into $N>2$ regions uses several active curves, can occur ambiguities when the interiors of two or more curves overlap. To guarantee that the curves define a partition of the image domain Ω, several solutions have been proposed in [15], [16], [17] and [18]. For instance, [17] uses:

$$\frac{\partial E_k}{\partial \gamma_k} = \left(\xi_k - \varphi_k + \lambda\kappa_{\gamma_k}\right)\mathbf{n}_k, \quad k=1,...,N-1 \tag{23}$$

with corresponding Euler-Lagrange descent equations

$$\begin{aligned}\frac{\partial\gamma_k}{\partial\tau}(s,\tau) &= -\left(\xi_k(\gamma_k(s)) - \varphi_k(\gamma_k(s)) + \lambda\kappa_{\gamma_k}(\gamma_k(s))\right) \\ &\quad \times \mathbf{n}_k(\gamma_k(s)), \quad k=1,...,N-1\end{aligned} \tag{24}$$

where κ_{γ_k} is the mean curvature of γ_k, $\mathbf{n}_k$ is the exterior unit normal function to curve γ_k, and function φ_k are defined by

$$\varphi_k(\gamma_k(s)) = \min_{i\neq k}\xi_i(\gamma_k(s)). \tag{25}$$

5 EXPERIMENTAL VERIFICATION

We ran several experiments to validate the approach and its implementation. We present four examples using two frames of each sequence. The use of a larger number of frames may improve results.

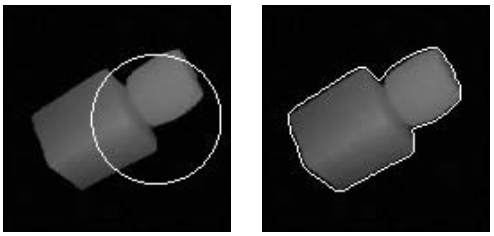
The first example uses the *Aftershaver* range image sequence which we constructed with a real range image of an after shave bottle from the Ohio State University (OSU/MSU/WSU Range Image Database) and synthetic motion. The object image motion causes one pixel motion in direction of X and one pixel in direction of Y. The derivatives were computed using parallel projection. There is no background motion. The first of the two frames used is shown in Figure 1(a) with the initial curve of segmentation; the final segmentation is shown in Figure 1(b).

The second example uses the *CubeRotating* range images sequence taking with the SR3000 camera (Mesa Imaging AG, Switzerland). This is a noisy real sequence taken by holding the camera in the hand. There were slight hand movements which caused slight camera pose movement in addition to the rotation of the cube. The first of the two frames used is shown in Figure 2(a) with the initial curve; the final segmentation is shown in Figure 2(b).

The third and fourth examples use the real *Conics* and *Chessmen* range images and synthetic motion. These images were taken with K2T structured light camera images in the Vision lab at the University of South Florida. In the first example we have four objects but only two objects move. Motion of the first object on the right (egg-like) image motion corresponds to two pixels in direction of Z. The motion of the third object from the right (cylinder) corresponds to an image motion of one pixel in direction of Y. There is no background motion. The first of the two frames used is shown in Figure 3(a) with the two initial curves; the final segmentation is shown in Figure 3(b). The segmentation delineates closely the two moving objects. In the second example we have six objects but only three objects move. Motion of the knight corresponds to two pixels in direction of Z. The bishop motion corresponds to one pixel in direction of X, that of the king to one pixel in direction of Y. There is no background motion. The first of the two frames used is shown in Figure 4(a) with the three initial curves, the final segmentation is shown in Figure 4(b). Here again the segmentation is accurate.

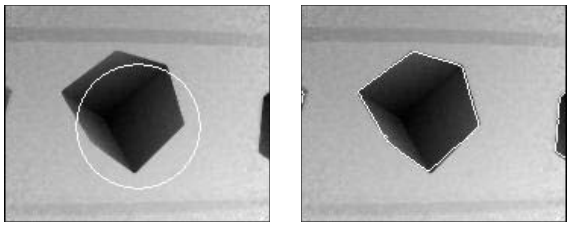
6 CONCLUSION

We investigated a variational, active curve evolution and level set method to segment multiple independent



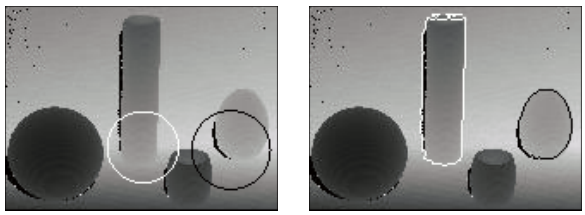
(a) (b)

Figure 1: (a) The first frame of *Aftershaver* range images sequence and initial level set; (b) The computed 3-D motion segmentation.



(a) (b)

Figure 2: (a) The first frame of *CubeRotating* sequence and initial level set; (b) The computed 3-D motion segmentation.



(a) (b)

Figure 3: (a) The first frame of *Conics* sequence and initial level set; (b) The computed 3-D motion segmentation.

(a) (b)

Figure 4: (a) The first frame of *chessmen* sequence and initial level set; (b) The computed 3-D motion segmentation.

3-D motions and simultaneously infer 3-D motion estimation in temporal sequences of range images. The corresponding EulerLagrange descent equations led to an algorithm which, after initialization, iterated two consecutive steps, namely, computation of rigid 3-D motion parameters by least squares, and curve evolution by level sets PDEs. The algorithm and its implementation have been validated on several range image sequences.

References

[1] P. J. Besl: "Active, optical range imaging sensors," *Machine Vision and Apps*, vol.1, pp.127-152, 1998.

[2] J. Beraldin and M. Rioux and F. Blais and L. Cournoyer and J. Domey: "Registered intensity and range image at 10 mega-samples per second," *Optical Engineering*, vol.31, no.1, Jan. 1992.

[3] R.A. Jarvis: "Range Sensing for Computer Vision," *Three-Dimensional Object Recognition Systems, Elsevier Science, Amsterdam*, pp.17-56, 1993.

[4] T. Oggier and M. Lehmann and R. Kaufmannn and M. Schweizer and M. Richter and P. Metzler and G. Lang and F. Lustenberger and N. Blanc: "An all-solid-state optical range camera for 3D-real-time imaging with sub-centimeter depth-resolution (SwissRanger)," *Proc. SPIE*, vol.5249, pp.634-545, 2003.

[5] T. Oggier and R. Kaufmann and M. Lehmann and P. Metzler and G. Lang and M. Schweizer and M. Richter and B. Bttgen and N. Blanc and K. Griesbach and B Uhlmann and K.-H. Stegemann and C. Ellmers: "3D-Imaging in Real-Time with Miniaturized Optical Range Camera," *Opto Conference Nurnberg*, 2004.

[6] K. Chang and K. Bowyer and P. Flynn: "An evaluation of multimodal 2D+3D face biometrics," *Trans. on Pattern Analysis and Machine Intelligence*, vol.27, no.4, pp.619-624, Apr. 2005.

[7] MESA Imaging AG : http://www.mesa-imaging.ch

[8] B. K. P. Horn and J. Harris: "Rigid body motion from range image sequences," *CVGIP: Image Understanding*, vol.53, no.1, pp.1-13, 1991.

[9] M. Yamamoto and P. Boulanger and J. Beraldin and M. Rioux: "Direct estimation of range flow on deformable shape from a video rate range camera," *Pattern Anal. Math. Intell*, vol.15, no.1, pp.82-89, 1993.

[10] H. Spies and B. Jahne and J. L. Barron: "Regularised range flow," *In editor, ECCV, Dublin, Ireland Lecture Notes in Computer Science*, vol.1843, Part2, pp.785-799, June 2000.

[11] H. Spies and J. L. Barron: "Evaluating the Range Flow Motion Constraint," *International Conference on Pattern Recognition*, vol.3, pp.30517, June 2002.

[12] G. Aubert and P. Komprobst: "Mathematical Problems in Image Processing," *Springer, New York*, 2002.

[13] S. Zhu and A. Yuille: "Region competition: unifying snakes, region growing, and bayes/mdl for multiband image segmentation," *IEEE trans. Pattern Anal. Mach. Intell.*, vol.18, no.9, pp. 884-900, Sep. 1996.

[14] J. Sethian: "Level Set Methos and Fast Marching Methods," *Cambridge Univ. Press, Cambridge, U.K.*, 1999.

[15] M. Rousson and R. Deriche: "A Variational Framework for Active and Adaptive Segmentation" *Proc. IEEE Workshop Motion and Video Computing*, pp. 56-61, 2002.

[16] T. Chan and L. Vese: "An Active Contour Model without Edges" *Proc. Intl Conf. Scale-Space Theories in Computer Vision*, pp. 141-151, 1999.

[17] C. Vazquez, A. Mitiche, and I. Ben Ayed: "Image Segmentation as Regularized Clustering: A Fully Global Curve Evolution Method" *Proc. Intl Conf. Image Processing*, pp. 3467-3470, 2004.

[18] A. Mansouri, A. Mitiche, and C. Vazquez: "Multiregion Competition: A Level Set Extention of Region Competition to Multiple Region Image Partioning" *Computer Vision and Image Understanding*, vol. 101, pp. 137-150, 2006.