MULTIPLE MOTION SEGMENTATION WITH LEVEL SETS WITHOUT PRIOR INFORMATION

Olfa Besbes, Ziad Belhadj

Ecole Supérieure des Communications de Tunis
Route de Raoued Km 3,5 - 2083 Ariana, Tunisie
olfa.besbes@cert.mincom.tn, ziad.belhadj@supcom.rnu.tn

ABSTRACT

Motion-based segmentation is an important task in object-oriented video applications. Different approaches to motion segmentation with level sets have been proposed. The key features of level sets representation are its ability to handle variations in the topology of the segmentation and its numerical stability. These approaches rely on certain prior information to perform motion segmentation. In this paper, we present a new algorithm for segmenting image into distinct regions of homogeneous motion. The two problems of motion estimation and motion segmentation are jointly solved without need of prior information. We provide experimental results on image sequences with synthetic and natural motion.

1. INTRODUCTION

We address the problem of segmenting the image plane into regions of homogeneous motion. Such a problem is of importance in applications ranging from region-based video compressing (such as MPEG-4), video database search and computer vision (scene analysis). Different approaches to motion-based image segmentation with level sets have been proposed. In [1], the segmentation is performed using a temporal change detection scheme. Thus, this approach does not extend to the cases of moving background and multiple motion. Furthermore, strong intensity edges at motion boundaries is required. These issues are resolved by formulating the problem as region competition [2], as recently proposed by Cremers [3, 4] and Mansouri [5, 6]. These approaches perform motion-based segmentation on the basis of certain prior information. In [3, 4], an iterative motion estimation/segmentation scheme with a prior on the number of distinct motion regions is developed. In [5], a complex clustering operation is performed before the segmentation in order to determine both the number of moving regions and their parameters. Although the above information are computed by the segmentation process in [6], sparse point correspondences are used as prior information to perform the motion segmentation.

In this paper, we propose a novel algorithm for multiple motion estimation and segmentation by minimizing a single energy functional. Unlike in [3, 4], our method is not iterative since a robust motion estimation is performed. To deal with large displacements and irregular intensity distributions, a robust incremental multiresolution algorithm is applied [7]. Inspired by recent work [6], we formulate the motion segmentation problem as a *sequential* segmentation with level set pursuit. Thus, an implicit clustering is computed through the segmentation itself. In contrast to [3, 4, 5, 6], the multiple motion segmentation is carried without any prior information (the number of distinct motion regions, the different motion parameters and sparse point correspondences). To increase accuracy, the motion weights are introduced in the level sets velocities functions. As a consequence, points conformity to a certain motion model is considered when performing the segmentation.

2. MULTIPLE MOTION IMAGE SEGMENTATION

2.1. Observation Model

Let $\Omega \subset \mathbb{R}^2$ denotes the images plane and let $\left\{I^n, I^{n+1}\right\}$ be the images at time instants n and $n+1$. The goal of motion segmentation is to estimate the partition $\{\Omega_i\}_{i=1}^N$ of the domain Ω such that each region Ω_i is characterized by a distinct motion transformation T_i. Assuming a constant velocity along motion trajectories, the observation model is defined as follows:

$$I^n(x) = I^{n+1}(T_i x) + \eta_i(x), \ \forall x \in \Omega_i, \forall i \in \{1, ..., N\},$$

where η_i are independent zero-mean stationary white noise processes with variance σ^2.

2.2. MAP Estimation

The maximum a posteriori estimate of $\{(\Omega_i, T_i)\}_{i=1}^N$ given I^n and I^{n+1} is expressed by:

$$\begin{aligned}\{\hat{\Omega}_i,\hat{T}_i\}_{i=1}^N &= \arg\max_{\{\Omega_i,T_i\}_{i=1}^N} P(\{\Omega_i,T_i\}_{i=1}^N|I^n,I^{n+1})\\ &= \arg\max_{\{\Omega_i,T_i\}_{i=1}^N} P(I^{n+1}|I^n,\{\Omega_i,T_i\}_{i=1}^N)\\ &\quad .\,P(\{\Omega_i,T_i\}_{i=1}^N|I^n).\end{aligned}$$

The motion-compensated intensity residual is defined as $\xi(T_i,x)=I^{n+1}(T_ix)-I^n(x)$. The $\aleph_i$ denotes the density function of the noise process η_i. Based on the observation model, we can express the first likelihood as follows:

$$P(I^{n+1}|I^n,\{\Omega_i,T_i\}_{i=1}^N)=\prod_{i=1}^N\prod_{x\in\Omega_i}\aleph_i(\xi(T_i,x)).$$

Since we aim a motion-based segmentation, we assume the independence of $\{\Omega_i,T_i\}_{i=1}^N$ and I^n. Furthermore, the independence of Ω_i and T_j, and uniform distributions for $\{T_j\}_{j=1}^N$ are assumed. Therefore, the *a priori* probability is expressed as:

$$P(\{\Omega_i,T_i\}_{i=1}^N|I^n)=\prod_{i=1}^N P(\Omega_i).$$

2.3. Energy Minimization

The MAP estimation problem is equivalent to the following energy minimization problem [5]:

$$\{\hat{\Omega}_i,\hat{T}_i\}_{i=1}^N=\arg\min_{\{\Omega_i,T_i\}_{i=1}^N} E\left(\{\Omega_i,T_i\}_{i=1}^N|I^n,I^{n+1}\right).$$

$$\begin{aligned}E\left(\{\Omega_i,T_i\}_{i=1}^N|I^n,I^{n+1}\right) &= \frac{1}{2\sigma^2}\sum_{i=1}^N\int_{\Omega_i}\xi^2(T_i,x)\,dx\\ &\quad+\frac{\lambda}{2\sigma^2}\sum_{i=1}^N\oint_{\partial\Omega_i}ds,\end{aligned}$$

where ds is the differential of arclength and λ is a positive constant. In fact, we choose the prior $P(\Omega_i)$ to be a function of boundaries $\partial\Omega_i$. Thus, our goal is to find the shortest closed curves $\{\partial\Omega_i\}_{i=1}^N$ which best separate moving regions with motion transformations $\{T_i\}_{i=1}^N$ from one another.

In order to carry out this minimization, the problem needs to be decomposed into interleaved minimizations with respect to $\{\Omega_i\}_{i=1}^N$ and $\{T_i\}_{i=1}^N$. In this work, we propose a sequential motion estimation/segmentation algorithm. So that, we assume in each step j that $\forall i\neq j$, $T_i=\hat{T}_i$ and $\Omega_i=\hat{\Omega}_i$. In this step, the above energy is only a function of Ω_j and T_j:

$$\begin{aligned}E\left(\{\Omega_i,T_i\}_{i=1}^N|I^n,I^{n+1}\right) &= \frac{1}{2\sigma^2}[\int_{\Omega_j}\xi^2(T_j,x)\,dx\\ &\quad+\sigma^2\int_{\Omega_j^c}dx+\lambda\oint_{\partial\Omega_j}ds].\end{aligned}$$

2.4. Robust Motion Estimation

The motion model T_j is estimated by minimizing the above energy (only its first term is considered). The 2D affine model is used for its relevance as motion descriptor. Given a spatial support, the T_j's parameters are computed only once according to a robust, incremental and multiresolution scheme described in [7]. Therefore, the robustness is ensured by minimizing an M-estimator criterion with a hard-redescending function. Adopting an incremental approach, we can handle large displacements. Given a current estimate $\hat{T}_j^k$ of the motion model, an increment ΔT_j^k is computed according to:

$$\Delta\hat{T}_j^k=\arg\min_{\Delta T_j^k}\int_{\Omega_j}W_j(x).r^2\left(\Delta T_j^k,x\right)dx,$$

with

$$\begin{aligned}r(\Delta T_j^k,x) &= \vec{\nabla}I^{n+1}(x+d_{\hat{T}_j^k}(x))d^t_{\Delta T_j^k}(x)\\ &\quad+I^{n+1}(x+d_{\hat{T}_j^k}(x))-I^n(x),\end{aligned}$$

is the optical flow residual obtained as a result of $\xi\left(\hat{T}_j^k,x\right)$'s first order expansion. $d_{\hat{T}_j^k}(x)$ is the displacement of point x according to the $\hat{T}_j^k$ model. W_j is a weight matrix of the motion model pixels conformities. $\vec{\nabla}I^{n+1}$ denotes the spatial gradient of the intensity function at time $n+1$. The incremental minimization is conducted within a multiresolution framework. For a detailed description of the RMR algorithm, the reader is referred to [7].

2.5. Motion Segmentation With Level Sets

Inspired by recent work [6], we formulate the motion segmentation problem as a sequential segmentation with level set pursuit. Thus, for the estimated $\hat{T}_1$ in the whole image support, the Euler-Lagrange descent equation of the defined energy functional with respect to $\hat{\Omega}_1$ yields the following level set evolution equation:

$$\frac{\partial u_1(x,t)}{\partial t}=-\left(w_1(x)\xi_1^2(x)-\sigma^2+\lambda\kappa_1\right)\left\|\vec{\nabla}u_1\right\|,$$

where κ_1 is the curvature of the zero-level set of u_1. The weight term $w_1(x) = \frac{1}{\lambda_w + W_1(x)}$ is introduced to take into account the motion model pixels conformities (where λ_w is a positive constant and W_1 is the weight matrix). At the convergence, a motion-based segmentation $\{\Omega_1, \Omega_1^c\}$ of the image domain Ω into one region and its complement is obtained. The estimated motion region, defined as $\hat{\Omega}_1 = \{u_1(.,\infty) > 0\}$, is characterized by the dominant motion model $\hat{T}_1$. At the next step, we estimate the dominant motion model T_2 in the Ω_1^c support. Since the previous segmentation is considered, a system of level set evolution equations is acquired:

$$\begin{cases} \frac{\partial u_1(x,t)}{\partial t} = -\left(w_1(x)\xi_1^2(x) - \zeta_1^2(x) + \lambda\kappa_1\right)\left\|\overrightarrow{\nabla} u_1\right\| \\ \frac{\partial u_2(x,t)}{\partial t} = -\left(w_2(x)\xi_2^2(x) - \zeta_2^2(x) + \lambda\kappa_2\right)\left\|\overrightarrow{\nabla} u_2\right\| \end{cases},$$

where

$$\zeta_1^2(x) = \chi_{\{u_2 <= 0\}}\sigma^2 + \chi_{\{u_2 > 0\}} w_2(x)\xi_2^2(x).$$
$$\zeta_2^2(x) = \chi_{\{u_1 <= 0\}}\sigma^2 + \chi_{\{u_1 > 0\}} w_1(x)\xi_1^2(x).$$

We define $\hat{\Omega}_2 = \{u_2(.,\infty) > 0\} \cap \Omega_1^c$. Similarly, this procedure is repeated until the k^{th} step, so that a motion segmentation $\{\hat{\Omega}_i, \hat{T}_i\}_{i=1}^k$ has been obtained. We then estimate the motion model T_{k+1} in the $\left(\hat{\Omega}_1^c \cup ... \cup \hat{\Omega}_k^c\right)$ support. Consider now the system of level set evolution equations

$$\begin{cases} \frac{\partial u_1(x,t)}{\partial t} = -(w_1(x)\xi_1^2(x) - \zeta_1^2(x) + \lambda\kappa_1)\left\|\overrightarrow{\nabla} u_1\right\| \\ \vdots \\ \frac{\partial u_k(x,t)}{\partial t} = -(w_k(x)\xi_k^2(x) - \zeta_k^2(x) + \lambda\kappa_k)\left\|\overrightarrow{\nabla} u_k\right\| \\ \frac{\partial u_{k+1}(x,t)}{\partial t} = -(w_{k+1}(x)\xi_{k+1}^2(x) - \zeta_{k+1}^2(x) \\ \qquad +\lambda\kappa_{k+1})\left\|\overrightarrow{\nabla} u_{k+1}\right\| \end{cases}$$

with $\zeta_h^2(x) = \min_{i\neq h, u_i>0} w_i(x)\xi_i^2(x)$, the min operator returns σ^2 whenever the defining condition is void. We define $\hat{\Omega}_{k+1} = \{u_{k+1}(.,\infty) > 0\} \cap \left(\cup_{i=1}^k \hat{\Omega}_i\right)^c$ the $k+1^{th}$ estimated region. The described pursuit process is achieved when $\bigcup_{i=1}^N \Omega_i = \Omega$ or no conforming motion model is computed to the remained region. The number of distinct motion regions is then defined to be N. Note that both the number of motion regions and their motion models are determined by the segmentation itself and therefore need not be known prior to segmentation as is the case with [5]. In contrast to [6], we perform our sequential motion-based segmentation without any prior information. In addition, we jointly solve the problems of motion estimation and motion segmentation.

3. EXPERIMENTAL RESULTS

We illustrate our motion estimation/segmentation algorithm on three image sequences. Figs. 1, 2 and 3 show segmentation results obtained on synthetic images. In the first experiment (Fig. 1), we demonstrate the capacity of our proposed approach to segment moving objects with same texture as the background. Different letters are moving with the same translational motion. Therefore, they constitute a single motion region. Due to level set representation, the topology change is handled. In the second example, multiple motion segmentation result is shown. Fig. (2d) shows the first motion region found by the algorithm, which corresponds to the static background. Figs. (2e) and (2f) show respectively the second and the third motion regions. The estimated motion field is shown in Fig. (2c). In particular, large displacements (24 pixels) are handled due to robust multiresolution motion estimation. Fig. 3 illustrates different level sets $\{u_i\}_{i=1}^3$ evolutions so that the principle of our motion segmentation algorithm is demonstrated. Fig. 4 shows motion-based segmentation result of a moving non-rigid object captured by a moving camera. In the last experiment (Fig. 5), we present the capability of our method to handle depth effect. The different layers corresponding to background (5d), flowerbed-house (5f) and tree (5h) are segmented. For each layer i, we show the weight matrix W_i corresponding to its estimated motion model. The estimated motion field and the level sets $\{u_i\}_{i=1}^3$ motion segmentation are respectively shown in figs. (5i) and (5j).

4. CONCLUSION

We presented a novel method for motion-based segmentation of images with multiple moving regions. A robust multiresolution motion models estimation is jointly performed. The main benefit of our proposed algorithm is that contrary to existing motion-based segmentation that use level sets, it does not require number of moving regions, motion parameters, rather other information be known prior to the segmentation. Experimental results, showing the validity of our algorithm and its potential, are illustrated. Present work focuses on generalization of the proposed method to multi-frame segmentation expressed in the spatiotemporal domain.

5. REFERENCES

[1] N. Paragios and R. Deriche, "Geodesic active contours and level sets for the detection and tracking of moving objects," *EEE Trans. on Pattern Anal. and Machine Intell.*, vol. 22, no. 3, pp. 266–280, Mars 2000.

[2] S. C. Zhu and A. Yuille, "Region competition: Unifying snakes, region growing, and bayes/mdl for multiband

image segmentation," *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 18, no. 9, pp. 884–900, September 1996.

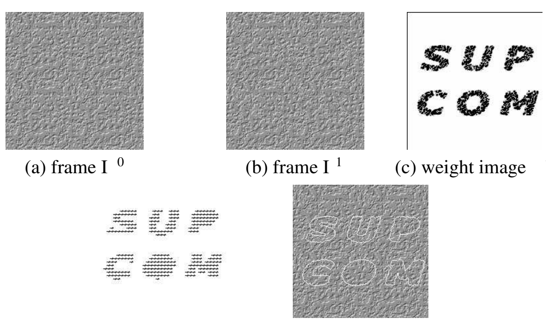
[3] D. Cremers, "A multiphase level set framework for motion segmentation," *4th Int. Conf. Scale Space Theories in Computer Vision*, vol. 2695, pp. 599–614, June 2003.

[4] D. Cremers, "A variational framework for image segmentation combining motion estimation and shape regularization," *IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 53–58, June 2003.

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

[6] A. R. Mansouri and A. Mitiche, "Spatial/joint space-time motion segmentation of image sequences by level set pursuit," *Proc. Int. Conf. on Image Processing*, vol. 2, pp. 265–268, September 2002.

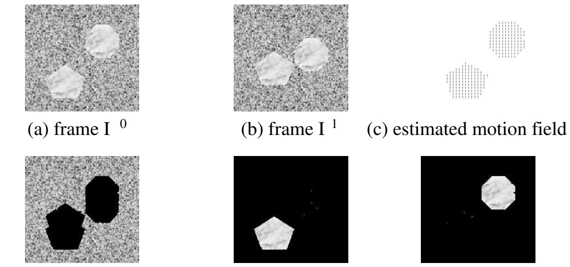
[7] J. M. Odobez and P. Bouthemy, "Robust multiresolution estimation of parametric motion models," *Int. Journal of Visual Communication and Image Representation*, vol. 6, no. 4, pp. 348–365, December 1995.



(a) frame I^0 (b) frame I^1 (c) weight image W_1

(d) estimated motion field (e) level set motion segmentation

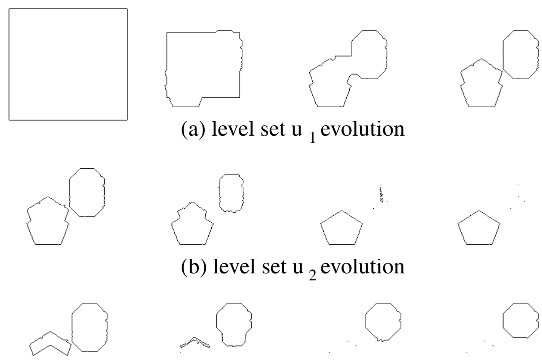
Fig. 1.



(a) frame I^0 (b) frame I^1 (c) estimated motion field

(d) background (e) moving object 1 (f) moving object 2

Fig. 2.

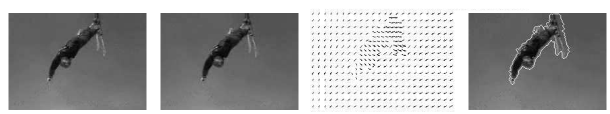


(a) level set u_1 evolution

(b) level set u_2 evolution

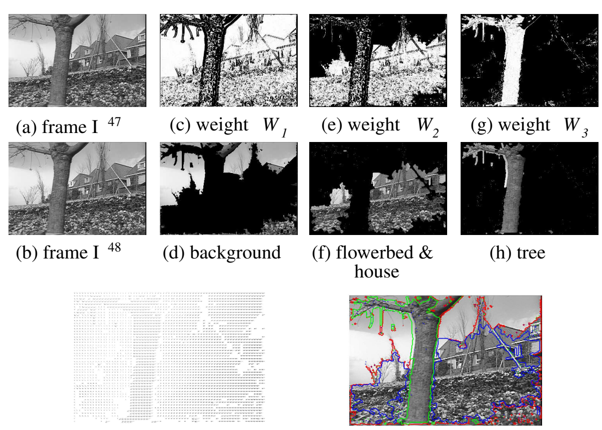
(c) level set u_3 evolution

Fig. 3.



(a) frame I^{35} (b) frame I^{36} (c) estimated motion field (d) level set motion segmentation

Fig. 4.



(a) frame I^{47} (c) weight W_1 (e) weight W_2 (g) weight W_3

(b) frame I^{48} (d) background (f) flowerbed & house (h) tree

(i) estimated motion field (j) level sets motion segmentation $u_{i=1,2,3}$

Fig. 5.