FAST MOTION ESTIMATION AND MOTION SEGMENTATION USING MULTI-SCALE APPROACH

C.Demonceaux[1,2] and D.Kachi-Akkouche[1]

[1]C.R.E.A., EA 3299
7, rue du moulin neuf
80000 Amiens, France
cedric.demonceaux@u-picardie.fr

[2]L.A.M.F.A., UMR 6140
33, rue Saint-Leu
80039 Amiens Cedex 1, France
djemaa.kachi@u-picardie.fr

ABSTRACT

The goal of this work is to find a fast method for motion estimation and motion segmentation. We chose to decompose the motion on a basis functions. That allows us to compute the velocity of each pixel by only solving a linear system. The motion segmentation is carried out using a markovian formulation. We minimize the associated energy by a determinist algorithm. But these choices induce two drawbacks : a temporal aliasing problem and a sensitivity of the optimization algorithm with the initialization. To overcome these two problems, we use a multiresolution method that successively computes the motion and the segmentation at each scale by motion compensation. Thanks to this procedure, we obtain a fast method of motion estimation and segmentation which does not require either initial spacial segmentation or dominant motion in the sequence.

1. INTRODUCTION

The motion estimation and motion segmentation in the image sequence are important problems in numerous applications : obstacles detection, video coding, content-based retrieval, video surveillance... The motion estimation assign a motion vector to each pixel in the image sequence. The motion segmentation divides the sequence into sets of pixels moving coherently. These two problems are closely related. Indeed, on the one hand, without good motion estimation, we cannot obtain a good motion segmentation and on the other hand a good motion segmentation allows to obtain better motion estimation.

Various approaches exist in literature about this problem. Let us classify these methods into three groups : methods that compute independently the motion and the segmentation, methods that compute recursively the motion and the segmentation and those ones that compute the motion and the segmentation together. In the first case, an initial spatial partition based on intensity, texture or color informa-

tion [6], [9], [1] is often used. A motion estimation is afterwards computed on each region. Then, the areas are merged by affine clustering [1], statistical tests [9], Markov random fields [6]. Considering these methods require an initial spatial segmentation, they are often very high cost in computational times. In the second case, the methods [10] estimate recursively the dominant motion and non-conform areas to this motion. The convergence of this method requires that the image sequence is governed by a dominant motion. The third methods minimize an energy which combines at one and the same time motion information and segmentation (from variational approaches [5], by bayesian framework [4], or robust estimators [8]). These methods generate difficult optimization problems and are costly in time.

In this article, we propose a fast method of motion estimation and segmentation. This method does not require an initial spatial segmentation, does not suppose that we can extract a dominant motion in the sequence. The motion estimation is computed using a basis of functions and the segmentation is obtained by a minimization of a markovian energy with a determinist algorithm. However, we are going to that to overcome the aliasing problem and to obtain a good initialization of the determinist algorithm, a multi-scale approach is necessary. In a first part, we explain our method in the mono-resolution case. In a second part, we extend the method in multi-resolution case. Then, we see how to initialize it at the coarsest scale. Finally, we validate this method on different sequences.

2. MOTION ESTIMATION AND SEGMENTATION USING FAST AND ROBUST METHOD

Let us consider image partition $\mathcal{P} = \{\mathcal{R}_1, \cdots, \mathcal{R}_p\}$ in p regions $\mathcal{P} = \{\mathcal{R}_1, \cdots, \mathcal{R}_p\}$ (we will see in the second part how to obtain this partition). The problem is to compute the velocity of each region R_i in order to obtain a best motion segmentation.

2.1. Motion estimation

To overcome the aperture problem and to obtain a fast motion estimation of each region, we project the Brightness Change Constraint Equation on basis functions.
Let us consider a basis of functions $(f^n)_{n=0\cdots N}$ of $L^2(\mathbf{R}^2)$ centered around $(0,0)$. We note for one point s in the image,

$$f_s^n(x) = f^n(x-s), \quad \forall n = 1 \cdots N.$$

To solve the Brightness Change Constraint Equation at point s is equivalent to solving :

$$\langle \overrightarrow{\nabla} I.\overrightarrow{v} + \frac{\partial I}{\partial t}, f_s^n \rangle = 0 \qquad \forall n = 1 \cdots N, \qquad (1)$$

with

$$\langle f, g \rangle = \int \int f(x)\overline{g(x)}dxdy.$$

Moreover, to estimate the displacement of each region $\mathcal{R}_i$, we modelize their motion by an affine model :

$$\overrightarrow{v}_{\mathcal{R}_i}(x,y) = B\Theta_i, \qquad (2)$$

with

$$B = \begin{bmatrix} x & y & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x & y & 1 \end{bmatrix},$$
$$\Theta_i = (a_i, b_i, c_i, d_i, e_i, f_i)^T.$$

This model is usually considered as a good trade-off between model complexity and model efficiency. It can take into account many kinds of motion (translation, rotation, scaling, deformation).
Let us note $\mathcal{N}_i = \{s = (x,y)|\text{supp}(f_s^n) \subset R_i, \forall n = 1..N \}$.
By substituting affine model (2) in (1), we obtain the following system :

$$\begin{array}{c} \forall s = (x,y) \in \mathcal{N}_i, \\ M_s\Theta_i = P_s \end{array} \qquad (3)$$

where M_s is a $6 \times N$ matrix and P_s is a vector of size N.
So, in order to estimate the motion of $\mathcal{R}_i$, we have an over-determined system, $(card(\mathcal{N}_i) \times N)$ equations and 6 unknowns. To avoid taking into account the points for which hypothesis (2) is not valid, we solve system (3) by a robust M-estimator of Tukey's biweight [7] :

$$\begin{array}{c} \widehat{\Theta}_i = \arg\min_{\Theta} \sum_{s \in \mathcal{N}_i} \sum_{j=1}^{N} \rho(r_s(j)), \\ r_s(j) = (M_s\Theta - P_s)(j). \end{array} \qquad (4)$$

Solution $\widehat{\Theta}_i$ of (4) is computed by an Iteratively Reweighted Least Squares (IRLS) :

$$\begin{array}{c} \widehat{\Theta}^j = \min_{\Theta} \sum_{s \in \mathcal{N}_i} \sum_{j=1}^{N} w_s(j)r_s(j)^2, \\ w_s(j) = \frac{1}{r_s(j)}\frac{\partial\rho}{\partial x}[r_s(j)]. \end{array} \qquad (5)$$

The solution of (3) by a robust M-estimator allows us to avoid the outliers points i.e. the points which are not valid for the assumption (2).
Equation (5) gives us a simple estimation of the motion of region $\mathcal{R}_i, i = 1\cdots p$.

2.2. Motion segmentation

In this section, we know the velocity of each region $\mathcal{R}_i$. The problem is now to segment the image considering this motion. To do that, we use a markovian formulation. Let S denote the set of sites s and $\mathcal{C}$ the set of cliques of two elements associated to a second-order neighborhood system $\mathcal{V}$. The issue of motion segmentation consists in searching a label field $e = \{e_s, s \in S\}$ with $e_s \in \{1, \cdots, p\}$. To solve this problem, we use the Maximum A Posteriori criterion. Due to the equivalence between MRF and Gibbs distribution, the problem can be formulated as the minimization of the energy :

$$\begin{array}{c} U(E) = \sum_{s \in S}\sum_{n=1}^{N}||\langle\overrightarrow{\nabla}I.\overrightarrow{v}_{\mathcal{R}_{e_s}} + \frac{\partial I}{\partial t}, f_s^n\rangle||^2 + \\ \lambda\sum_{<s,s'>\in\mathcal{C}}(1 - \delta(e_s, e_{s'})) \end{array} \qquad (6)$$

where

$$\delta(p,q) = \begin{cases} 1 & \text{if} \quad p = q, \\ 0 & \text{else.} \end{cases}$$

In order to obtain a fast method, we use a determinist algorithm (I.C.M. [3]) for the minimization of (6).
So, thanks to the projection of Brightness Change Constraint Equation (1) and to the determinist minimization of (6), we achieve a fast method to estimate the motion and the segmentation. However, this solution generates two problems. Indeed, the solution of (6) by an I.C.M. algorithm is closely dependent on the initialization. The second problem is due to the temporal aliasing. Because the sequence is sampled in time, we must estimate the temporal derivative with a finite difference :

$$\frac{\partial I}{\partial t} \simeq I(t+1) - I(t) \qquad (7)$$

But, in [2], it is proved that this approximation is valid if the optical flow satisfies :

$$||\overrightarrow{v}|| < \text{Cst.size of } f \text{ support.} \qquad (8)$$

Consequently, we have to do a trade-off between the number of equation $(card(\mathcal{N}_i))$ and the velocity of the regions. To overcome these problems, we generalize our method with a multi-scale approach. The sensitivity of I.C.M. is solved by the projection of the label field from the coarse scale to the finer scale. The aliasing problem is overcome by motion compensation between each scale.

3. MULTI-SCALE APPROACH

Using wavelets as basis functions for optical flow computation makes the multi-scale analysis easier. Let us introduce $(\Psi^n)_{i=1\cdots N}$ a wavelet basis in $L^2(\mathbf{R}^2)$ and :

$$\Psi_{jk}^n(x,y) = 2^{-j}\Psi^n(2^{-j}x - k_1, 2^{-j}y - k_2),$$

where $k = (k_1, k_2)$ and j is the index of resolution.
We suppose that we have obtained partition $\mathcal{P}^{j+1}$ at coarse scale $j+1$ in p_{j+1} regions $\mathcal{R}_i^{j+1}, i = 1\cdots p_{j+1}$:

$$\mathcal{P}^{j+1} : (x,y) \mapsto i, \quad (x,y) \in \mathcal{R}_i^{j+1}.$$

with the motion $\Theta_i^{j+1} = \left(a_i^{j+1}, b_i^{j+1}, c_i^{j+1}, d_i^{j+1}, e_i^{j+1}, f_i^{j+1}\right)^T$.
For each point in $\mathcal{R}_i^{j+1}$, we can set :

$$\overrightarrow{v}(x,y) = \overrightarrow{v}_{\mathcal{R}_i^{j+1}}(x,y) + \overrightarrow{v}_{\epsilon_i}(x,y)$$
$$\text{with } ||\overrightarrow{v}_{\epsilon_i}(x,y)|| < ||\overrightarrow{v}_{\mathcal{R}_i^{j+1}}(x,y)||$$

Let us introduce $\widetilde{I}(t+1)$ as

$$\widetilde{I}((x,y), t+1) = I((x,y) + \overrightarrow{v}_{\mathcal{R}_{\mathcal{P}^{j+1}(x,y)}^{j+1}}, t+1).$$

For each point (x,y), the motion between $I(t)$ and $\widetilde{I}_j(t+1)$ is exactly of $\overrightarrow{v}_{\epsilon_{\mathcal{P}^{j+1}(x,y)}}(x,y)$. This motion is estimated at scale j thanks to the resolution of the following system :

$$\forall (x,y) = 2^j(k_1, k_2) \in \mathcal{R}_i^{j+1},$$
$$\langle \overrightarrow{\nabla}I.\overrightarrow{v}_{\epsilon_i} + \frac{\partial \widetilde{I}}{\partial t}, \Psi_{jk}^n \rangle = 0 \qquad \forall n = 1\cdots N, \qquad (9)$$
$$\text{with } \frac{\partial \widetilde{I}}{\partial t} = \widetilde{I}(t+1) - I(t).$$

If we set :

$$\overrightarrow{v}_{\epsilon_i}(x,y) = B\Theta_i^j,$$

the motion is computed using the robust resolution of the same system as (3), where $I(t+1)$ is replaced by $\widetilde{I}(t+1)$.

$$\forall (x,y) = 2^j(k_1, k_2) \in \mathcal{N}_i^{j+1},$$
$$M_{jk}\Theta_i^j = P_{jk}. \qquad (10)$$

Then, we obtain the segmentation at scale j by the minimization of the following energy :

$$U^j(E) = \sum_{s=2^j k \in S} \sum_{n=1}^N ||\langle \overrightarrow{\nabla}I.\overrightarrow{v}_{\epsilon_{e_s}} + \frac{\partial \widetilde{I}}{\partial t}, \Psi_{jk}^n \rangle||^2$$
$$+ 2^j\lambda \sum_{<s,s'>\in \mathcal{C}^j} (1 - \delta(e_s, e_{s'})),$$
$$U^j(E) = \sum_{s=2^j k \in S} V_1^j(e_s) + 2^j\lambda \sum_{<s,s'>\in \mathcal{C}^j} V_2^j(E).$$
$$(11)$$

that we initialize with the segmentation found at scale $j+1$ and where $\mathcal{C}^j$ is the set of cliques at scale j. We add to label set $\{1, \cdots, p_{j+1}\}$ a waiting label ρ with an energy $V_1^j(\rho) = \mathcal{V}_\rho$ allowing new region creation at scale j.
This multi-scale method therefore allows us to solve the aliasing problem and the sensibility of the I.C.M., as wells as to accelerate the computation. What remains to do is to segment the sequence at the coarsest scale in order to initialize the algorithm.

4. INITIALIZATION

At coarsest scale J, as we do not know *a priori* the number of regions in the image, we have two possibilities : to initialize using only one region and to use waiting label ρ recursively, or to initialize the method with as many regions as pixels. The first method used by [10] is equivalent to a moving object detection at the coarsest scale. This method is valid only if a dominant motion exists in the sequence. For this reason, we chose the second idea.
So, at scale J, we have as many regions $\mathcal{R}_i^J$ as pixels. The optical flow at each pixel and consequently at each region is computed from the following system :

$$\forall s = 2^J(k_1, k_2),$$
$$\begin{cases} \langle \overrightarrow{\nabla}I.\overrightarrow{v}_{\mathcal{R}_s} + \frac{\partial I}{\partial t}, \Psi_{Jk}^n \rangle = 0 & \forall n = 1\cdots N, \\ \overrightarrow{v}_{\mathcal{R}_s^J} = B\Theta_s^J \end{cases}$$
$$(12)$$

A first segmentation is carried out with the minimization of U^J. To avoid an over-segmentation of the image, a motion segmentation by regions is performed considering the label field $E' = \{e_{\mathcal{R}_1}', \cdots, e_{\mathcal{R}_p}'\}$, $e_{\mathcal{R}_i}' \in \{1, \cdots, p\}$. The new segmentation is obtained from the minimization of :

$$U_R^J(E') = \sum_{i=1}^p \sum_{s\in \mathcal{R}_i} V_1^J(e_s') + 2^J\lambda \sum_{<s,s'>\in \mathcal{C}^J} (1 - \delta(e_s', e_{s'}')).$$
$$(13)$$

Then, a new motion estimation is computed on each region and we segment the image pixel by pixel (minimization of (11)) and region by region (minimization of (13)). These three steps are performed until there is no union of one region with another one. So, we obtain a segmentation at the coarsest J to initialize our method at scale $J-1$.

5. EXPERIMENTS AND RESULTS

The proposed method has been tested on many sequences. We here present results obtained on two sequences. The first one, Yosemite synthetic sequence (Fig 1) which is generated by flying through the Yosemite valley. Calendar (Fig 2) is a TV sequence involving large displacements. It includes several different moving objects and a horizontal camera motion. The calendar translates vertically, the train pushes a ball and the pendulum rotates. We apply our method on 4 scales with the same parameter of penalization $\lambda = 60$ on the two sequences. In a general way, the motion and the segmentation are correctly estimated. The proposed method has been implemented on Matlab with a PC-Pentium III 933 Mhz, the motion estimation and the segmentation take less than 2 minutes on an image of size 256×256.
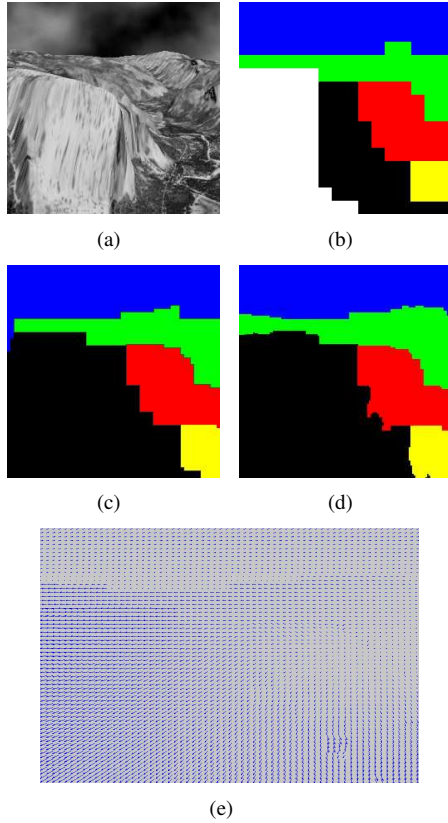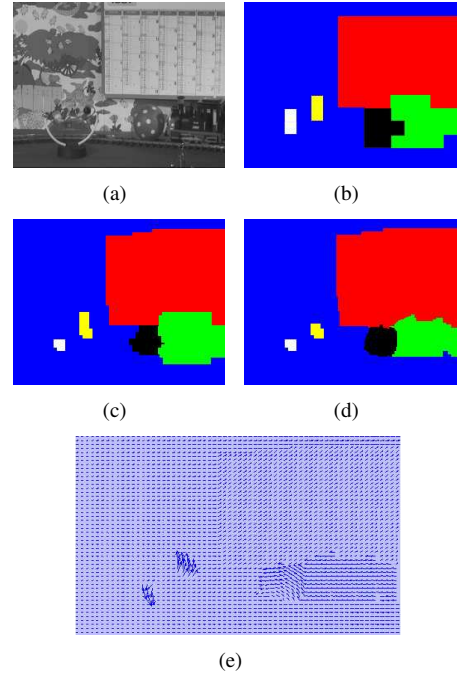
Fig. 1. (a) Frame of Yosemite sequence, (b) Initialization of the algorithm $J = 4$, (c) motion segmentation at $j = 2$,(d) final motion segmentation, (e) optical flow estimation.

Fig. 2. (a) Frame of Calendar sequence, (b) Initialization of the algorithm $J = 4$, (c) motion segmentation at $j = 2$,(d) final motion segmentation, (e) optical flow estimation.

6. CONCLUSION

We have introduce a method which performs the motion and the segmentation in an image sequence. This method uses a wavelet analysis of the optical flow and a markovian modelization of the segmentation. To overcome the aliasing problem and the sensitivity of the deterministic algorithm of optimisation, we introduce a multi-scale method. Thanks to the fast algorithm of wavelets decomposition of Mallat and the multi-scale approach, we obtain a fast method of motion estimation and segmentation. However, we can see a block effect in the segmentation which can be corrected in our future works by adding a temporal dependence between the current segmentation and the previous ones.

7. REFERENCES

[1] Y. Altunbasak, P.E. Eren, and A.M. Tekalp. Region-based parametric motion segmentation using color information. *GMIP*, 60(1) :13–23, January 1998.

[2] C. Bernard. Discrete wavelet analysis : A new framework for fast optic flow computation. In *ECCV98*, pages 354–368, 1998.

[3] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *RoyalStat*, B-36(2) :192–236, 1974.

[4] M.M. Chang, A.M. Tekalp, and M.I. Sezan. Simultaneous motion estimation and segmentation. *IP*, 6(9) :1326–1333, September 1997.

[5] D. Cremers. A variational framework for image segmentation combining motion estimation and shape regularization. In *CVPR*, pages 53–58, 2003.

[6] M. Gelgon and P. Bouthemy. A region-level motion-based graph representation and labeling for tracking a spatial image partition. *Pattern Recognition*, 33(4) :725–740, March 2000.

[7] P.W. Holland and R.E. WelschHsu. Robust regression using iteratively reweighted least-squares. *Commun. Statist.-Theor. Meth.*, A6 :813–827, 1977.

[8] E. Mémin and P. Pérez. Joint estimation-segmentation of optic flow. In *ECCV98*, pages 563–580, 1998.

[9] F. Moscheni, S. Bhattacharjee, and M. Kunt. Spatiotemporal segmentation based on region merging. *PAMI*, 20(9) :897–915, September 1998.

[10] J.M. Odobez and P. Bouthemy. Direct incremental model based image motion segmentation for video analysis. *Signal Processing*, 66(2) :143–155, April 1998.