Motion detection using wavelet analysis and hierarchical markov models

Cédric Demonceaux[1,2] and Djemâa Kachi-Akkouche[1]

[1] C.R.E.A, E.A. 3299,
7, rue du moulin neuf, 80000 Amiens, France

[2] L.A.M.F.A., UMR 6140,
33, rue Saint-Leu, 80000 Amiens, France

{cedric.demonceaux,djemaa.kachi}@u-picardie.fr

Abstract. This paper deals with the motion detection problem. This issue is of key importance in many application fields. To solve this problem, we compute the dominant motion in the sequence using a wavelet analysis and robust techniques. So, we obtain an estimation of the dominant motion on several image resolutions. This method permits to define a hierarchical Markov model in a natural way. Thanks to this modelization, we overcome two problems: the solution sensibility in relation to the initial condition with a Markov random field, and the temporal aliasing. Moreover, we obtain a semi-iterative algorithm faster than using the multi-scale techniques. Thus, we introduce a fast and robust algorithm in order to compute the motion detection in an image sequence. This method is validated on real image sequences.

1 Introduction and previous works

Motion detection is an important problem in many applications: obstacle detection, video coding, content-based retrieval, video surveillance... Motion detection consists in separating the image in two regions : stationary region relatively with the camera motion and moving region. With a fixed camera, a difference between two consecutive images can be sufficient to get a bi-partition of the picture in a stationary zone and a motion zone. However, this type of method is very sensitive to the change of illumination conditions. Many authors have proposed an improvement of this method. Several methods use statical tests such as Hsu and *al* [9] who assume the intensity locally affine, Rosin [17] who modelizes the noise in the sequence by a normal distribution. Aach and *al* [1], Sifakis [18] define the inter-frame difference with a probability density function. Another approach consists in regularizing the masks of motion detection by a Markov model [4], [11].
With a mobile camera, the problem is more difficult. We can use some constraints on the apparent displacement field of the camera supposed known [19]. Another way consists to compensate the sequence with the dominant motion. It allows to return to the simpler case of a fixed camera. Then, classical technics of thresholding [10], or bayesian methods [14], [15] are used. An initial spatial partition based on intensity, texture or color information can be used too [13]. Moreover, some techniques compute simultaneously motion and detection [5], [6].
In many methods, the motion computation in the image is performed by the measure of the displaced frame difference ($DFD(x,t) = I(x+v, t+1) - I(x,t)$). Indeed, this method is often preferred to the differential method because it allows to estimate great displacements in image sequence $I(x,t)$. But, as the DFD is not linear, the motion estimation by DFD is computed by gradient descent. This method is more expensive in computing times than the proposed approach in this work. In this article, we estimate the motion in the sequence from the Brightness Change Constraint Equation (B.C.C.E.). The problem of the great displacement estimation, i.e the aliasing problem, is overcomed thanks to a wavelet analysis of this equation and a motion compensation between each scale. This method enables to obtain a motion estimation at each scale by the resolution of an over-determined linear system. The study of B.C.C.E at each scale gives a measurement characterizing the conformity with the dominant motion for each point. To solve the motion detection problem from the obtained multi-scale data, we suggest to modelize the problem by a hierarchical markovian model. This model combines an *a priori* dependence in space and in scale with a Markov random field on the coarsest level of the image. Chardin [16], who used these fields to texture segmentation, established that these models are in fact a generalization of the multi-scale approach of [7] and permit to obtain better results in shorter cpu time. Moreover, contrary to the multi-scale approach which only uses observations at the finer scale, the hierarchical definition of the problem uses observations at each scale of the image. This allows to overcome the temporal aliasing problem generated by the motion estimation with differential

methods. Thanks to the wavelet analysis and to the hierarchical modelization, the method presented here is a fast method to detect moving objects in an image sequence.
This paper is organized as follows: In a first part, we compute the dominant motion in the scene by a wavelet analysis of the Brightness Change Constraint Equation and a robust M-estimator. This study enables us to obtain a fast estimation of the dominant motion at each resolution of the image. Then, in a second part, thanks to a hierarchical model, we see how to overcome the temporal aliasing problem inherent to the optical flow estimation by the differential method and how to solve the problem of motion detection. Finally, we see results on synthetic and real sequences.

2 Robust dominant motion estimation

Let us consider image sequence $I((x,y),t)$. The problem is to estimate dominant motion $\overrightarrow{v}((x,y),t)$. To do that, we assume that the intensity constant in time for each physical point. Thus, by derivation, we obtain the well known Brightness Change Constraint Equation (B.C.C.E.):

$$\overrightarrow{\nabla} I((x,y),t).\overrightarrow{v}((x,y),t) + \frac{\partial I((x,y),t)}{\partial t} = 0 \tag{1}$$

To estimate the dominant motion in the sequence, we solve the B.C.C.E. (1) by a wavelet analysis and an M-robust estimator.

Let us consider the wavelets basis $(\Psi^n)_{i=1\cdots N}$ in $L^2(\mathbf{R}^2)$ centered around the origin $(0,0)$, and let us consider the N functions centered around point $(2^j k_1, 2^j k_2)$ defined as :

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

where $k = (k_1, k_2)$ and j is the index of resolution. Taking the inner product of (1) with Ψ^n_{jk}, we obtain the following system :

$$\langle \overrightarrow{\nabla} I.\overrightarrow{v} + \frac{\partial I}{\partial t}, \Psi^n_{jk}\rangle = 0 \qquad \forall n = 1\cdots N, \tag{2}$$

where

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

We assume that dominant motion due to the camera movement $\overrightarrow{v}$ is affine, i.e:

$$\overrightarrow{v}(x,y) = B.\Theta^j, \quad (x,y) = (2^j k_1, 2^j k_2) \tag{3}$$

with

$$B = \begin{bmatrix} x & y & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x & y & 1 \end{bmatrix}$$
$$\Theta^j = \left(a^j, b^j, c^j, d^j, e^j, f^j\right)^T$$

This model is a good tradeoff between complexity and representativeness. It can take into account many kinds of camera motion (translation, rotation, scaling, deformation).
By substituting affine model (3) in (2), we obtain for each j resolution the following system :

$$\begin{array}{l} \forall n = 1..N, \forall (x,y) = (2^j k_1, 2^j k_2) \\ a^j \langle x \frac{\partial I}{\partial x}, \Psi^n_{jk} \rangle + b^j \langle y \frac{\partial I}{\partial x}, \Psi^n_{jk} \rangle + c^j \langle \frac{\partial I}{\partial x}, \Psi^n_{jk} \rangle + \\ d^j \langle x \frac{\partial I}{\partial y}, \Psi^n_{jk} \rangle + e^j \langle y \frac{\partial I}{\partial y}, \Psi^n_{jk} \rangle + f^j \langle \frac{\partial I}{\partial y}, \Psi^n_{jk} \rangle + \\ \langle \frac{\partial I}{\partial t}, \Psi^n_{jk} \rangle = 0 \end{array} \tag{4}$$

and by integrating by parts:

$$\begin{array}{l} \forall (x,y) = (2^j k_1, 2^j k_2), \forall n = 1..N \\ a^j \left[\langle xI, \frac{\partial \Psi^n_{jk}}{\partial x} \rangle + \langle I, \Psi^n_{jk} \rangle \right] + b^j \langle yI, \frac{\partial \Psi^n_{jk}}{\partial x} \rangle + c^j \langle I, \frac{\partial \Psi^n_{jk}}{\partial x} \rangle + \\ d^j \langle xI, \frac{\partial \Psi^n_{jk}}{\partial y} \rangle + e^j \left[\langle yI, \frac{\partial \Psi^n_{jk}}{\partial y} \rangle + \langle I, \Psi^n_{jk} \rangle \right] + f^j \langle I, \frac{\partial \Psi^n_{jk}}{\partial y} \rangle = \\ \langle \frac{\partial I}{\partial t}, \Psi^n_{jk} \rangle \end{array} \tag{5}$$

We obtain an over-determined system, $N \times 2^{p+q-2j}$ equations for an image of size $2^p \times 2^q$ and 6 unknowns at each j resolution. To avoid taking into account the points where hypothesis (3) is not valid, we solve system (5) by a robust M-estimator of Tukey's biweight [8].

2.1 Robust motion estimation

We note system (5):

$$M^j \Theta^j = P^j. \tag{6}$$

Where M^j and P^j are respectively of dimension $6 \times 2^{p+q-2j} N$ and $1 \times 2^{p+q-2j} N$.

We solve this system (6) by a robust M-estimator of Tukey's biweight ρ. i.e. we look for Θ^j solution of:

$$\begin{array}{l} \widehat{\Theta}^j = \arg\min\limits_{\Theta} \sum\limits_i \rho(r_i, C) \\ r_i = M^j \Theta^j(i) - P^j(i). \end{array} \tag{7}$$

Solution $\widehat{\Theta}^j$ of (7) is computed by an Iteratively Reweighted Least Squares (IRLS):

$$\widehat{\Theta}^j = \min_{\Theta} \sum_{i=1}^{N \times 2^{p+q-2j}} w_i r_i^2, \tag{8}$$

where $w_i = \frac{1}{r_i} \frac{\partial \rho}{\partial x}(r_i)$. The resolution of (6) by a robust M-estimator allows us to avoid the outliers points i.e. the points which are not valid for assumption (3). Equation (8) gives an estimation of the motion at each j scale. However, in order to overcome the aliasing problem, we have to do motion compensation between each scale.

2.2 Motion compensation

Considering the image sequence is sampled in time, we have to estimate the temporal derivative of I with a finite difference formula:

$$\langle \frac{\partial I}{\partial t}, \Psi^n_{jk} \rangle \simeq \langle I(t+1) - I(t), \Psi^n_{jk} \rangle$$

[2] proves that this approximation is valid if the optical flow verifies :

$$||\overrightarrow{v}|| < \text{K.size of wavelets support},$$

where K is a constant. So, dealing with a fine scale, only minor displacements can be estimated. Consequently, to compute large displacements, we have to compute recursively an estimation of the flow at a coarse scale and estimate the residual between this value and the real flow at a finer scale.
Let us suppose that at coarsest scale $j = J$, the solution of (6) is Θ^J. At scale $J-1$, we can split vector Θ^J as follows :

$$\Theta^{J-1} = \mathcal{P}_{J \to J-1}(\Theta^J) + \epsilon^{J-1},$$

where $\mathcal{P}_{j \to j'}$ is the projector from scale j to scale j':

$$\mathcal{P}_{j \to j'}(\Theta) = diag(1 \quad 1 \quad 2^{j-j'} \quad 1 \quad 1 \quad 2^{j-j'}).\Theta$$

Let us introduce $\widetilde{I}_{\Theta^{J-1}}(t+1)$ as

$$\widetilde{I}_{\Theta^{J-1}}((x,y),t+1) = I((x,y) + \overrightarrow{V}_{\mathcal{P}_{J \to J-1}(\Theta^J)}, t+1).$$

The motion between $I(x,t)$ and $\widetilde{I}_{\Theta^J}(x,t+1)$ is exactly $\overrightarrow{V}_{\epsilon^{J-1}}$.
ϵ^{J-1}_p is the solution of system (6) where we replace $\langle \frac{\partial I}{\partial t}, \Psi^n_{(J-1)k} \rangle$ by $\langle \widetilde{I}_{\Theta^{J-1}_p}(t+1) - I(t), \Psi^n_{(J-1)k} \rangle$.
Then, we compute iteratively the optical flow by motion compensation from the coarsest scale to the finest scale.
Thus, we obtain an estimation of the dominant motion at each j resolution.

3 Moving objects detection with hierarchical model

We have performed the dominant motion by a multi-scale method. This computation allows us to define in a natural way a hierarchical Markov model in all the resolutions of the image in order to solve moving object detection. This modelization has two advantages: it overcomes the problem of the sensibility to the initial conditions that we meet with a classical Iterated Conditional Modes [3] (I.C.M.), and it can solve the aliasing problem by compensation contrary to multi-scale methods such as [7], which only uses observations at the finer scale.

Let us note $S = \cup_{n=0}^{J} S^n$ where S^i indicates the level i of the image resolution, $E = \{E_s, s \in S\}$ and $Y = \{Y_s, s \in S\}$ respectively the random field of the labels

of the detected motion and the random field of the observations. E_s can take two values, 0 or 1, where 1 corresponds to a site that does not conform with the dominant motion and 0 to a site that conforms with this motion. Let us note E^n the whole of the labels at level n, *i.e.* $E^n = \{E_s, s \in S^n\}$ and in the same way $Y^n = \{Y_s, s \in S^n\}$. Finally, let us note $\overline{i}$ the parent of site i, $\underline{i}$ the whole of the children of i and $\underline{\underline{i}}$ the whole of the sites forming the tree of root i (fig 1). With this graphical structure and with some hypotheses described in [16], the distribution of (X, Y) can be written as :

$$P(E = e, Y = y) \propto exp - [\sum_{<i,j>\in S^J} v_{i,j}(e_i, e_j) + \sum_{i \notin S^J} w_i(e_i, e_{\overline{i}}) + \sum_{i \in S} o_i(e_i, y_i)] \tag{9}$$

where $< i, j >$ designates pairs of neighbors in S^J, $v_{i,j}$ and w_i are local functions

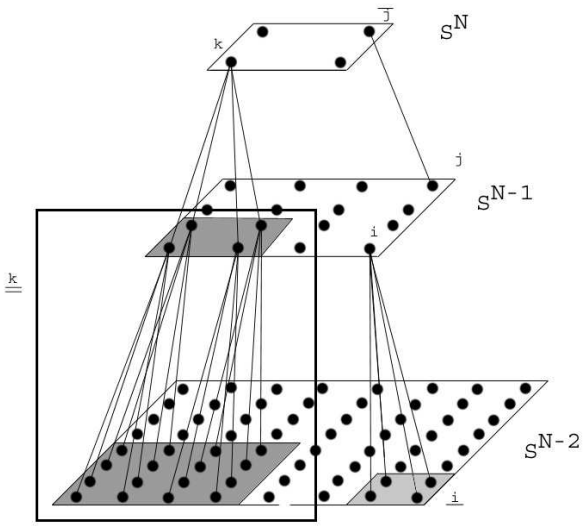


Fig. 1. Hierarchical structures. $\overline{j}$ the parent of the site j, $\underline{i}$ the whole of the children of i and $\underline{\underline{k}}$ the whole of the sites forming the tree of root k

capturing respectively the spatial *a priori* and the hierarchical *a priori*, and o_i expresses the point-wise relation between observed variable y_i and unknown x_i. The associated MAP estimator to this distribution :

$$\widehat{e} \in argmax_e P(e|y) = argmax_e P(e, y) \tag{10}$$

is computed from the following fast semi-iterative algorithm [16].
For the space and hierarchical potentials, we choose *a priori* usual functions of

Potts type :

$$v_{i,j}(e_i, e_j) = \alpha[1 - \delta(e_i, e_j)],$$
$$w_i(e_i, e_{\overline{i}}) = \beta[1 - \delta(e_i, e_{\overline{j}})].$$

These potential functions enforce the homogenization of the labels field in space and in scale.

Let us define observations o_i which express the point-wise relation between observed variables y_i.
To assess the error of the motion estimation at point $(2^j k_1, 2^j k_2)$ we set at a j fixed scale :

$$E(j,k,\Theta^j) = \frac{\sum_{n=1}^{N} ||\langle \overrightarrow{\nabla} I, \Psi^n_{jk} \rangle || C}{\sum_{n=1}^{N} ||\langle \overrightarrow{\nabla} I, \Psi^n_{jk} \rangle||^2} \quad \text{where } C = |\langle \overrightarrow{\nabla} I . \overrightarrow{V}_{\epsilon^j} + I_{\Theta^j}(t+1) - I(t), \Psi^n_{jk} \rangle|. \tag{11}$$

This measurement features an interesting property. Indeed, we can establish that

$$\begin{array}{l} \forall \epsilon > 0, \\ E(j,k,\Theta^j) \leq \frac{\lambda_1}{\lambda_1+\lambda_2}\epsilon \Rightarrow ||\overrightarrow{V}_{\epsilon^j} - \overrightarrow{V}_{jk}|| \leq \epsilon, \\ E(j,k,\Theta^j) \geq \sqrt{\frac{\lambda_2}{\lambda_1+\lambda_2}}\epsilon \Rightarrow ||\overrightarrow{V}_{\epsilon^j} - \overrightarrow{V}_{jk}|| \geq \epsilon, \end{array} \tag{12}$$

where $\overrightarrow{V}_{jk}$ is the real flow at scale j between images $I(t)$ and $\widetilde{I}_{\Theta^j}(t+1)$ at point $(2^j k_1, 2^j k_2)$ and where λ_1, λ_2 are respectively the smallest and the greatest eigenvalue of

$$A = \sum_{n=1}^{N} \begin{pmatrix} |\langle I, \frac{\partial \Psi^n_{jk}}{\partial x} \rangle|^2 & \langle I, \frac{\partial \Psi^n_{jk}}{\partial x} \rangle \overline{\langle I, \frac{\partial \Psi^n_{jk}}{\partial y} \rangle} \\ \langle I, \frac{\partial \Psi^n_{jk}}{\partial x} \rangle \overline{\langle I, \frac{\partial \Psi^n_{jk}}{\partial y} \rangle} & |\langle I, \frac{\partial \Psi^n_{jk}}{\partial y} |\rangle^2 \end{pmatrix}$$

We define $\forall \epsilon > 0$ the numbers

$$l_{jk} = \frac{\lambda_1}{\lambda_1+\lambda_2}.\epsilon \text{ and } L_{jk} = \sqrt{\frac{\lambda_2}{\lambda_1+\lambda_2}}.\epsilon. \tag{13}$$

According to (18), $E(j,k,\Theta^j)$ translates the error made by approximating the real optical flow at point $(2^j k_1, 2^j k_2)$ by V_{ϵ^j} (velocity between the image $I(t)$ and the image $\widetilde{I}_{\Theta^j}(t+1)$).
If $E(j,k,\Theta^j) \leq l_{jk}$, we can assert that the error is smaller than ϵ and if $E(j,k,\Theta^j) \geq L_{jk}$, the error is larger than ϵ. This error will enable us to decide if point $(2^j k_1, 2^j k_2)$ follows the dominant motion up to ϵ.

Thanks to this measurement error, we define o_i as follows:

$$\begin{array}{l} \forall x_i = (2^j k_1, 2^j k_2) \in S^j \\ o_i(e_i, y_i) = \begin{cases} f(E(j,k,\Theta^j_{e_i}), l_{jk}) & \text{if } e_i = 0 \\ 1 - f(E(j,k,\Theta^j_{e_i}), L_{jk}) & \text{if } e_i = 1 \end{cases} \end{array} \tag{14}$$

where

$$f(x,d) = \frac{1}{1 + e^{(\frac{-4}{d}(x-d))}}.$$

These potential functions enforce the sites to take label 0 (conform to motion) in the case where $E(j,k,\Theta_D^j) \leq l_{jk}$ and to take label 1 (not conform to motion) if $E(j,k,\Theta_D^j) \geq L_{jk}$.
So, we have defined a hierarchical Markov model in order to obtain a fast algorithm for the motion detection problem as we can see in the following section.

4 Experimental results

The proposed method has been tested on many sequences. We here present results obtained on four sequences of size 256×256. In the results, the black areas represent the areas true to the dominant motion.
Synthetic sequence:
This sequence is a synthetic sequence created to check that the motion compensation permits to detect the moving objects with large displacements. It is composed of a mobile bottom in translation of 13 pixels towards the left between the two images, of a circle in translation towards the right of 20 pixels and of a circle in translation to the bottom of 10 pixels (fig 3(b)). We estimated the dominant motion over 6 levels of resolution (fig 3(c)). Figure (2) shows the cpu times in seconds of various methods (on a PC Pentium III 933 Mhz). The results show the method allows to estimate large displacements in the scene thanks to the motion compensation between each scale of resolution. The moving object detection based on a hierarchical model is compared with two different resolutions (J=2 and J=3). Moreover, we compare these results with a traditional mono-resolution I.C.M. algorithm and without motion compensation (fig 3(f)). The modelization of the moving object detection problem by a hierarchical model gives good results. The results are almost similar except for the detection of two small zones in case J=2, normally in conformity with the dominant motion due to the temporal aliasing. A simple I.C.M. algorithm does not allow to detect correctly moving objects because of its sensitivity to initialization and of aliasing temporal and is moreover much longer than our method. In a general way, on the whole of the sequence, the modelization of problem on 3 levels of resolution gives good results.
Traffic sequence: In the sequence, the camera is motionless (fig 4(a)), 9 vehi-

	cpu times
Dominant motion estimation	1,8s
Motion Detection with J=3	1,2s
Motion Detection with J=2	2,9s
I.C.M mono resolution without motion compensation	10,8s

Fig. 2. cpu times for different methods.

cles turn left and a vehicle in the top right advances towards the left. We detect

correctly the vehicles, except the black car where the spatial gradient is too small to detect a motion.

Ping Pong sequence: The Ping-pong sequence has a Nil dominant motion (fig 5(a)). The arm of the player, his left hand, the ball and the racket are mobile and correctly detected.

Car sequence: The car sequence has a dominant motion from the right to the left caused by the camera displacement. Two cars move with independent motions and we can see a minor motion of the foliage due by the wind (fig 6(a)). We have tested our method with two different ϵ. With $\epsilon = 2$, as the foliage moves with minor motion, only the two cars are detected. Besides, with $\epsilon = 1$ a part of the tree is detected too.

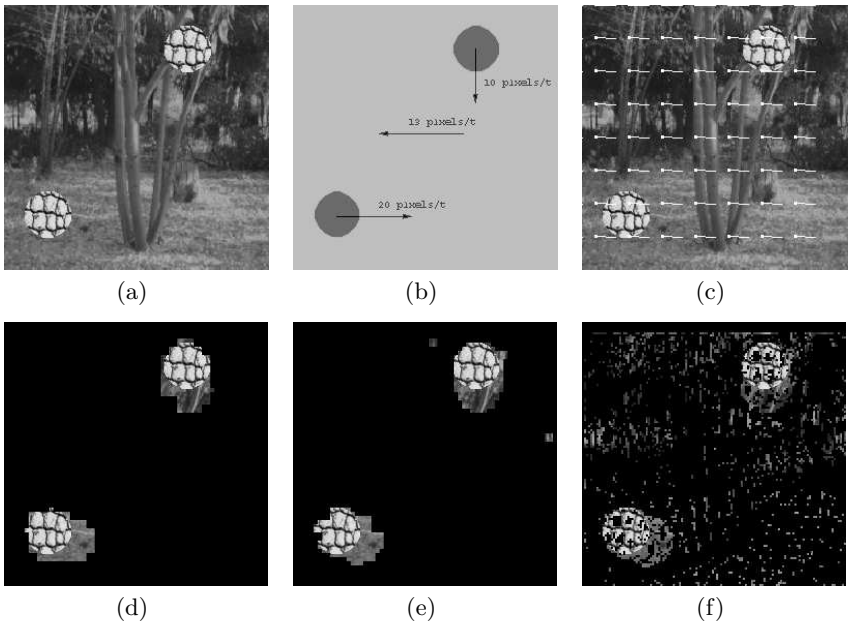


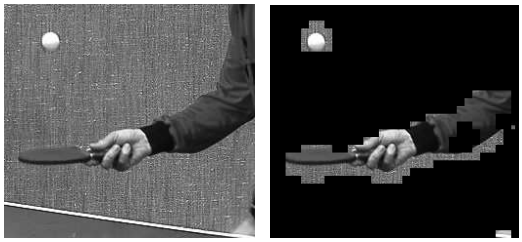
Fig. 3. $\epsilon = 1.2$, $\alpha = 3$, $\beta = 50$, (a) image test, (b) real motion, (c) dominant motion estimation, (d) motion detection with J=3, (e) J = 2 , (f) I.C.M. algorithm monoresolution without motion compensation.

5 Conclusion

We have introduced a new method of motion detection in an image sequence. This method estimates the dominant motion in the scene thanks to a wavelet analysis. This analysis enables us to obtain an estimation of the dominant motion at various scales by motion compensation. From these various data, we have

(a) (b)

Fig. 4. $\delta = 1$, $\alpha = 5$, $\beta = 50$, (a) original image, (b) motion detection with J=3.



(a) (b)

Fig. 5. $\epsilon = 1$, $\alpha = 1$, $\beta = 50$, (a) original image, (b) motion detection with J=3.

(a) (b) (c)

Fig. 6. $\alpha = 5$, $\beta = 100$, (a) original image, (b) motion detection with J=3, $\epsilon = 1$, (c) motion detection with J=3, $\epsilon = 2$.

defined a Markovian hierarchical model. Thanks to the fast algorithm of wavelets decomposition of Mallat [12], and to this hierarchical modelization, the method presented here is a fast method to detect moving objects in a sequence.
However, as we can see in the results, the hierarchical modelization generates blocks effects inherent to the structure. It would be interesting to add new interactions between the levels. Moreover, we will modelize the problem using a temporal data. So, we will use the label field previously estimated $(t-1)$ to compute the next label field (t). This method will make it possible to avoid the detection of outliers due to the noise in sequences and to the motion estimation errors.

Appendix:

Let us note V_{jk} the real velocity at the scale j at point $(2^{-j}k_1, 2^{-j}k_2)$.
$\exists(\lambda_1, \lambda_2)$ such as:

$$\frac{\lambda_1}{\lambda_1+\lambda_2}||\overrightarrow{V}_{\epsilon^j} - \overrightarrow{V}|| \leq E(j,k,\Theta^j) \leq \sqrt{\frac{\lambda_2}{\lambda_1+\lambda_2}}||\overrightarrow{V}_{\epsilon^j} - \overrightarrow{V}_{jk}|| \tag{15}$$

with λ_1 and λ_2 respectively the smallest and the greatest eigenvalues

$$A = \begin{pmatrix} \sum_{n=1}^{N} |\langle I, \frac{\partial \Psi_{jk}^n}{\partial x}\rangle|^2 & \sum_{n=1}^{N} \langle I, \frac{\partial \Psi_{jk}^n}{\partial x}\rangle \langle I, \frac{\partial \Psi_{jk}^n}{\partial y}\rangle \\ \sum_{n=1}^{N} \langle I, \frac{\partial \Psi_{jk}^n}{\partial x}\rangle \langle I, \frac{\partial \Psi_{jk}^n}{\partial y}\rangle & \sum_{n=1}^{N} |\langle I, \frac{\partial \Psi_{jk}^n}{\partial y}\rangle|^2 \end{pmatrix}$$

Proof:

$$\begin{aligned} \sum_{n=1}^{N} ||\langle \overrightarrow{\nabla} I, \Psi_{jk}^n\rangle||^2 &= \sum_{n=1}^{N} [|\langle I, \frac{\partial \Psi_{jk}^n}{\partial x}\rangle|^2 + |\langle I, \frac{\partial \Psi_{jk}^n}{\partial y}\rangle|^2] \\ &= \text{trace}(A) \\ &= \lambda_1 + \lambda_2 \end{aligned}$$

and as A is a symmetric positive definite matrix,

$$\lambda_1 ||X||^2 \leq X^T A X \leq \lambda_2 ||X||^2 \quad \forall X \tag{16}$$

$\overrightarrow{V}_{jk}$ verifies:

$$\overrightarrow{\nabla} I . \overrightarrow{V}_{jk} = -\frac{\partial I}{\partial t}. \tag{17}$$

Consequently,

$$E(j,k,\Theta^j) = \frac{\sum_{n=1}^{N} ||\langle \overrightarrow{\nabla} I, \Psi_{jk}^n\rangle||.|\langle \overrightarrow{\nabla} I.(\overrightarrow{V}_{\epsilon^j} - \overrightarrow{V}_{jk}), \Psi_{jk}^n\rangle|}{\sum_{n=1}^{N} ||\langle \overrightarrow{\nabla} I, \Psi_{jk}^n\rangle||^2}$$

And thanks to the Cauchy-Schwartz inequality,

$$E(j,k,\Theta^j) \leq \sqrt{\frac{\sum_{n=1}^{N} |\langle \overrightarrow{\nabla} I.(\overrightarrow{V}_{\epsilon^j} - \overrightarrow{V}_{jk}), \Psi_{jk}^n \rangle|^2}{\sum_{n=1}^{N} ||\langle \overrightarrow{\nabla} I, \Psi_{jk}^n \rangle||^2}}$$

Moreover,

$$\begin{aligned} \sum_{n=1}^{N} |\langle \overrightarrow{\nabla} I.(\overrightarrow{V}_{\epsilon^j} - \overrightarrow{V}_{jk}), \Psi_{jk}^n \rangle|^2 &= \sum_{n=1}^{N} |\langle I, \tfrac{\partial \Psi_{jk}^n}{\partial x} \rangle (V_{\epsilon^j} - V_{jk})_x + \langle I, \tfrac{\partial \Psi_{jk}^n}{\partial y} \rangle (V_{\epsilon^j} - V_{jk})_y|^2 \\ &= (\overrightarrow{V}_{\epsilon^j} - \overrightarrow{V}_{jk})^T A (\overrightarrow{V}_{\epsilon^j} - \overrightarrow{V}_{jk}) \\ &\leq \lambda_2 ||\overrightarrow{V}_{\epsilon^j} - \overrightarrow{V}_{jk}||^2 \end{aligned} \tag{18}$$

Thus:

$$E(j,k,\Theta^j) \leq \sqrt{\frac{\lambda_2}{\lambda_1 + \lambda_2}} ||\overrightarrow{V}_{\epsilon^j} - \overrightarrow{V}_{jk}||$$

And, according to the Minkowski inequality,

$$\sum_{n=1}^{N} ||\langle \overrightarrow{\nabla} I, \Psi_{jk}^n \rangle||.|\langle \overrightarrow{\nabla} I.(\overrightarrow{V}_{\epsilon^j} - \overrightarrow{V}_{jk}), \Psi_{jk}^n \rangle| \geq ||A.(\overrightarrow{V}_{\epsilon^j} - \overrightarrow{V}_{jk})|| \geq \lambda_1 ||\overrightarrow{V}_{\epsilon^j} - \overrightarrow{V}_{jk}||$$

Consequently,

$$E(j,k,\Theta^j) = \frac{\sum_{n=1}^{N} ||\langle \overrightarrow{\nabla} I, \Psi_{jk}^n \rangle||.|\langle \overrightarrow{\nabla} I.(\overrightarrow{V}_{\epsilon^j} - \overrightarrow{V}_{jk}), \Psi_{jk}^n \rangle|}{\sum_{n=1}^{N} ||\langle \overrightarrow{\nabla} I, \Psi_{jk}^n \rangle||^2} \geq \frac{\lambda_1}{\lambda_1 + \lambda_2} ||\overrightarrow{V}_{\epsilon^j} - \overrightarrow{V}_{jk}||$$

References

1. T. Aach, L. Duembgen, R. Mester, and D. Toth. Bayesian illumination-invariant motion detection. In *ICIP01*, pages 640–643, 2001.
2. C. Bernard. Discrete wavelet analysis: A new framework for fast optic flow computation. In *ECCV98*, pages 354–368, 1998.
3. J. Besag. On the statistical analysis of dirty pictures. *RoyalStat*, B-48(3):259–302, 1986.
4. P. Bouthemy and P. Lalande. Detection and tracking of moving objects based on a statistical regularization method in space and time. In *ECCV90*, pages 307–311, 1990.
5. M.M. Chang, A.M. Tekalp, and M.I. Sezan. Simultaneous motion estimation and segmentation. *IP*, 6(9):1326–1333, September 1997.
6. D. Cremers. A variational framework for image segmentation combining motion estimation and shape regularization. In *CVPR*, pages 53–58, 2003.
7. F. Heitz, P. Perez, and P. Bouthemy. Multiscale minimization of global energy functions in some visual recovery problems. *CVGIP*, 59(1):125–134, January 1994.

8. P.W. Holland and R.E. WelschHsu. Robust regression using iteratively reweighted least-squares. *Commun. Statist.-Theor. Meth.*, A6:813–827, 1977.
9. Y.Z. Hsu, H.H. Nagel, and G. Rekers. New likelihood test methods for change detection in image sequences. *CVGIP*, 26(1):73–106, April 1984.
10. M. Irani, B. Rousso, and S. Peleg. Detecting and tracking multiple moving objects using temporal integration. In *ECCV92*, pages 282–287, 1992.
11. F. Luthon, A. Caplier, and M. Lievin. Spatiotemporal mrf approach to video segmentation: Application to motion detection and lip segmentation. *SP*, 76(1):61–80, July 1999.
12. S.G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *PAMI*, 11(7):674–693, July 1989.
13. H.T. Nguyen, M. Worring, and A. Dev. Detection of moving objects in video using a robust motion similarity measure. *IEEE Trans. on Image Processing*, 9(1):137–141, January 2000.
14. J.M. Odobez and P. Bouthemy. Detection of multiple moving objects using multiscale markov random fields, with camera compensation. In *ICIP94*, pages 257–261, 1994.
15. N. Paragios, P. Perez, G. Tziritas, C. Labit, and P. Bouthemy. Adaptative detection of moving objects using multiscale techniques. In *Proc. 3rd IEEE Int. Conf. Image Processing, ICIP'96*, Lausanne, Switzerland, September 1996.
16. P. Perez, A. Chardin, and J.M. Laferte. Noniterative manipulation of discrete energy-based models for image analysis. *PR*, 33(4):573–586, April 2000.
17. P.L. Rosin. Thresholding for change detection. *CVIU*, 86(2):79–95, May 2002.
18. E. Sifakis and G. Tziritas. Moving object localisation using a multi-label fast marching algorithm. *Signal Processing: Image Communication*, 16(10):963–976, 2001.
19. W.B. Thompson and T.C. Pong. Detecting moving objects. *IJCV*, 4(1):39–58, January 1990.