MOTION ESTIMATION PROCESS FOR LARGE RANGE AMPLITUDE OF MOTION

Rémy Leconge, Olivier Laligant, Fréderic Truchetet, Alain Diou
Le2i, Université de Bourgogne, France
tel (33)-03-85-73-10-90 Fax (33)-03-85-73-10-97 e-mail: r.leconge@iutlecreusot.u-bourgogne.fr

***Abstract*- For a medical application, we are interested in estimation of optical flow on the area around the patient eyes. Generally the main method of motion estimation can only be applied for small displacements. Our approach is based on a Markov random field model combined with an algorithm of block matching in a multiresolution scheme. The motion tracking is achieved by a block matching algorithm. This method gives the optical flow whatever the amplitude of the motion is, if included in the range defined by the multiresolution approach. The results clearly show the complementarity of Markov random fields estimation and block matching across the scales.**

1.INTRODUCTION

This research was carried out within the framework of a medical application, in which we are particularly interested in the movement of the eyelid. The motion characterization must be accurate because not entirely structured areas with not homogenous speed are studied. The movement observation is done on a sequence of gray levels images. In addition the movements present great differences in amplitude and the low speed of acquisition makes difficult the motion estimation of large amplitudes. A motion is considered as a motion of large amplitude when it exceeds three or four pixels by frame. There are a lot of techniques of movement estimation, but generally they are classified in three main categories: differential methods (methods of Horn and Schunk, Nagel…), mapping methods (block matching), and transform methods (Gabor transform, Fourier transform…). The first methods give good results provided that only movements of small amplitude are considered. The second techniques operate correctly on structured elements and motions of large amplitudes. The last ones are exploitable only for elementary movements. In our case, movements have various amplitudes. So a simple and efficient method to compute motion estimation for a large range of speed is proposed. The complementarity of differential method and block matching allows to characterize all the movements in image sequences. The cooperation of these two methods is achieved by a multiresolution strategy. Section 2 summarizes the results of the theory of Markov random fields model and the principle of block matching applied to motion estimation. Components of the multiresolution strategy are explained in detail as a preamble to the algorithm in section 3. In the last section some results are presented.

2. MOTION ESTIMATION

A. Markov random fields (MRF)

In this section, some results on MRF [1][2][3] applied in motion estimation are presented. These results are obtained from Gibbs fields model and Hammersley Clifford theorem. The central assumption is that the luminance of pixel is constant in two successive images, although some works take into account variation luminance[4]. Therefore, MRF modeling is equivalent to an energy function and the motion estimation problem is reduced to the minimization of this energy function.
Classically the energy function is composed of two parts. The first one is the data energy:

$$\left(f\left(s+p_s^t,t+dt\right)-f(s,t)\right)^2$$

where *f* is an image sequence.

The latter is the regularization energy based on the Tikhonov regularization model usually used for motion estimation:

$$V(s,s_j,p,p_j)=\beta\left\|p_s^t-p_{s_j}^t\right\|^2$$

where
V( ) is the potential function
s is the current site
s_j is a set of sites in the neighborhood of *s*
p_s^t is the motion vector on the site i at the time t. Its components are u_s and v_s .
β is the parameter of regularization, varying from zero to infinity

Concerning the choice of β, small values lead to accurate solution for image data but results are sensitive to noise. On the other hand, high value implies smooth solution but results do not match the data luminance. Nevertheless, detection of movement of subpixel amplitude require small value for β. Best compromise between the detection of the small movements of the eyelid and the elimination of the noise of acquisition system has to be found.
The total energy function is equal to :

$$E(s,t)=\left(f\left(s+p_s^t,t+dt\right)-f(s,t)\right)^2+\sum_{\{s,s_j\}\in C_2}\beta\left\|p_s^t-p_{s_j}^t\right\|^2$$

where C_2 is the second order clique.
The minimum of the energy function is obtained when the derivative form u_s and v_s is equal to zero, assuming that only small displacement is taken into account.

The solution is $p_s=(u_s, v_s)$
With

$$us = \frac{-2\left(2f_x\dot{f} + \sum_{\{s,s_j\}\in C_2} -2\beta us_j\right)\times\left((f_y)^2+4\beta\right) + \left(2f_y\dot{f} + \sum_{\{s,s_j\}\in C_2} -2\beta vs_j\right)\times(2f_x f_y)}{4\left((f_x)^2+4\beta\right)\times\left((f_y)^2+4\beta\right)-(2f_x f_y)^2}$$

$$vs = \frac{-2\left(2f_x\dot{f} + \sum_{\{s,s_j\}\in C_2} -2\beta vs_j\right)\times\left((f_y)^2+4\beta\right) + \left(2f_y\dot{f} + \sum_{\{s,s_j\}\in C_2} -2\beta us_j\right)\times(2f_x f_y)}{4\left((f_x)^2+4\beta\right)\times\left((f_y)^2+4\beta\right)-(2f_x f_y)^2}$$

where
u_s and v_s are the components of p_s^t

$$f_x = f(x+dx, y, t) - f(x,y,t);$$
$$f_y = f(x, y+dy, t) - f(x,y,t);$$
$$\dot{f} = f(x,y,t+1) - f(x,y,t);$$

The expression of $p_s(u_s, v_s)$ is a function of space luminance, temporal gradients and speed of neighbors of the considered pixel. The value of $p_s(u_s, v_s)$ is obtained by an iterative process which converges towards the final value. The estimation of movement by MRF is finally reduced to u_s and v_s computing. All small movements are estimated with this simple calculation. Unfortunately this method leads to incoherent results for great amplitude movements.

B. Block matching process (BMP)

The estimation of motion by block matching [5] consists, choosing a block in an image, in finding the best similar block with respect to a criterion in another reference image. The displacement vector is deduced from the positions of the two blocks. The criterion of similarity between two blocks is generally the quadratic error or difference in absolute value. In our case the criterion is defined as following:

$$C = \sum_B \left|\left(I1 - \overline{I1}\right) - \left(I2 - \overline{I2}\right)\right|$$

where $I1$ and $I2$ are the pixel luminances of the blocks respectively in the previous image and in the current image. $\overline{I1}$ and $\overline{I2}$ are the average pixel luminances of the blocks respectively in the previous image[6] and in the current image. B is the subset of pixels of the block.

The search for the reference block is done within a window whose dimensions are selected according to the detected displacement between current image and reference image.
The strategies of search are numerous, we can quote the binary searching, searching in spiral, or the hierarchical searching.
The block matching allows the estimation of great amplitude movement if initialization and correlation criterion are correctly chosen. Moreover the size of search area must be in relation to motion amplitude.

3. METHOD

It has been shown in a preceding section that estimate of the optical flow by Markov random fields allows only a detection of small movements of amplitude less than 3 pixels, and that block matching detects movements of greater amplitudes if the process of search is well initialized. A multiresolution approach is used to merge these two methods[7][8]. Indeed if a movement is significant on scale 0 (initial image) it will become a movement of low amplitude in scales of coarser resolution. The MRF is used to characterize this small displacement and motion tracking across the scale is achieved by BMP. BMP initialization is obtained with MRF. This Section is organized as follows. First, the pyramidal framework is described, then the propagation of movement estimation across scales and BMP parameters are exposed. Finally our algorithm is presented.

A. Components of pyramidal approach

a. Pyramid construction:

A lot of method like Gaussian[9] pyramid, multiresolution analyze using Spline[10][11] allow to achieve an image pyramid. Many multiresolution schemes have been developed over the last years. These methods are mainly optimized in regard of details needed to obtain the scale j from the scale j+1. We can quote criteria like entropy and aliasing. In this work, no study has been conducted to choose the best multiresolution method. However precaution to limit aliasing has been taken by smooth filtering. The construction of the image pyramid is a fine to coarse process. A pixel at scale j+1 is simply the mean of its 4-neighborhood at scale j. The scale image j+1 of the pyramid is obtained by subsampling. If m*m is the number of pixels of the image 0, the size of the image j is $(m/2^j)* (m/2^j)$ (figure 1).

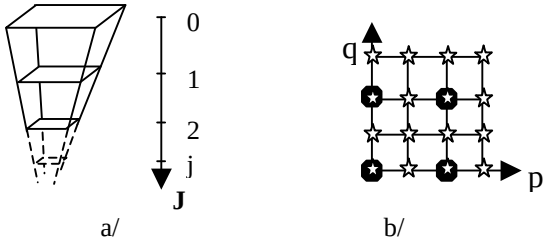

Figure1: (a) Images pyramid construction. (b) Scale j+1 is obtained from scale j. Scale j is composed of 16 pixels ✯. The four black reference mark pixels • (p even and q even) represent the scale j+1.

b. Propagation of movement estimation :

The movement estimation is carried out in a coarse to fine approach. The difficulty of the approach "coarse to fine" is to associate information points of scale j to scale j-1. Indeed, we need to initialize movement image j-1 on the base of movement estimation at scale j.
Between scales j and j-1, the correspondence of motion vectors for the pixels illustrated in black on figure 2 is directly obtained with a scale factor two. All the other points at scale j-1 are deduced from linear combinations as shown on figure 2. Before presenting algorithm, let us clarify the parameters of BMP.

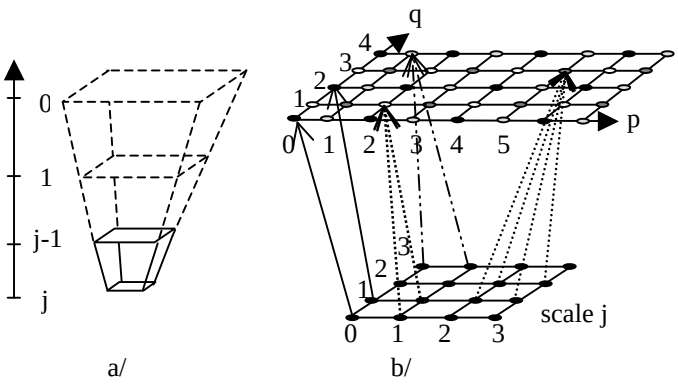

Figure 2: (a) Movement pyramid creation. (b) movement propagation: relation between pixels at scale j and pixels at scale j-1. Black reference mark : p et q are even, white reference mark : p is even and q is odd, light gray reference mark: p is odd et q is even, dark gray reference mark: p et q are odd.

c. Parameters of Block Matching :

As previously shown BMP consists in linking a block of an image to another one in another image. In this work the chosen block size is 3*3 pixels. The search direction for matching is initialized at scale j-1 by propagating motion vector estimated by MRF at scale j. The size of search area is hierarchical and depends on the motion vector determined at scale j. Indeed, if a displacement of one pixel at scale *d* is detected, displacement at scale zero will be of 2*d* pixels. As defined above the correspondence criterion is the mean of luminance pixel to pixel between the two blocks. The block minimizing this difference is selected. Figure 3 represents an example of evolution of the process of BMP if the pyramid is made up on only three scales.

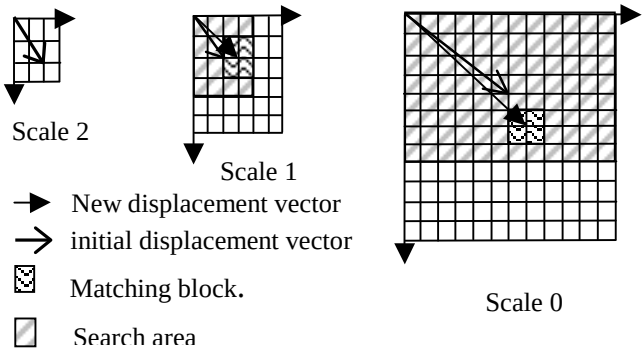

Figure 3: Process of block matching across scales

The size of research area is defined by the displacement vector estimated on the coarser scale. This area is centered on the pixel which luminance is calculated with information of the lower scale. Its height and its width are four time the value found on the coarse scale.

B. Algorithm :

Let be a temporal sequence of (n+1) images

$$\{I^0, I^1,\ldots,I^k,\ldots\}_{k\in[0,n]}$$

where k denote the time
First of all, a multiresolution pyramid of $(P^{k-1}, P^k)_{k\in[0,n]}$ is built following the process described in 3.A.a. All calculations will be carried out on a pair of successive pyramids noted (P^{k-1}, P^k) where:
$P^{k-1}=\{I_0^{k-1}, I_1^{k-1},\ldots,I_J^{k-1}\}$ and $P^k=\{I_0^k, I_1^k,\ldots,I_J^k\}$with J+1 the number of scale.
The task of the movement estimation algorithm is to build gradually the pyramid of movement :

$$\{(U^k_J,V^k_J), (U^k_{J-1},V^k_{J-1}), ..., (U^k_0,V^k_0)\}$$

where (U^k_0,V^k_0) is the expected movement vector image.

The algorithm is organized as follows:

- For each pyramid image, movement estimation is processed by MRF on the coarsest scale (I_J^{k-1}, I_J^k) and produces vector image$(U^k_{j=J},V^k_{j=J})$
- Vectors are propagated on the next finer scale as explained in section 3.A.b to initialize (U^k_j,V^k_j) with j=j-1.
- At the finer scale, two configurations are possible to refine (U^k_j,V^k_j):

The initial vector is null and in this case motion vector is determined by MRF.
A no-null initial vector is found, it is used to initialize the BMP whose parameters are defined in paragraph 3.A.c.

- The algorithm is iterated from step 2 until scale 0. The final result is the vector image (U^k_0 ,V^k_0)

To improve the process, a modification in step 3 is introduced. If motion is detected for a given point at scale j+1, the initial vectors at the scale j will not be null in the 4-neighborhood. Consequently, all these points are checked to know if they correspond to real movement. So, before estimation by the block matching, the propagation of movement is refined by MRF. If a point does not correspond to movement, its initialization is forced to 0. Figure 4 illustrates the utility of control before the application of block matching on the same example. Figure 5 shows an example of two image pyramids where complementarity of MRF and BMP is clear.

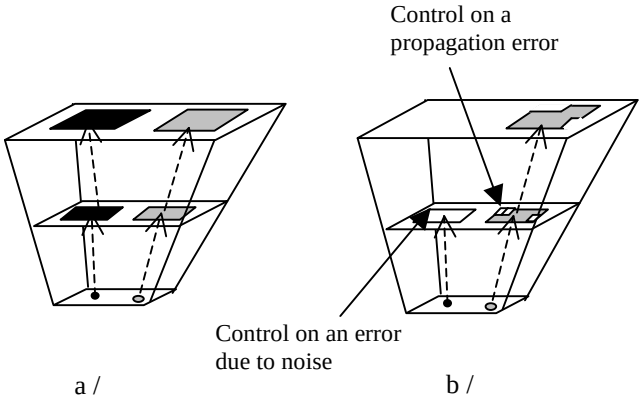

Figure 4: Interest of control. (a) propagation of information without control: the error is propagated. (b) propagation of information with control, to eliminate error. ●pixel where motion is detected by Markov process. ●pixel where false motion due to noise is detected by MRF. ■ set of pixels where block matching is applied, ■set of pixels where block matching is applied, □ set of pixels forced to 0.

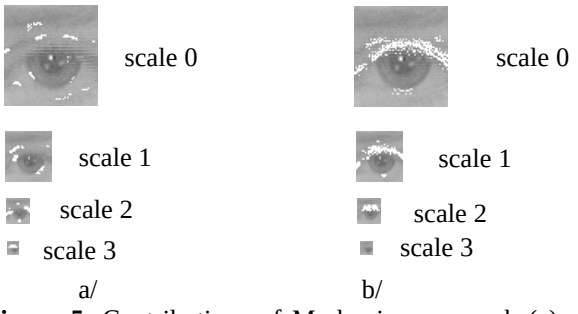

Figure 5: Contributions of Markovian approach (a) and of block matching (b). The white points show pixels where motion is detected.

4. SOME EXPERIMENTAL RESULTS

Results are presented for synthetic and real sequences. In particular, the robustness of the method against the noise is shown, then its behavior on movements of various amplitudes is illustrated and finally some quantitative movement estimates are presented.

A. Noise effect:

Real images are corrupted by the noise of the acquisition system. At first, the noise of the acquisition system is estimated. This system includes optic, CCD and electronic. This noise is classically modelized by a Gaussian distribution and the estimation of the mean variance is achieved by images sequence of homogenous luminance with different gray levels. Measure gives a value of 3 for the norm variance. Subsequently, the robustness of the algorithm is evaluated on noisy synthetic images. This test permits to adjust the parameters, mainly β. Pair of images extracted of the test sequence are shown in figure 6.

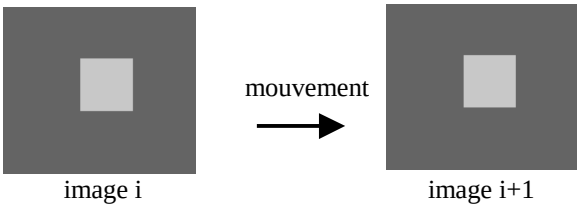

Figure 6: Two successive images of the sequence.

The sequence of images represents a homogeneous square of luminance moving in translation on a homogeneous background (whose luminance is different to the luminance of the square). The movement is a translation of one pixel to the right according to the horizontal axis. The results are consigned in figure 7.
The result on the figure 7a shows that the detection of movement is correct. For noisy sequence with Gaussian noise of variance similar to the noise of acquisition system some vectors present false directions and wrong norms. Obviously the overall movement estimation is satisfactory (figure 7b). If the noise variance increases the result is degraded and can not be easily interpreted. We still recognize the dominating movement, but it seems unreasonable to calculate the norm of the motion vectors in order to determine the speed of the object. Experiment and setting have been conducted and show that the algorithm will not be sensitive to noise added by acquisition system.

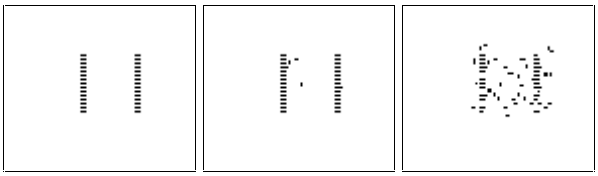

Figure 7: Displacement vectors (a) on the initial sequence without noise, (b) on the sequence with noise (noise variance = 4), (c) on the sequence with noise (noise variance = 9)

B. Speed effect:

An homogeneous disc with a colored sector has been designed to study the behavior of our algorithm for movements of various amplitudes along the image sequence (corresponding to various speed). Then this disc in uniform rotation (angular velocity $\pi/3$ rd/s) has been shot at the rate 25 images per second.
The figure 8c shows that norm of vectors decreases near the center of the disc.

For the visibilitv. one vector out of four vectors is shown.

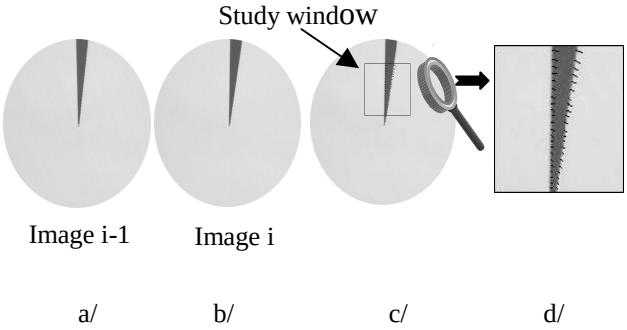

Figure 8: Pair of images extracted from the sequence (a, b) and displacement vectors field associated (c, d).

In figure 9, the contribution of MRF and BMP across the scale is clearly shown.

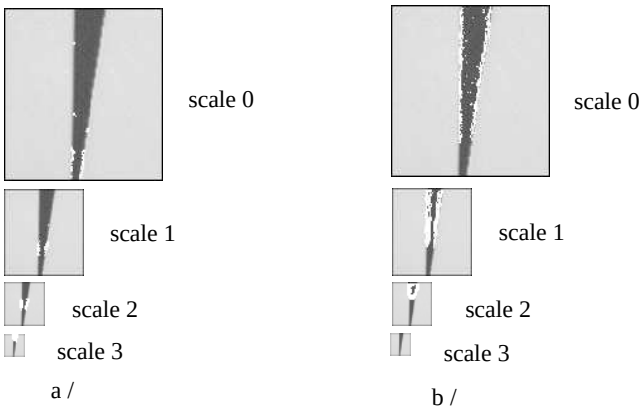

Figure 9: Contribution at different scales of Markovian approach (a), of block matching (b). The white points show pixels over which motion is detected.

Movements with large amplitude are detected in the coarsest scales. As the amplitude of movement decreases (near the center of the disc), the estimation is initialized at finest scales. The efficiency of our algorithm is evaluated using the comparison between experimental and theoretical angular speed (figure 10).

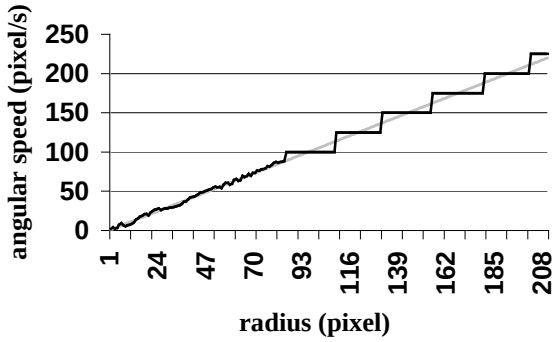

Figure 10: Comparison between theoretical and experimental angular speed. The curve is composed of two parts corresponding to the Markov method and the block matching process. The speed increases by one pixel/s when the radius increase by twenty-three pixels. Motion estimation by Markov random fields is subpixel consequently the experimental curve is similar to the theoretical one. The motion estimation using block matching process is constant at intervals of twenty-three pixels because block matching process has a resolution of one pixel.

C. Examples of application on deformable area:

Figure 10 presents two pairs of images showing the beginning and the closing respectively of eyelid. The algorithm seems reliable for the determination of the direction and norm of the motion vectors. Some estimation errors are due to the noise and the aliasing which will be eliminated in future experimentation with full frame camera. A solution to cope with noise problem would be to increase the value of β, that would imply the disappearance of false vectors. However as specified in paragraph 2.A, if β is too much increased, noise but also small movement vectors will be removed.
The speed of the eyelid is more significant at its beginning than at its end. For each person the speed of the eyelid is different, and can even vary for the same person according to her tiredness or her nervousness and the voluntary or involuntary nature of the movement, so we give just a speed range.
These values were obtained with sequences test in laboratory. It is probable that this interval should be modified if experiences were made with large number of persons. Finally, these values, in their current units (pixels/s) depend on the experimental conditions of acquisition (resolution of the sensor, lens, object/ camera distance, video rate). Our study shows two phases in the movement of the eyelids (see table1).

PHASES OF MOTION	SPEED
Beginning of eyelid movement	150 to 240 pixels per second
End of eyelid movement	30 to 60 pixels per second

Table 1: Eyelid movement estimation. Experimental Conditions: Resolution of video sensor: 768*572 pixels, optics focal distance : 16 mm, object / system distance: 50 centimeters, video rate: 30 images par second.

Beginning of the movement of the eyelid: speed values vary between 5 and 8 pixels per image. With a video rate of 30 image per second, this range corresponds to 150 and 240 pixels per second .
End of the movement of the eyelid: range 1 and 2 pixels per image (30 and 60 pixels per second).
Within these conditions, the speed of eyelid varies between 0.9 centimeter per second and 7.2 centimeters per second.

5. CONCLUSION

A multiresolution scheme for the cooperation of two different approaches in motion estimation has been presented. These two approaches are the Markov fields and block matching. The resulting method is able to estimate various amplitudes of motion in the image sequence. The minimum of motion depends on the trade off between motion and noise while the maximum of motion detected depends on the depth of the multiresolution pyramid. Presented examples clearly show the contribution of each of the two detection methods. The multiresolution method can be applied to objects motion estimation as well as to estimation of motion on deformable surfaces. In this work, application of the movement of the eyelid. Automation of the measures of the movement of the eyelid closing and opening are possible. Perspectives concern the study of the value of the regularization parameters and the construction of the pyramid.

6. REFERENCES

[1] F. Heitz, P. Perez, P. Bouthemy "Constrained Multiscale Markov Random Fields and the Analysis of Visual Motion" *publication interne de l'IRISA* n° 627, (1992).
[2] F. Heitz, P. Bouthemy "Multimodal estimation of discontinuous optical flow using Markov random fields" *IEEE transaction on Pattern Analysis an Machine Intelligence*, pp 1217-1232, (1993).
[3] M. Kardouchi , "contribution à l'étude du mouvement avec les champs de Markov : Régularisation et prise en compte des discontinuité", *Thèse de l'université de Bourgogne*, Dijon (1998).
[4] J.C Pesquet "Motion estimation in the presence of illumination variations", *Signal Processing: Image Communication* vol 16 pp 373-381, (2000).
[5] C.Alexandre and H. Vu Thien, " Fast motion estimation algorithm" *Traitement du signal* , vol. 13, no 4, pp 351-359, (1996).
[6] P. Fua " Combining stereo and monocular information to compute dense depth maps that preserve depth maps discontinuites", *International Joint Conference on Artificial Intelligence*, pp 1292-1298, Sydney (1991).
[7] P. Anandan, " A computational Framework and an algorithm for the measurement of visual motion", *International Journal of Computer Vision*, vol. 2, pp 283-310, (1989).
[8] S. Zafar, Y. Zhang, B. Jabbari "Multiscale Video Representation Using Multiresolution Motion Compensation and Wavelet Decomposition" *IEEE Journal on selected areas in communications* vol. 11, no 1, pp 24-35, (1993).
[9] P. J. Burt, E. H. Adelson, "The Laplacian Pyramid as a compact image code" *IEEE Trans communication* vol COM 31 pp 532-540, (1983).
[10] A. Aldroubi, M. Unser. "Families of Multiresolution and Wavelet Spaces with Optimal Properties". *Numer. Funct. Anal. And Optimiz*. 14(5), pp 417-446, (1993).
[11]M. Unser, A. Aldroubi, and M. Eden . A Family of Polynomial Spline Wavelet Transforms. *Signal Processing*, 30(2), pp 141-162, (1993).
[12] R. Leconge, O. Laligant, F. Truchetet , " Markov random fields and block matching for multiresolution motion estimation " *Proceeding Electronic Imaging* vol 3968, (2000).