Markov random fields and block matching for multiresolution motion estimation

Rémy Leconge, Olivier Laligant, Fréderic Truchetet
Le2i, Université de Bourgogne, France

ABSTRACT

For a medical application, we are interested in estimation of optical flow on the face and particularly on area around the eyes. Among the methods of optical flow estimation, gradient estimation and block matching are the main ones. However the gradient based approach can only be applied for small displacements (one or two pixels). Generally the process of block matching gives only good result if the searching strategy is judiciously selected. Our approach is based on a Markov random field model combined with an algorithm of block matching in a multiresolution scheme. The multiresolution approach leads to detect a large range of displacement amplitude. The large displacements are detected on the coarse scales and the small ones will be detected successively on finer scales in a coarse to fine strategy. The Markov random field allow the initialization and the control of motion estimation across all scales. The tracking of motion 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 show clearly the complementarity of Markov random fields estimation and block matching across the scales.

Keywords: Motion, Block Matching, Markov Random Fields, Multiresolution

1. INTRODUCTION

This research was carried out within the framework of a medical application, in which we are interested particularly in the movement of the eyeball. The characterization of the motion must be precise because not entirely structured area with not homogenous speed are studied. The observation of the movement is done on a sequence of images in gray levels. In addition the movements present great differences in amplitude and the low speed of acquisition makes difficult the estimation of the motion of strong amplitudes. There are a lot of techniques of estimation of movement, but generally they are classified in three main categories: differential methods (methods of Horn and Schunk, Nagel…),methods by mapping (block matching), and methods by transform (Gabor transform, Fourier transform…). The first methods give good results provided that only movements of low amplitude are considered. The seconds operate correctly on structured elements and motion of strong amplitudes. The last are exploitable only for elementary movements. In our case, movements have various amplitudes. Differential method and block matching should be used together. The complementarity of these approaches allows to characterize all the movements in images sequences. A multiresolution strategy to operate cooperatively these two methods is proposed. Section 2 present the bases of the theory of Markov random fields applied to the motion. Then we will present in section 3 the process of block matching. The multiresolution strategy, used the method will be developed in section 4. In the last section we will show our results.

2. MOTION ESTIMATION BY MARKOV RANDOM FIELDS :

In the following action, we point out the theoretical framework of Markov random fields [3.4.6.]and we recall the definition of Gibbs fields used to obtain the distribution of probability of Markov random fields by equivalence. The concept notion of energy function is introduced. The motion estimation is achieved by minimization of this energy function.

2.1. Definition of Markov random fields

A random variable field $\mathrm{P} = (\mathrm{P}s_1, \mathrm{P}s_2, ..., \mathrm{P}s_n)$ defined on a grid S is a Markov random field associated with a neighborhood V if and only if:

- Prob (P = p)>0.

- $\text{Prob}\left(P_{s_k} = p_{s_k} / P_{s_l} = p_{s_l},\ k \neq l,\ s_l \in S - \{s_k\}\right) = \text{Pr ob}\left(P_{s_k} = p_{s_k} / P_{s_l} = p_{s_l},\ k \neq l,\ s_l \in V(s_k)\right)$

  où $p = \left(p_{s_1}, p_{s_2}, ..., p_{s_n}\right)$ is a realization of P_{s_i}

2.2. Definition of Gibbs fields

Definition of the potential: Let Ω a space of configuration on S and A a part of S. A potential on A is an application defined on Ω and depending on random variables x_s ,$s \in A$.

$$V_A : \Omega \to \mathbf{R}$$
$$x \mapsto V_A(x)$$

A random variable field P is a Gibbs field if and only if a neighborhood g on S and a family $V = \{V_c, c \in C\}$ of potentials on the cliques verify

$$\text{Prob}(p) = \frac{1}{Z} e^{(-U(p))} \qquad \text{where Z is a normalization constant}$$

$$Z = \sum_{p \in \Omega} e^{-U(p)}$$

U(p) is called energy function:

$$U(p) = \sum_{c \in C} V_c(p_c)$$

C is the set of cliques c of S, and p_c is the restriction of p to C.

A Markov random field has local proprieties, on the contrary a Gibbs field has global proprieties. Nevertheless the theorem d'Hammersley Clifford establishes an equivalence between these two fields.

Theorem of Hammersley Clifford: X is a Markov field associated to a neighborhood g if X is a Gibbs field. ▪

The consequences of this theorem are important; it allows to obtain the distribution of probability Prob(p) associated with a Markov random field.

2.3. Estimator of the probability of Markov random fields:

Definition: **observation field**

Let an image sequence *f*, each image is defined by :

$f_i = \{ f(r, i),\ r \in R \}$ where *f*(r, i) is the gray level on the site r of R at the time i.

If we suppose that there are *nb* images in the sequence we obtain :

$$\forall r \in R, \forall i \in \{t, t+dt, ..., t+nb \times dt\}, \quad f_i(r,i) \in \{0,255\}$$

the motion is estimated on a pair of successive images, we can consider this pair as the observation field. ▪

Notation : the observation field at time t is : $O^t = \{O_r^t, r \in R\}$

the realization of O^t : $o^t = \{o_r^t = (f(r,t), f(r,t+dt), r \in R)\}$

f(r, t) et *f*(r, t+dt) are respectively the gray level on the pixel r at time t and t+dt.

Ω the set of all possible realizations of O^t : $\Omega = \{0,...,255\}^{\text{card } R} \times \{0,...,255\}^{\text{card } R}$

Definition: **primitive field**

The realization of the primitive field is equivalent to the displacement vector field. ▪

Notation : the primitive field at time t is : $P^t = \{P_r^t, r \in R\}$

the realization of P^t is equal to : $p^t = \{p_r^t, r \in R\}$

The estimator associated with the Markov random fields is

$$\tilde{p} = \text{argmin}_{p \in \Omega}(U(p/o)) \quad \text{with } U(p/o) = U(o/p) + U(p)$$

2.4. Energy function and its minimization :

The motion estimation problem is therefore reduced to the minimization of energy function.
The energy function $U(o^t/p^t)$ is the sum of all local energy functions $U_s(o_s^t/p_s^t)$ on all site s.

$$U(o^t/p^t) = \sum_{s \in S} U_s(o_s^t/p_s^t)$$

The luminance of one pixel is assumed steady during two successive images:

$$f(s + p_s^t, t + dt) - f(s,t) = 0$$

In fact, we minimize on each pixel $s \in S$: $\left(f(s + p_s^t, t + dt) - f(s,t)\right)^2$

This term will be the first expression of $U_s(p_s/o_s)$. The second is the regularization energy and is based on the Tikonov regularization model usually used for motion estimation.

$$V_2(s_i, s_j, p_i, p_j) = \beta \left\| p_{s_i} - p_{s_j} \right\|^2$$

Where β is the parameter of regularization, varying from zero ad infinitum. Small value leads 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.
There are small movements among displacements to detect. To take into account these small displacements, we choose a small value for β. We have to find a good compromise between the detection of the small movements of the eyelid and the elimination of the noise of acquisition system.
The energy function is equal to :

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

The minimization of the energy function $U(p_s^t/o_s^t)$ leads to the estimation of displacement p_s^t :

$$p_s^t = \arg\min_{p_s^t \in \Omega_s} \left(U(p_s^t/o_s^t)\right)$$

We find the global minimum of the energy function $U(p_s^t/o_s^t)$.

The second term of the energy function is convex but not the first one, due to the variations of $f(s + p_s^t, t + dt)$ during the minimization. So the energy function $U(p_s^t/o_s^t)$ is not convex.
To avoid local minimum during the minimization stage, only small displacement are assumed:

$$f(x + u, y + v, t + 1) = f(x, y, t) + f'_x u + f'_y v + \dot{f}$$

Thus, the first term becomes convex:

$$U_s(o_s^t/p_s^t) = \left(f(x + u, y + v, t + 1) - f(x, y, t)\right)^2 = (f'_x)^2 u^2 + (f'_y)^2 v^2 + (\dot{f})^2$$

where $f'_x = f(x + dx, y, t) - f(x, y, t)$;

$f'_y = f(x, y + dy, t) - f(x, y, t)$;

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

The regularization energy U(p) is :

$$U\left(p^t\right)=\sum_{s\in S}\sum_{\{s,s_j\}\in C_2} U_s\left(p_s^t\right)=\sum_{\{s,s_j\}\in C_2}\beta\left\|p_s^t-p_{s_j}^t\right\|^2$$

recall $p_s^t=(u_s,v_s)$ and $p_{s_j}{}^t=(u_{s_j},v_{s_j})$

$$U_s\left(p_s^t\right)=\sum_{\{s,s_j\}\in C_2}\beta\left(u_s^2+v_s^2+u_{s_j}^2+v_{s_j}^2-2u_{s_j}u_s-2v_{s_j}v_s\right)$$

$$U_s\left(p^t\right)=4\beta u_s^2+4\beta v_s^2-2\beta u_s\sum_{\{s,s_j\}\in C_2}u_{s_j}-2\beta v_s\sum_{\{s,s_j\}\in C_2}v_{s_j}+\beta\sum_{\{s,s_j\}\in C_2}\left(v_{s_j}^2+u_{s_j}^2\right)$$

Finally :

$$U\left(p_s^t/o_s^t\right)=\left[\left(f_x'\right)^2+4\beta\right]u_s^2+\left[\left(f_y'\right)^2+4\beta\right]v_s^2+2f_x'f_y'u_sv_s+\left(2f_x'f_y'+\sum_{\{s,s_j\}\in C_2}-2\beta u_{s_j}\right)u_s$$
$$+\left(2f_x'f_y'+\sum_{\{s,s_j\}\in C_2}-2\beta v_{s_j}\right)v_s+\sum_{\{s,s_j\}\in C_2}\beta\left(u_{s_j}^2+v_{s_j}^2\right)+\left(\dot f\right)^2$$

The minimum of $U\left(p_s^t/o_s^t\right)$ is obtained when the derivative is equal to zero.
The solution is :

$$u_s=\frac{-2\left(2f_x'\dot f+\sum_{\{s,s_j\}\in C_2}-2\beta u_{s_j}\right)\times\left(\left(f_y'\right)^2+4\beta\right)+\left(2f_y'\dot f+\sum_{\{s,s_j\}\in C_2}-2\beta v_{s_j}\right)\times\left(2f_x'f_y'\right)}{4\left(\left(f_x'\right)^2+4\beta\right)\times\left(\left(f_y'\right)^2+4\beta\right)-\left(2f_x'f_y'\right)^2}$$

$$v_s=\frac{-2\left(2f_y'\dot f+\sum_{\{s,s_j\}\in C_2}-2\beta v_{s_j}\right)\times\left(\left(f_x'\right)^2+4\beta\right)+\left(2f_x'\dot f+\sum_{\{s,s_j\}\in C_2}-2\beta u_{s_j}\right)\times\left(2f_x'f_y'\right)}{4\left(\left(f_x'\right)^2+4\beta\right)\times\left(\left(f_y'\right)^2+4\beta\right)-\left(2f_x'f_y'\right)^2}$$

The expression of p(u, v) is a function of the space and temporal gradients of luminance and the speed of neighbors of the pixel considered. The value of p(u, v) is obtained by an iterative process which converges towards the final values. The estimation of movement by Markov random fields is finally reduced to calculation of u_s and v_s. All small movements of images are detected with this simple calculation. Unfortunately this method can not be applied to movements of great amplitude and leads to incoherent results.

3. THE BLOCK MATCHING PROCESS

The estimation of motion by block matching[1.] consists, knowing a block in an image, in finding the best similar block with respect to a criterion in another image reference (figure 1). The vector displacement is deduced from the position of the two blocks. Form the position of the two blocks, a vector displacement is deduced. The criterion of similarity between two blocks are generally the following: the quadratic error or the difference in absolute value (it is the sum of the absolute values of the difference of luminance between each pixel). The search for the reference block is done within a window whose dimensions are selected according to the detecting displacement between the current image and the image of reference. The strategies of search are numerous, we can quote the binary search, search in spiral, or the hierarchical search.

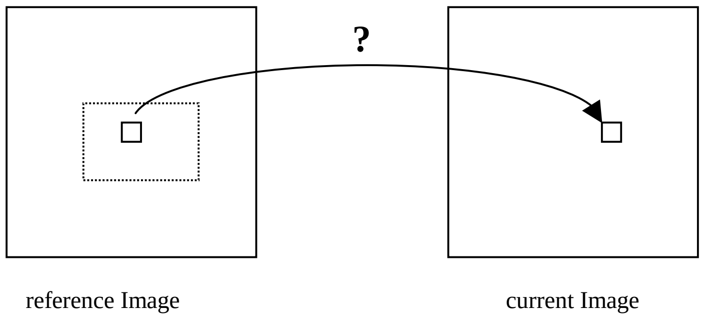


Figure 1 : Principle of the block matching process

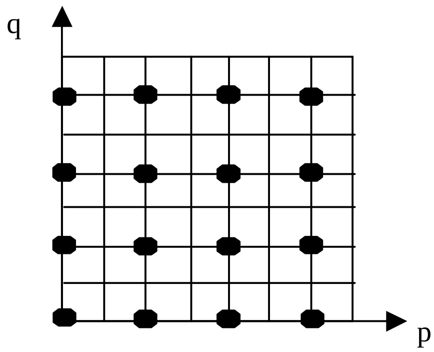
The block matching allows the movement estimation of great amplitude, provided that the initialization and the criterion of correlation are correctly chosen. Moreover the size of search area must correspond to motion amplitude.

4. METHOD

It has been shown in a preceding paragraph that estimate of the optical flow by Markov random fields allows only a detection of small movements, 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 [2.5.] . Indeed if a movement is significant on scale 0 (initial image) it will become a movement of low amplitude in the scales of coarser resolution. We use the Markov Model to characterize this small displacement. Then the process of block Matching is used in the different scales for the motion tracking. Initialization of the Matching Block is obtained with Markov Model.

4.1. Pyramid construction:

The construction of the image pyramid is a fine to coarse process. Pixel at scale i+1 simply the mean of its 4-neighborhood at scale i. The scale image i+1 of the pyramid is obtained by sub sampling. If n*n is the number of pixel of the image 0, the dimension of the image i is $(n/2^i)*(n/2^i)$ (figure 2).



scale i is composed of 64 pixels
The sixteen black reference mark pixels (p even et q even) represent the scale i+1

Figure 2 : creation of the scale i+1 from the scale i

4.2. Propagation of movement estimation :

As shown in section 4.4, the estimation of movement is carried out with a coarse to fine approach. The difficulty of the approach "coarse to fine" is to associate information points of scale i+1 to scale i. Indeed, we need to initialize movement image i on the base of movement estimation of the scale i+1.

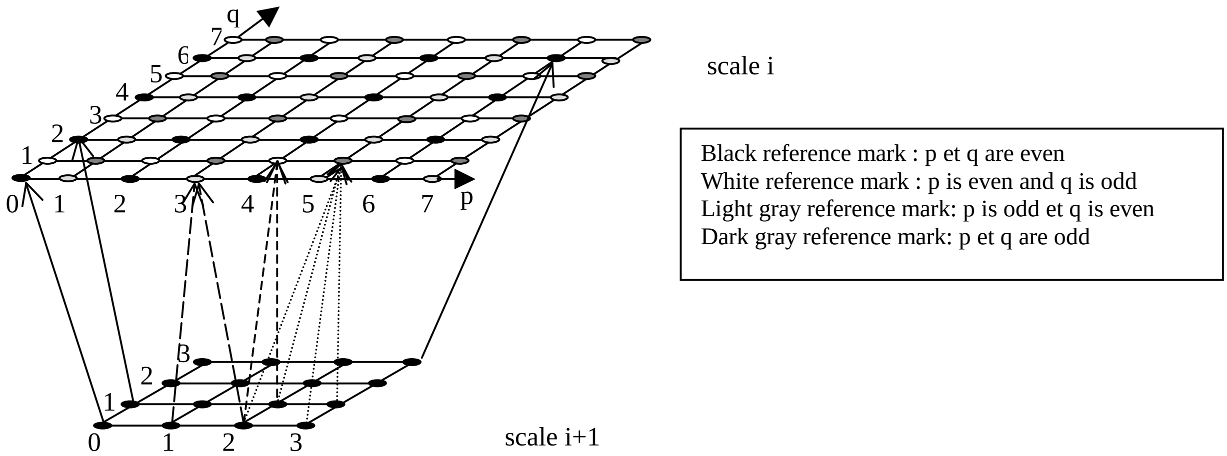


Figure 3 :relation between pixels of scale i+1 and pixels of scale i

Between the scales i+1 and the scale i, the correspondence of the motion vectors for the pixels illustrated in black is obtained with a scale factor two. All the other points of scale i are obtained from linear combinations as shown on figure 3
Before presenting algorithm, let us clarify the parameters of the algorithm of block matching.

4.3. Parameters of Block Matching :

As we have seen previously the block matching consists in linking a block of an image to another image. The block size chosen is 3*3 pixels. The search direction of matching is initialized at the scale i by the propagation of the motion vector estimated by Markov random fields at the scale i+1. The area of search is hierarchical and depends on the motion vector. Indeed, if a displacement of a pixel on scale N is detected, displacement on scale zero will be of 2*n pixels. The criterion of correspondence is the difference of luminance pixel to pixel between the two blocks. The block which is retained minimizes this difference. Figure 4 represents an example of evolution of the process of block matching if the pyramid is made up only on three scales.

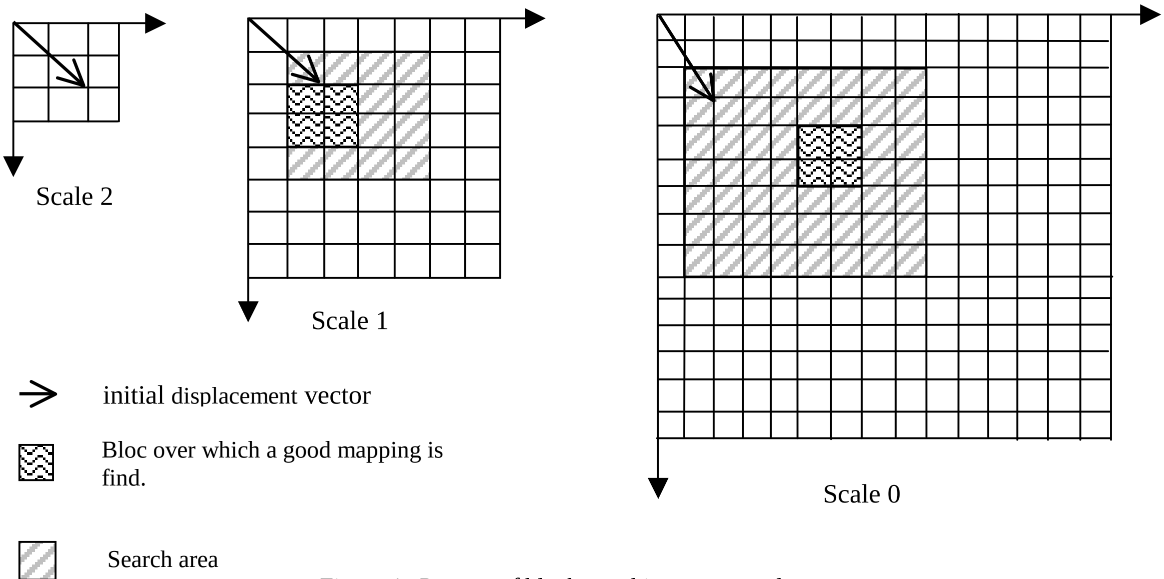


Figure 4 : Process of block matching across scales

4.4. Algorithm :

We work on a sequence of (n+1) images $\{I^0, I^1,\ldots,I^n\}_{i \in [0,n]}$, but all calculations will be carried out on a pair of successive images noted (I^{i-1}, I^i) to clarify the explanations.

First of all, a multiresolution pyramid of $(I_k^{i-1}, I_k^{i})_{k \in [0,N]}$ where N is the index of the coarsest scale is built. This multiresolution pyramid is in conformity with the explanations given in paragraph 4.1

The movement estimation algorithm permits to build the pyramid of movement $(U^i_k, V^i_k)_{k \in [0,N]}$ where (U^i_0, V^i_0) is the expected movement image vector.
The algorithm is organized as follow:

1. For each pyramid image, movement estimation is processed by Markov fields on the coarsest scale (I_N^{i-1}, I_N^{i}) and produces vector image$(U^i_{k=N}, V^i_{k=N})$
2. Vectors are propagated on the next finer scale as explained in section 4.2 to initialize (U^i_k, V^i_k) with k=k-1.
3. At the upper scale, two configurations are possible to refine (U^i_k, V^i_k):
   - The initial vector is null and in this case motion vector is determined by Markov random fields.
   - A no-null initial vector is used to initialize the search of the block matching whose parameters are defined in paragraph 4.3.
4. The algorithm is iterated from step 2 until scale 0. The final result is the vector image(U^i_0, V^i_0)

a : Contribution of Markovian approach, the white points show the pixels over which Markovian approach detect motion.

b : Contribution of block matching, the white points show the pixels over which the block matching detect motion.

Figure 5 : Contribution of the two methods for various scales

A stress has been added to the step 3. If motion is detected for a given point at the scale k+1, the initial vectors at the scale k will be not null in the 4-neighborhood. Consequently, we have to check if all these points correspond to real movement. So, before the application of the block matching, the propagation of movement is refined by Markov random fields. If a point does not correspond to movement, its initialization is forced to 0. Figure 5 shows an example of two image pyramids where complementarity of Markov fields and block Matching is clear. Figure 6 illustrates the utility of control before the application of block matching on the same example.

a: propagation of information with control
The white points show the pixels over which we apply the block matching after control.

b: propagation of information without control
The white points show the pixels over which we apply the block matching without control.

Figure 6 : interest of control

Le principle of the algorithm is presented in the figure 7.

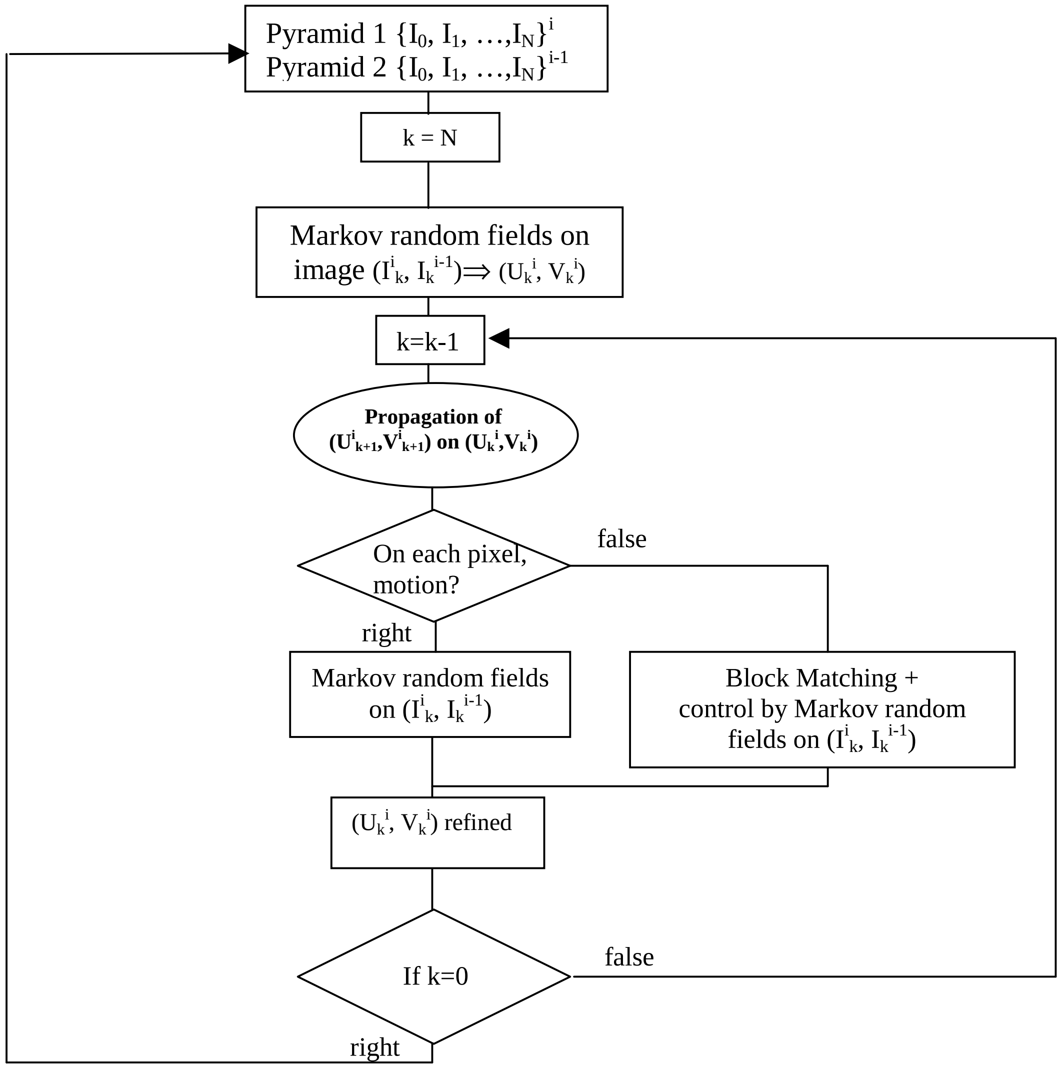


Figure 7 : algorithm

5. RESULTS OF THE METHOD

Results are presented for sequences of different images. In particular, the robustness of the method opposite the noise is shown, then its behavior on movements of various amplitudes, finally we will give measurements.

5.1. Noise effect:

Our algorithm on noisy synthetic images is applied to evaluate its robustness of the method on real images. For this reason the acquisition noise variance has been estimated. Then Gaussian noise has been added to synthetic images sequence. Pair of images extracted of the test sequence are shown in figure 8. 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 a pixel to the right according to the horizontal axis. The results are consigned in figure 9.

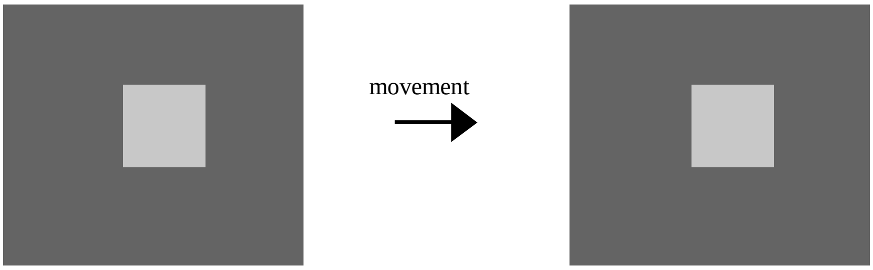
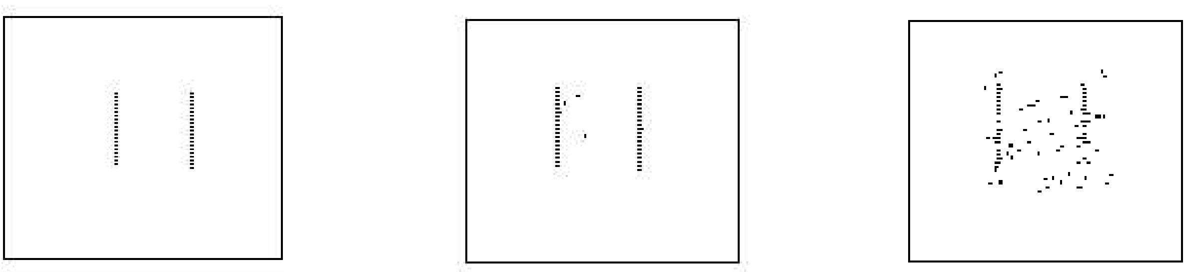


image i　　　　　　　　image i+1

Figure 8 : Pair of images extracted from the sequence



a: Displacement vectors associated with the initial sequence
Without noise

b: Displacement vectors associated with the sequence with noise
noise variance =4

c: Displacement vectors associated with the sequence with noise
noise variance =9

Figure 9 : displacement vectors field according to the acquisition noise variance

The result on the figure 9a 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 norms. Obviously ,the overall movement estimation is satisfactory(figure 9b). If the noise variance increases the results 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. This test show that the algorithm will not be sensitive to noise added by acquisition system. The noise variance (Gaussian model) of the system is about 3.

5.2. Result obtained opposite movements of various amplitudes:

To know the behavior of our algorithm for movements of various amplitudes on image sequence, an homogeneous disc with a sector of different color has been designed. Then this disc has been shot (25 images per second) in uniform rotation. Its angular velocity was about $\pi/3$ radiant per second (figure 10).

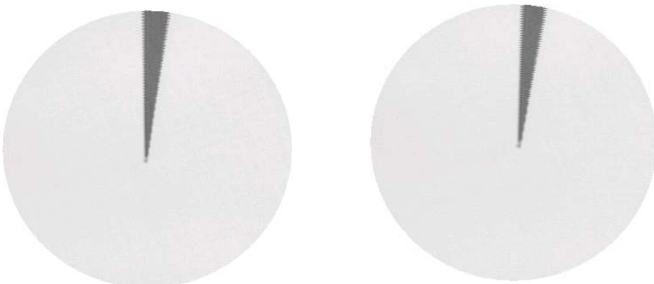


Image t　　　　　　　　Image t+1

Figure 10 :Pair of images extract from the sequence

The algorithm has been applied on an observation windows as illustrated in figure 11. The result is given on figure 11:

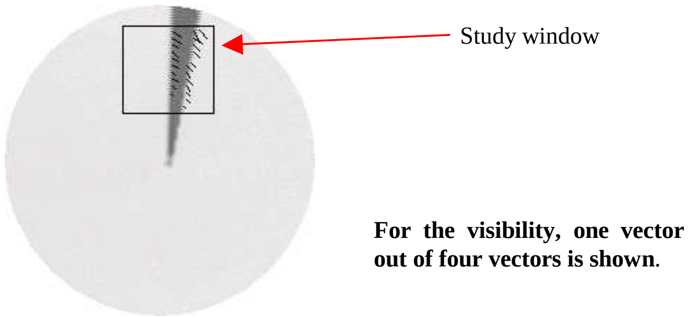


Figure 11 :displacement vectors field

The figure 11 shows that norm of vectors decrease near the center of the disc. In figure 12, the contribution of Markow field and block matching across the scale is shown.

a: Contribution of Markovian approach in the different scales. The white points show the pixels over which Markovian approach detect motion.

b: Contribution of block matching in the different scales The white points show the pixels over which the block matching detect motion.

Figure 12 : Contribution of two methods at different scales

This figure shows clearly that small movements near the center of the disc are detected on the coarsest scales. As the amplitude of movement increases near the border of the disc, the estimation is initialized at finest scales.

5.3. Examples of application on deformable area:

Figure 13 presents two pair of images of the beginning and the closing respectively of the eyelid.

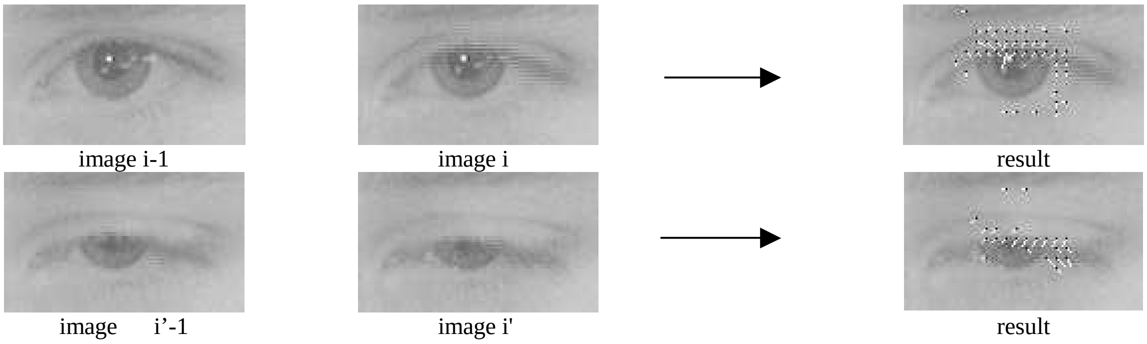


Figure 13 : images and results for blinking of eyelid, For visibility. Only one vector out of four vector is shown. Horizontal aliasing is due to the interlacing

Our algorithm seems relatively 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 by using full frame camera. A solution with noise problem would be to increase the value of the parameter β, that would involve the disappearance of these vectors. But as specified in paragraph 2.4, if we increase too much β noise but also small movement vectors will be removes.

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, and the range only is given.

These values were obtained with sequences in the 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, objective, object/ camera distance, rate video). Our study shows two phases in the movement of the eyelids:

- Beginning of the movement of the eyelid: speed values varies between 5 and 8 pixels per image. With a rate video of 30 image per second, this range correspond to 150 and 240 pixels per second .
- End of the movement of the eyelid: range 1 and 2 pixels per image and corresponds to 30 and 60 pixels a second.

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

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

Within these conditions, the speed of eyelid varies between 0.9 centimeter par second and 7.2 centimeters per second.

6. CONCLUSION

The method presented is able to estimate simultaneously the small movements as well as the movements of larger amplitude. In our case, it is applied to the movement of the eyelid, but it could be used for any application where there are small and great displacements. The prospects for this study are to improve the criterion of mapping, to determine in a rigorous way the value of the regularization parameter and to try the wavelet transform in construction of the pyramid of the images

7. REFERENCES

1. C.Alexandre and H. Vu Thien, " Fast motion estimation algorithm" traitement du signal , vol. 13, no 4, pp 351-359, 1996.

2. 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.
3. 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
4. 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.
5. S. Zafar, Y. Zhang, B. Jabbari "Multiscale Video Representation Using Multiresolution Motion Compensation and Wavelet Decomposition" IEEE vol. 11, no 1, pp 24-35, 1993
6. M. Kardouchi , Thèse de l'université de Bourgogne, "contribution à l'étude du mouvement avec les champs de Markov : Régularisation et prise en compte des discontinuité", Dijon 1998.