Asian Journal of Information Technology 6 (3): 296-302, 2007

Basic Methods for Motion Detection in Images Sequence

[1]T. Bouden and [2]N. Doghmane
[1]Department of Engineering, Automatic Faculty, NDT Laboratory,
BP°98 Ouled Aissa Jijel University, 18000, Algeria
[2]Department of Electronic, Campus Sidi Ammar 23000,
Annaba University, Algeria

Abstract: In the physical world, motion segmentation of images sequences is based on visual motion perception. This does not depend on prior interpretation or recognition of shape and form. However, it does depend on motion information (spatiotemporal object-environment relations). It is generally recognized that the analysis of moving objects proceeds in four stages: The first is the detection of variations in intensity over time in the environment. The second is the segmentation of moving areas and objects masks building. The third is the estimation of motion parameters. The fourth one is the 3D motion interpretation. In the study, we are dealing with detection and region-based segmentation methods. These methods may easily extend to estimate motion parameters. Here we are mainly concerned with comparing studies using determinist and stochastic modelling (images difference, maximum likelihood detector and Markov random field model) to detect the moving objects masks.

Key words: Images sequence, segmentation, detection, moving objects, Markov model, maximum likelihood, site, clique, determinist and stochastic relaxation

PROBLEM POSITION

The problem here is to discriminate starting from the signal intensity obtained in each pixel of the images sequence, the pixels where a change of this last occurred and to be able to rebuild all the area of the changes due to the same movement; who can be homogeneous or not?

DETECTION BY IMAGES DIFFERENCE

To be able to detect a temporal change of the function intensity I at the pixel (p,t), we must be interested in the temporal difference (FD: Frames Difference) between two images at time t and t-dt where dt is the temporal sampling step:

$$\forall p \, FD(p,t) = I(p,t) - I(p,t-dt) \qquad (1)$$

We can detect the changes as moving areas, generally called the masques by a simple thresholding (Jain *et al*., 1979; Wiklund and Grunlund, 1987; Lalande, 1990; Bouthemy and Odobez, 1995; Bouden, 1995; Alice, 1997; Thomas, 2005; Qiang *et al*., 2005) defined by:

$$\forall p \; CT(P) = \begin{cases} 1 & \text{si } |FD(p)| \geq \lambda \\ 0 & \text{ailleurs} \end{cases} \qquad (2)$$

With λ is a fixed threshold follows the thresholding procedure.

MAXIMUM LIKELIHOOD DETECTION

They are founded on the tests of probability developed by Yakimovsky (1976) for the segmentation of fixed images; they were adapted to the temporal changes detection by Jain and Jian (1988) and were also largely developed in (Lalande, 1990; Bouthemy and Odobez, 1995; Bouden, 1995; Alice and Franck, 1997; Thomas, 2005; Qiang *et al*., 2005).

A_1 and A_2 are two same windows centred respectively in (p0,t) (pixel (i0, j0) of the image t of the sequence) and (p0, t-dt) (pixel (i0, j0) of the image t-dt of the sequence), the Likelihood test probability takes into account the following assumptions:

1/ H0: The two windows have not the same distributions of I ==> manifestation of non temporal change (the stationary situation of I).

Corresponding Author: T. Bouden, Department of Engineering, Automatic Faculty, NDT Laboratory, BP°98 Ouled Aissa Jijel University, 18000, Algeria

2/ H1: The two windows have the same distributions of I ==> manifestation of temporal change.

For each assumption a likelihood function is associated. The report/ratio as of these functions, maximized and then compared with a threshold; allows to detect or not, a temporal change. The distribution of the function I can be modelled by several models:

Constant area modelling: For each area, the intensity is supposed to follow the following model:

$$I_{area} = \text{Const}(\mu) + \text{noise} \quad (3)$$

The noise is supposed to be Gaussian and centred $N(\mu,\sigma_2)$. The Likelihood function is given by:

$$L = \left(\frac{1}{2\pi\sigma_2}\right)^{\frac{N}{2}} \exp - \frac{\sum_{x=1}^{N}(I(p_x)-\mu)^2}{2\sigma^2} \quad (4)$$

$I(p_x)$: The observed intensity at the pixel p_x.

We consider two windows A_1 and A_2 with size N = nxn, respectively centred at (p_0, t) and $(p_0, t+dt)$. The two passed hypotheses can be written as:

1/ H_0: All pixels of A_1 and A_2, have the same distribution (same parameters $N(\mu_0,\sigma_0^2)$).
2/ H_1: Pixels of A_1 have the distribution $N(\mu_1,\sigma_1^2)$; and pixels of A_2, have the distribution $N(\mu_2,\sigma_1^2)$; with $\mu_1 \neq \mu_2$. The decision here applies only to the centred pixel p_0; the likelihood ratio is so given by:

$$R_1 = \frac{K_1 \exp - \frac{\sum A_1 I_{\mu 1}}{2\sigma_1^2} K_2 \exp - \frac{\sum A_2 I_{\mu 2}}{2\sigma_1^2}}{K_0 \exp - \frac{\sum A_1 \cup A_2 I_{\mu 0}}{2\sigma_0^2}} \quad (5)$$

$$\text{With: } I_{\mu i} = (I(p) - \mu_i)^2$$

This expression can be simplified by considering the equality of the variances ($\sigma_0^2=\sigma_1^2=\sigma$). After development and simplification of calculations (Lalande, 1990; Bouthemy and Odobez, 1995; Bouden, 1995; Alice and Franck, 1997; Fulvio, 2001; Thomas, 2005; Qiang *et al*., 2005) we will lead to the criterion of detection according to:

$$\forall\, p \quad CT(p) = \begin{vmatrix} 1 & \text{Si } H_1 \\ 0 & \text{Si } H_0 \end{vmatrix} \quad (6)$$

This mode of detection is not very sensitive to the noise and remains from the point of view time real, easy with the establishment.

Linear modelling: For each area, the intensity varies according to the polynomial law of order 1 (i.e.,: bilinear):

$$I_{area} = \phi_1 + \phi_2\Delta_i + \phi_2\Delta_j + \text{noise} \quad (7)$$
Δ_i and Δ_j Spatial variations.

This modelling authorizes a continuous variation of the intensity function in a given area. We will proceed in the same way that previously, we have obtained the same Eq. 4 and 5. The sensitivity of this detector is related to the complexity of the model. The results showed, that this method offers a good compromise (sensitivity to the noise and cost of calculation).

MARKOV RANDOM FIELD DETECTION

In this part, we will point out the broad outline of the Markov Random Field (RMF) modelling used for motion detection. This modelling is associated to the optimization algorithms which remain expensive in times computing in unquestionable applications (Lalande, 1990; Bouthemy and Odobez, 1995; Bouden, 1995; Alice and Franck, 1997; Fulvio, 2001; Thomas, 2005; Qiang *et al*., 2005). The observations are the differences between two consecutive frames (Fig. 1):

$$o(i,j,t) = |I(p,t)\text{-}I(p,t\text{-}dt)| \quad (8)$$

The labels take their values in {a,b}=Λ, indicate the absence or the presence of movement of each site S (pixels p at the moment t):

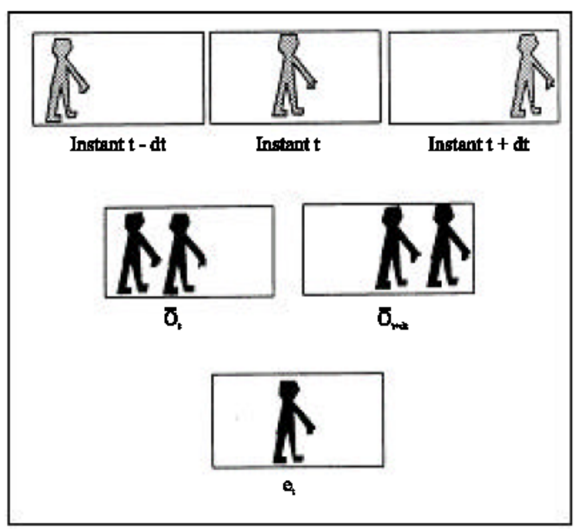


Fig. 1: Realization of the relation between E and O

$$e(p,t)=\begin{cases} a \Rightarrow \text{Pixel.}(p,t)\in \text{moving area} \\ b \Rightarrow \text{Pixel.}(p,t)\in \text{Static area} \end{cases} \quad (9)$$

We consider here a frame S at time t with the site s there (S (s = (p, t)); we note now:

E = {E(s), s∈S}: Labels Field at time t.
E = e : Particular realisation of E.
O = {O(s), s∈S}:Observations field at time t.
O = o Particular realisation of O.
Ω = {a_i, b}: Labels set.

The relation between the observations and labels, takes the following generic form:

$$o = \psi(e) + n. \quad (10)$$

ψ :derived from I(p, t) at t.

$$\psi(e_t,e_{t-dt})=\begin{cases} 0 \ \text{ si } e_t = e_{t-dt} = b\text{..object statique} \\ ml > 0 \text{ si } e_t = e_{t-dt} = a\text{..objet mobile} \\ m2 >> ml \text{ si } e_t \neq e_{t-dt}\text{..transition} \end{cases} \quad (11)$$

n: Gaussian noise of zero mean and variance σ^2 constant for the image.
η is a statiotemporal neighborring for s and C.

The set of all the cliques in S associated to η are illustrated in the Fig. 2:

$$C = C_S + C.$$

With:
C_S : Set of spatial cliques (08 neighbours pixels).
C_T: : Set of temporal cliques (02 per pixel).

The neighbour pixel r(i+v, j+w, t+dt) ∈η, r≠s and the values of (v, w, dt) are in {-1, 0, +1}.

In this modelling, we suppose that the labels field is a Markov field associated to the neighbouring system η. Therefore, using Hammersley and Clifford theorem (Lalande, 1990; Bouthemy and Odobez, 1995; Bouden, 1995; Alice and Franck, 1997; Fulvio, 2001; Thomas, 2005; Qiang *et al*., 2005) it follows a Gibbs distribution:

$$P(E=e)=\frac{1}{Z}\exp\{-U_m(e)\} \quad (12)$$

Z : Normalisation constante.
$U_m(e)$: Energy spatiotemporal function, ensures the space-time homogeneity of the masks and eliminates the isolated pixels due to the noise:

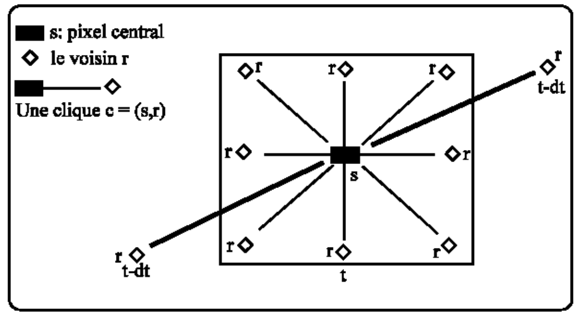


Fig. 2: Neighbouring system

$$U_m(e)=\sum_{c\in C_s} V_c(e_s,e_r)+\sum_{c\in C_t} V_c(e_s,e_t) \quad (13)$$

With:

$$V_C(e_s,e_r)=\begin{cases} -\text{cont.}b_s\text{..Si..}e_s = e_r \\ +\text{cont.}b_s\text{..Si..}e_s \neq e \end{cases} \quad (14)$$

Cont = √2 Vertical spatial cliques.
Cont = 1 Diagonal spatial cliques.

For a simple temporal passed clique:

$$V_C(e_s,e_r)=\begin{cases} -b_p\text{..Si..}e_s = e_r \\ +b_p\text{..Si..}e_s \neq e \end{cases} \quad (15)$$

For a simple temporal future clique:

$$V_C(e_s,e_r)=\begin{cases} -b_f\text{..Si..}e_s = e_r \\ +b_f\text{..Si..}e_s \neq e \end{cases} \quad (16)$$

As for the energy of adequacy between the observations and the labels, which plays the part of a constraint of smoothing of the fields of labels, is defined by:

$$U_a(e,o)=\frac{1}{2\sigma^2}\sum_{s\in S}\left([o_t-\psi(e_t,e_{t-1})]^2+[o_t-\psi(e_t,e_{t-1})]^2\right) \quad (17)$$

The global energy is given by:

$$U = U_m(e) + U_a(e,o) \quad (18)$$

To solve our problem; we use the criterion of the Maximum A Priori (MAP) drift of the theorem of Bayes (11-15). From Eq. 12, 18 we can easily have:

$$\text{Max } P[E = e/O=o] \Leftrightarrow \text{MaxP } P[E = e,O = o]$$

$$\Leftrightarrow \text{MinU}.$$

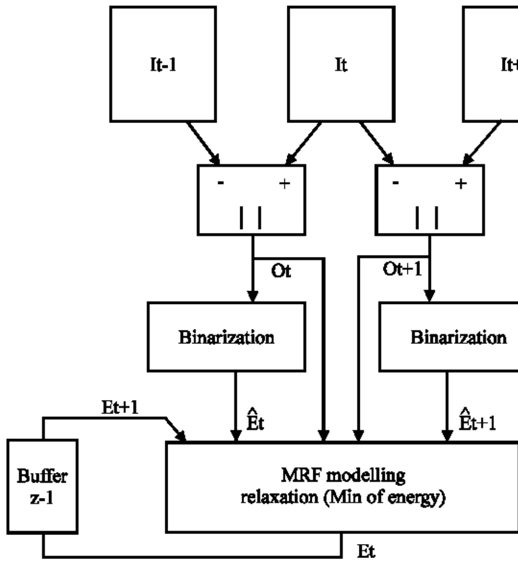


Fig. 3: Optimization algorithms

To find the minimum of total energy, we use the relaxation algorithms which are deterministic such as ICM (Iterated Conditional modes) or stochastic such as simulated annealing (Lalande, 1990; Bouthemy and Odobez, 1995; Bouden, 1995; Alice and Franck, 1997; Fulvio, 2001; Thomas, 2005; Qiang *et al.*, 2005).

Determinist relaxation algorithm RD: Consist in seeking the configurations sites and labels (s, e) corresponding to optimal energies

(Find (s, e)? ➔ MinU(e,o)).

These algorithms (RD) can be trapped by local minimas. Their unfolding is as follows (ICM):

- To choose a site s.
- Calculation of U for any E possible of site s;
- To retain the label e which leads to the weakest energy?

Stochastic relaxation algorithm RS: The stochastic methods (RS) of simulated type reheated are combinations of sampling according to the distribution Gibbs and a downward procedure in temperature. The factor of temperature is introduced into the calculation of the function energy. Theoretically the algorithm converges towards a configuration of minimal energy if the temperature decrease. We will describe below the algorithm of Metropolis, used to study balance at low temperature of the very great systems:

- At the iteration k: e_k = e.
- At the iteration k+1: choose e' near e.
  $\Delta U = U(e') - U(e)$.
  if $\Delta U<0$ alors e_{k+1} = e'.
  Selseif e_{k+1} = e' with probabilité $e^{-\Delta U}$.
  e_{k+1} = e with probabilité $1-e^{-\Delta U}$.
  Endif.
- End.

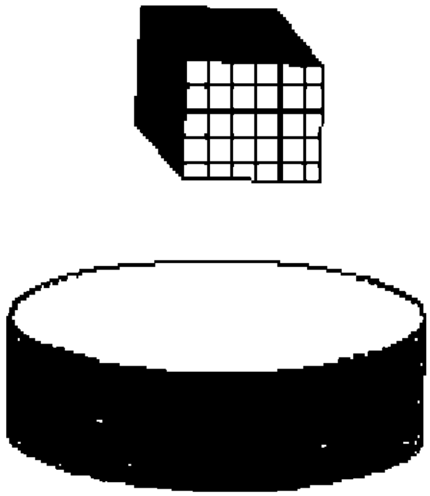
La figure ci-dessous résume l'algorithme d'optimisation utilisé pour la détection d'objets mobiles.

The Fig. 3 summarizes the algorithm of optimization used for the motion detection (Lalande, 1990; Bouthemy and Odobez, 1995; Bouden, 1995; Alice and Franck, 1997; Fulvio, 2001; Thomas, 2005; Qiang *et al.*, 2005).

RESULTS

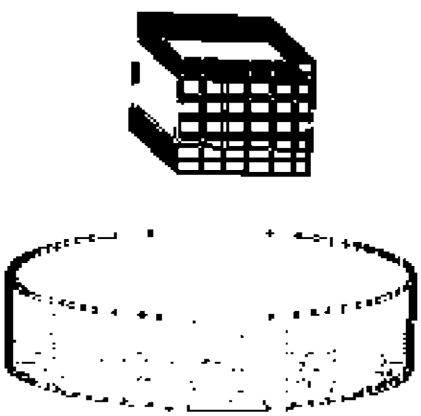
We will present some obtained results by the various studied methods applied to the following sequences images: Sequence aqua, is a natural sequence of 4 images. And on the synthetic cubic sequence of 12 images.

The frame difference method is the better in the computing time and gives noisy masks of moving area (Lalande, 1990; Bouden, 1995; Alice and Franck, 1997). But in the case of synthetic sequences without noise it's the better one (Fig. 4). The other methods (Bouden, 1995) take enough computing time and give the best and higher compromise mask quality and noise sensitivity and time computing (Fig. 4 and 5). These studied methods



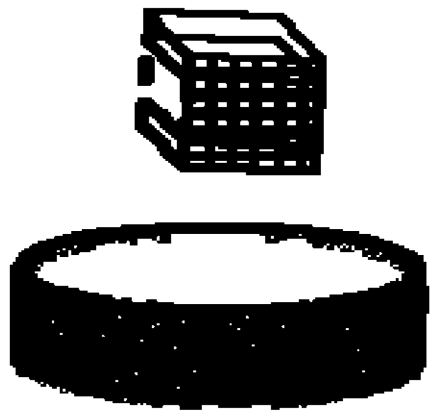
(a) 1th frame of synthetic sequence

(b) 7th frame of synthetic sequence

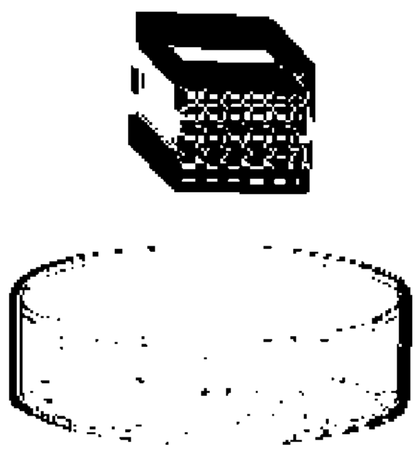


(c) Frame difference detection with ? =10

(d) Likelihood detection constant model with ? = 10



(e) Likelihood detection linear model with ? = 75



(f) Markov model with (RD) and parameters values: (m2=2m1=20, b_{fs}=30,b_p = 20 Et b_f=10)

Fig. 4: Results for the Synthetic cubic sequence

(a) natural aqua sequence

(b) Likelihood detection constant model with ? = 10

(c) Likelihood detection linear model with λ =75

(d) Markov model with (RD) and parameters values: (m2=2m1=20, b_{fs}=30,b_p = 20 Et b_f=10)

(e) Markov model with (RS) and parameters values: (m2=2m1=20, b_{fs}=30,b_p = 20 Et b_f=10)

Fig. 5: Results for the natural aqua sequence

have the network realization with hypothesises on the scene and its conditions lighting.

CONCLUSION

In general, the use of these various methods depends on the type of application considered. The limitation of the two last studied methods is the time computing. The likelihood detection and Markov model detection methods, realize the best compromise (sensitivity to the noise and cost of calculation). But if the notion of time is of primary importance and that the quality of the segmentation is not interesting a simple frames difference can be carried out.

REFERENCES

Alice, C. and L. Franck, 1997. Approche spatio-temporelle pour l'analyse de séquences d'images. Application en detection des mouvement. Traitement du Signal, Vol. 14.

Bouden, T., 1995. Detection des objets mobiles et estimation de leurs mouvement. Master Thesis, Electronic Departement, Annaba University Algeria.

Bouthemy, P. and J.M. Odobez, 1995. Détection et segmentation du mouvement par une approche robuste et markovienne. Journées sur les nouvelles techniques pour la compression et la représentation des signaux audiovisuels, CCETT, pp: 47-53.

Fulvio, M., 2001. A statistical approach to motion estimation. Ph.D. Thesis Lausanne EPFL.

Jain, R., D. Militzer and H.H. Nagel, 1979. Separating non stationary scene components in a sequence of real world tv-image. IEEE. PAMI., 1: 206-214.

Jain, R. and K. Jain, 1988. Displacement measurement and its applications in interframe image coding. IEEE. Trans. Comm., 29: 1788-1808.

Lalande, P., 1990. Détection du mouvement dans une séquence d'images selon une approche markovienne: Application à la robotique sous-marine. Ph.D Thesis University of RENNES-1.

Qiang, L., J. Sclabassir, Li Ching-Chung and S. Mingui, 2005. An application of MAP-MRF to change detection in image sequence based on mean field theory. EURASIP. J. Applied Signal, 13: 1956-1968.

Thomas, V., 2005. Détection et analyse de mouvement dans des séquences d'images par une approche probabiliste a contrario. Équipe d'accueil: VISTA-IRISA/INRIA; Composante universitaire: SPM; Ph.D Thesis Rennes1 University France.

Wiklund, J. and G.H. Grunlund, 1987. Image sequence analysis for object tracking. Procc. 5th Scandinavian Conf. Image Analysis, pp: 641-648.

Yakimovsky, Y., 1976. Boundary and object detection in real images. J. Ass. Comp. Mach., 23: 599-618.