Genetic Algorithms Based Motion Estimation

M. Zaim, A. El ouaazizi & R. Benslimane

Université Sidi Mohamed Ben Abdellah
LTTI , Ecole Supérieure de **T**echnologie
Route d'Imouzzer BP. 2427, Fès, Morroco
Tel: 212 5 60 05 85 – Fax: 212 5 60 05 88
E-mail: mzaim@eudoramail.com

ABSTRACT

We derive a robust algorithm for estimating motion parameters of rigid features resulted from a segmentation of dynamic scenes into several differently moving objects. Thus each objects is characterized by a transform h (x, T) with a parameter vector T which implicitly describes the surface shape and the three-dimensional motion of the objects in the scene. In this paper, a split and merge technique for the segmentation is used. Moreover to estimates the vector parameter T we propose to use a genetic algorithm.

Keywords: Moving object, Segmentation, Estimation, Genetic algorithms.

1 Introduction

Motion estimation and motion-based segmentation are among the important tasks in the field of computer vision [15,8,10,11], since the image components thereby extracted generally correspond to meaningful entities. Provided, they can be obtained for a whole image sequence, such partitions can serve as data input for region based coding schemes, tracking procedures or interpretation stages of the dynamic content of the observed scene.

The most algorithms for motion estimation use the displacement field [7,11], which is based on the well known optical flow constraint (OFC) equation [2]. This differential equation, issued from a linearization of the brightness constancy assumption, links the spatio-temporal gradients of the luminance to the unknown velocity vector. Due to the differential nature of the OFC, this standard modeling does not hold for large displacements and demands two requirements:

1. The optical field should vary smoothly, so each flow vector should be closed to the average of its neighbors.
2. The edge motion should be compatible with the spatial and temporal gray- level gradient.

In this paper, we present a new algorithm for estimating parameters of motion. Thus, the corresponding features of images $I_n(x)$ and $I_{n+1}(x)$ are extracted from the background, Then they are modellede and their motion is estimated by the use of genetic algorithm[5].

The region detector used is a region-growing algorithm, which is based on the Split- and- Merge algorithm [17]. The first step of the algorithm splits the image until obtaining a partition S such as all the regions satisfy a homogeneity criterion [12]. The second step merges adjacent regions by considering new homogeneity criteria. Then, the very small regions are grouped with the nearest adjacent large ones.

This paper is organized as follows: section 2 gives the model of motion and surface of a moving object. In section 3, the problem statement is discussed. In section 4, we will give the principle of Split- and-Merge technique for the segmentation. Section 5, focuses on the implementation of genetic algorithm in the motion estimation and finally in section 6, we give experimental results.

2 Basic model and equations

The connection between two images I (x) and I (x'), and the connection between the coordinate system {x} and {x'} in the image plane is given by a transform x'= h(x,T) with the vector T. The elements of the vector T are called mapping parameters because they describe the connection between the two images if the structure of the transform h(x, T) is known. This later must be unique and revertible in the image region under consideration. Furthermore, we should have x'= h (x, T=0)=x.

Transform h(x, T) depends on three factors:

- The three-dimensional motion of the object which is described by a rotation matrix R and a translation d:

$$X'=RX+d \tag{1}$$

Or more precisely

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} \tag{2}$$

$X=(X,Y,Z)^T$ et $X'=(X',Y',Z')^T$ are the three-dimensional coordinates.

- The mathematical model, which describes the projection from three-dimensional space onto the camera plane:

$$x' = \frac{(1+a_1)x + a_2 y + a_3}{a_7 x + a_8 y + 1} \qquad y' = \frac{a_4 x + (1+a_5)y + a_6}{a_7 x + a_8 y + 1} \tag{3}$$

describes the connection of the coordinates $\{x=(x, y)^T\}$ and $\{x'=(x', y')^T\}$ in the camera plane if central projection is utilized.

- The approximation of the moving object surface, for instance planar surface:

$$(Z = a X + b Y + c) \tag{4}$$

The vector parameters is $T=(a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8)^T$ implicitly contain motion parameters (R, d) and the surface information (a, b, c).

The situation is illustrated in figure 1:

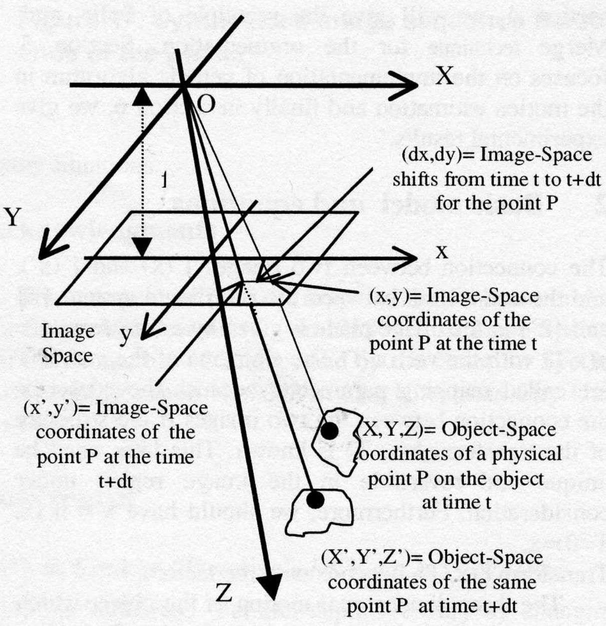


Fig. 1. Basic geometry for three-dimensional motion estimation.

In this paper, we restrict our study to the 2D rigid motion, with rotation angle θ and translation vector Tr, it maps a point (x, y) onto a point (x', y'), as follows:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} Tr_x \\ Tr_y \end{pmatrix} = \begin{pmatrix} a_1 x + a_2 y + a_3 \\ a_4 x + a_5 y + a_6 \end{pmatrix} \tag{5}$$

The vector parameters this time is $T=(a_1, a_2, a_3, a_4, a_5, a_6)^T$ implicitly contains motion parameters (θ, Tr). Furthermore, the description is not capable of handling occlusion effects. It is only valid for those parts objects, which can be actually seen in the two images $I_n(x)$ and $I_{n+1}(x)$. Parts which are uncovered cannot be described without information.

3 Problem Statement

To obtain an appropriate estimate $\hat{T}$ of the parameter vector T of the transformation h (x, T), model adaptive parameter estimation techniques are used. For this purpose, the objects of image $I_{n+1}(x)$ which results from the objects in $I_n(x)$ through the transform h (x, T) with the unknown motion vector T is almost rebuilt with a model image $I_m(x,\hat{T})$. This model image results from the segmented image $S_n(x)$ through the same transform h(x, T) as $I_{n+1}(x)$ using a model vector $\hat{T}$. The difference $e(\hat{T})$ between the predicted feature $S_m(x,\hat{T})$ and $S_{n+1}(x)$ is described by an error function $J\{e(\hat{T})\}$ which can be minimized by modified Newton algorithm combined with quasi-Newton method [11]. In additional to the computational requirement, the practical use of this algorithm faces serious problems:

- The motion estimation may not work well if the motion is too heavy, i.e. if the starting value of $\hat{T}$ is too far away from the true value T. Thus, the algorithm might converge to a wrong minimum of the error function, namely to a local one instead of the global one.
- When the Hessian has no inverse, not all parameters can be estimated because the error function doses not have a unique minimum.

To overcome this disadvantage we propose a minimization by a genetic algorithm to obtain the optimal $\hat{T}=T^*$.

With this minimization the object $S_m(x,\hat{T})$ is moved towards $S_{n+1}(x)$ according to the transform $h(x,\hat{T})$. Thus $S_m(x,\hat{T})$ is becoming closer to $S_{n+1}(x)$. At the optimum of the error function when the estimated vector $\hat{T}=T^*$ equal the true vector T, $I_m(x,\hat{T})$ and $I_{n+1}(x)$ will be similar. As $S_m(x,\hat{T})$ moved toward $S_{n+1}(x)$, it is possible to say that $I_m(x,\hat{})$ is the predicted image when trying to predict $I_{n+1}(x)$ from $I_n(x)$. Thus, it is possible to predict $I_{n+1}(x)$ exactly from $I_n(x)$ if the correct transform h(x,T) is chosen and $\hat{}$ is estimated correctly.

4 Segmentation

4.1 Formal Definition of Region

Homogeneity is an important property of regions and is used as the main segmentation criterion in region growing. The criteria for homogeneity can be based on grey level, texture, model using semantic information, etc.

Let a region G be defined as a maximal homogeneous and connected subset of the image I. It is also necessary to define function that can evaluate the homogeneity of a region. [6] Defines the region segmentation of an image I as a partition S=(G_1, G_2,.. G_n) with the following properties:

1. $I = \bigcup_{i=1}^{n} G_i$
2. $\forall i \in \{1,..n\},\ G_i\ \ connected$
3. $\forall i \in \{1,..n\},\ \forall j \in \{1,..n\}, i \neq j, G_i \cap G_j = \Phi$
4. $\forall i \in \{1,..n\},\ P(G_i) = true$
5. $\forall i \in \{1,..n\},\ \forall j \in \{1,..n\}, i \neq j, G_i\ \ and$
   $G_j\ \ adjacent,\ \ P(G_i \cup G_j) = false$

P is a predicate function defining the region homogeneity. Each region is subset of pixels with closed boundary. However, generally, the boundaries are wrongly localized due to homogeneity criterion and therefore are not very reliable.

4.2 Region Growing Algorithm

Our algorithm takes into account the edges that are supposed to be detected more precisely and more reliably than the region boundaries.

The first process is a median filter in order to eliminate the noise on the image. Then, after having computed the gradient image with adequate operator, an edge extractor provides the edge segmentation.

The first part of the region-growing algorithm has two steps and is based on the Split- and- Merge algorithm [18]. The first step splits the image until obtaining a partition S such as all the region satisfy a homogeneity criterion; the second step merges adjacent regions by considering the edge segmentation and a new homogeneity criterion.

Image Splitting: A homogeneity criterion P_{ampl} is used to determinate if an area G_i must be splitted into four different sub- areas with the same size.

$$G_i \xrightarrow{split} \left\{ \begin{array}{ll} \bigcup_{j=1}^{4} G_{ij} & if \quad P_{Ampl}(G_i) = false \\ G_i & \qquad Otherwise \end{array} \right\}$$

The homogeneity criterion P_{Ampl} is defined as follows:

$$P_{Ampl}(G_i) = True \Leftrightarrow \underset{(x,y)\in G_i}{Max}\ f(x,y) - \underset{(x,y)\in G_i}{Min}\ f(x,y) \leq \xi_1$$

This process is repeated recursively on every new area and ends when all the areas satisfy the homogeneity criterion

Image Merging: After the previous step, the image is divided in homogenous areas according to the criterion P_{Ampl}. The Merging process selects two adjacent regions belonging to different areas. If the region resulting from the merging of these two regions does not meet homogeneity criterion $P_{min-max}$, or, if one or several edges are detected on the common boundaries, then the two region are not merged. Otherwise, the two regions are merged.

For each pair (G_i, G_j) of adjacent regions which the border B(G_i , G_j) is defined as :

$$B(G_i, G_j) = \left\{ \begin{array}{l} (p_i, q_j)\ \ as\ \ p_i \in G_i, q_j \in G_j \\ and\ \ d_4(p_i, q_j) = 1 \end{array} \right\}$$

$$with\ \ d_4(p_i, q_j)\ \ is\ \ the\ \ forth\ \ neighborg\ \ of\ \ p_i$$

We compute:

<u>The length of the border:</u>

$$Bl(G_i, G_j) = card(B(G_i, G_j))$$

<u>The mean gradient in the border:</u>

$$B_g(G_i, G_j) = \frac{\sum_{(p_i, q_j)\in B(G_i, G_j)} \left| I(p_i) - I(q_j) \right|}{Bl(G_i, G_j)}$$

if

$$\left. \begin{array}{l} G_i,\ G_j\ \ neighboring\ \ regions \\ And \\ P_{\min\max}(G_i \cup G_j) = True \\ And \\ B_g(G_i, G_j) \leq \xi_2 \end{array} \right\} \xrightarrow{merge} G_i \cup G_j$$

The homogeneity criterion $P_{min-max}$ is defined as follows:

$$\left|Max(G_i \cup G_j) - Min(G_i \cup G_j) \leq \xi_3\right|$$

5 Estimating Motion Parameters

5.1 Method of differentials (optical flow)

Many Methods have been proposed for the computation of optical flow during the last decades. Generally, these methods can be classified into four categories, i.e. matching-based methods, gradient-based methods, energy-based methods and phase-based methods.

Among numerous optical flow methods, gradient-based methods, which compute optical flow from spatial-temporal derivatives of image intensity, are the most used [3]. Because gradient-based methods try to attach a motion vector to each point in the image plane, they are able to provide dense optical flow field, which is necessary in cases of the interpretation of three-dimensional motion parameters and shape of objects.

Most gradient-based methods start from gradient constraint equation, which relates image intensity's gradients to the two components of velocity vector

$$f_x u + f_y v + f_t = 0 \qquad (6)$$

where f_x, f_y and f_t are first-order spatial and temporal gradients of intensity $f(x, y, t)$; u and v are image velocity vector's two components along x and y axes.

A great advantage of the approach is that it is easy to measure spatial and temporal gradients, by combining smoothing and differencing operations, thus it will be relative easy to build it in real-time systems. However, it requires tow assumption:

1. The flow field should vary smoothly, so each flow vector should be close to the average of its neighbours.
2. The edge motion should be compatible with the spatial and temporal grey-level gradients.

Horn & Schunck [2] proposed an iterative algorithm, which yields flow fields that satisfy these conditions. Although the algorithm was derived using a sophisticated analysis, it amounts in effect to adjusting an estimated flow field to satisfy each of the conditions in turn more closely. On each iteration, the algorithm:

1. Replaces each flow vector with an average of itself and its neighbours; (This is just like smoothing an image using local averaging, except that there are two components and we have to smooth both of them).
2. Changes the component of each flow vector along the local grey level gradient to make it closer to the temporal to spatial grey-level gradients.

5.2 The genetic approach

Genetic Algorithms (GAs)[5][19], are pseudo-stochastic search methods whose derive their fundamental ideas and terminology from the Darwinian ''Natural selection'' theory, according to which individuals that are better fit to a given environment are more likely to survive.

While solving an optimization problem using GAs, each solution is usually coded as an alphabet string of finite length called chromosome. Each string or chromosome is considered as an individual. A collection of M individuals is called population. GAs start with a randomly generated population of size M, and in each iteration, a new population of the same size is generated from the current population by applying operators, termed selection, crossover and mutation [16], that mimic the corresponding processes of natural selection.

To estimate motion parameters with GAs, we encode the vector $T=(a_1, a_2, a_3, a_4, a_5, a_6)^T$ in a way that allows manipulation by genetic work. Therefore, we consider the chromosome representation individuals as a binary string of finite length.

The phenotype of the k[th] individual is defined by:

$a_1(k)$	$a_2(k)$	$a_3(k)$	$a_4(k)$	$a_5(k)$	$a_6(k)$

The corresponding genotype (chromosome representation):

$\alpha_1\alpha_2 \ldots \alpha_{N/9}$	$\alpha_{(N/9)+1} \ldots \alpha_{2N/9}$	...	$\alpha_{(8N/9)+1} \ldots \alpha_N$

α_i is a binary value, $i=1 \ldots N$ *bits.*

Fitting Function: Each individual is evaluated by his fitness value. The evaluation function specifies the quality of the estimate. To this end, knowledge of the signal and noise statistics should be incorporated. The most general approach is given by a maximum-likelihood estimate. If the noise is not too strong, a simpler error is given by:

$$J\{e(\hat{T})\} = \frac{1}{2} E\{(S_m(x,\hat{T}) - S_{n+1}(x))^2\} \qquad (7)$$

i.e. the variance of the model error is sufficient. The expectation value E{.} is obtained by summing over the object and dividing by the number of pixel within the image region. For stationary signals, J{e(.)}

equals the negative cross correlation function of $S_m(x, \hat{T})$ and $S_{n+1}(x)$ plus an additive constant.

Stopping Criteria: There exists no criterion in the literature[5] [19], which ensures the convergence of GAs to an optimal solution. Usually, two stopping criteria are used in Genetic Algorithms. In the first, the process is executed for a fixed number of iterations and the best individual obtained is taken to be the optimal one. In the second, the algorithm is terminated if no improvement in the fitness value of the best individual for a fixed number of iterations, and the best chromosome is taken to be the optimal one.

6 Experimental Results

The default parameters values, that are used in all our experiment, i.e. the population size is $NI=10$, the crossover rate is $p_c=0.8$, the mutation rate is $p_m=0.045$. In this paper, there are two differently moving objects in front of a uniform background. In this experiment, a synthetic images is used (Fig.2). The computer transformed a part of each object with the affine transform to produce the image $I_{n+1}(x)$. Figure 2(a_1) and 2(a_2) shown the two images $I_n(x)$ and $I_{n+1}(x)$, 2(b_1) and 2(b_2) shown the segmented images $S_n(x)$ and $S_{n+1}(x)$ and figure 2(c_1) an 2(c_2) shown the difference images. Using the proposed genetic algorithm discussed before, we tried to estimate The vector parameters T utilizing just these two images features $S_n(x)$ and $S_{n+1}(x)$. For this purpose the objects in $S_n(x)$ moved towards the objects in $S_{n+1}(x)$ until the error function calculated from the predicted object $S_m(x, \hat{T})$ and $S_{n+1}(x)$ was minimum. After many generations, this minimum was reached and the parameter vector T is obtained. Therefore, the predicted image is very similar to the original one and the original parameter vector T is estimated very accurately.

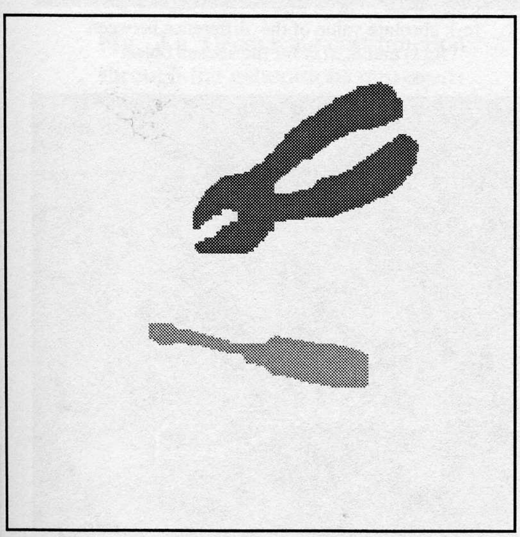
Figure.3: shows the behavior of cost function through the generations. Indeed, the process was terminated when no improvement in the fitness value was obtained, the best chromosome is taken to be the optimal one.

Figure.4: indicates the estimate optical velocity if the Horne and Schunck Method is used.

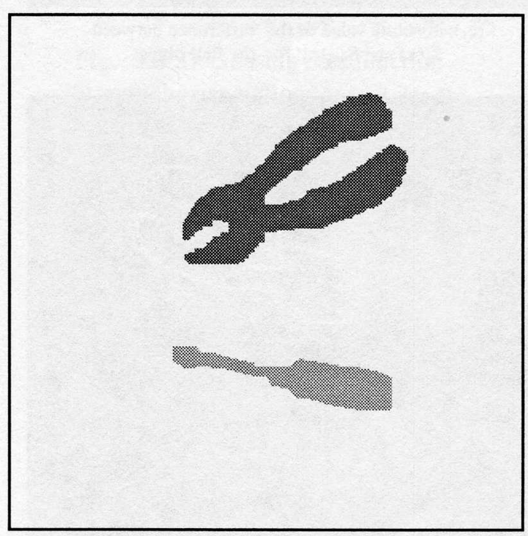
Table.1: shows the efficiency of our approach, especially when the assumption of small motion is not respected.

	Motion Parameters	Real Parameters	by Horne Method	by GA
Object: 1	T_{rx1}	6	0.57	6.41
	T_{ry1}	-10	-1.39	-10.32
	o_1	-8	-3.86	-8.01
Object: 2	T_{rx2}	12	4.66	12.13
	T_{ry2}	7	0.17	6.98
	o_2	10	1.94	10.20

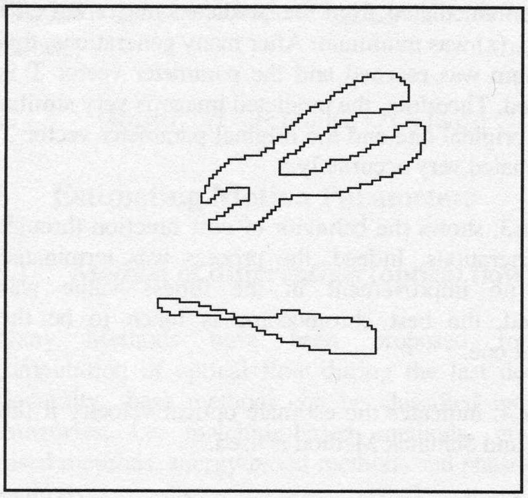
Tab. 1. Motion estimation using both methods differential method, genetic approach



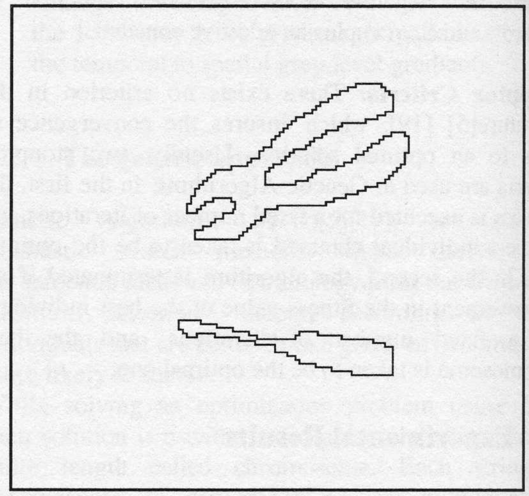
(a_1): image $I_n(x)$



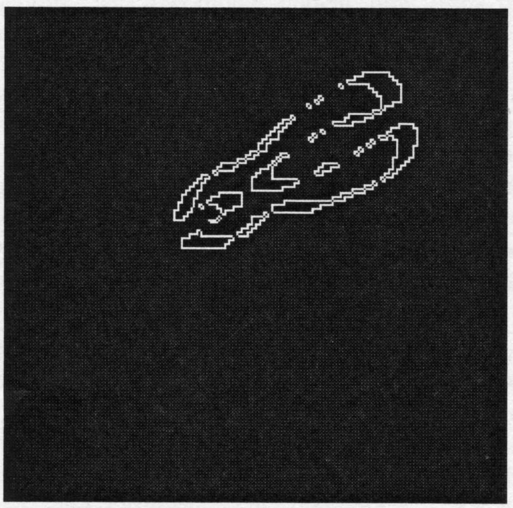
(a_2): mage $I_{n+1}(x)$



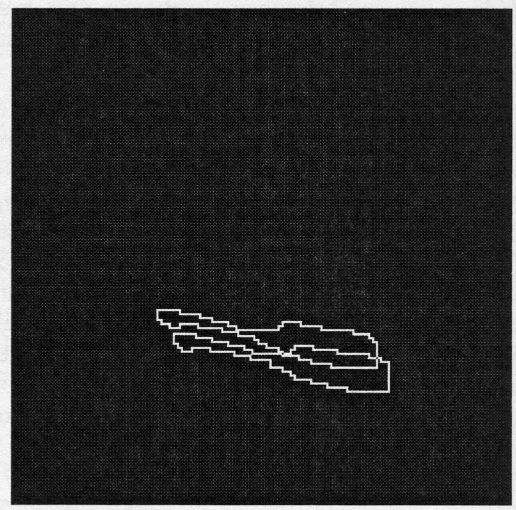
(b_1): segmented image $S_n(x)$



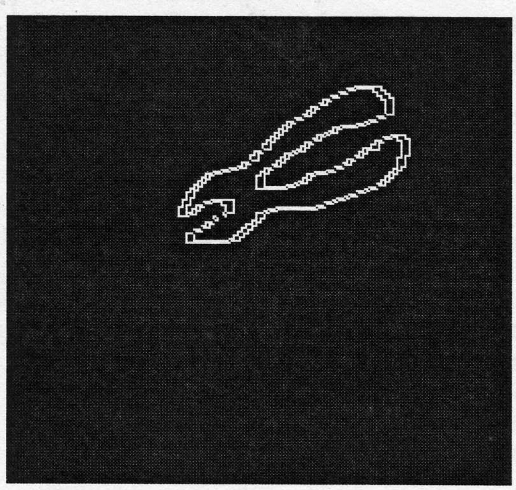
(b_2): segmented image $S_{n+1}(x)$



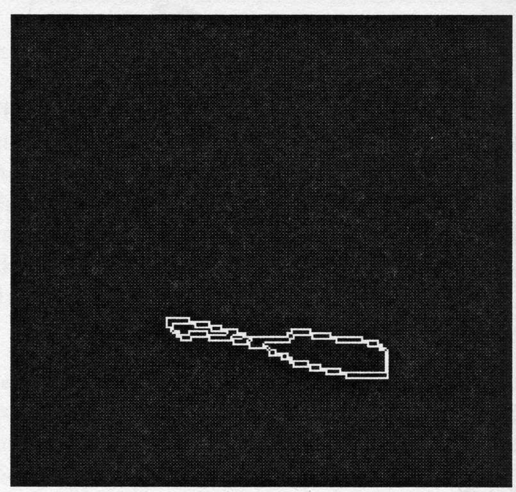
(c_1): absolute value of the difference between $S_n(x)$ and $S_{n+1}(x)$ for the first object



(c_2): absolute value of the difference between $S_n(x)$ and $S_{n+1}(x)$ for the second object

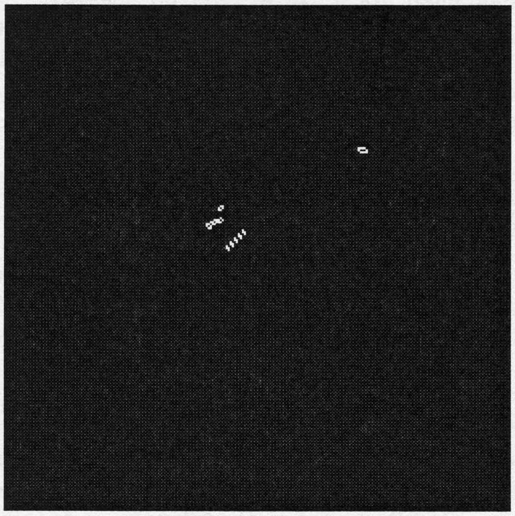


(d_1): difference between $S_{n+1}(x)$ and $S_m(x, \hat{\ })$ after 70 iterations of GA (first object)

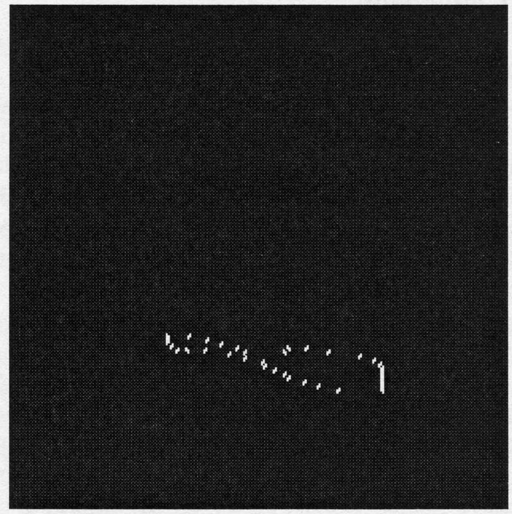


(d_2): difference between $S_{n+1}(x)$ and $S_m(x, \hat{\ })$ after 70 iterations of GA (second object)

Fig.2



(e_1): after 400 iterations of GA



(e_2): after 400 iterations of GA

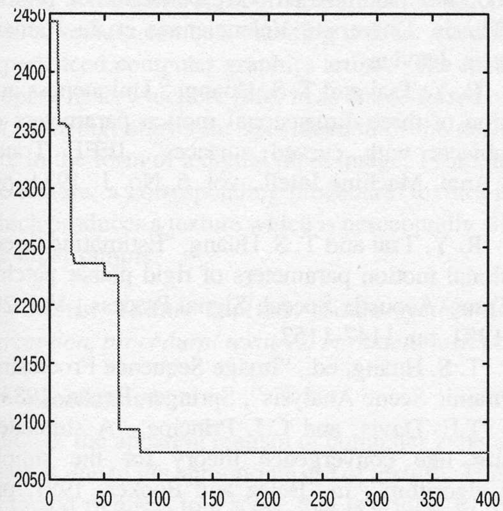


Fig. 3.a Decreasing cost function through the generations (first object)

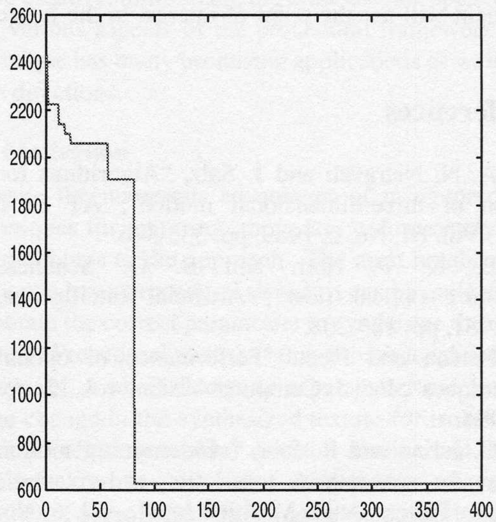


Fig. 3.b Decreasing cost function through the generations (second object)

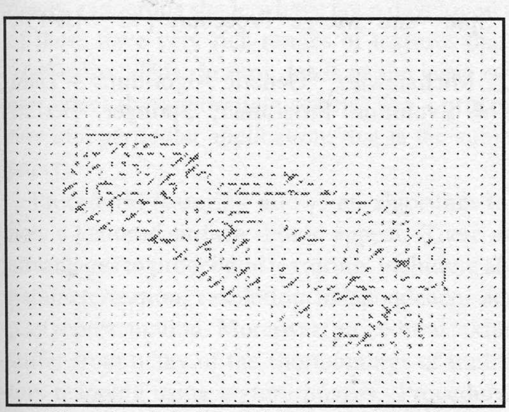


Fig. 4.a Motion field obtained by the Horne and Schunck Method (first object)

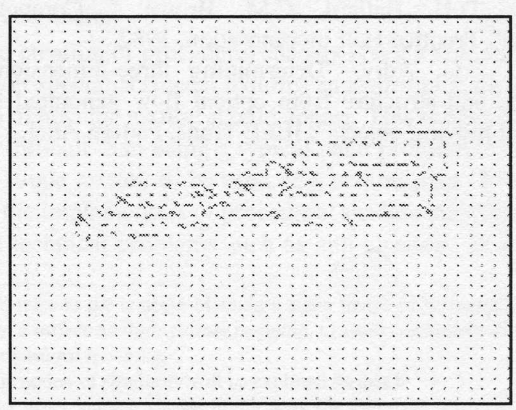


Fig. 4.b Motion field obtained by the Horne and Schunck Method (second object)

7 Conclusion

To obtain a good description of a scene, it must be segmented into different regions, each of these objects is characterized by its surface, and motion information.
In this contribution, we have presented a new algorithm for estimating the motion parameters. The procedure is based on minimizing an error function through the generations, Thus the segmented image can be described and tracked within the scene; so that an accurate movement compensated prediction can be reached by using a genetic algorithm. Nevertheless this algorithm it is not restricted to special assumptions about object motion or shape but can deal with quite general geometrical connections between two successive image regions. However, the algorithm has relatively a high time consuming, to overcome this inconvenient will use the point of interest in the future work.

8 References

[1] A. N. Netravali and J. Salz, "Algorithms for estimation of three-dimensional motion", AT & T Techn. J., Vol. 64, No. 2, 1989, pp. 335-346.
[2] B. K. P. Horn and B. G. Schunck, "Determining optical flow", Artificial Intelligence, Vol.17, 1981, pp. 185-203.
[3] Barron and Fleet. "Performance of optical flow techniques", Int'l J. Computer Vision, vol. 12, pp. 43-77, 1994.
[4] C. Jerian and R. Jain, " Determining motion parameters for scenes with translation and rotation', IEEE Trans. Pattern Anal. Machine Intell., vol. 6, No. 4, 1984, [5] D.E. Goldberg, "Genetic Algorithm in Search, Optimization, and Machine Learning", Addison Wesley, 1989.
[6] D.H. Ballard, C.M. Brown " Computer Vision", Prentice Hall Incorporation, 1982.
[7] G. Adiv, "Determining three-dimensional motion and structure from optical flow generated by several moving objects", IEEE Trans. Pattern Anal. Machine Intell, vol. 7, No. 4, 1985, pp. 384-401.
[8] H. H. Nagel, "Image sequence-ten (octal) yers-from phenomenology towards a theoritical foundation", Proc. 8th International Conference on Pattern Recognition, 1986, pp. 1099-1102.
[9] J. K. Aggarwal and N. Nandhakumar, " On the computation of motion from sequences of images-A review", Proc. IEEE, Vol. 76, No. 8, 1981, pp. 917-935.
[10] M. Hötter and J. Ostermann, "Object-oriented analysis-synthesis coding of moving images", Signal processing: Image communication, Vol. 1, No. 2, 1989, pp. 117-138.
[11] N. Diehl, "Object-oriented motion estimation and
segmentation in image sequences", Signal processing: Image communication, Vol. 3, 1991, pp. 23-56.
[12] N. Benamrane, Contribution à la vision stéréoscopique par mise en correspondance de région. Ph.D. Thesis, Université Valenciennes et du Hainant Cambresis, Janvier.
[13] R. Y. Tsai and T. S. Huang, " Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces", IEEE Trans. Pattern Anal. Machine Intell., vol. 6, No. 1, 1984, pp. 13-27.
[14] R. Y. Tsai and T. S. Huang, "Estimating three-dimensional motion parameters of rigid planar patch", IEEE Trans. Acoustic Speech Signal Process., Vol. 29, No. 6, 1981, pp. 1147-1152.
[15] T. S. Huang, ed., "Image Sequence Processing and Dynamic Scene Analysis", Springer, Berlin, 1983.
[16] T.E. Davis, and C.J. Principe, "A simulated annealing like convergence theory for the simple genetic algorithm", In Belez and Booker, 1991, pp. 174-181.
[17] T. Pavlidis, 1977. Structural Pattern Recognition. Springer, Berlin.
[18] T.Pavlidis, H. Horowitz " Picture Segmentation by Directed Split- and Merge Procedure", 2nd IJCPR, 1974.
[19] Z. Michalewiez, "Genetic Algorithms + Data Structure = Evolution Programs", Springer, Berlin, 1992.