THE QUALITATIVE KINEMATICAL DESCRIPTION IN DYNAMIC SCENE ANALYSIS

Edouard FRANCOIS and Patrick BOUTHEMY
IRISA/INRIA, Campus Universitaire de Beaulieu,
35042 Rennes Cedex, France

Abstract

A very challenging task in dynamic scene analysis is the recovery of informations about 3D kinematical behaviour of objects in the scene. This paper is concerned with a qualitative approach which enables to get such descriptions from the apparent motion in the image sequence. The interpretation based on qualitative symbols is probably easier and more robust than on 3D quantitative measurements. The process is in fact two-fold : the partition of the image in areas comprising an unique motion and the kinematical description of the motion of each area. Kinematical description has to provide two information classes : the motion type and the trajectory type. To this end, first, cues of the apparent velocity vector field in the 2D image are defined through a first order development, each of them describing a particular aspect of this field. Second, we establish the relation between these cues and the 3D motion parameters, which allows to determine a set of labels associated with different kinematic configurations. Third, the label validation step is solved using a statistical approach. Numerous experiments on different sequences have been carried out.

Résumé

En analyse de scène dynamique, un des objectifs principaux est l'obtention d'informations concernant le comportement cinématique 3D des objets dans la scène observée. Cette étude est une première approche vers une interprétation qualitative de comportements cinématiques dans la scène à partir du mouvement apparent dans la séquence d'images. Le procédé nécessite deux étapes : la partition de l'image en zones de mouvements apparents différents et la description qualitative du mouvement de chaque zone. Cette description doit fournir deux sortes d'informations : le type de mouvement et le type de trajectoire. A cette fin, nous introduisons, via un développement linéaire, des descripteurs du champ de vitesse apparent décrivant chacun un aspect particulier de ce champ. Ensuite, nous établissons les relations entre ces descripteurs et les paramètres 3D du mouvement, ce qui permet de déterminer un ensemble d'étiquettes associées à différentes configurations cinématiques 3D. L'étape de validation des étiquettes est résolue par une approche statistique. De multiples expérimentations sur diverses séquences ont été effectuées.

Keywords : Image sequence, velocity field, symbolic description, information criteria, motion interpretation.

1. Introduction

A very challenging task in dynamic scene analysis is the recovery of informations about 3D motion and structures in the scene. A solution, which has been largely investigated, is to quantitatively estimate the 3D parameters, using for instance optical flow based methods, [6] . Substantial works have been devoted to this topic, in particular to the recovery of quantitative 3D motion and structure parameters in the scene from optic flow, [1] . Because of noisy measurement of apparent motion, problems of numerical instability and estimation errors of the 3D measures appear. Then it becomes attractive to follow a qualitative approach, in order to obtain stable and robust descriptions, which still keeps enough richness of information in many situations. Indeed theoretical studies have pointed out that the geometry of the apparent velocity field contains by itself significant useful information, [4] [7] .

Thomson et al primarily proposed another approach for dynamic scene analysis, which emphasizes a qualitative way of reasoning and modeling, [14] . They suggested that practical operations such as obstacle avoidance could be driven using only qualitative cues, and not necessary an explicit quantitative reconstruction of the world. Burger and Bhanu addressed this problem by multiple possible qualitative descriptions of the scene, simultaneously maintained, instead of calculating the single quantitative description, [3] . Nagel presents several approaches to provide conceptual descriptions of the scene, based on different typical situations, [9] . More recently, Nelson and Aloimonos solved the obstacle avoidance problem by a qualitative analysis, based on the divergence of the optic flow, [10] .

These qualitative approaches were chosen to obtain robust descriptions of the scene. The method described in [10] seems to provide good and stable results. Yet, it gives a partial description of the scene. In this paper, we try to provide a complete qualitative kinematical description of the scene, using the image sequence acquired by a camera. This camera can be stationary or mobile. Moreover, to cope with this problem we have defined a statistical approach. The interpretation based on qualitative symbols is probably easier than on quantitative estimations. We do not need to explicitely compute any 3D measurements as depth maps for instance, nor to extract the focus of expansion (F.O.E.). Moreover, as we aim at obtainig a qualitative description, and not the exact estimation of 3D parameters, a complete camera

calibration is not required.

The process is in fact two-fold :
- the partition of the image in areas comprising an unique motion.
- the kinematical description of the motion of each area.

The first problem is solved in [2] . We deal here with the second one. Kinematical description has to provide two information classes :
- the motion type (e.g. rotation or translation).
- the trajectory type (e.g. perpendicular or parallel to the optical axis).

The problem has to be investigated according to different aspects, which can be described as follows (the two first aspects concern the theoretical analysis of the "physical" problem, the third one is the explicit way of solving it) :

- first, cues on the apparent velocity vector field in the 2D image are defined through a first order development, each of them describing a particular aspect of this field.
- second, we establish the relation between these cues and the 3D motion parameters, which enable to determine a set of labels associated with different kinematic configurations.
- third, the proper label (model) validation problem is solved using a statistical approach.

Sections 2 and 3 present respectively the introduction of relevant terms describing a vector field, and the relation between the 3D projected motion and these terms. Section 4 shows, starting from the introduction of qualitative models, how the qualitative interpretation is addressed. Section 5 explicitely describes the statistical tools we use to achieve the model validation. Results and conclusion are developed in Sections 6 and 7.

2. 2D vector field description

The basic idea of our approach is that the geometry of the 2D velocity vector field in the image contains information about the 3D motion, which can be significant and sufficient to the interpretation. The first step is then to introduce relevant terms, describing specific forms of the field, and which could be easily and naturally interpreted. Some studies and experiments allowed us to conclude that a first order development of a vector field can be sufficient to the interpretation. We then consider the first order development of the velocity vector $\vec{\omega}$ around a point $g\,(x_g, y_g)$ in the image :

$$\vec{\omega}\,(x,y) = \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} a_g \\ b_g \end{pmatrix} + \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} \begin{pmatrix} x - x_g \\ y - y_g \end{pmatrix}$$

with $\alpha = \frac{\partial u}{\partial x}, \beta = \frac{\partial v}{\partial x}, \gamma = \frac{\partial u}{\partial y}, \delta = \frac{\partial v}{\partial y}$ (1)

We can use the following decomposition of any matrix M, [13] :

$$M = \frac{1}{2}(\text{trace } M)\, I + \frac{1}{2}\left(M - M^T\right) + \frac{1}{2}\left[M + M^T - (\text{trace } M)\, I\right] \quad (2)$$

Considering $M = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}$, this allows the following formulation :

$$M = \frac{1}{2}\, div\, [D] + \frac{1}{2}\, rot\, [R] + \frac{1}{2}\, hyp1\, [H_1] + \frac{1}{2}\, hyp2\, [H_2]$$

$$\text{where} \quad [D] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, [R] = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

$$[H_1] = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, [H_2] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad (3)$$

Hence,

$$\vec{\omega}\,(x,y) = \begin{pmatrix} a_g \\ b_g \end{pmatrix} + \begin{pmatrix} \frac{1}{2}dx.div - \frac{1}{2}dy.rot + \\ \frac{1}{2}dx.hyp_1 + \frac{1}{2}dy.hyp_2 \\ \\ \frac{1}{2}dy.div + \frac{1}{2}dx.rot - \\ \frac{1}{2}dy.hyp_1 + \frac{1}{2}dx.hyp_2 \end{pmatrix} \quad (4)$$

with $dx = x - x_g, dy = y - y_g$

The four terms whose coefficients are

- $div = \alpha + \delta$ (for divergence)
- $rot = \gamma - \beta$ (for rotational) (5)
- $hyp1 = \delta - \alpha$
- $hyp2 = \gamma + \beta$ (for hyperbolic terms)

correspond to fields which belong to four orthogonal supplementary sub-spaces. Four different fields of these four independant sub-spaces are represented in Fig. 1. Any vector field can be approximated by a linear combination of a divergent field, a rotational field, and two hyperbolic fields. The interest of these four first order terms in comparison with α, β, γ and δ, is that the interpretation of any vector field, even non linear, can be made more easily based on them. When the field is the projection of 3D velocity vectors in the image plane, a particular kind of motion in the scene will imply a particular combination of these basic fields. Indeed, a divergent velocity vector field is the result of an axial motion, i.e. along the optical axis of the sensor. A rotational field appears as the result of a rotation around this axis. We will establish in the next section the link between 3D motion parameters and each of the four terms introduced above.

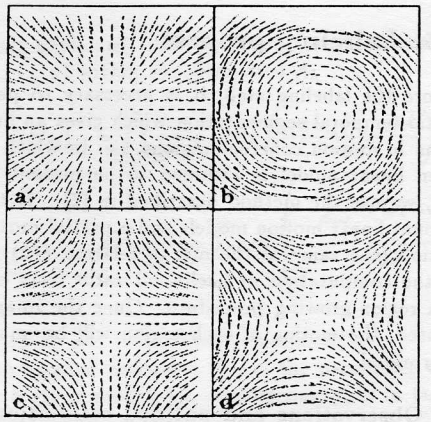

Figure 1 : *(a) divergent field, (b) rotational field, (c) and (d) hyperbolic fields*

3. Relations 2D-3D

We consider the relative motion of an object, with respect to the camera, consisting of translational velocity $T = (U, V, W)^T$, and rotational velocity $\Omega = (A, B, C)^T$, in the coordinate system shown in Fig. 2.

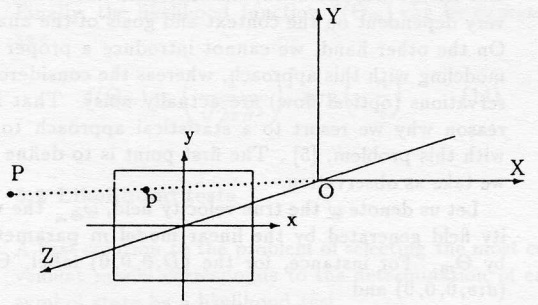

Figure 2 : *Coordinate system*

We define as the point $p(x, y)$ the perspective projection in the image plane of a point $P = (X, Y, Z)^T$. We get (assuming focal length f is equal to 1)

$$x = X / Z \quad , \quad y = Y / Z \qquad (6)$$

The instantaneous velocity of P is given by the formula $\vec{V} = (\dot{X}, \dot{Y}, \dot{Z})^T = \vec{T} + \vec{\Omega} \wedge \vec{OP}$. Deriving (6) with respect to time and after some developments, we obtain the following relations between the apparent 2D motion and the 3D motion parameters, [8] :

$$\begin{cases} u = \frac{U}{Z} + B - Cy - x\frac{W}{Z} - Axy + Bx^2 \\ v = \frac{V}{Z} - A + Cx - y\frac{W}{Z} + Bxy - Ay^2 \end{cases} \qquad (7)$$

Given the above remark upon the relevance of first order terms for the interpretation, we only consider the first order terms of the depth function Z.
By combining with (6), this implies :

$$\frac{1}{Z} = \frac{1}{Z_0}(1 - \gamma_1 x - \gamma_2 y)$$

Taking into account this development and considering expressions (7), for point P and reference point G, we can show that the following relations can be derived :

$$\begin{cases} u = a_g + (-\frac{\gamma_1}{Z_0}U - \frac{W}{Z_0})dx + \\ \qquad (-\frac{\gamma_2}{Z_0}U - C)dy + o^2 \\ v = b_g + (-\frac{\gamma_1}{Z_0}V + C)dx + \\ \qquad (-\frac{\gamma_2}{Z_0}V + \frac{W}{Z_0})dy + o^2 \end{cases} \qquad (8)$$

where $g = (x_g, y_g)$ is the projection of G, $(a_g, b_g)^T$ is the velocity vector of point g, and again $dx = x - x_g, dy = y - y_g$. By identifying in (8) the linear terms of (1), and using (5), we obtain

$$\begin{cases} div = -2\frac{W}{Z_0} - \gamma_1\frac{U}{Z_0} - \gamma_2\frac{V}{Z_0} \\ rot = 2C - \gamma_1\frac{V}{Z_0} + \gamma_2\frac{U}{Z_0} \\ hyp1 = -\gamma_1\frac{U}{Z_0} + \gamma_2\frac{V}{Z_0} \\ hyp2 = -\gamma_1\frac{V}{Z_0} - \gamma_2\frac{U}{Z_0} \end{cases} \qquad (9)$$

These relations will be used to settle the qualitative interpretation of the motion, based on the velocity vector field. Indeed, the qualitative analysis of the 3D motion will rely on these relations (9).

4. Definition of the qualitative interpretation

Briefly, the way of solving the interpretation problem is the following :
- segmentation of the velocity field
- estimation of linear parameters (or cues)
- selection of the significant models and interpretation

We propose a decomposition of the qualitative interpretation in different levels, from the quantitative data which is the velocity field, to the global qualitative description.

Since we are concerned with qualitative interpretation, what is important is the comparison of the linear parameters $div, rot, hyp1$ and $hyp2$ to zero, and not the estimation of each parameter of the 3D motion. Accordingly, we associate to each quantitative term $div, rot, hyp1, hyp2$ a qualitative (boolean) variable $V_{div}, V_{rot}, V_{hyp1}, V_{hyp2}$, equal to 0 if its associated quantitative variable is non significant, and respectively symbols D, R, H_1, H_2 otherwise.

Once the different variables, and their possible states have been defined, the crucial point is the choice of the models. Given four independant parameters taking two possible states, we consider as the set of models S_{model} the complete set of all the possible combinations of these variable states. There are therefore $2^4 = 16$ possible models in the $(V_{div}, V_{rot}, V_{hyp1}, V_{hyp2})$ base (Fig 3). The

set of these sixteen models S_{model} constitutes the first level of the qualitative interpretation.

The actual qualitative description is introduced in the second level of the interpretation. It takes into account the physical reality of the motion, explicitely explained in the relation (5). Indeed, from this relation, it can be seen that the only symbol associations that are physically possible are the following : $(0,0,0,0)$, $(D,0,0,0)$, $(0,R,0,0)$, $(D,R,0,0)$ and (D,R,H_1,H_2). If hyp_1 and hyp_2 are non-zero, then U or V are non-zero, and consequently div and rot are non-zero (unless very particular cases). Then the $(D,R,0,H_2)$ model for instance is not physically feasible. This defines a sub-set of five physical models out of the set S_{model}, that we call set of labels S_{label}, Fig. 3. From the relations (5), it is clear that several typical dynamic situations can be expressed. For instance, in the $(V_{div}, V_{rot}, V_{hyp1}, V_{hyp2})$ basis, the $(D,0,0,0)$ association or label is the qualitative description of a motion along the optical axis (it is besides usual to base the obstacle detection on a divergence cue). $(0,R,0,0)$ describes a rotation around this axis, and so on.

The third level of the interpretation depends on the considered application. Let us consider the example of a car driving situation. As far as obstacle detection is concerned, the qualitative evaluation of the kinematic behaviours of the others objects in the scene relative to the vehicle of interest is crucial. Three main kinematic classes are introduced :
- object getting closer to the vehicle
- object moving away from it
- object moving across in front of it

These classes are directly tied to the labels described above.

MODELS	
$(0,0,0,0)$	$(0,0,0,H_2)$
$(0,0,H_1,0)$	$(0,R,0,0)$
$(D,0,0,0)$	$(0,0,H_1,H_2)$
$(0,R,0,H_2)$	$(D,0,0,H_2)$
$(0,R,H_1,0)$	$(D,0,H_1,0)$
$(D,R,0,0)$	$(0,R,H_1,H_2)$
$(D,0,H_1,H_2)$	$(D,R,0,H_2)$
$(D,R,H_1,0)$	(D,R,H_1,H_2)

LABELS	
(0 , 0 , 0 , 0)	
(0 , R , 0 , 0)	
(D , 0 , 0 , 0)	
(D , R , 0 , 0)	
(D , R , H_1 , H_2)	

CLASSES

Figure 3 : *the qualitative interpretation levels:*
1) models related to vector fields
2) labels related to velocity fields
(and their representative textures for results display)
3) classes related to 3D motion interpretation

5. Labeling process

Before the labeling step, it is necessary to performe a spatio-temporal segmentation which gives an image partition in motion coherent regions, i.e. in which only one given linear model is assumed to be present. To this end, we use the method described in [2] . It takes into account linear motion models and a partial motion information (the same as the one in relation (12)); it relies on the computation of likelihood ratio tests embedded in a region-growing procedure. Once the image is segmented, the motion of each region will be qualitatively interpreted.

5.1 Observations and optimal estimation of the parameters

Now, the key point is to properly achieve the numerical-to-symbolic step : that is deriving symbols from numerical data. Given an area in the image, we must determine the more convenient label, i.e. the model which is the best fit to the 3D motion of the object. The simple comparison of the magnitudes (or function of the magnitudes) of the quantitative terms $div, rot, hyp1, hyp2$ to a threshold, as initialy described in [14] , remains very difficult and tricky. On one hand, the threshold choice would be very dependent on the context and goals of the analysis. On the other hand, we cannot introduce a proper noise modeling with this approach, whereas the considered observations (optical flow) are actually noisy. That is the reason why we resort to a statistical approach to cope with this problem, [5] . The first point is to define what we take as observation.

Let us denote $\underline{\omega}$ the true velocity field, $\underline{\omega}_{\Theta_m}$ the velocity field generated by the linear model m parametrized by Θ_m. For instance, for the $(D,0,0,0)$ label, $\Theta_m = (div,0,0,0)$ and

$$\underline{\omega}_{\Theta_m} = \begin{pmatrix} a_p & + & \frac{1}{2}.div.x \\ b_p & + & \frac{1}{2}.div.y \end{pmatrix}$$

$I(x,y)$ is the image intensity function, $\underline{\nabla} I$ and I_t its spatial and temporal derivatives. $\underline{\omega}$ and I are linked by the well-known image flow constraint equation, [1] :

$$\underline{\omega}.\underline{\nabla} I + I_t = 0 \quad (10)$$

If the true field has been estimated, we consider as observations the vectorial random variables :

$$\underline{e}_{\Theta_m}(x,y) = \underline{\omega}_{\Theta_m}(x,y) - \underline{\omega}(x,y) \quad (11)$$

If the true velocity vector field is not available, we consider the following scalar random variables :

$$e_{\Theta_m}(x,y) = (\underline{\omega}_{\Theta_m}(x,y) - \underline{\omega}(x,y)).\underline{\nabla} I(x,y)$$

i.e. using relation (10),

$$e_{\Theta_m}(x,y) = \underline{\omega}_{\Theta_m}(x,y).\underline{\nabla} I(x,y) + I_t(x,y) \quad (12)$$

In both cases, the considered variables are supposed to be independant gaussian zero-mean variables. The likelihood function $f(\Theta_m)$ of a model m can therefore be

easily explained. It is the product of the densities of each $e_{\Theta_m}(x,y)$ within the given area A_k

$$f(\Theta_m) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{e_{\Theta_m}(x_i,y_i)^2}{2\sigma^2}\right)$$
$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n exp\left(-\frac{\sum_{i=1}^{n} e_{\Theta_m}(x_i,y_i)^2}{2\sigma^2}\right)$$

where n is the number of points of A_k.
The optimal estimator $\hat{\Theta}_m$ of the parameter vector Θ_m for a given model, i.e. maximizing f, is in fact derived according to :

$$\hat{\Theta}_m = \arg\min_{\Theta_m} \sum_{i=1}^{n} e_{\Theta_m}(x_i,y_i)^2 \qquad (13)$$

The unknown variance σ^2 is supposed to depend on the given model and is a posteriori estimated by:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_{\hat{\Theta}_m}(x_i,y_i)^2$$

Finally, the likelihood function $f(\hat{\Theta}_m)$ can be expressed as :

$$f(\hat{\Theta}_m) = \left(\frac{1}{\sqrt{2\pi\hat{\sigma}^2}}\right)^n exp\left(-\frac{n}{2}\right) \qquad (14)$$

5.2 Likelihood tests

A first solution to the problem of selecting the most convenient model corresponds to the determination of each symbol state by a likelihood test.
For instance, V_{div} will be set by testing hypothesis $\underline{H_0}$: (D,R,H_1,H_2) against hypothesis $\underline{H_1}$: $(0,R,H_1,H_2)$. Practically, we first estimate the optimal parameter vector $\hat{\Theta}_0 = (\hat{div},\hat{rot},\hat{hyp1},\hat{hyp2})$ for the (D,R,H_1,H_2) model, i.e. which maximizes the likelihood function $f(\Theta_0)$ as expressed in (13) and (14).
In the same way, we determine $\hat{\Theta}_1 = (0,\hat{rot},\hat{hyp1},\hat{hyp2})$ for the $(0,R,H_1,H_2)$ model. Then we compare $\frac{f(\hat{\Theta}_1)}{f(\hat{\Theta}_0)}$ (in fact the logarithm of this ratio : $L(V_{div}) = \log\left(\frac{f(\hat{\Theta}_1)}{f(\hat{\Theta}_0)}\right)$) to a threshold λ. If the ratio is lower than this preset threshold, the V_{div} symbol is equal to D, 0 otherwise:

$$L(V_{div}) \begin{array}{c} (0) \\ H_1 \\ > \\ < \\ H_0 \\ (D) \end{array} \lambda \qquad (15)$$

This must be done for each symbol V_{div}, V_{rot}, V_{hyp1}, V_{hyp2}, in a predetermined order, which takes into account the physical possibilities of models, and the particular application of the analysis. For instance, if the main goal of the analysis is obstacle detection, this tree will be built up from the V_{div} symbol. Once each symbol state is determined, the optimal label is defined as their association.

5.3 Statistical information criterion

A more attractive method consists to consider the significant sub-set of labels S_{label}, and to test *together* all the models associated with these labels. The optimal label m is found using a statistical information criterion, of the kind, [12] :

$$\hat{m} = \arg\min_{m} \left[-\log[f(\hat{\Theta}_m)] + \Psi(n).dim(\Theta_m)\right] \quad (16)$$

where f is again the likelihood function of the model m, $\hat{\Theta}_m$ denotes the optimal parameter vector of the model m; $\dim(\Theta_m)$ is the model dimension (for instance, the dimension of the $(D,R,0,0)$ label is 2); and Ψ a given function. Before comparing the models, the optimal parameter vector $\hat{\Theta}_m$ has to be found for each model. The second term of the criterion acts as a penalization term on the model complexity. Hence, if the likelihood of a given model is not really better than a simplier one, the more complete model will be eliminated because of the penalization term. This means that in fact, it does not bring any significant additional information. After preliminary tests, it appeared that the most interesting criterion for this application is the Rissanen criterion (RIC). In this case, $\Psi(n) = \frac{1}{2}\log(n)$. By the way, this version of Ψ can also be derived from a Bayesian approach, [5] . Taking into account the relation (14), $\hat{m}$ is given by :

$$\hat{m} = \arg\min_{m} \left[n \log \sum e^2_{\hat{\Theta}_m} + \log n.\dim(\hat{\Theta}_m)\right] \quad (17)$$

6. Experimental results

These two approaches have been tested on three kinds of data : real images with synthetic motions (Fig. 4a, 4b), sequences acquired by a camera attached to a robot arm moving in a static environment (Fig. 5a, 5b), and sequences in road context (Fig. 7a, 7b). The aim of these tests are of course to validate the global approach, and also to compare the different possible observations and the different labeling techniques. A priori, the complete observation, i.e. the velocity vector should be preferable, but of course requires significative supplementary computations. Indeed, we can think that the gradient based observation, that is only a partial motion information on the velocity vector, would penalize the parameter estimation, and consequently the labeling. Therefore, we have tested these two kinds of observations (introduced in (11) and (12)). Velocity vectors have been estimated using the method developped in [11] .

For notation convenience, we denote the velocity vector observation (11) O_V, the gradient based observation (12) O_G, the likelihood test method $GLR - Method$ and the Rissanen criterion method $RIC - Method$.

In order to validate the method principle, we have first considered real data with simulated motions (example

1 - Fig. 4). We have created two successives images which are composed of four real sub-images undergoing specific synthetic motions (scaling, translation, rotation, and a composition of scaling and rotation). These motion should be labeled respectively $(D,0,0,0)$, $(0,0,0,0)$, $(0,R,0,0)$ and $(D,R,0,0)$. The images and the different results are presented in Fig. 4. In this example, we have tested the two methods $GLR-Method$ and $RIC-Method$ with the O_G observations.

The $GLR-Method$ gives very good results for the four different motions in the image, since it chooses for each one of the four regions the right label. The threshold does not depend on the size of the analysed region, and the kind of motion.

The $RIC-Method$ gives the right label for the three motions $((D,0,0,0)$, $(0,R,0,0)$ and $(D,R,0,0)$. For the motion $(0,0,0,0)$, this method selects the complete label (D,R,H_1,H_2).

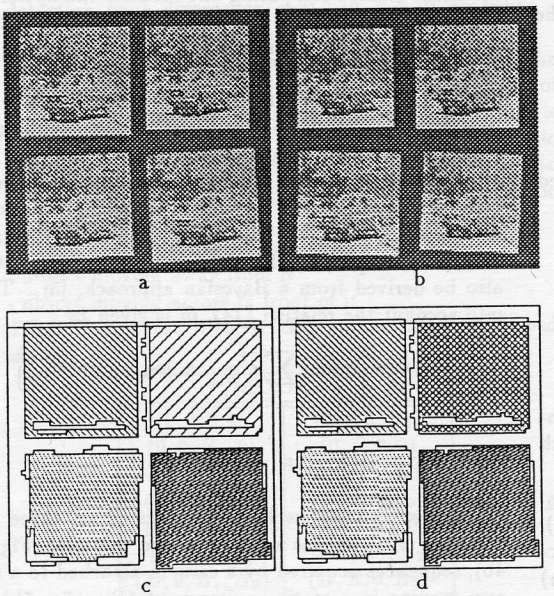

Figure 4 : *real images with synthetic motions*
(a) and (b) : the two successive images.
(c) : Labeling with the GLR-Method threshold = 0.0025
(d) : Labeling with the RIC-Method

The second set of experiments has been made on sequences acquired in the laboratory. The scene is static, and the camera is moving (example 2 - Fig. 5 and 6). In this example, the scene is composed of a scale model representing an abbaye, and the camera is attached to the arm of a moving robot. We have dealt with two different camera motions : translation along its optical axis (that should be labeled $(D,0,0,0)$), and rotation around its optical axis (that should be labeled $(0,R,0,0)$). In this case, it is not necessary to segment the image since the same motion is present in the whole image. However, we have arbitrarily performed the labeling in small squared areas.

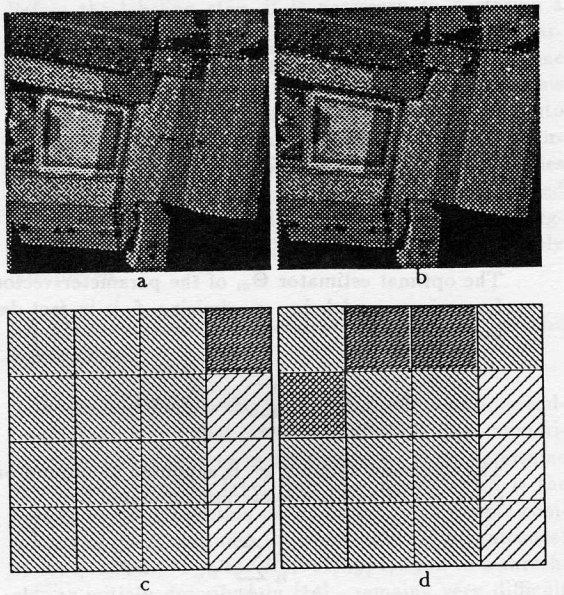

Figure 5 : *static scene and moving camera translation along the optical axis* $(D,0,0,0)$.
observations : O_G *- window size : 64*64*
(a) and (b) : two successive images.
(c) : Labeling with the GLR-Method threshold = 0.01
(d) : Labeling with the RIC-Method

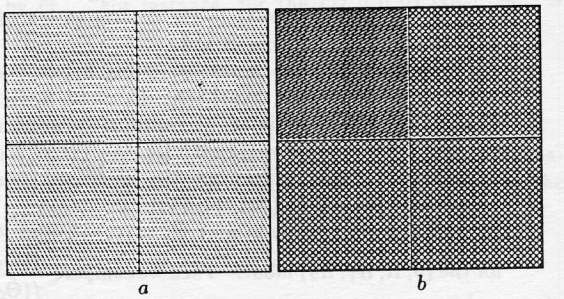

Figure 6 : *static scene and moving camera rotation around the optical axis (0,R,0,0).*
observations : O_V *- window size : 128*128*
(a) : Labeling with the GLR − Method threshold = 0.5
(b) : Labeling with the RIC − Method

Fig. 5 presents results of labeling for the first example. Two successives images out of the sequence where the camera is translating along its optical axis (theoretical label : $(D,0,0,0)$). The labeling has been undertaken on square windows of size 64*64 pixels, with the O_G observations. In the second example (theoretical label :

$(0, R, 0, 0))$, we have used 128*128 windows and the O_V observations. The aim is to test each method in different configurations.

The results obtained with the $GLR - Method$ (Fig.5 and 6 and other complementary results not reported in this paper) induce the following remarks. First, the threshold can be set to the same predetermined value whatever the kind of motion and the size of the analysed region. Second, this threshold is different for the O_G observations and for the O_V observations. Moreover, it is easier to set it in this last case. Globaly, results are quite good with this method.

Results obtained with the $RIC - Method$ are much less satisfactory than with the $GLR - Method$. It depends quite largely on the size of the window. Indeed, it seems that when this size is too big, the criterion often validates the complete label (D, R, H_1, H_2). This occurs with both kinds of observations. A surprising behaviour of this method is that the results are better with the O_G observations than with the O_V observations, contrary to the $GLR - Method$.

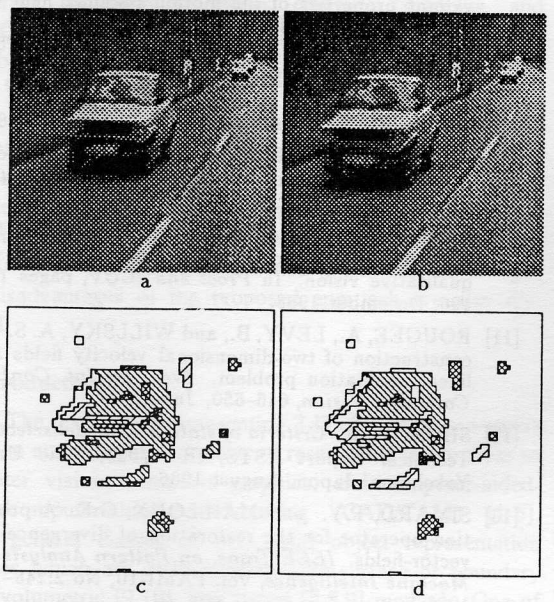

Figure 7 : *road context sequence - labeling on O_G*
(a) and (b) : two successive images.
(c) : Labeling with the $GLR - Method$ threshold = 0.01
(d) : Labeling with the $RIC - Method$

Finally, we have treated a real sequence depicting outdoor scenes in a road context (example 3 - Fig. 7). The vehicles in the scene are moving toward the camera, which is itself attached to a moving car. All the movements in the scene are parallel to its optical axis. The corresponding regions in the image should be labeled as $(D, 0, 0, 0)$. We focuse on the nearest vehicle in the scene. The considered observations are the O_G observations.

Results obtained with the $GLR - Method$ are again very good. The nearest car is labeled $(D, 0, 0, 0)$. The threshold is equal to 0.01. This threshold is the same on the whole sequence, and at each time of the sequence, the right label is chosen.

The RIC_{Method} gives the following results : in the sequence of eight images, the right label is chosen five times, the complete label twice, and a totaly wrong label $((0, R, 0, 0))$ one time. These results are much better on this sequence than for the previous example.

The global behaviour of the two methods, with a quite complete set of data, can be summarized in several points :

- the $GLR-Method$ gives very good results whatever the kind of observations, the size of the analysed area and the kind of motion.
- however, the threshold is usually different according to the sequence, and the kind of observations. It is very robust when the observations are the O_V, it is more sensible with the O_G.
- the $RIC-method$ seems to give better results with the O_G than with the O_V observations, contrary to what could be predicted.
- it often chooses the complete label $((D, R, H_1, H_2))$ even if the right one is a simplier one.

The tendency of the $RIC - method$ to validate the complete label can be explained : as soon as the estimation of a pure motion (e.g. $(D, 0, 0, 0)$) is too noisy, the resulting motion field can be seen as a complete motion one $((D, R, H_1, H_2))$. Indeed, in the motion space, pure motions could be represented as "Dirac distributions", all the remaining being the complete motion one. Thus, the distance between any pure motion class (for instance $(D, 0, 0, 0)$) and the (D, R, H_1, H_2) class is very narrow. As we do not directly deal with an acquired digital signal but with derived estimates, it may occur that the parameter estimation may be inaccurate. In such a case, as the criterion is relatively flat, i.e. the criterion values for several models among the set are not greatly different, the criterion may validate another model. Nevertheless the interest of such criterion lies in the fact that it does not require any threshold definition and does not compare models in pair but together. We are investigating the way of improving the quality of the parameter estimation by preprocessing the observation field.

Another alternative which is currently studied to improve the performance of the $RIC-Method$ is as follows. Instead of considering only the S_{label} set, we actually consider the S_{model} set. In this case, it may happen that an unappropriate model (i.e. physically impossible) is selected. This has led to extend the $RIC-Method$, by taking into account every model of S_{model} to choose the right label. Each model of S_{model} contributes in this case to the likelihood of each label, of course at different degrees according to the label. For the while, we have described these contributions very roughly : for instance, if the criterion has selected the $(D, 0, H1, 0)$ model, we decide that the true label is $(D, 0, 0, 0)$. The results shown in Fig. 8 prove that this approach can be interesting. However, we have to formulate more properly the contribution of

each model of S_{model} to each label of S_{label}. Indeed, the choice of the $(D, 0, H1, 0)$ model can also in certain cases signify that the true label is (D, R, H_1, H_2). Presently, the $(D, 0, H1, 0)$ model only contributes to the $(D, 0, 0, 0)$ label. This modified version of the $RIC - Method$ is still in progress.

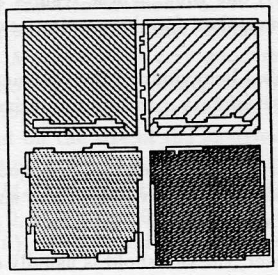

Figure 8 : *preliminary results on the example 1 (cf Fig.4a-b) with the extended $RIC - Method$*

7. Conclusion

This paper has dealt with the derivation of a qualitative description of the kinematic behaviours of the objects in the scene. It represents a real alternative to quantitative estimations of the 3D motion and structure parameters. The interests of this study are three-fold. First, we have explicitly and analytically linked the description cues to the 2D apparent motion and the 3D motion. Second, we have described model-based statistical decision methods to achieve the numerical-to-symbolic step. They enable to address any determination of qualitative motion information in a non ad-hoc and unified manner. Third, motion interpretation even if the velocity field is not beforehand estimated.

The two methods present advantages and drawbacks. The first one ($GLR - Method$) yields very good results, but requires the determination of a threshold value. The second one ($RIC - Method$) allows to test *together* the defined labels, but results are not yet completely satisfying. Several attempts are currently pursued to improve this last method (enhancement of observations, direct consideration of the S_{model} set). As a matter of fact, the main extension should be the following. Up to now, labeling is achieved instantaneously (i.e. considering motion information between only two successive images). We are now integrating the temporal axis in the labeling process, which should enlarge the range of the qualitative description and then the efficiency of the $RIC - Method$.

Acknowledgments

This work is supported by MRT (French Ministery of Research and Technology) in the context of the EUREKA european project PROMETHEUS, under PSA-contract VY/85241753/14/Z10.
We thank Dr Enkelman for providing the image sequence of Fig.7a-b.

References

[1] AGGARWAL, J.K. and NANDHAKUMAR, N. On the computation of motion from sequences of images- a review. *Proc. of the IEEE*, Vol. 76, No 8:917–935, August 1988.

[2] BOUTHEMY, P. and SANTILLANA RIVERO, J. A hierarchical likelihood approach for region segmentation according to motion-based criteria. *Proc. 1st ICCV*, 463–467, 1987.

[3] BURGER, W. and BHANU, B. Dynamic scene understanding for autonomous mobile robot. *Proc. CVPR*, 736–741, 1988.

[4] CARLSSON S. Information in the geometric structure of retinal flow field. In *Proc. 2nd ICCV*, pages 629–633, December 1988.

[5] FRANCOIS, E. and BOUTHEMY, P. *Vers une interprétation qualitative de comportements cinématiques dans la scène à partir du mouvement apparent.* Technical Report 1081, INRIA-Rennes, August 1989.

[6] KOENDERINK, J.J. Optic flow. *Vision Research*, Vol. 26, No 1:161–180, 1986.

[7] KOENDERINK, J.J. and VAN DOORN, J.J. Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta*, Vol. 22, No 9:773–791, 1975.

[8] LONGUET-HIGGINS, H.C. and PRAZDNY, K. The interpretation of a moving retinal image. In *Proc. Roy. Soc. Lond.*, pages 385–397, April 1980.

[9] NAGEL, H.H. From image sequences towards conceptual descriptions. *Image and vision computing*, Vol. 6, No 2:59–74, May 1988.

[10] NELSON, R. C. and ALOIMONOS, J. Using flow field divergence for obstacle avoidance: towards qualitative vision. In *Proc. 2nd ICCV*, pages 188–196, December 1988.

[11] ROUGEE, A., LEVY, B., and WILLSKY, A. S. Reconstruction of two-dimensional velocity fields as a linear estimation problem. *Proc. 1st Int. Conf. on Computer Vision*, 646–650, June 1987.

[12] SHIBATA, R. *Criteria of statistical model selection.* Technical Report KSTS/RR-86/009, Keio Univ., Yokohama, Japon, August 1986.

[13] SIMARD, P.Y. and MAILLOUX, G.E. A projection operator for the restoration of divergence-free vector-fields. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-10, No 2:248–256, 1988.

[14] THOMPSON, W.B., BERZINS, V.A., and MUTCH, K.M. Analyzing object motion based on optical flow. *Proc. ICPR*, 791–794, 1984.