Statistical filtering of motion field from image sequences

Jean-Pierre CASSOU †
and Antoine CORNUEJOLS ††

† Arcanset Group, 38 bd. R. Poincarré 92380 Garches.
and
Laboratoire de Météorologie Dynamique, Ecole Normale Supérieure, F-75005 Paris

†† Laboratoire de Recherche en Informatique, UA410 du CNRS Universite de Paris-sud
Batiment 490, 91405 ORSAY (France)

Abstract :

Techniques for the estimation of the motion vector field from image sequences may be subdivided into two main different approaches. Either the motion field is calculated from tracking shape or feature correspondences between images, or it is estimated using the optical flow vector field method and its variations.

The latter approach is very alluring because it doesn't require the previous identification of remarkable points in the images, but so far it has stumbled over the complexity of interpreting, i.e. solving, the gigantic set of interdependent constraints generated. Furthermore, it has proven difficult to introduce and express in such a framework the notion of model of what is looked for in the images.

We treat both of these problems in this paper. First, we identify the interpretation of an optical flow vector field, obtained from a first order Taylor approximation of the spatial variation of the gray gradient [Schunk85], to a complex optimization problem. And we give it a form which make it amenable to new techniques based on energy minimization recently introduced [Kirk83, Carn85]. Second we introduce the notion of models of what is of interest in the images. Particularly, it was found interesting to focus on three particular types of motion: uniform motion, rotation and contraction/ dilatation. Together with a definition of the types of shapes to be looked for, this provides a parametric model of the patterns that are to be detected or extracted in the vector field.

We then are able to run a simulated annealing process which tends to minimize the difference between the energy asssociated with the "observed" vector field , and the energy associated with a vector field that would result from the parametric model. At the final equilibrium stage, this yields the optimal parameters of the model that would produce the noisy image sequence observed, and therefore provides the user with a high level description of the image sequence in terms of the phenomena he/she is looking for.

We illustrate this method with the detection of objects in uniform motion in a sequence of images.

Problems of performances of this scheme and perspectives for further improvements are also addressed in the course of this paper.

1. The problem of identifying moving objects.

The problem we study in this paper is, very classically, the identification of objects in motion through an image sequence. It implies both the description of the objects and the characterization of their motion. This is not a trivial task because in general there are infinitely many two-dimensional velocity fields that are consistent with the changing image, and additional constraints, based on assumptions about the physical world, must be made.

We chose to take as a constraint that the image sequence resulted from a parametric model involving solids in translation, rotation or compression/dilatation. And, considering it as an optimization problem, we solved it with a recently developped technique: Simulated Annealing.

2. The classical approaches and previous works.

The analysis of image sequences to detect moving objects in dynamical scenes is full of applications, for instance in robotics, medicine, transports, or meteorology, and has been an active on-going research axis for a long time. Several approaches have been tried.

Leaving aside the so-called *global approaches* [Nagel,1986] that postulate that the differences between two frames in a image sequence can be described by a function depending only on a few parameters, the most widely used techniques are *local*. They emphasize the study of differences at the level of the pixels between two images. Among them one approach consists in the determination and extraction of "change regions" by establishing first a parameterized description of structural gray value perceptions and then applying statistical tests to isolate the significant regions of variations. This approach meets some difficulties in exploiting the resulting map and in disentangling the factors that may have contributed to the detected changes without calling for highly application-specific knowledge. There remains two main approaches.

One is to analyse separately the images of the sequence using classical segmentation techniques in order to isolate the elementary patterns. Then one tries to track localisable features through the images and to measure their displacement. This method works well for 2D images of rigid objects but is prone to failure in case of 3D scenes of plastic patterns.

The other attempts to directly evaluate the displacement vector **u** from the spatio-temporal gray values I(x,y,t) available from at least two images. Actually, only the optical flow, which denotes the image plane shift of the gray structure between the images, can be derived from this information. This approach has been investigated by many authors and is more thoroughly described in the following. It is promising because it is in essence independant of any application-specific information and may soon justify the implementation of special purpose processing architectures. Its use requires however some additional assumptions which are the basis of the differences between the authors. We ourselves emphasize certain hypotheses that seem natural, and, above all, demonstrate the use of a new computational procedure to solve the corresponding problem. Besides, it is noteworthy that this computational tool may be applied almost as well to other optical flow models.

3. Statistical filtering of motion field.

Section 2 gave an overview of the different possible paths to the estimation of motion in an image sequence, mainly tracking feature correspondences, or interpreting the optical flow vector field. The approach taken in this work belongs to the second category. We consider that it is more easy and natural to extract the information concerning movement from the low-level and local data corresponding to spatial and time gradients in the image sequence, rather than by involving high level expectations about the scene and complex non-local computations. But our own approach doesn't preclude the possibility of providing the vision system with a model of what is looked for in an image. This is actually, as we shall see, the basis of the novelty of our own approach compared to the one of [Horn & Schunk,1981]. However the model lays on domain-independent constraints and the computations remain completely local giving rise to interesting possibilities for rapid processing.

A brief account of the optical flow concept is first given in order to make this paper self-contained and to explain the [Horn & Schunk,1981] assumptions and method. Then the simulated annealing technique is described together with our implementation for the analysis of image sequences. And finally the results obtained using this method, and plans for future improvements are given.

3.1 The optical flow concept.

The idea of optical flow rests on the constatation that a real time vision system, such as biological ones, observes images that vary essentially continuously in their characteristics. This leads to an interesting method of motion perception based on local comparison of brightness.

Let characterize each point (x,y) at time t by its brightness I(x,y,t). If during the period dt, the point has moved in the direction (dx,dy) then we get:

$$I(x,y,t) = I(x+dx,\ y+dy,\ t+dt)$$

Then, expanding the right hand side of this equation using a first order Taylor expansion:

$$I(x,y,t) = I(x,y,t) + \frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt$$

which yields:

$$-I_t = I_x v_x + I_y v_y \qquad (3.1\text{-}1)$$

That means that the *time* rate of change in intensity of the image is the product of the *spatial* rate of change of the brightness region by the *velocity* of that region.

The equation (3.1-1) constrains the velocity (v_x,v_y) to lie on the line it expresses.

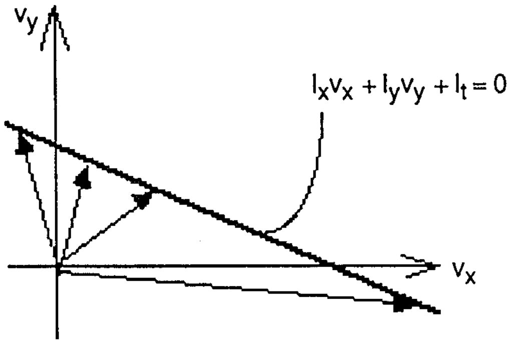

Fig. 3.1-1: Constraints on the velocity of the point (x,y)

As [Horn & Schunk,1981] note, "optical flow cannot be computed locally since only one independent measurement is available from the image sequence at a point, while the flow velocity has two components. A second constraint is needed." [Horn & Schunk,1981] postulate that the spatial variation of the optical flow components should be as small as compatible with the system of Euler-Lagrange partial differential equations associated with the optimization problem specified by the equation:

$$\iint dx\,dy\,\left([\vec{\nabla I}^{T} \cdot \vec{u} + I_t]^2 + \alpha^2\,[u_x^2+u_y^2+v_x^2+v_y^2]\right) \rightarrow \text{minimum}$$

They then employ the iterative Gauss-Siedel method to obtain an approximation to the system expressed above.

With our technique, we are able to get rid of the smoothness hypothesis and to replace it with a natural spatial constraint: pixels which are part of the same object should have the same velocity. We then give a very general model of neighbourhood that can cover and segment the image into objects of different velocities. The segmentation process and the corresponding determination of the velocity characteristics are obtained very simply using a technique derived from statistical physics, and recently applied to several optimization problems: simulated annealing.

3.2 Simulated annealing and image sequences.

The method of simulated annealing is a technique that has recently attracted significant attention as suitable for optimization problems of very large scale. Its underlying principle rests on the observation of Nature's strategy to reach states of minimum energy for slowly cooled systems. Thus, liquid metal sufficiently slowly cooled reach a quasi crystalline structure corresponding to their minimum energy state. To do so Nature employs a minimization algorithm based on the so-called Boltzmann probability distribution,

$$\mathrm{Prob}(E) \sim e^{(-E/kT)}$$

This equation expresses the idea that a system in thermal equilibrium at temperature T has its energy probabilistically distributed among all different energy states E. In other words a system described by its degrees of freedom x_i has all of its possible configurations $\{x_i\}$ weighted by their respective probability factor, $e^{(-E(\{xi\})/kT)}$, where $E(\{x_i\})$ is the energy of the configuration. The "miracle" is that at low temperatures, the otherwise extremely rare ground state configurations among all the possible configurations, are the ones that dominate the Boltzmann distribution. However low temperature is not a sufficient condition for finding ground states of a system, if it imposed too rapidly, before the system can in a way make a rough exploration of the phase space, the system is likely to get trapped in a local minimum configuration. In practical contexts, a careful annealing is required.

In 1953, Metropolis et al. [Metropolis,1951] first incorporated this kind of process into numerical calculations. The idea is that, accordingly to thermodynamics, a system sometimes goes uphill as well as downhill; but the lower the temperature, the less likely is any significant uphill excursion. At each step of the calculation the system can make a choice between several options, if, say, it considers an option that leads to a change DE in the energy, it will actually take it and change the state of the system with a probability $p=e^{(-DE/kT)}$. Notice that if DE<0, i.e. this option leads to a decrease in energy, the probability is greater than unity, in such cases the change is arbitrarily assigned a probabilty p=1, that is the system always takes such an option. This general scheme, of always taking a downhill step while sometimes taking an uphill step, is known as the Metropolis algorithm. By repeating this step many times with a given temperature T, one simulates the thermal motion of atoms in contact with a heat bath at temperature T. The whole optimization process involves lowering the temperature by slow stages until the system "freezes" and no further changes occur.

Provided with this wonderful "natural" mechanism the art of the researcher is to provide the energy function describing the system together with the options opened to the system, its possible configurations and the control parameter T.

In this research we applied simulated annealing to the optimization problem related to the finding of motion field from the optical flow obtained from image sequences. In this scheme, the states of low energy correspond to the "ideal" optical flow stemming from some typical movements of objects in the scene. For instance, we first tried to identify objects in uniform motion. Given this model of what is of interest, it is possible to derive a corresponding energy function that is minimum when the state of the system segments the scene into regions of uniform speed. In that way, starting from any initial state (segmentation with attached speeds), and given noisy data, the algorithm will converge toward the ideal scene that would have produced such data.

The constraints that allow us then to extract the necessary information from the optical flow are a spatial constraint: for instance, the pixels of an object are adjacent and have the same velocity; and a constraint on the type of velocity field that are looked for: for instance, uniform motion, rotation, contraction/dilatation with respect to a fixed point. This guides the determination of an appropriate energy function.

As an example we give here an illustration of the simulated annealing process applied to the determination of objects in uniform motion from two images.

The images are square ones. They are multi-leveled gray (256 levels in our images). And there is no assumption on the scene observed other than the objects are connected (this can also be changed easily), and their respective motions are uniform. There can be a large amount of noise, and that was actually the case in the images we analyzed.

The first point is to see what that means for a collection of points to be in uniform motion (or for that matter in any other type of motion as soon as it is well identified). The optical flow hypothesis allows us, as we saw in section 2, to draw the line on which the velocity (u_i,v_i) of a point i must lie. In the ideal case, then, a collection of points animated with the same uniform speed, should have their respective lines intersecting at a unique place in the (u,v) plane corresponding to their speed (given that all these lines are not a single one which is highly unlikely). Finding this place or speed is equivalent to finding the parameters u and v that minimize the following systems of equations, (3.2-1) corresponding to a L^1 norm, and

$$\sum_i \left| \frac{\partial I(x_i,y_i,t)}{\partial x} u_i + \frac{\partial I(x_i,y_i,t)}{\partial y} v_i + \frac{\partial I(x_i,y_i,t)}{\partial t} \right| \qquad (3.2\text{-}1)$$

$$\sum_i \left[\frac{\partial I(x_i,y_i,t)}{\partial x} u_i + \frac{\partial I(x_i,y_i,t)}{\partial y} v_i + \frac{\partial I(x_i,y_i,t)}{\partial t} \right]^2 \qquad (3.2\text{-}2)$$

(3.2-2) to a L^2 norm.

A possible configuration of the system consists therefore in a segmentation of the image at time t, for which the corresponding global energy is the sum of the energies of each segment calculated with the formula (3.2-1) or (3.2-2). For instance, applying the equation (3.2-1) to look for the velocity (u_s,v_s) minimizing the energy of the segment s, we find that

$$\sum_s \left(\min_{u_s,v_s} \left\{ \sum_{i\varepsilon s} \left| \frac{\partial I(x_i,y_i,t)}{\partial x} u_i + \frac{\partial I(x_i,y_i,t)}{\partial y} v_i + \frac{\partial I(x_i,y_i,t)}{\partial t} \right| \right\} \right)$$

the global energy is given by (3.2-3):

At each step of the algorithm, the system can either (1): split an existing segment or region into two pieces, that therefore can have different values (u,v) minimizing their respective energy, or (2): replace a segment by the union of itself and one of its neighbour, yielding a new value (u,v) for all the points of this new region. Each time the system consider an option, SPLIT or MERGE, it computes the attached change in energy DE and acts accordingly to the decision of an oracle computing the expression $e^{(-DE/kT)}$. Of course, if the energy was exactly the one given by expression (3.2-3), the option SPLIT would never be rejected, because even in the worst case where the two new regions were in fact part of the same object and having therefore the same velocity, the corresponding DE would simply be 0. It is thus necessary to add a coefficient to (3.2-3) penalizing great numbers of segments in the scene. That way a split would be profitable to the system only if it brings effectively a better segmentation of the image. Accordingly one obtains a definition of E which is the sum of the energy given in (3.2-3) and of the term: κn_s, where n_s denotes the number of segments or objects in the scene.

4. Conclusion.

We have cast the classical problem of motion identification into an optimization framework well suited for the method of Simulated Annealing. Accordingly the expression of an energy function was given, which minimum corresponds to the ideal parametric model associated with the image sequence observed.

This new technique, in addition to solve elegantly the problem, allows one to "play" with variations on the formulation of the model, for instance the type of energy function to minimize or the kind of possible spatial re-arrangements giving a segmentation of the image. But above all, one of the main advantages of this approach is to allow the user to define and then obtain high-level descriptions of the phenomenon under study, e.g. in terms of polygons or whirlpools.

It was experimented with success on uniform motion cases using an "augmented" PC-AT with FORTRAN programs. More experiments are planned to determine the efficiency of this method on complex scenes. Its robustness to noise is inherently very large. Schemes to improve the computing speed are projected.

References :

_P. CARNEVALLI, L. COLETTI, S. PATARNELLO (85): **"Image processing by simulated annealing"** in *IBM J. Res. Develop.* vol.29, No.6, nov. 1985.

_B.K.P. HORN and B.G. SCHUNK (81): **"Determining Optical Flow"**, *Artificial Intelligence* 17, 185-203.

_T.S. HUANG, ed. (83): **Image Sequence Processing and Dynamic Scene Analysis,** *NATO ASI Series* F 2, Springer Verlag 1983.

_S. KIRKPATRICK, C.D. GELATT, M.P. VECCHI (83): **"Optimization by Simulated Annealing"** in *Science* vol.220,13 may 1983, pp.671-680.

_H-H. NAGEL (86) : **"Image Sequences - Ten (octal) Years - From Phenomenology towards a Theoretical Foundation."** Proc. of the *ICPR-86*. Paris, France.

_B.G. SCHUNCK (85): **"Image Flow: Fundamentals and Future Research."** in Proc. *IEEE Conference on Computer Vision and Pattern Recognition CVPR-85*, June 19-23, 1985, San Francisco, CA, pp.560-571.