

A COMPUTER VISION METHOD FOR MOTION DETECTION USING COOPERATIVE KALMAN FILTERS

Florence GERMAIN and Thomas SKORDAS

ITMI

11, Chemin des Prés, BP 87

F-38243 Meylan cédex

France

e-mail: f.germain@itmi.fr, t.skordas@itmi.fr

ABSTRACT :

In this paper, we describe a method for the computation of the optical flow in a sequence of images acquired by a moving camera. The method is guided by the simultaneous satisfaction of two opposite constraints introduced by the regularization process used to remove the underdetermination related to the aperture phenomenon. Two cooperative Kalman filters are used which allow us to compute the motion information present in the images. Such a method is particularly well suited to deal with outdoor scenes and some real experiments are presented which highlight its validity.

KEY WORDS: Computer Vision, Motion detection and estimation, Kalman filtering.

1 Introduction

In the context of a rapidly growing demand for a computer based interpretation of images, motion analysis plays a major role. Image motion is mainly used:

1. as a meaningful information on objects motion in the context of scene behavioral analysis [4];
2. as a model of the space-time redundancy in an image sequence, in the context of coding for image transmission [3].

Motion analysis is commonly based on a pixel per pixel estimation of the instantaneous displacement of the underlying physical points. The resulting dense field, estimated from each pair of consecutive images in a sequence of images acquired through a single CCD camera, is commonly named the *optical flow*.

1.1 Explicit motion information

The optical flow estimation relies on the local information which is intrinsically part of the image. In the following, we will refer to it as the explicit motion information.

The explicit motion information has a local nature. The extraction process, at a given pixel, of a motion information is conditioned by the existence of a non zero spatial gradient. Therefore, the explicit information is available only inside a non homogeneous area of an image.

Moreover, explicit information, when it does exist, is incomplete. It only provides a partial view of the underlying motion, depending on the spatial gradient orientation at the considered pixel. Most of the time, it amounts to the projection of the searched displacement along the local spatial gradient of the intensity. This projection is commonly known as the *orthogonal displacement*. The underdetermination

which comes from the lack of a second projection illustrates a physical phenomenon known as the *aperture phenomenon* [5]. The ensued ambiguity summarizes the difficulties which are faced when estimating the motion in image sequences.

1.2 The regularization operation

Due to the incomplete nature of the explicit motion information, the optical flow estimation implies a spatial integration of the local information, in order to remove, when this is possible, the underdetermination related to the aperture phenomenon. This operation, usually named *regularization*, requires an accurate determination of the spatial scope of integration. Intuitively, this integration must respect the boundaries between two homogeneously moving image areas. Therefore, every qualitative change in the displacement vector field must be effectively detected.

The quality of a regularization operator will be judged on both its integration capability and its ability to detect a qualitative change of the estimated vector field. The antagonistic nature of these two constraints plays a major role in the choice of a regularization method.

We propose an estimation of the optical flow to be used for the analysis of moving scenes. Our approach is guided by the simultaneous satisfaction of the opposed constraints put by the regularization. It is based on the use of a parametric estimator built on a Kalman filtering process which is similar to the one proposed by Stuller and Krishnamurthy [1].

2 A parametric Kalman model for the motion estimation

The parametric Kalman model which has been introduced in [1] generates a parameterized family of optical flow estimators. The motion is estimated using a line-based scanning

of the image. The spatial integration of explicit information is due to the process memory which is derived from the model state equation. The main interest of this approach comes from its ability to remove the underdetermination bound to motion estimation.

In order to use this model, three main hypotheses must be satisfied:

1. the illumination is supposed to be constant between two consecutive images of the sequence;
2. the displacement between two consecutive images can be locally assimilated to a translation;
3. the image noise is supposed to be white and of known variance.

2.1 Model description

The model equation of the filter provides the relation between the displacement vectors associated to two consecutive pixels. It can be expressed as a function of a transition matrix Φ and of a white noise (w_i):

$$D(i) = \Phi D(i-1) + w_{i-1}$$

The model makes the assumption of a parametric transition matrix whose type is:

$$\Phi = \lambda I_d$$

with $0 \leq \lambda \leq 1$ and I_d being the identity matrix. The choice of λ determines the estimator's *memory*, therefore it strongly influences the trade-off between the integration capability and the adaptability when being faced to a sudden change in the estimated process.

In the following, E_λ will denote the estimator parameterized by λ .

In every pixel, the measure is equal to the component of the searched displacement along the spatial intensity gradient. The fact that this component is the only information to be intrinsically available provides an intuitive justification.

2.2 Estimator running

Between two consecutive pixels, the estimation is updated both in the direction of the spatial intensity gradient and in the directly orthogonal direction, under the premise that there exists a significant variation of the gradient along the current image line. When the preceding condition is satisfied, and according to the operator memory, the estimation process progressively removes the underdetermination due to the aperture phenomenon. The estimation error remains minimum along the spatial intensity gradient direction and is maximum along the orthogonal direction. Associated to each pixel, the value of this second component of the error is a measure of the residual underdetermination.

The parametric Kalman model represents a continuum between two extreme estimators (resp. $\lambda = 0$ and $\lambda = 1$) (refer to figure 1). These two extreme estimators are radically opposed from the regularization point of view. Finally, each instance of the model is a trade-off between a good integration capability and a good adaptability to the underlying estimation process.

3 Using two cooperating Kalman filters

3.1 Extreme models

Closely looking at the two extreme cases generated by the extrema of λ in the $[0, 1]$ range is extremely valuable. The E_0 and E_1 estimators are well suited to the motion estimation of a rigid object (E_1) and to the motion estimation of a totally random motion (white motion process) (E_0).

E_0 estimator: As the process memory is non-existent, no spatial integration is performed. In this case, it is easily shown that the gain vector is oriented in the direction of the intensity spatial gradient. The model converges toward the estimate of the local orthogonal displacement vector field.

E_1 estimator: As the process memory is total the spatial integration is maximum. A significant change in the gradient direction between two consecutive pixels leads to an optimal update of the gain, in the direction perpendicular to the gradient one. In this case the model converges towards the estimate of the true displacement vector field.

From the above, one may infer that:

- The E_1 estimator allows an optimal removal of the underdetermination bound to motion estimation. But, as a consequence, the estimator's memory is detrimental to its adaptability.
- The E_0 estimator does not allow any underdetermination removal at all. But its adaptability, that is its ability to quickly adapt itself to a sudden change of the estimated signal is excellent.

The principle of our method is based on a cooperative usage of the E_1 and E_0 models to provide an accurate motion estimation. It is justified by the simultaneous exploitation of the good integration characteristics of E_1 and of the good adaptability of E_0 .

3.2 Cooperation principle

We have highlighted the E_1 estimator capability to, according to the parametric family, optimally remove the underdetermination bound to the motion estimation. The major problem related to the usage of E_1 is its lack of ability to adapt itself to a qualitative change in the estimated system.

The basic principle of the cooperation relies on the parallel activation of the E_1 and E_0 estimators in order to use the adaptability of the second to the benefit of the first.

The useful information of E_0 resides in the innovation term. In this specific case, the innovation - the difference between a measure and the predicted value of this measure - is easily expressed as a function of the spatial gradient $G(i)$, the searched displacement $D(i)$ and a white noise v_i according to the equation:

$$z_i = G(i)^T D(i) + v_i$$

As a consequence, every qualitative modification in the "intensity spatial gradient" vector field G_i , as well as in the "true displacement" vector field $D(i)$ induces a change of the E_0 estimator innovation.

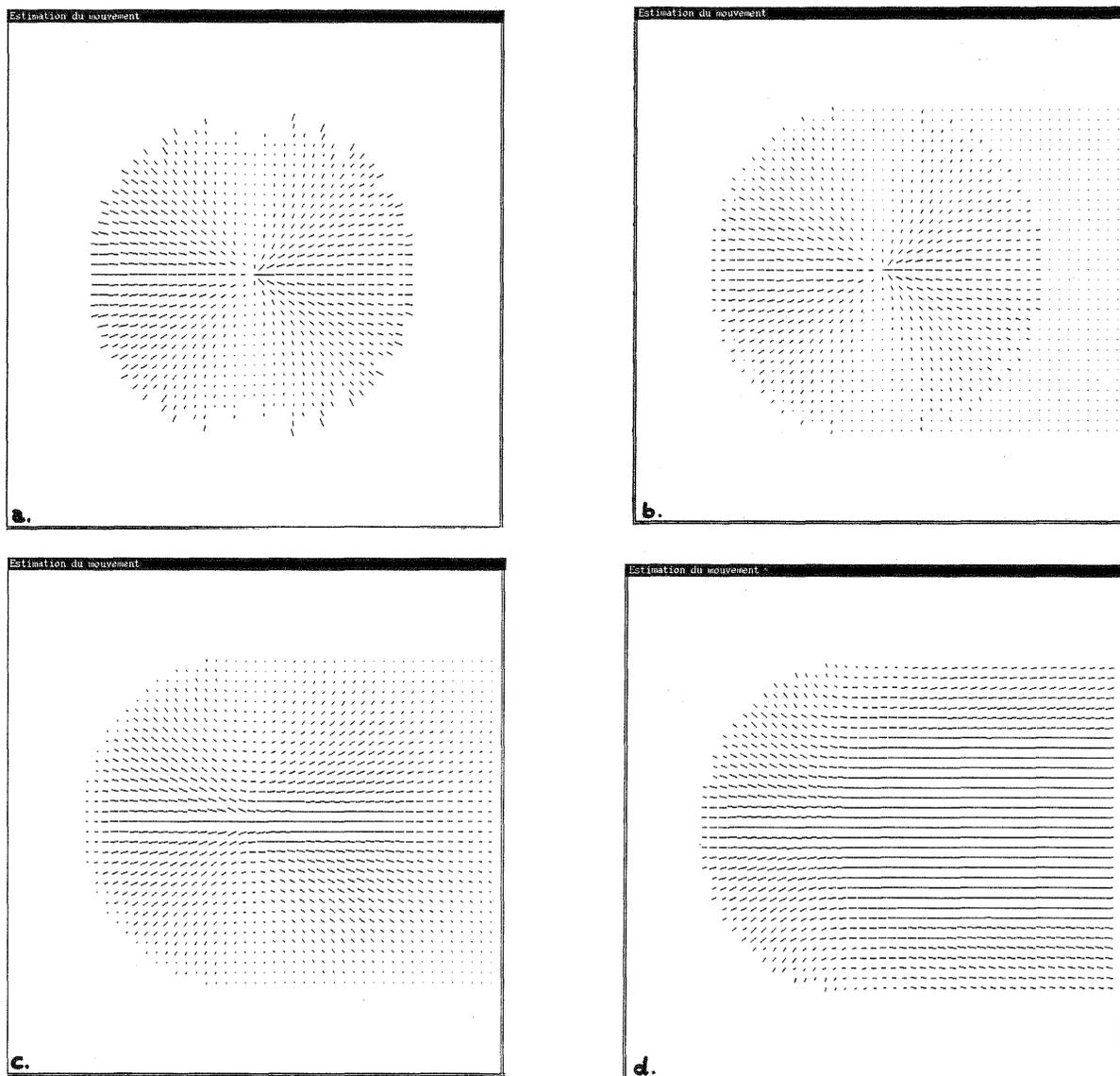


Figure 1: The parametric Kalman model, as a continuum between two extreme estimators. Three motion estimators, respectively $E_{\lambda=0.1}$, $E_{\lambda=0.8}$ and $E_{\lambda=1.0}$, are applied to the case of an horizontally moving synthetic 2D gaussian form, with a displacement of one pixel per frame. (a) Exact orthogonal motion field. (b) Result of an $E_{\lambda=0.1}$ estimation. The model converges toward the estimate of the orthogonal

motion field. (c) Result of an $E_{\lambda=0.8}$ estimation. The estimation of the exact motion field (horizontal motion field) begins to appear. In this case however, one may notice that the right boundary tends to vanish. (d) Result of an $E_{\lambda=1.0}$ estimation. The exact motion field is estimated, but the right boundary of the form is now completely lost.

The idea is to test, while estimating, if there is a change in the innovation output of the E_0 estimator. A PAGE-HINKLEY cumulated sum test is used for the detection of abrupt changes [2]. When a change is detected, the E_1 model is re-initialized. This is to insure the non integration of information which is related to two different homogeneously moving areas of the image.

4 Experimental results

This section describes the results obtained when applying the ideas described in this paper to a sequence of real Infra Red images. Such images have been acquired using the thermic camera "ATHOS" of the SAT company. A pair of

two consecutive images of the sequence is shown on figure 2. The images show a traffic road of Paris and were taken with a camera mounted on a moving vehicle. The size of the raw images is 256×512 pixels.

The two cooperative Kalman filters have been applied to this pair of images. In order to better show the found changes of the motion field, we give the detection result only in a small window corresponding to the left car of the raw images (refer to Figure 2). The orthogonal motion field related to this car is estimated using the $E_{\lambda=0}$ estimator (refer to Figure 3). The innovation of the Kalman filter, associated to the dotted line of Figure 3, is represented in Figure 4. A detection of abrupt changes in the innovation signal is done, using a PAGE-HINKLEY cumulated sum test. Four changes are found, directly associated to the limits of the

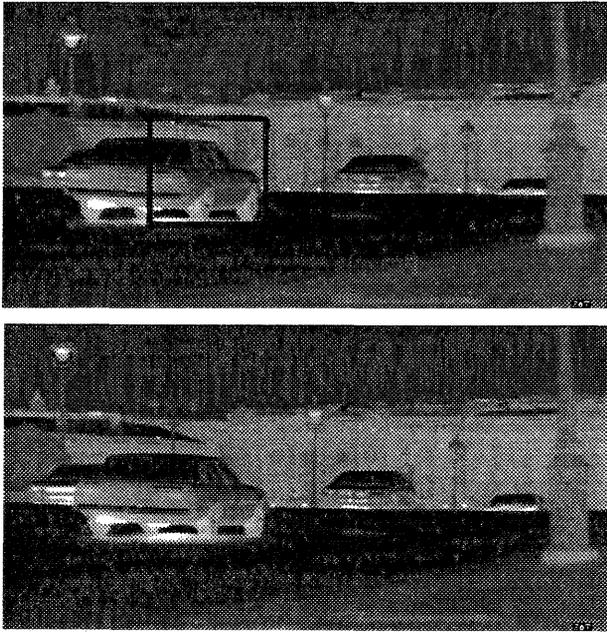


Figure 2: A pair of IR raw images.

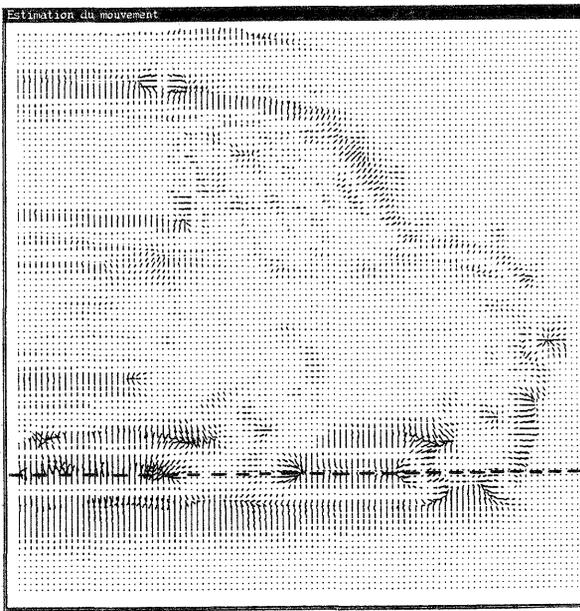


Figure 3: Detection of changes in the orthogonal motion field.

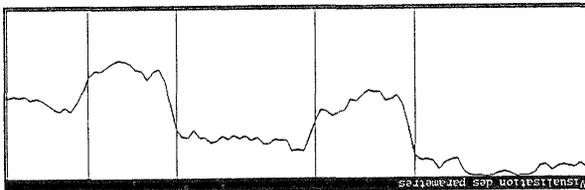


Figure 4: The innovation signal of the Kalman filter corresponding to the dotted line shown in the previous figure. One may notice that the four detected changes correspond to the frontiers of the car wheels.

car wheels.

5 Conclusion

In the introduction we have shown that the explicit information is characterized by its local and incomplete nature. The proposed regularization method allows one to complete the explicit information where it is available.

In the case of images of outdoor scenes, which are often well textured, the explicit motion information is available everywhere in the images. A problem arises when processing images of indoor scenes which are very structured and where texture is very often not available. Let us consider, for example, the case of an object moving over an homogeneous background. The motion is only effectively perceived on the object boundaries. The information provided by the boundaries must be propagated inside the object. Therefore, it is no more an information processing issue - as the regularization is - but a topological identification problem aiming at the determination of structural areas in the image.

The next step of our study will put the emphasis on how to take into account the structural content of an image in the context of motion estimation.

Acknowledgments

We gratefully acknowledge the kindness of the SAT company (*Société Anonyme de Télécommunications, 41 rue Cantagrel, F-75013 Paris Cedex*) for providing us with sequences of IR images, some of which are presented in this article.

References

- [1] J. Stuller, G. Krishnamurty, "Kalman Filter Formulation of Low-Level Television Image Motion Estimation", *Computer Vision, Graphics and Image Processing*, Vol. 21, No. 2, pp 169-204.
- [2] M. Basseville, "Detecting Changes in Signals and Systems - A Survey", *Automatica*, Vol. 24, No. 3, pp 309-326, 1988.
- [3] A. N. Netravali, J. D. Robins, "Motion-Compensated Television Coding: Part 1", *The Bell System Technical Journal*, March 1979.
- [4] H. H. Nagel, "On The Estimation of Optical Flow: Relations between Different Approaches and some New Results", *Artificial Intelligence*, 33, pp. 299-324, 1987.
- [5] E. C. Hildreth, "Computations Underlying the Measurement of Visual Motion", *Artificial Intelligence*, 23, pp. 309-354, 1984.