A NEW PEL-RECURSIVE KALMAN-BASED MOTION ESTIMATION METHOD

G. TZIRITAS

Laboratoire des Signaux et Systèmes (CNRS)
ESE, Plateau de Moulon, 91192 Gif-sur-Yvette, France

ABSTRACT

We present in this paper a new pel-recursive algorithm for estimating the displacement vector field in image sequences. Firstly we use a brightness-offset term in the motion constraint equation. We follow the Kalman approach to formulate the estimation problem. The state vector of dimension 3 is composed of the displacement vector and the brightness offset. The extended Kalman filter is used to estimate this state vector. The new algorithm is applied on two normalized TV sequences and it is shown that the new algorithm is always better than the commonly used algorithms. The mean square displaced frame difference obtained is about 20% less in comparison with the commonly used algorithms.

1. INTRODUCTION

Motion estimation is very important for image analysis and image coding. It is based on measures of spatiotemporal variations in image sequences. Generally it is assumed that the image brightness is only displaced, but not changed, with only an additive measurement noise. This is expressed in the so-called motion constraint equation. The image brightness $I(x,y)$ at point (x,y) at moment t is then given by

$$I(x,y;t) = I(x-u,y-v;t-\Delta t) + b(x,y;t) \tag{1}$$

where (u,v) is the displacement vector in the interval Δt, and $b(x,y;t)$ is an additive noise, which in the commonly used motion estimation algorithms is assumed to be a white noise in both spatial and temporal sense. This assumption may not be justified for many real image sequences.

In effect, we must distinguish between the optical flow and the 2D velocity vector field. The optical flow is defined as the 2D vector field which satisfies the motion constraint equation (1), where $b(x,y;t)$ is only a measurement noise. The 2D velocity (or displacement) vector field results from the projection of the real 3D motion vector field into the image plane. This distinction is clearly made by B.K.P.Horn [6] in a qualitative sense. A.Verri and T.Poggio [13] have given a quantitative expression of the difference between the optical flow and the 2D velocity vector field. In a simplified formulation they obtain the same equation (1), but the process $b(x,y;t)$ contains some terms which depend on the 3D velocity vector and the geometrical and photometric characteristics of the 3D objects projected into the image plane. The case where $b(x,y;t)$ may be considered as a white measurement noise, is when the 3D motion is translational and the 3D object surface is lambertian.

Changes in the brightness patterns are taken in consideration for the motion measurement problem by N.Cornelius and T.Kanade [3]. They use a regularization approach to solve the estimation problem, where brightness changes are allowed to vary smoothly. C.Moloney and E.Dubois [9] also use the same iterative approach, where illumination variation is also considered. C.S.Fuh and P.Maragos [5] give a method of image matching to estimate displacement, brightness change and illumination variation.

In this paper we propose a pel-recursive algorithm which takes into account the above considerations. This motion estimation algorithm is principally aimed at compensating for motion in image coding, but may be also useful for image analysis tasks. The new algorithm is based on Kalman formulation of the estimation problem.

2. THE MOTION ESTIMATION PROBLEM

We firstly rewrite the motion constraint equation (1) in a form more suitable for estimation purpose and we consider the discrete case

$$I(i,j;k)=I(i-v(i,j),j-u(i,j);k-1)+d(i,j)+n(i,j) \qquad (2)$$

where (i,j) are the row and column indices, k is the time index, $(v(i,j),u(i,j))$ is the corresponding velocity (or displacement) vector, $d(i,j)$ is the predictable part of the random process which expresses the dissension between 2D motion field and the optical flow, and $n(i,j)$ is an unpredictable (white) noise. We omit the dependence of $v(i,j)$, $u(i,j)$, $d(i,j)$ and $n(i,j)$ on time index because it is irrelevant for the estimation method used here. Equation (2) is the measurement equation we shall use in the following. What is to estimate is the velocity vector $(v(i,j),u(i,j))$ and the brightness offset $d(i,j)$. We then form a state vector as following

$$\boldsymbol{x}(i,j) = \begin{bmatrix} u(i,j) \\ v(i,j) \\ d(i,j) \end{bmatrix} \qquad (3)$$

The state vector equation is assumed to have recursion in only one direction according to the raster scan. Then we obtain

$$\boldsymbol{x}(i,j) = \boldsymbol{F}(i,j)\,\boldsymbol{x}(i,j-1) + \boldsymbol{q}(i,j) \qquad (4)$$

We assume the stationarity of $\boldsymbol{x}(i,j)$, in the sense that the transition matrix $\boldsymbol{F}(i,j)$ is space-invariant and the statistical characteristics of the noise vector $\boldsymbol{q}(i,j)$ are also space-invariant. More precisely we suppose that

$$\boldsymbol{F}(i,j) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \rho \end{bmatrix}, \text{ with } 0 \le \rho \le 1 \qquad (5)$$

$$E\{\boldsymbol{q}(i,j)\,\boldsymbol{q}^T(l,m)\} = \begin{bmatrix} \sigma_v^2 & 0 & 0 \\ 0 & \sigma_v^2 & 0 \\ 0 & 0 & \sigma_d^2 \end{bmatrix} \delta(i,l)\,\delta(j,m) = \boldsymbol{Q}\delta(i,l)\,\delta(j,m) \qquad (6)$$

where $\boldsymbol{q}(i,j)$ is zero-mean. J.Stuller and G.Krishnamurthy [11] also used such a model, with only two components, those of the displacement vector, and with a transition factor slightly less than 1. Concerning the measurement noise $n(i,j)$ we suppose that it is zero-mean and space-invariant as well. We have

$$E\{n(i,j)n(l,m)\} = R\delta(i,l)\,\delta(j,m) \tag{7}$$

Having such a formulation for the estimation problem we can use the extended Kalman filter to solve it. The model we use is essentially one-dimensional in space, because the transition in (4) concerns only the horizontal coordinate. A two-dimensional model is considered by J.N.Driessen and J.Biemond [4] for a state-vector composed of several motion vectors.

3. A PEL-RECURSIVE KALMAN-BASED ESTIMATION

We present a new pel-recursive algorithm according to theorem 8.1 of A.Jazwinski [7] on the extended Kalman filter. The *a priori* estimation of the state vector at point *(i,j)* is given by

$$\left.\begin{array}{l} \hat{\mathbf{x}}^{\mathbf{0}}(i,j)=\mathbf{F}\hat{\mathbf{x}}(i,j-1); j>0 \\ \hat{\mathbf{x}}^{\mathbf{0}}(i,0)=\mathbf{F}\hat{\mathbf{x}}(i-1,N-1); i>0 \\ \hat{\mathbf{x}}^{\mathbf{0}}(0,0)=\mathbf{0} \end{array}\right\} \tag{8}$$

where N is the number of pixels per line. The above recursion may be advantageous in the case of stationarity of the state vector. The case of non-stationarity is taken into account by the detector of discontinuities introduced below. The covariance matrix of this *a priori* estimation is given by

$$\left.\begin{array}{l} \mathbf{P}^{0}(i,j)=\mathbf{F}\mathbf{P}(i,j-1)\mathbf{F}^{\mathbf{T}}+\mathbf{Q}; j>0 \\ \mathbf{P}^{\mathbf{0}}(i,0)=\mathbf{F}\mathbf{P}(i-1,N-1)\mathbf{F}^{\mathbf{T}}+\mathbf{Q}; i>0 \\ \mathbf{P}^{\mathbf{0}}(0,0)=c\mathbf{Q} \end{array}\right\} \tag{9}$$

where $c>>1$.

Taking into account the image brightness at point *(i,j)*, the *a posteriori* estimation is obtained by

$$\hat{\mathbf{x}}(i,j)=\hat{\mathbf{x}}^{\mathbf{0}}(i,j)+\boldsymbol{K}(i,j)\big(I(i,j;k)-\hat{I}(i,j;k)\big) \tag{10}$$

with $\hat{I}(i,j;k)=I(i-\hat{v}^{0}(i,j),j-\hat{u}^{0}(i,j);k-1)-\hat{d}^{0}(i,j)$, where $\boldsymbol{K}(i,j)$ is the Kalman gain given by

$$\boldsymbol{K}(i,j) = \frac{\mathbf{P}^{0}(i,j)\mathbf{H}(i,j)}{\mathbf{H}^{\mathbf{T}}(i,j)\mathbf{P}^{\mathbf{0}}(i,j)\mathbf{H}(i,j)+R} \tag{11}$$

and where $\boldsymbol{H}(i,j)$ is the following vector

$$\boldsymbol{H}(i,j) = -\begin{bmatrix} \mathrm{I}_{\mathrm{x}}(i-\hat{v}^{0}(i,j),j-\hat{u}^{0}(i,j);k-1) \\ \mathrm{I}_{\mathrm{y}}(i-\hat{v}^{0}(i,j),j-\hat{u}^{0}(i,j);k-1) \\ -1 \end{bmatrix} \tag{12}$$

The covariance matrix of the *a posteriori* estimation is given by

$$\boldsymbol{P}(i,j) = \big(\boldsymbol{I}-\boldsymbol{K}(i,j)\boldsymbol{H}^{T}(i,j)\big)\boldsymbol{P}^{0}(i,j)\big(\boldsymbol{I}-\boldsymbol{K}(i,j)\boldsymbol{H}^{T}(i,j)\big)^{T}+ R\boldsymbol{K}(i,j)\boldsymbol{K}^{T}(i,j) \tag{13}$$

In this standard structure we associate a detector of discontinuities, like that presented in [12]. Without discontinuity, we suppose that the variable $I(i,j;k)-\hat{I}(i,j;k)$ is zero-mean and distributed according to a

Laplacian low. In the case of discontinuity, we make the same hypothesis concerning the frame difference. Under these assumptions, a discontinuity is detected using a Bayesian criterion, if

$$|I(i,j;k)-\hat{I}(i,j;k)|-|I(i,j;k)-I(i,j;k-1)|>s \tag{14}$$

4. RELATED METHODS

We present in this paragraph two known methods, which are related to that presented here, in order to make comparisons. For both methods no brightness offset is considered; this means that $d(i,j) = 0$. Therefore, only two parameters must be estimated, the components of the displacement vector. The dimension of the state-vector is equal to two.

The first method (named here Algorithm B) was introduced by J. Stuller and G. Krishnamurthy [11], and also studied by M. Mijiyawa [8]. This method can be obtained directly from the formulas of the precedent paragraphs with $\rho = 0$ and $\sigma_d^2 = 0$.

The second method (named here Algorithm C) is derived from Algorithm B after relaxation of the hypothesis that the covariance matrix of the reduced state-vector $\boldsymbol{q}(i,j)$ is space-invariant. It is assumed that

$$\boldsymbol{Q}(i,j) = \frac{\sigma_v^2}{\frac{R}{\sigma_v^2}+\mathbf{H}^T(i,j)\mathbf{H}(i,j)}\ \boldsymbol{H}(i,j)\boldsymbol{H}^T(i,j) \tag{15}$$

where $\boldsymbol{H}(i,j) = -\begin{bmatrix} I_x(i-\hat{v}^0(i,j), j-\hat{u}^0(i,j);k-1) \\ I_y(i-\hat{v}^0(i,j), j-\hat{u}^0(i,j);k-1) \end{bmatrix}$. J.C.Pesquet [10] has demonstrated that this algorithm is similar to that first introduced by C.Cafforio and F.Rocca [2], and also presented in a different way in [1,12]. Effectively equations (9)-(13) and (15) give

$$\hat{\mathbf{x}}(i,j)=\hat{\mathbf{x}}^0(i,j)+\frac{\mathbf{H}(i,j)}{\frac{R}{\sigma_v^2}+\mathbf{H}^T(\mathrm{i,j})\mathbf{H}(\mathrm{i,j})}\left(I(i,j;k)-\hat{I}(i,j;k)\right) \tag{16}$$

5. SIMULATIONS

In the following, results of applying the three algorithms on two TV sequences: "Car" and "Mobile and Calendar" are given. We use a reduced format with 264 lines of 674 pixels per line for each of the 60 fields. The first two fields are used as references, thus we give results over 58 fields.

In Fig.1 (resp.2) we give the *a priori* displaced frame difference for the "Car" (resp."Mobile and Calendar") sequence versus field number. The mean square Displaced Frame Difference is defined as following

$$e(i,j) = \begin{cases} \mathrm{I(i,j;k)-I(i-\hat{v}^0,j-\hat{u}^0;k-1)-\hat{d}^0} & \text{for no discontinuity pels} \\ \mathrm{I(i,j;k)-I(i,j;k-1)} & \text{for discontinuity pels} \end{cases} \tag{17}$$

$$\text{DFD} = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} e^2(i,j) \tag{18}$$

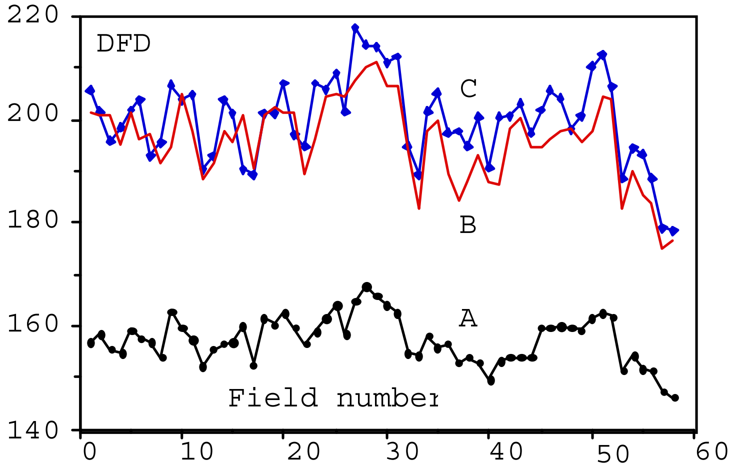

Fig.1. The displaced frame difference for the "Car" sequence using the three algorithms.

The gain obtained on mean square DFD is about 20% in comparison with algorithm C. Algorithm B gives for the "Car" (resp."Mobile and Calendar") sequence approximately the same results as the algorithm C (resp.A). This is due to the fact that for the "Mobile and Calendar" sequence the hypothesis of translational motion and Lambertian surfaces may be admitted. Thus brightness offset does not give any improvement. But due to the translational motion a gain is obtained in comparison with the algorithm C, which is always the worst among the three algorithms, simulated on the above sequences.

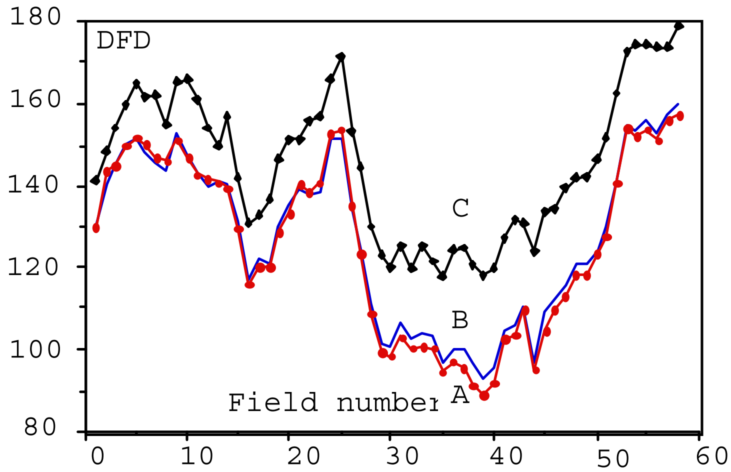

Fig.2. The displaced frame difference for the "Mobile and Calendar" sequence using the three algorithms.

The proposed motion estimation method is also used in a predictive coding structure. On the "Car" sequence we used a 5-level quantizer {0, ±19, ±38}, and we obtained a mean distortion over 10 fields of about 44, or 31.7 db, as Peak SNR, for an entropy of 1.59 bits/pel. On the "Mobile and Calendar" sequence we used a 7-level quantizer {0, ±19, ±38, ±57}, and we obtained a mean distortion over 56 fields of about 34, or 32.8 db, for an entropy of 1.1 bits/pel. Detected discontinuities must be transmitted; the percentage of those points is 0.53 (resp. 0.04) for the "Car" (resp. "Mobile and Calendar") sequence.

6. CONCLUSION

A new algorithm is presented, based on the extended Kalman filter, for displacement vector field estimation in image sequences. This algorithm is suitable for coding applications. It is simulated on two real image sequences and it is shown to give better results on mean square displaced frame difference, in comparison with other commonly used algorithms. Some preliminary results concerning the performance of this algorithm in a predictive coding scheme are given. A suitable encoding technique, with memory, might give a rate better than the one-dimensional entropy.

REFERENCES

[1] J.Biemond, L.Looijenga, D.E.Boekee “A pel-recursive Wiener-based displacement estimation algorithm”, *Signal Processing*, Dec. 1987, pp. 399-412.

[2] C.Cafforio and F.Rocca "The differential method for motion estimation" in *Image Sequence Processing and Dynamic Scene Analysis*, T.S.Huang, Ed., New York, Springer-Verlag, 1983, pp. 104-124.

[3] N.Cornelius and T.Kanade "Adapting optical flow to measure object motion in reflectance and X-ray image sequences" *ACM SIGGRAPH/SIGART Interdiscipl. Workshop on Motion*, April 1983, pp. 50-58.

[4] J.Driessen and J.Biemond "Motion field estimation by 2-D Kalman filtering" *Proc. of EU.SI.P.CO. 1990*, pp. 975-978.

[5] C.S.Fuh and P.Maragos "Affine models for image matching and motion detection" *Proc. IEEE Int. Conf. on ASSP*, 1991, pp. 2409-2412.

[6] B.K.P.Horn, *Robot Vision*, MIT Press, 1986.

[7] A. Jazwinski, *Stochastic processes and Filtering Theory*, Academic Press, New York, 1970.

[8] M.M.Mijiyawa "Utilisation de l'information du mouvement pour amélioration de la définition des images de télévision couleur" Thèse de Docteur-Ingénieur, Paris, 1987.

[9] C.Moloney and E.Dubois "Estimation of motion fields from image sequences with illumination variation" *Proc. IEEE Int. Conf. on ASSP*, 1991, pp. 2425-2428.

[10] J.C.Pesquet "Méthodes adaptatives en codage prédictif des images numériques" Th_se de l'Universit_ Paris-Sud, Sept.1990.

[11] J.Stuller and G.Krishnamurthy "Kalman filter formulation of low-level television image motion estimation" *Computer Vision, Graphics and Image Proccessing*, Vol.21, 1983, pp. 169-204.

[12] G.Tziritas "Recursive and/or iterative estimation of the two-dimensional velocity field and reconstruction of three-dimensional motion" *Signal Processing*, Vol.16, Jan. 1989, pp. 53-72.

[13] A.Verri and T.Poggio "Motion field and optical flow: qualitative properties" *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.PAMI-11, No.5, May 1989, pp.490-498.