DISPLACEMENT ESTIMATION FOR IMAGE PREDICTIVE CODING AND FRAME MOTION-ADAPTIVE INTERPOLATION

Georges TZIRITAS

Laboratoire des Signaux et Systèmes (C.N.R.S.), Ecole Supérieure d'Electricité, Plateau du Moulon
91190 GIF-sur-YVETTE, FRANCE

ABSTRACT

We present and investigate methods of displacement vector estimation, which can be used in predictive image coding or motion-adaptive frame interpolative coding. In both cases, a differential approach is adopted. This means that the displacement vector estimation is based on the measurement of the spatiotemporal gradient of the image sequence.

In the case of predictive coding the motion estimator is composed of three parts. The first is a spatial predictor of the displacement vector, which is based on a spatial autoregressive relation of the velocity field. The coefficients of this relation are made intensity-dependent. The presence of discontinuities inherent to the motion and to the instabilities of the estimation algorithm makes necessary a stage of detection of all type of discontinuities. Finally, an a posteriori estimator achieves the task of displacement vector estimation. This last stage is of iterative form.

The application of this algorithm in a very noisy image sequence has permitted to obtain a gain of about 40% in the absolute value of the difference with the predicted displacement vector and 65% with the estimated one after two iterations.

The same structure of displacement field estimation can be used to make a frame interpolation motion-adaptive. We present such an estimator, which is slightly different from that proposed in predictive coding, knowing that a non-causal a priori estimation can be realized. We also propose a completely different algorithm, which uses a hierarchical representation of the displacement field. A dynamic hierarchical 4-trees estimation algorithm is presented. The motivation for such a proposition comes from the fact that large areas in a great number of image sequences are no moving or have homogeneous two-dimensional motion, that is to say translational. Considering this redundancy in the displacement field a significant gain in complexity of calculations can be obtained, if the estimation is carried out rather over blocks than pel-by-pel. We have applied the last algorithm in the case of a sequence with moderate motion and some numerical results are given in this article.

1. INTRODUCTION

Motion estimation from a sequence of images arises in many application areas, principally in scene analysis and image coding. We present in this article an estimator, which takes into account the constraints of predictive coding. We consider that the prediction error after movement compensation is transmitted. We also present a temporal interpolator using an hierarchical motion estimator.

In both cases, that is in frame prediction and interpolation, a differential approach is adopted. The image intensity is considered as a continuous spatio-temporal function. The estimation of the displacement vector is based on the spatial and temporal gradient of the intensity. The equation used is that of motion constraint, which is obtained under hypothesis that the intensity variation is uniquely due to the motion. Hence using a first order Taylor development, one can write

$$\Delta I = I(x+\Delta x, y+\Delta y; t+\Delta t) - I(x, y; t) = (u\, I_x(x, y; t) + v\, I_y(x, y; t) + I_t(x, y; t))\, \Delta t = 0,$$

$I(x, y; t)$ being the intensity and u (resp. v) the horizontal (resp. vertical) component of the velocity vector. In real images this equation is not exactly verified and minimizing some quadratic error on this displaced frame difference and using some ideas of spatial coherence of the displacement vector carry out the estimation. A review of methods proposed in the literature to solve this problem, can be found in [4].

In the case of predictive coding the estimator is composed of three parts. The first part is an a priori estimator, which is a recursive predictor of the velocity vector, using the estimates of previous image points. The second part is a discontinuity detector. These discontinuities are due to either occluding edges of objects under different 3D motions, or edges of different surfaces of the same object, or algorithmic discontinuities. The detection criterion is based on the prediction error. The last part is an a posteriori estimator, which is an adaptive iterative filter. Figure 1 gives the estimator structure. The same form of structure is used in [1,3,6]. The displacement estimation in the context of predictive coding is the object of Section 2.

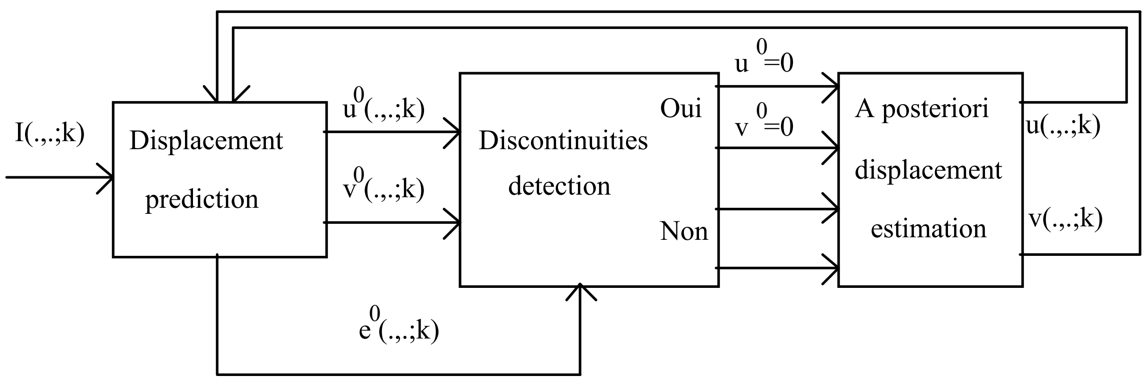

Fig.1. Structure of the displacement estimator

We discuss the frame motion-adaptive interpolation in Section 3. Two approaches are considered; the first one, is a direct application of the above structure in the context of interpolation, whereas the second one is a dynamic hierarchical estimator of the displacement field. We discuss here briefly the idea, which motivates this last approach. The computational complexity of pel-by-pel estimation is excessively large for image sequences presenting large areas of constant velocity vector, when block estimation could give the same accuracy.

Some applications of the studied methods in real image sequences are presented in the last Section.

2. MOVEMENT ESTIMATION FOR PREDICTIVE CODING

2.1. A priori estimation of the velocity vector

Let us consider the problem of estimating the velocity vector at point $A \equiv (i,j)$ (Fig. 2), where i is the horizontal coordinate and j the vertical coordinate. Let $[\,u \quad v\,]^T$ be the velocity vector at this point. We have to estimate this vector knowing the velocity at "precedent" points, that is in the domain

$$\{ (i-m,j) : m = 1,\dots,i \} \cup \{ (l,j-n) : n = 1,\dots,j \;;\; \forall\, l \}$$

We are limited to a causal neighborhood, which contains three points

$$\{ B \equiv (i-1, j),\ C \equiv (i, j-1),\ D \equiv (i-1, j-1) \}$$

given in Figure 2.

Let $[u_B \;\; v_B]^T = [u(i-1,j) \quad v(i-1,j)]^T$ be the velocity vector of point B, and in the same way for C and D. The estimation is made by minimizing the following quadratic form

$$Q_0(u,v) = [u-u_B \quad u-u_C \quad u-u_D]K^{-1}[u-u_B \quad u-u_C \quad u-u_D]^T + [v-v_B \quad v-v_C \quad v-v_D]K^{-1}[v-v_B \quad v-v_C \quad v-v_D]^T$$

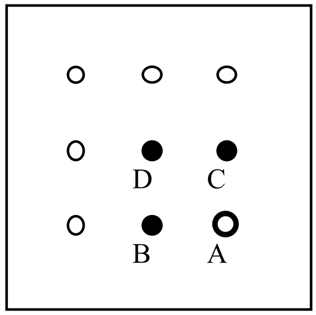

Figure 2. The causal domain of displacement vector prediction

We can write $Q_0(u,v) = (u\mathbf{1}-\mathbf{a})^T K^{-1}(u\mathbf{1}-\mathbf{a}) + (v\mathbf{1}-\mathbf{b})^T K^{-1}(v\mathbf{1}-\mathbf{b})$,
where $\mathbf{1} = [\,1 \quad 1 \quad 1\,]^T$, $\mathbf{a} = [\,u_B \quad u_C \quad u_D\,]^T$ and $\mathbf{b} = [\,v_B \quad v_C \quad v_D\,]^T$. Matrix K may be considered as the covariance of the vector $u\mathbf{1}-\mathbf{a}$. This matrix is spatially variant and unknown. We have therefore to make a heuristic choice.

We give the solution of the minimization of the form $Q_0(u,v)$, before exposing our choice. We obtain the following solution

$$u^0 = \frac{\mathbf{1}^T K^{-1}\mathbf{a}}{\mathbf{1}^T K^{-1}\mathbf{1}} \text{ and } v^0 = \frac{\mathbf{1}^T K^{-1}\mathbf{b}}{\mathbf{1}^T K^{-1}\mathbf{1}}.$$

Matrix K depends on the intensity, and particularly on the spatial gradient of the intensity, which measures the spatial activity of the image. We attribute to the covariance a value proportional to the inner product of the intensity gradients in the respective directions and to the product of the distances of the corresponding points. This conjecture conducts to the following matrix

$$K = \begin{bmatrix} \mu + I_x^2 & 0 & I_x^2 \\ 0 & \mu + I_y^2 & I_y^2 \\ I_x^2 & I_y^2 & \mu + I_x^2 + I_y^2 \end{bmatrix}$$

I_x (resp. I_y) being the horizontal gradient (resp. vertical). Scanning the frame I(.,.;k) of order k in the sequence, one disposes the spatial gradient at the precedent frame I(.,.;k-1). A possible choice to calculate the gradient in the matrix K is the gradient at point (i-1-u(i,j) , j-v(i,j) ; k-1), which is the precedent point in the scanning displaced in the previous frame. After inversing matrix K, and after some approximations, we obtain the following predictor

$$u^0 = f_x u_B + f_y u_C - f_x f_y u_D$$

and in the same way for v^0, where the coefficients f_x and f_y depend on the gradients I_x, I_y and on μ ($0 < f_x , f_y < 1$) in the following way

$$f_x = \frac{\mu + I_y^2}{\mu + I_x^2 + I_y^2} \quad \text{and} \quad f_y = \frac{\mu + I_x^2}{\mu + I_x^2 + I_y^2}$$

If $I_x^2 + I_y^2$ is small in comparison with μ, then $f_x = f_y = 1$, and the prediction relation becomes

$$u^0 = u_B + u_C - u_D$$

This relation can be connected to the spatial relation on the velocity field under a hypothesis of orthographic projection and locally planar surface of the 3D object [5]. In fact, under this hypothesis the velocity components are expressed locally as linear functions of the coordinates and therefore the above spatial relation is obtained.

2.2. Detection of discontinuities in the motion estimation

In natural images exist independent 3D rigid movements, so the global motion of the scene is no rigid. There also exist edges between different surfaces, which are subject to the same 3D rigid motion, but they have different characteristics, and therefore a discontinuity results in the 2D velocity field. The estimation method being based on the intensity, the algorithm may causes false displacements, so a new initialization is necessary.

An efficient estimation of the velocity vector necessitates the conjoint detection of all these discontinuities. This detection have to be based on the correctness of the a priori estimation and consequently on the prediction error of the intensity. It must be based on the a priori estimation, because this expresses the spatial coherence of the velocity vector.

Two approaches are possible in the context of the transmission of image sequences. Before presenting them, let us define the prediction error

$$e^0(i,j) = I(i,j;k) - I(i-u^0, j-v^0;k-1)$$

1. The prediction error is compared with the difference between the two frames. If a discontinuity is detected, the difference between the two frames is transmitted in the place of the prediction error, as well as the information that a discontinuity is detected. The a priori estimation is then reset to zero ($u^0 = v^0 = 0$).

2. A causal neighborhood is considered and the predicted displacement is applied. This operation gives the following prediction error

$$e^0(i,j) = \sum_{(m,n)\in D} |I(i-n,j-m;k) - I(i-n-u^0, j-m-v^0;k-1)|$$

(for example, D = {(0,1),(1,0)}). This prediction error is compared with the difference between the two frames in the same neighborhood. If a discontinuity is detected, the a priori estimation is reset to zero, as in the first approach. But in this case the transmission of the detection information is not necessary, because the decoder can also carry out the detection.

2.3. A posteriori estimation of the velocity vector

The a priori estimation initializes the estimation algorithm of the velocity vector. The a posteriori estimation utilizes an algorithm of iterative form.

The estimation error is expressed by

$$e(i,j) = I(i,j;k) - I(i-u, j-v;k-1)$$

Let $[\, u^{n-1} \quad v^{n-1} \,]^T$ the estimation of the velocity vector at n-1 iteration. The optimization is carried out by minimizing a quadratic criterion

$$Q(u,v) = e^2 + \lambda[(u-u^{n-1})^2 + (v-v^{n-1})^2]$$

The second part of this form can be considered as a regularization or stabilization constraints. The form to minimize is not quadratic in report to the unknown variables. To obtain a quadratic form, we have to linearize the estimation error. The Taylor development of order 1 of $I(i,j;k) - I(i-u, j-v;k-1)$ gives

$$I(i-u, j-v;k-1) = I(i-u^{n-1}, j-v^{n-1};k-1) - I_x(i-u^{n-1}, j-v^{n-1};k-1)\,(u-u^{n-1}) - I_y(i-u^{n-1}, j-v^{n-1};k-1)\,(v-v^{n-1})\,.$$

Thus we obtain

$$e(i,j) = e^{n-1}(i,j) + I_x(i-u^{n-1}, j-v^{n-1};k-1)\,(u-u^{n-1}) + I_y(i-u^{n-1}, j-v^{n-1};k-1)\,(v-v^{n-1})$$

Using this approximation, we obtain the following solution

$$u^n(i,j) = u^{n-1}(i,j) - \frac{I_x(i-u^{n-1}, j-v^{n-1};k-1)e^{n-1}(i,j)}{\lambda + I_x^2(i-u^{n-1}, j-v^{n-1};k-1) + I_y^2(i-u^{n-1}, j-v^{n-1};k-1)}$$

and in the same way for v. The number of iterations is either fixed in advance, or with a stopping test on estimation error.

3. MOTION-ADAPTIVE IMAGE INTERPOLATION

A technique to reduce the rate of transmission is to skip frames at the transmitter and interpolate the skipped frames at the receiver. To avoid either jerkiness or blurring of the interpolated images, we have to adapt the interpolator to the motion. We shall study here the case of interpolating one frame between every two frames, that is to say interpolating the frame of order k between the frames of order k-1 and k+1. If $[\hat{u} \quad \hat{v}]^T$ is the velocity vector at the point (i , j), then this point is interpolated by

$$\hat{I}(i,j;k) = 0.5*(I(i+\hat{u}, j+\hat{v};k+1)+I(i-\hat{u}, j-\hat{v};k-1))$$

The displacement vector is estimated using these same frames of order k-1 and k+1. The estimation is based on the error

$$\begin{aligned} e(i,j) &= I(i+u,j+v;k+1) - I(i-u, j-v;k-1) \\ &= (I_x(i, j;k+1)+I_x(i, j;k-1))\,u + (I_y(i, j;k+1)+I_y(i, j;k-1))\,v + I(i,j;k+1) - I(i, j;k-1) \end{aligned}$$

The estimator may have the same structure, as in the case of predictive coding. We develop it in the Section 3.1. In the Section 3.2 we present another approach, which is motivated by the fact that large areas in many image sequences are not changed for a long time. A multiresolution hierarchical estimator is then proposed.

3.1. Iterative estimation

The structure of the estimator of Figure 1 is also used here. The difference is that the a priori estimator may be non-causal. For the first component of the velocity vector we propose the following a priori estimator

$$u_\alpha^{n-1}(i,j) = \alpha u_\alpha^{n-1}(i,j) + 0.25(1-\alpha)(u_\alpha^{n-1}(i-1,j) + u_\alpha^{n-1}(i+1,j) + u_\alpha^{n-1}(i,j-1) + u_\alpha^{n-1}(i,j+1))$$

The a posteriori estimation is given by

$$u^n(i,j) = u_\alpha^{n-1}(i,j) - \frac{(I_x(i,j;k+1) + I_x(i,j;k-1))e_\alpha^{n-1}(i,j)}{\lambda + (I_x(i,j;k+1) + I_x(i,j;k-1))^2 + (I_y(i,j;k+1) + I_y(i,j;k-1))^2}$$

In the same way we obtain the other velocity component.

3.2. Hierarchical estimation

For no-moving areas or for areas homogeneous in the sense of velocity vector, one can significantly reduce the amount of calculations to estimate the motion by dividing the frame into blocks and by realizing the estimation in an hierarchical way. Thus we introduce a dynamic multiresolution on the velocity field using 4-trees. If, for example, the first resolution grid is composed from blocks of 16*16 points, a block, which is not homogeneous in the sense of the motion, is divided into four blocks of 8*8 points. We need then a criterion of homogeneity. The criterion we use verifies if the motion of a block can be considered as a two-dimensional translation. This test is realized on the estimation error.

Considering that the movement of a block is translational, we can use a linear regression estimator for the displacement vector components [2]

$$\hat{u} = \frac{\sum I_y^2 \sum I_x I_t - \sum I_x I_y \sum I_y I_t}{\sum I_x^2 \sum I_y^2 - (\sum I_x I_y)^2}$$

$$\hat{u} = \frac{\sum I_x^2 \sum I_y I_t - \sum I_x I_y \sum I_x I_t}{\sum I_x^2 \sum I_y^2 - (\sum I_x I_y)^2}$$

where $I_x = I_x(i,j;k+1)+I_x(i,j;k-1)$, $I_y = I_y(i,j;k+1)+I_y(i,j;k-1)$ and $I_t = I(i,j;k+1)-I(i,j;k-1)$, the summation carried out over a block.

Having this estimation we test the homogeneity hypothesis. The criterion used here is based on the power of the intensity estimation error. The cost function is defined as follows

$$Q(\hat{u},\hat{v}) = \sum \;(I(i+\hat{u},j+\hat{v};k+1) - I(i-\hat{u},j-\hat{v};k-1))^2)$$

The decision concerning homogeneity is obtained by comparing the above quantity to a threshold. If the block is classified non-homogeneous, it is divided into four blocks and the estimation-detection procedure is recommenced.

4. APPLICATIONS

We have applied the algorithms presented in this article, in two sequences of real images. The first one, named "CAR", represents a car in movement, when the camera pans the scene. A strong additional noise disturbs the intensity and different types of spatial details are present in the scene. The second one, named "GIRL", represents a girl speaking and moderately moving. Both sequences contain only the luminance, coded in 8 bits.

The horizontal gradient is obtained by a filter with a finite impulse response given in the following matrix

$$F_x = \frac{1}{80}\begin{bmatrix} -3 & -5 & 0 & 5 & 3 \\ -5 & -8 & 0 & 8 & 5 \\ -3 & -5 & 0 & 3 & 3 \end{bmatrix}$$

The transposed matrix gives the vertical gradient. This filter is composed of a low-pass filter to smooth the derivative, and of a bilateral derivation.

We have applied the algorithm of Section 2 in the case of the "CAR" sequence, with $\mu = 30$ and $\lambda = 200$. These values have to be compared to the spatial activity of the image, which can be measured by the mean of the square of the spatial gradient, which is about 150. For the eleven first frames of the sequence the mean of the absolute value of the difference between two successive frames were 18.618. The mean of the absolute value of the displaced frame difference using the predicted displacement vector was 11.281; using the a posteriori estimated displacement vector after two iterations it was 6.860. The percentage of detected discontinuity points was 2.739 %.

The algorithm of Section 3.2 is applied in the case of "GIRL" sequence. For a mean of the absolute value of the difference between two images equal to 3.253, the mean of the absolute value of the displaced frame difference was 1.820 and the mean of the absolute value of the interpolation error was 1.395. Three levels of resolution were used, the initial consisting of 16*16 blocks.

5. REFERENCES

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

2. C.Cafforio, F.Rocca "Methods for measuring small displacements of television images", *IEEE Trans. on Information Theory*, Vol. IT-22, No. 5, pp. 573-579, Sept. 1976.

3. R.Moorhead, S.Rajala "Motion-compensated interframe coding", *Proc. ICASSP*, pp. 347-350, 1985.

4. H.G.Musmann, P.Pirsch, H.-J.Grallert "Advances in Picture Coding", *Proc. of the IEEE*, Vol. 73, No. 4, pp. 523-548, Apr. 1985.

5. G.Tziritas "Estimation of motion and structure of 3D objects from a sequence of images",*1st IEEE Conf. on Computer Vision*, pp. 693-697, London, June 1987.

6. D.Walker, K.Rao "Improved pel-recursive motion compensation", *IEEE Trans. on Communications*, Vol. COM-32, No. 10, pp. 1128-1134, Oct. 1984.