MOTION COMPENSATION IN BLOCK DCT CODING BASED ON PERSPECTIVE WARPING

L. Capodiferro, S. Puledda*, G. Jacovitti***

* Fondazione Ugo Bordoni c/o ISPT, Viale Europa 190, 00149 Rome, Italy
Tel: +39-6-54802132; Fax: +39-6-54804401; email: licia@fub.it
** INFOCOM Dpt., University of Rome “La Sapienza”, via Eudossiana 18, 00184 Rome, Italy
Tel: +39-6-44585838; Fax: +39-6-4873300; email: gjacov@infocom.ing.uniroma1.it

ABSTRACT

In this paper we present a technique for bit-rate reduction in the H263 coder by the introduction of perspective transformation in the advanced prediction option of motion compensation. It is based on the use of the available displacement vectors for estimating the image warping. Since block matching gives not reliable estimates of the warping an adaptive technique discarding inconsistent transformations has been adopted.

1. INTRODUCTION

Motion compensation is an important function of most video coding schemes because it allows to take into account the high degree of correlation usually existing between consecutive frames. Conventional motion compensation methods are based on block matching (BM) techniques. BM techniques consist of searching in the past frame the block most similar (in a mean square or absolute error sense) to the current one in the present frame. The relative position of the two blocks (displacement vector) and their difference (residual) are separately encoded for representing the current block. BM is an effective technique in medium-high bit rate coding environment, where the differences among consecutive frames are relatively small. In low bit rate applications, such as ISDN videoconference or videotelephony, the BM becomes less efficient because of the interframe correlation loss and of the simultaneous need of limiting the residual information rate.

For this reason, some attempts are being made in order to improve the interframe prediction with more elaborate techniques. In the H263 standard environment for low bit rate coding of CIF and quarter CIF formats an "Advanced Motion Compensation" (AMC) option is defined [1]. The AMC decomposes the current block of the present frame in four subblocks, and applies a separate BM to each subblock. Basically, the prediction of a given subblock from the previous frame is made by considering the displacement vector of the current block containing the subblock and the two displacement vectors of the blocks contiguous to the subblock. The prediction is obtained with a linear combination of the pixels pointed by the three displacement vectors. The weights are proportional to the distance of the pixel from the center of the current block. Doing so, the residuals amount decreases on the average. This approach can be justified by considering an affine transformation defined by the three displacement vectors [2], even if the AMC is not a geometric transformation (warping) because it performs linear combinations of pixel values rather than positions.

Many recent attempts to achieve higher compression rates have been focused on the use of more complex motion compensation techniques. Instead of searching for pure translational BM, these techniques aim to motocompensate simple geometric transformations such as rotations, scaling, etc.. Geometric transformations are employed in morphing techniques, and constitute the basic operators for iterative contractive coders [3]. Recently, the use of blocks shearing has been proposed in [4].

In this contribution we introduce perspective transformations into the H263 scheme, based on the use of the available displacement vectors for estimating the image warping. Since BM gives not reliable estimates of the warping due to ambiguities in presence of uniform or linear patterns, an adaptive technique discarding inconsistent transformations has been adopted.

In section 2 the geometric transformations for digital image warping applications are introduced. Then the perspective prediction algorithm for motion compensation is described in section 3, and some experimental results in terms of bit rate and SNR are reported in section 4.

2. SPATIAL TRANSFORMATIONS

A spatial transformation is a mapping function that establishes a geometric relationship between all points in an image and its warped counterpart [5].
The mapping can be expressed, using homogeneous coordinate system as:

$$[x,y,w]=[u,v,w']\cdot\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix} \qquad (1)$$

where [u,v,w’] refers to the input image, and [x,y,w] to the warped image.
Normalizing the matrix coefficients so that $a_{33}=1$, the remaining eight degrees of freedom allow to express the affine and perspective classes of transformations respectively related to the generic mapping of triangles and quadrilaterals.
In particular, in this work we are interested in mapping a generic quadrilateral into another. The eight independent coefficients of the matrix may be derived by establishing a correspondence between four points in the reference and in the observed images and by solving the resulting system of equations.

3. PERSPECTIVE MOTION COMPENSATION ALGORITHM

The perspective motion compensation algorithm, proposed in this work, is directly introduced in the Advanced Motion Compensation option of the H263 coder.
It is based on the search, for any generic current block, of a quadrilateral source in the previous frame, that takes into account the possible geometric transformations occurred for that block.
For this purpose we define a “warping area” which includes the current block. On the basis of the motion vectors relative to the points that delimit this area, we define a perspective warping (PW) of the current block into a resulting quadrilateral pattern in the previous frame. By the correspondence between the said points and their counterparts we evaluate the coefficients of the transformation matrix characterizing this warping. Each pixel of the current block is predicted from the corresponding point in the previous frame after bilinear interpolation from the nearest pixels.
The residuals of the prediction are calculated and compared to the standard BM residuals. If the energy of the perspective residuals does not exceed the energy of the translational BM residuals, then the block is labeled as "Perspective Predictable" (PP).
The coding is performed adopting the conventional AMC interpolation for non-PP blocks, and the PW correspondence for PP blocks. The PP-ness of the blocks is signaled separately with a run length codification of the lexicographically ordered sequence of the blocks.
At the decoder side, the reconstruction is performed according to the PP-ness nature of the block.
No extra motion vectors calculation is required for this algorithm. This allows for simple implementation and gives the possibility of mixing the perspective and the advanced prediction technique adaptively.

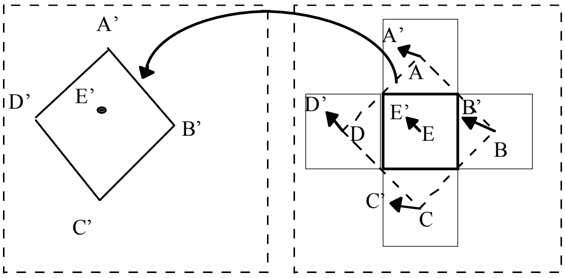

Figure 1. First and second strategies for warping area choice.

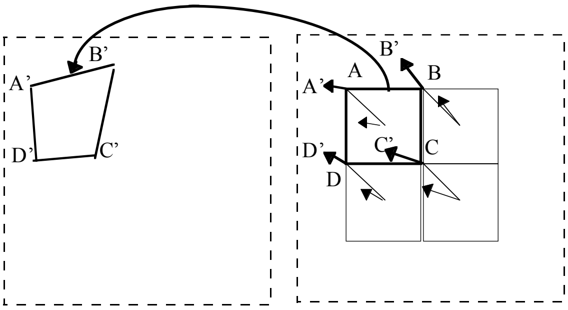

Figure 2. Third strategy for warping area choice.

We examined three different strategies for choosing the warping area for each block of the current frame.
In the first strategy (PW1), we considered a squared area (including the current block) delimited by the four points placed in the middle of the adjacent blocks in the two principal directions (fig.1). The correspondence between these points and their

counterparts in the previous frame (located by the motion vectors) defines the transformation matrix and so the perspective warping. This strategy doesn't take into account the motion information of the current block itself.

In the second strategy (PW2), we considered the point placed in the middle of the coding block too. Now, the correspondence between the five pairs of points do not uniquely identify the matrix transformation, which is found by solving an overdetermined system under a minimum square criterion.

In the third strategy (PW3), we restricted the warping area to the current block itself associating to its four vertices the motion vectors of the adjacent botton-rigth blocks (fig.2). This method results more accurate though there is an offset in motion vectors assignment.

4. EXPERIMENTAL RESULTS

Some experiments were carried out using typical videothelephone QCIF (176x144 pels) test sequences: "Carphone", "Foreman", "Mother and Daughter" and "Salesman".

In fig. 3a) a sample frame, from carphone sequence, is displayed. The reconstructed frame obtained via H263 algorithm, and via our third strategy are respectively reported in fig. 3b) and 3c). There is no visually appreciable difference between the two reconstructed images, and the SNR is almost the same (37.83 dB in H263 - 37.85 in our method).

On the other hand we obtained here a prediction gain of about 6.5% (9767 bits in H263 A.M.C.- 9176 in our method). However this gain does not take into account the side information necessary for the individuation, at the received side, of about 40 pp-blocks.

A detail of the frame n.81 is reported in fig 4a). The predicted detail obtained via the H263 Advanced Motion Compensation, and our algorithm are respectively reported in figs 4b) and 4c). The corresponding residuals are reported in figs 4d) and 4e).

Global results are reported in tables 1 and 2. In table 1 the SNR averaged over 150 frames of the four sequences is shown together with the overall bit rate associated to the prediction methods.

Figure 3a. the frame n. 81 of carphone sequence

Figure 3b. The H263 AMC reconstruction of the frame n. 81

Figure 3c. The H263 PW3 reconstruction of the frame n. 81

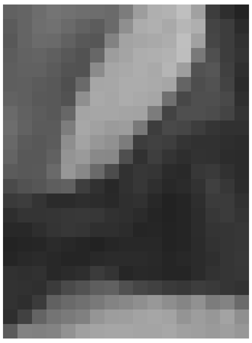

Figure 4a. A detail of the frame n. 81 of carphone sequence

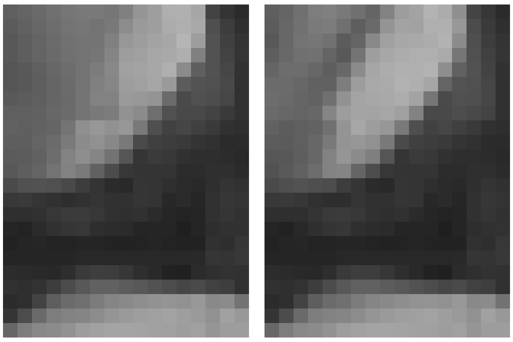

Figure 4b,c. Reconstructed details: b) AMC prediction, c) PW prediction.

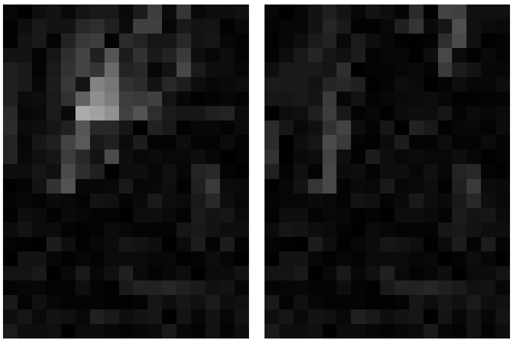

Figure 4d,e. Prediction residuals: d) AMC prediction, e) PW prediction.

In table 2 the net bit-rates obtained by including the side information necessary to communicate the PP-ness of the blocks is reported

From these results we observe that the third strategy shows the best performance. In practice, the technique gives no saving at very low bitrates (in the range below 32 kbit/s) and a gain of about 1 kbit/s in the range of 55-60 kbit/s with respect to the AMC prediction technique, without SNR loss.

	H263 AMC	H263 PW1	H263 PW2	H263 PW3
Foreman	34.30 61.07	34.33 59.63	34.30 59.64	34.34 58.99
Carphone	37.71 68.2	37.73 66.06	37.71 66.47	37.75 66.01
Mother & Daughter	37.18 60.67	37.19 58.44	37.18 59.00	37.21 58.17
Salesman	37.80 57.17	37.81 55.85	37.81 55.86	37.81 55.96

Table 1 SNR and prediction bit-rates comparison for the four sequences.

	H263 AMC	H263 PW1	H263 PW2	H263 PW3
Foreman	61.07	60.97	60.65	60.41
Carphone	68.2	67.22	67.31	67.20
M & D	60.67	59.60	59.76	59.21
Salesman	57.17	56.52	56.42	56.56

Table 2. Net overall bit-rates comparison.

6. REFERENCES

[1] International Telecommunication Union "Draft Recommendation H.263", 25 May 1995.

[2] A. Neri, S. Colonnese, "Affine Warping Using H263 Overlapped Motion Compensation With Proper Weighting Coefficients", IS&T/SPIE'S, Conference on Digital Video Compression, Algorithms and Technologies, S. Jose (CA), 28 Jan.- 2 Feb. 1996.

[3] A.E. Jacquin, "Fractal image coding: a review", Proc IEEE, vol.81, n.10, pp.1451-1465, Oct. 1993.

[4] S.S. Skrzypkowiak, V.K. Jain, "Affine Motion Estimation Using a Neural Network", Proc. ICIP-95 IEEE Int. Conf. on Image Processing, Oct. 23-26 1995 Washington DC, Vol I pp. 418 - 421.

[5] G. Wolberg, “Digital Image Warping”, Los Alamitos, CA, IEEE Computer Society Press, 1990