JB 91-02

Paper Entitled:

"Predictive Block-Matching Motion Estimation
Schemes for Video Compression --
Part II: Inter-frame Prediction of Motion Vectors

(with Sohail Zafar and Ya-Qin Zhang)

Southeastcon '91 Conference
pg 1093-1095
Williamsburg, Virgina
April 7-10, 1991

Predictive Block-Matching Motion Estimation Schemes for Video Compression -- Part II
Inter-frame Prediction of Motion Vectors

Ya-Qin Zhang* Sohail Zafar John S. Baras

* Contel Technology Center
15000 Conference Center Drive
Chantilly, VA 22021

Dept. of Electrical Engineering
University of Maryland
College Park, MD 20742

ABSTRACT

A new predictive block-matching motion estimation scheme is proposed in this paper for efficient video codec design. The proposed scheme is based on the "inertia" effect of natural video scenes and takes advantage of the motion vectors obtained in the previous frames. Not only motion searching time and overhead are significantly reduced, the proposed scheme also provides realistic motion tendency, which is very useful for frame skipping/interpolation techniques. These advantages are also demonstrated by some experimental results obtained on Abekas video system.

1. INTRODUCTION

Motion compensation is a common technique used for low-bit rate video compression such as the proposed CCITT px64 coding standard for videophone and video-teleconferencing applications [1]. Recently, it has been applied to compress full-motion video as well. Motion-compensated coding systems for TV/HDTV at different quality are under development. Various coding schemes for full-motion video with motion compensation at 1 Mbps with VCR quality is also currently under study by the joint CCITT/ISO MPEG committee [2]. Motion compensation not only significantly improves the compression efficiency compared to a simple DPCM, it also provides the possibility for frame skipping/interpolation without sacrificing the picture quality if the motion speed and path can be correctly estimated.

Various motion estimation/compensation schemes have been developed and implemented for interframe video compression. In the accompanying paper [3], an inter-block predictive pattern searching scheme has been proposed. It is shown that the proposed scheme significantly increases the searching efficiency as compared to full-motion search. In this paper, the predictive pattern searching scheme is extended to inter-frame estimation, i.e. the motion information of a block is estimated from that of the corresponding block in the previous frames.

It is well known from the classic mechanics that an object, once in motion, will persist in that motion due to the inertia property. By the same token, we expect that a moving object in a video sequence (say TV sequence) will have certain "inertia". Indeed, it is very unlikely that an object only moves in a few consecutive frames and then stops unless there is a scene change. Therefore, we can estimate the motion vectors of consecutive frames based on the motion vector information obtained from the past frames. Our experimental results in the following section also support this assumption that consecutive video frames have very similar motion activities including the speed and tendency unless there is a scene change. This is the principle of the proposed inter-frame predictive motion estimation scheme. Similar techniques have been proposed for pel-based motion estimation [4][5][6]. To the best of our knowledge, this is the first time that a predictive concept is used for block-based motion estimation.

The most important variable in a block-based searching/matching scheme is the search area Ω which governs the accuracy of the match and the computational burden. Most of the simplified search approaches, such as the 2-D logarithmic search, the 3-step search, and the conjugate direction search, assume that correlation function has a single peak and decreases monotonically as the search deviates from the peak in a search region [6]. Experiments show that the correlation function depends strongly upon the contents of scene, and the search algorithm sometimes slides to a wrong direction or is trapped by a false peak. These situations often arise when the motion region contains highly detailed and complicated images such as narrow, strip-shaped or net-shaped figures. Hence, the best match for motion vectors can only be obtained from full search. In the proposed predictive searching scheme, the "inertia" property of the motion is utilized to obtain more realistic motion vectors. This property is very useful in frame skipping/interpolation applications.

2. INTER-FRAME PREDICTIVE MOTION COMPENSATION

In the block-based interframe predictive motion estimation scheme, a video frame is divided into many blocks. Motion vectors of the initial frame are calculated either by full-search or any suitable technique. A buffer is used to store the vertical and horizontal components of the motion vectors $\mathbf{V}_i(m, n) = (\mathbf{V}_{x,i}(m, n,), \mathbf{V}_{y,i}(m, n))$. The motion vectors for a number of consecutive frames are then estimated from that of the corresponding blocks and a small neighborhood of the block in the previous frames. Thus the motion vector $\mathbf{V}_{i+1}(m, n)$ for block $B_{i+1}(m, n)$ in frame i+1 is predicted based on motion vectors $\mathbf{V}_i(m'', n'')$ where m'' and n'' $\in \Delta$. Δ is a neighborhood around block $B_i(m, n)$ in the frame i. Thus the motion vectors for $B_{i+1}(m, n)$ is given by,

$$\mathbf{V}_{i+1}(m, n) = E[\mathbf{V}_{i+1}(m, n)] + d\mathbf{V}_{i+1}(m, n)$$

where $|d\mathbf{V}_{i+1}(m, n)| \le |\Delta|$, the differential motion vector calculated with a reduced search area Δ. The initial estimate $E[\mathbf{V}_{i+1}(m, n)]$ can be calculated using an autoregressive (AR) prediction model given by,

$$E\{ \mathbf{V}_{i+1}(m,n) \} = \sum_{p,q \in \Theta} \alpha_{p,q} \mathbf{V}_i (m - p, n - q)$$

Where $\{\alpha_{p,q}\}$ is a set of prediction coefficients and Θ can be a causal, semicausal or noncausal set defined by:

$$\Theta = \{ p, q: q > 0 \ \forall \ p \} \cup \{ p, q: \ q=0, p > 0 \} : \text{Causal Set}$$
$$\Theta = \{ p, q: q > 0 \ \forall \ p \} \cup \{ p, q: \ q=0, \forall \ p \neq 0 \} : \text{SemiCausal Set}$$
$$\Theta = \{ p, q: \ \forall \ (p,q) \in \Delta \} : \text{Noncausal Set}$$

The estimation process is refreshed periodically by a full search strategy since a moving object only retains its speed and direction for a certain period. The benefits from predictive estimation are three-fold. First, the searching area is greatly reduced (e.g. a quarter of the searching area for refreshed frames), and so is the computational complexity. Secondly, the motion vector overhead information is reduced since motion vectors are decorrelated by the prediction process. Finally, the motion vectors estimated from this procedure are more realistic since it reflects the real physical phenomena. It is understood that most motion estimation schemes attempt to minimize the output data rate, which doesn't necessarily give the realistic motion vectors because of the complexity and randomness of the real-world imagery.

Figure 1 shows the inter-frame predictive motion compensation scheme. In the figure frame i is the refreshed frame whose motion vectors are calculated by full-search with searching area Ω. The motion vectors for frame i+1 are then found using the proposed predictive scheme using a reduced search area Δ for the differential motion vectors. A very simple model for estimation was used in our simulations,

$$E[\mathbf{V}_{i+1}(m, n)] = \mathbf{V}_i(m, n)$$

i.e., the initial estimate for a block was chosen to be the motion vector for the corresponding block in the previous frame.

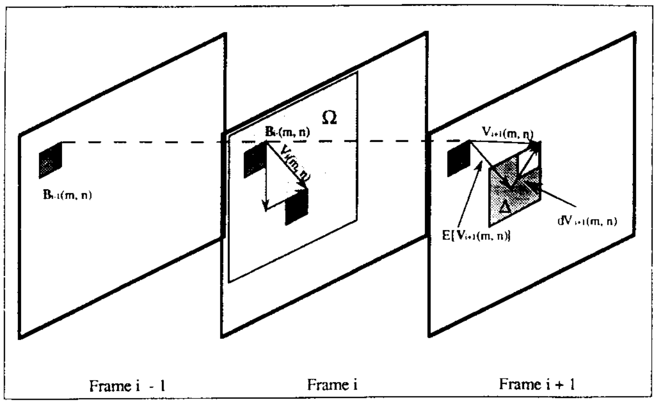

Figure 1: Inter-frame predictive block pattern searching scheme

Combined inter-block/inter-frame predictive searching can also be used to further reduce the searching complexity. In this case, the motion vectors of the refreshed frame are estimated by inter-block prediction scheme proposed in [3] and the motion vectors of subsequent frames are estimated by inter-frame prediction.

3. Inter-Frame Predictive Motion Compensated Compression System

The inter-frame motion prediction scheme is tested for a "CAR" sequence. The video compression system we use for testing the proposed motion estimation scheme is shown in Figure 2. It is basically an inter-frame hybrid DPCM/DCT compression scheme. Due to its robustness and good performance, this scheme has been used in the CCITT H.261 codec for video telephony [1], in the CCITT MPEG video compressor for full-motion video [2], and in the General Instrument's Digicipher video codec for HDTV compression simulation [7]. The motion compensator is defined to be block-based, the specific motion searching scheme and error criterion are left open to implementors in the proposed draft standards. In most simulation studies, a full motion searching scheme is employed. In this paper, the proposed inter-frame predictive searching scheme is implemented and compared with the "optimal" full search scheme.

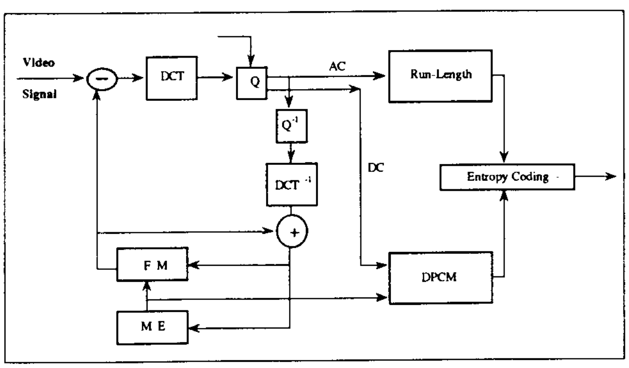

Figure 2. Block Diagram of Inter-frame motion-compensated hybrid DCT/DPCM video compression scheme

After motion compensation, the displaced difference frame is block DCT transformed and uniformly quantized. The quantization process takes advantage of human visual characteristics and coefficients are weighted prior to the uniform quantization. Then, the AC coefficients are zig-zag scanned and run-length coded. The DC coefficients of adjacent blocks are further DPCM coded to reduce the inter-block redundancy. Motion vectors are noiselessly coded and transmitted. All quantities are entropy-coded prior to transmission.

The "CAR" is a full-motion color video sequence in CCIR 601 format with 720 by 480 per frame, 16 bits per pixel in YUV format (or 24 bits in RGB format) and 60 fields interlaced per second. It is basically a fast camera panning sequence which is ideal for testing different motion compensation algorithms. Motion estimation is only implemented for the Y component, assuming that the U and V components have the same motion activity.

4. Experimental Results

Motion activities in both vertical and horizontal directions for the sequence "CAR", which are measured by the full-motion searching scheme and the minimum absolute difference matching criterion, are shown from Figure 3(a) and 3(b) respectively. A block size of 8 by 8 and searching area of 16 by 16 are used for the measurement.

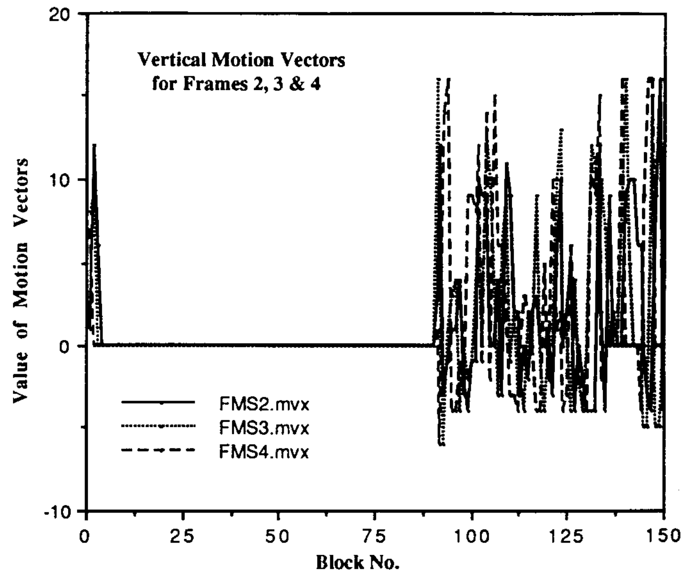

Figure 3(a): Motion Activities in the vertical direction for frame 2, 3 & 4 measured by full search.

It can be easily observed from Figure 3(a) & 3(b) that there is very slight motion in the vertical direction and very heavy motion is associated with the horizontal direction. This is consistent with the actual video observation. The measurement results also indicate that consecutive video frames tend to have similar motion activities. In another word, the motion activity in one frame can be well predicted from the previous frame, which is the motivation for our inter-frame predictive motion estimation scheme.

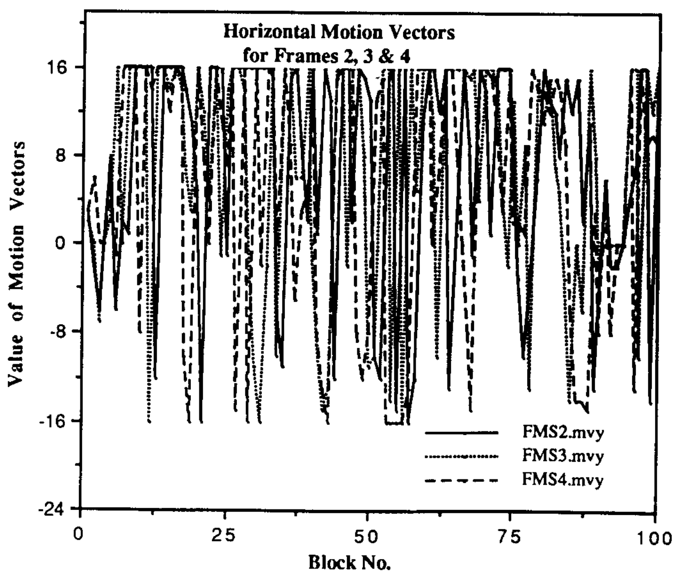

Figure 3(b): Motion Activities in the horizontal direction for video frame 2, 3 & 4 measured by full search.

As mentioned earlier a very simple model for prediction was used where the prediction for motion vector of a block is just the motion vector of the corresponding block in the previous frame. Figure 4 shows the first-order entropy of the displaced frame difference (DFD) obtained with full motion search and predictive search and is compared with the entropy of the simple frame difference with no motion compensation.

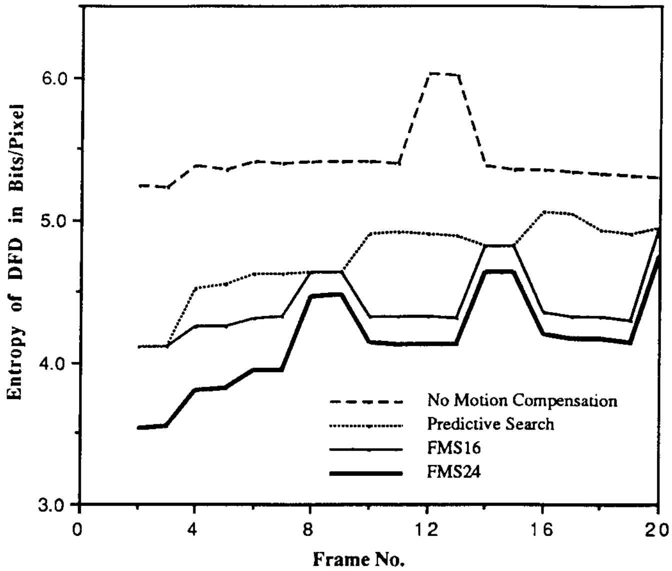

Figure 4: First-order Entropy of Differential Frame Difference (DFD) for no motion compensation, full search with Ω of 24x24 (FMS24), 16x16 (FMS16) and Predictive search Ω is 16x16 with Δ 4x4.

Although the entropy of the DFD image is higher in case of predictive search, the entropy of the motion vectors themselves is greatly reduced as shown in Figure 5. The effect of periodic refresh and interlacing is also visible in the figure. It can be seen that the entropy of the motion vectors in predictive search is greatly reduced for both horizontal and vertical directions.

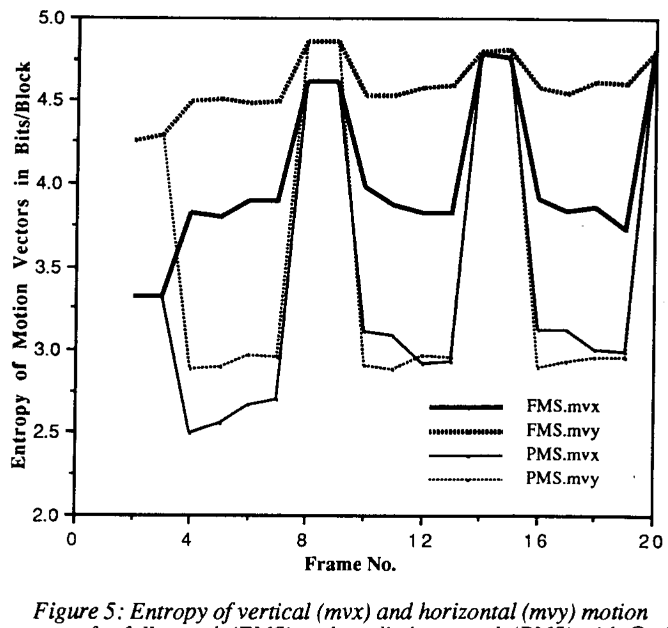

Figure 5: Entropy of vertical (mvx) and horizontal (mvy) motion vectors for full search (FMS) and predictive search (PMS) with Ω of 16x16.

There is no doubt that full search is optimal, but we can see from Figure 6 that the signal-to-noise ratio for the reconstructed video signals in case of predictive search follows closely the full search performance. The performance of the system can be greatly enhanced using a more complex prediction model.

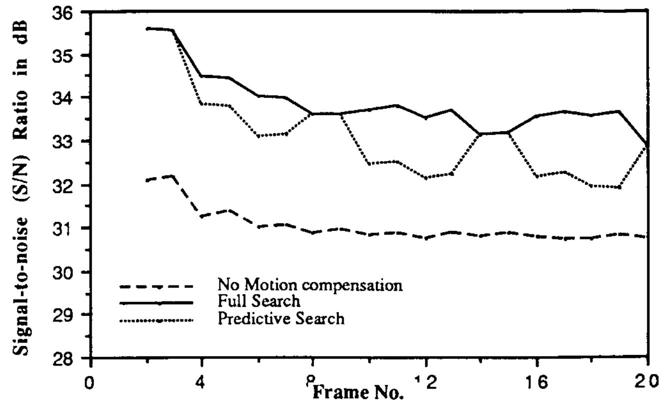

Figure 6: Signal-to-noise ratio of the reconstructed image sequence using no motion compensation, full search and interframe predictive search for motion compensation.

5. Conclusions

A new predictive block-matching motion estimation scheme was implemented in this paper for efficient video codec design. The proposed scheme is based on the "inertia" effect of natural video scenes and takes advantage of the motion vectors obtained in the previous frames. The benefits from this prediction process are three-fold. First, the searching area is greatly reduced, and so is the computational complexity. Secondly, the motion vector overhead information is reduced since motion vectors are decorrelated by the prediction process. Finally, the motion vectors estimated from this procedure are more realistic since it reflects the real physical phenomena, which is a very useful property in frame skipping/interpolation techniques. These advantages were also demonstrated by simulation results including the coded data rate, DFD entropy, motion vectors and reconstructed signal-to-noise ratio. It should be pointed out that only a simple prediction model is implemented in this paper, further results in more general AR models are currently under study.

References

[1] *Draft Revision of Recommendation H.261*, Document 572, CCITT SG XV, working party XV/1, Spec. Grp. on Coding for Visual Telephony.

[2] *MPEG Video Simulation Model Three*, International Organization for Standardization Coded Representation of Picture and Audio Information, 1990

[3] S.Zafar, Y.Zhang and J.Baras, "Predictive Block-Matching Motion Estimation Schemes for Video Compression --- Part I: Inter-block Prediction," *Proceedings of IEEE Southeastcon'91*

[4] R.Moorhead, S.Rajala and L.Cook, "Image Sequence Compression Using a Pel-Recursive Motion-Compensated Technique," *IEEE J. selected Areas in Commu.*, Vol.5, No.7, August 1987

[5] K.Xie, L.Eycken, A.Oosterilinck, "Some New Techniques in Displacement Estimation Algorithms," *SPIE Visual Communications and Image Processing*, Vol.1199, pp. 18-29, 1989

[6] H.Musmann, P.Pirsch and H.Grallert, "Advances in picture coding," *Proc. of IEEE*, Vol.73, No.4, pp.523-548, April 1985

[7] *Digiciper™ for HDTV System*, Submitted by General Instrument Corp., Videocipher division, San Diego, CA, June 8, 1990