Telecommunication Standardization Sector
Study Group 15
Experts Group for ATM Video Coding
(Rapporteur's Group on Part of Q.2/15)

Document AVC-529
Version 1
July, 1993

INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO-IEC/JTC1/SC29/WG11
CODED REPRESENTATION OF PICTURE AND AUDIO INFORMATION

ISO-IEC/JTC1/SC29/WG11
MPEG 93/561
New York, July 1993

Title: **On the B-frame design complexity of spatio-temporal weighting**

Source: **D. Schinkel, A. Koster, A. Hekstra - PTT Research**

Purpose: **Information**

Introduction

In this contribution, we recapitulate the following well known video coding techniques,

- B-frames as a way to have two prediction blocks for each block in an image,
- compatible video coding based on reconstructed signal prediction.

Reconstructed signal prediction corresponds to switch position B of the harmonized hybrid scalable solution (see TM5). In this paper, we consider spatio-temporal weighting (ST weighting) as a combination of the two aforementioned techniques, i.e., application of the bi-prediction machinery required for B-frames to compatible coding.

Applications that do not in any mode support B-frames, may not a priori have the memory and computational resources available required for simultaneous prediction out of two pictures. Therefore, in judging the additional complexity due to ST weighting we distinguish between M=3 and M=1 applications.

The bi-prediction machinery required for B-frames

The essence of bidirectional prediction, is that a block in a picture can have a prediction from blocks in two other pictures. Within the MPEG syntax, some restrictions apply to the order of different types of pictures. In an implementation, this means that in both the encoder and the decoder,

- two reference picture memories must be available,
- the memory bandwith must allow for two prediction pixels to be retrieved per coded pixel.

At this point, the order of I-, P- and B- pictures is of little concern for the required hardware cost of the bi-directional prediction engine. In the encoder, on a block by block basis a decision which of four coding modes gives the best performance must be made:

- intra coded block,
- prediction out of reference picture 1,
- prediction out of reference picture 2,
- average of both prediction blocks.

This decision can e.g. be based on an error energy criterion and refinements are possible with respect to the bit cost of motion vectors associated with the use of prediction blocks. Similarly, in the decoder, depending on the macro block mode, a maximum of two prediction pixels must be retrieved per output pixel.

Compatible coding based on reconstruced signal prediction

The reconstructed signal prediction method is well-known, and has been extensively investigated. The essence is, that the current side information picture forms an additional reference frame from which prediction is possible on a block by block basis. These predictions can serve as alternatives to motion compensated predictions, and can be used for I-pictures. This method has been incorporated into the harmonized hybrid scalable solution of Test Model 5 as switch position B. Option b) of harmonized hybrid scalability is widely advocated, in particular because it allows for a somewhat higher SNR than option a). The gain amounts to a few tenths of a dB. The largest reduction in bitrate due to the use of side information is obtained for I-pictures and scene cuts.

Use of bi-prediction machinery for compatible coding

Spatio-temporal weighting was proposed by PTT Research as a way to improve the coding performance of reconstructed signal prediction. The mechanisms available in the encoder and decoder to form a prediction out of two reference pictures is combined with the idea of reconstructed signal prediction, i.e., the side information is an additional reference picture. In previous documents, PTT Research has reported SNR gains of several tens of a dB for the ST weighting technique for M=3. It is emphasized that new applications and new methods for video coding or format conversion can be envisaged that substantially increase the predictive value of the side information. A higher SNR coding gain results from,

- all applications that use M=1, as for P-frames the predictive value of side information is higher as for B-frames,
- nonspectral upconversion methods e.g. based on neural nets,
- compatible stereoscopic coding (side information image has full resolution).

Conclusions

In any case, the use of ST-weighting remains an option. If this coding mode is not selected there is *no* memory size, memory bandwidth, processing power or bit cost associated with it.

ST weighting gives a gain a performance which adds to the SNR performance of the coding schemes. This is especially true for M=1, and the gain may increase as future applications or upconversion methods may increase the predictive value of the side information.

For M=3 applications, when one chooses to restrict the values of the weights to (0, 1/2, 1) and ST weighting is used only for I- and P-pictures, no additional memory size, memory bandwith or processing power results from the use of ST weighting.

Only for M=1, a higher memory bandwidth is required just as if B-frames would have to be supported.