EFFICIENT VIDEO CODING INTEGRATING MPEG-2 AND PICTURE-RATE CONVERSION

F.J. de Bruijn, W.H.A. Bruls, D. Burazerovic, and G. de Haan
Philips Research Laboratories, Eindhoven, The Netherlands

ABSTRACT
We present an MPEG-2 compatible video codec that uses picture-rate up-conversion during decoding. The up-conversion autonomously regenerates major parts of frames without vectorial or residual data. Consequently, the bit rate is greatly reduced.

INTRODUCTION
The MPEG-2 video compression standard allows forward predictive and bidirectional predictive encoding, resulting in the generation of P- and B-frames respectively. Motion-compensated predictive coding exploits the temporal correlation between consecutive frames. In practice, the efficiency gain over intra encoding is often not more than a factor of 4 compared to the bit rate of an intra (I) frame in the same group of pictures (GOP).

Intuitively, we consider this factor of 4 to be somewhat disappointing given the visual similarity between consecutive frames and the quality of another motion-compensated picture interpolation application known as “Natural Motion” (NM) [1]. The NM-algorithm, developed by Philips for its high-end 100 Hz televisions, removes motion judder from film-originated video material. To this end, it generates additional intermediate pictures between the ones registered on the film, instead of simply repeating earlier ones. This interpolation process shows a clear similarity with the generation of B-frames in MPEG, but NM does not require the transmission of vectorial and residual data.

The autonomous operation of the NM-process makes it an interesting addition to a video-compression system. We have designed an application in which a NM-based algorithm is integrated with an MPEG-2 scheme. The NM-process is set up to generate alternative ‘B-frames’ based on an input of MPEG I- and P-frames, both during encoding and decoding. In the encoder, each NM output frame is compared with an original B-frame. A criterion, specifically designed for this task, decides whether it is necessary to locally fall back on the original B-frame content in order to prevent visible errors. In this case, the vectorial and residual data of the original B-frame data is preserved in the MPEG-stream. The addition of a proprietary extension to a standard MPEG-2 chain may affect compatibility, so the combination may particularly be useful for consumer applications that require local video storage, e.g. on hard disk.

NATURAL MOTION VS. MPEG
The limited efficiency gain of predictive coding over intra coding has two main causes: the motion-estimation process and the criterion to evaluate each locally predicted picture.

In most MPEG-2 implementations, motion compensation is based on a computationally efficient derivation of full-search block matching (FSBM). The motion vectors resulting from FSBM minimize the blockwise difference between two blocks. The blockwise difference is often calculated as the sum of absolute differences (SAD) or as the mean squared error (MSE). In either case, the difference criterion does minimize the local residual data amount, but does not result in a true-motion estimate. Consequently, MPEG motion vectors may not necessarily describe the true object motion, and tend to be temporally and spatially inconsistent. In practice, transmission of residual (difference) data appears to be vital for artifact-free reconstruction.

In contrast with MPEG, the temporally interpolated frames produced by NM have to result in high-quality motion-compensated predictions without the addition of residual information. To this end, the interpolated frames generated by the NM-process are based on estimates of the true motion, which are generated using a three-dimensional recursive search (3DRS) block matcher instead of a full-search block matcher [1, 2]. The motion-vectors estimated using 3DRS exhibit a high degree of spatial and temporal consistency.1

COMBINED SOLUTION
An obvious solution to improve on standard MPEG is to skip frames during encoding, followed by NM after decoding. In particular, skipping B-frames which are not re-used in the prediction process avoids error accumulation. Unfortunately, when NM is applied to interpolate frames over large temporal distances, which is the case when several B-frames are predicted from (locally) decoded I- or P-frames, visible errors may occur since 3DRS fails to track small objects correctly.

To reliably regenerate ‘B-frames’ with NM during decoding, the occurrence of visible errors must be detected in the encoder. In practice, however, a simple pixelwise comparison with MPEG B-frames causes an abundance of high SAD-values in almost any detailed area, even in the absence of visual errors. However, the application of 3DRS to frame rate conversion shows that small deviations from the true motion are not perceived in consistently moving areas.

1. Previous work has already shown that application of 3DRS enables a computationally more efficient MPEG-2 encoder implementation with better compression results compared to FSBM[3, 4, 5].
To quantify vector inconsistency (VI) we determine the maximum absolute componentwise difference of a vector with surrounding motion vectors,

$$\text{VI}(x, y) = \max_{(\xi, \eta) \in \Omega} \left| \tilde{D}(x, y) - \tilde{D}(x + \xi, y + \eta) \right|, \quad (1)$$

where \(\tilde{D}\) is a 2-dimensional motion vector that describes the displacement between two consecutive frames, and \(\Omega\) is the spatial area of evaluation. In areas where the inconsistency is low, we choose to disregard large local SAD-errors, and in inconsistent areas we choose to consider even small local SAD-errors as visible. Each SAD-value is, therefore, thresholded with a value \(T_{\text{SAD}}(x, y)\) which varies as a function of the local VI-value,

$$T_{\text{SAD}}(x, y) = \begin{cases} T_{\text{SAD, high}}, & \text{VI}(x, y) < T_{\text{VI}} \\ T_{\text{SAD, low}}, & \text{otherwise}, \end{cases} \quad (2)$$

with \(T_{\text{VI}}\) some threshold. The value of \(T_{\text{SAD, high}}\) is such that seriously large match errors are not neglected, even if the vectors are consistent. The value of \(T_{\text{SAD, low}}\) is such that the noise is ignored. The decision to locally fall back on the original MPEG B-frame content is taken in case the local SAD-error exceeds the locally varying threshold,

$$S_{\text{fallback}}(x, y) = \begin{cases} 1, & \text{SAD}(x, y) > T_{\text{SAD}}(x, y) \\ 0, & \text{otherwise}. \end{cases} \quad (3)$$

If \(S_{\text{fallback}}(x, y) = 1\) the original B-frame data is inserted in the MPEG-stream. The binary map of blockwise decisions \(S_{\text{fallback}}(x, y)\) can be sent as private data. Figure 1 depicts an MPEG-2 encoder combined with Natural Motion in which the error criterion is incorporated. Options to establish the threshold values will be discussed in our full paper.

### RESULTS

Experiments conducted on various video sequences show that in many cases less than 10% of the original B-frame DCT-blocks are required as fall-back data. In an MPEG-2 GOP structure of \(N = 24\) and \(M = 4\), we measured a decrease of the bitrate up to 30%, even in the case of complex or fast movements. A more elaborate evaluation will be given in our full paper.

### CONCLUSIONS

A new bidirectional predictive video coding method has been introduced as an improvement on the MPEG-2 method, and achieves a significant bitrate reduction at comparable visual quality. A new error criterion has been proposed to enable the application of true-motion vectors in motion-compensated predictive coding.

### REFERENCES


