Motion Interpolation Prediction Through Motion Field Analysis

R. Leonardi†and H. Bheda‡
AT&T Bell Labs
†Visual Communications Reserch Dept., rm. 4C-516
‡VLSI Systems Research Dept., rm. 4E-532
Holmdel, NJ 07733

Recently, digital storage of video for CD-ROM applications has lead to new insights for using long-term correlation in image sequences. High quality coding is obtained through a sophisticated compression scheme allowing to reduce the original bandwidth of 168Mbit/s to a compressed bit-stream of 1.1Mbits/s. Such a performance is essentially achieved on the basis of the so-called conditional motion-compensated interpolation technique (CMCI) [1]. This result emerged from the Motion Picture Expert Group (MPEG) work, appointed by the International Organization for Standardization (ISO). CMCI suggests to code distant frames through motion-compensated prediction, whereas neighboring frames are motion-compensated interpolated. Block based motion estimates are used to compensate previous frames resulting in errors that are then quantized for coding together with the aformentioned motion estimates. Interpolation errors are quantized more coarsely than prediction errors as they do not feed back in the temporal DPCM loop that ties predicted frames. Moreover, temporal masking effects limit the visual impairments that appear on poor reconstruction of interpolated frames.

For fast random access requirements, the MPEG coding strategy partitions the video signals into groups of frames (GOFs), containing one intraframe, and several predicted frames separated by interpolated frames. The intraframe does not involve any temporal prediction, allowing this way to rapidly decode the corresponding bit-stream for reconstruction.

In the MPEG simulation model [2], three different types of prediction have been considered for motion compensated interpolation: "forward", "backward" and "interpolative" motion prediction. In the first case, only previous frames are compensated using a block-based motion estimate between the current interpolated frame and a previous predicted frame or intraframe. The second case is identical to the first one, except that it uses the future frame as a reference. Finally, "interpolative" motion prediction uses both a past and future reference in the compensation process. The interpolative compensation involves the averaging of both frames proportionally to their distance to the current frame. for notation purposes, we consider a set of M interpolated frames preceeded by a reference frame N. Explicitely, the interpolation error associated with a block $B_{N+i}(x,y)$ at location (x,y) in frame $N+i$ is obtained by:

$$\begin{aligned} E_{N+i}(x,y) \; = \; & B_{N+i}(x,y) - 1/(M+1) \times [i \times B_N(x + d^{(x)}_{N,N+i}, y + d^{(y)}_{N,N+i}) + \\ & (M+1-i) \times B_{N+M+1}(x + d^{(x)}_{N+i,N+M+1}, y + d^{(y)}_{N+i,N+M+1})] \qquad (1) \end{aligned}$$

where $[d^{(x)}_{a,b} d^{(y)}_{a,b}]$ represents the motion estimate in both directions between frames a and b.

Apart from the situations in which a non-translational motion or deformable body motion are present, this overall prediction scheme works relatively well. "Forward" motion prediction is entitled to be used in areas of the frame for which there exists no matching in the previous frame, i.e. areas that were covered at time N. Respectively, "backward" motion prediction will take care of areas that have no reference in the future, i.e. areas that will be covered at time $N + M + 1$. Finally, "interpolative" motion prediction should work optimally when there is a smooth motion of the object between the two reference frames with no deformation and covered/uncovered objects. The averaging process used in the computation of the interpolation error (see (1)) allows to compensate for the effect of noise present in a single frame.

We disregard the effect of scene cuts in what follows, and assume that each GOF contains only one frame. In this context, we study how to make use of the distant motion field (DMF) structure computed between predicted frames to minimize the information to be sent for interpolated frames. For every interpolated frame, a complete correspondence exists in the previous and/or future reference frame. This correspondence can be estimated from the structure of the DMF. Whenever the DMF is smooth, it is likely that the motion will remain smooth along the temporal dimension as well. This way, if a motion estimate exists between two predicted frame, it can be used to estimate very accurately displacements for interpolated frames. This allows now to neglect the corresponding interpolation errors, resulting in a cost reduction of the overhead and error coding for interpolated frames. On the other hand, a discontinuity in the DMF suggests the presence of different moving objects, corresponding to covered/ uncovered areas in the interpolated frames. Unless a clear segmentation of each object can be obtained in the DMF, it is impossible to estimate the displacement associated to interpolated frames. We defer this issue to a further study that would make use of motion segmentation prediction [3]. In this case, we prefer to send an "interpolative", "forward" or "backward" motion estimate together with an interpolation error.

In our experiments, we found that for complicated scenes such as the ones used for MPEG material, at least 75% of the motion field could be considered smooth. This results in a significant reduction of the coding cost for interpolation frames. Together with this result, we present ways to estimate the smoothness of the motion field and how to get a good motion estimate for interpolation of smooth areas.

References

- [1] A. Puri, R. Aravind, B.G. Haskell and R. Leonardi, "Video Coding with Motion-Compensated Interpolation for CD-ROM Applications", *Signal Processing: Image Communication*, Vol. 2, No. 2, August 1990.
- [2] "MPEG Video Simulation Model Three (SM3)", Doc. ISO-IEC/JTC1/SC2/WG8 N/ MPEG90.
- [3] M. Orchard and R. Leonardi, "Motion Segmentation Prediction for Hybrid Video Coding", *Proceedings of the 1990 PCS*, Cambridge, pp. 1.3.1-1.3.3., March 1990.