Proceeding of the 1990 Picture Coding Symposium, Cambridge, 26-28 marzo, 1990.

Motion Segmentation Prediction for Hybrid Video Coding [1]

Michael Orchard† and Riccardo Leonardi‡
†Department of Electrical Engineering
Princeton University, Princeton, NJ 08544
‡AT&T Bell Labs, Visual Communications Research Dept.
Holmdel, NJ 07733

Hybrid predictive/transform techniques are used by most image sequence coding schemes currently under consideration (e.g. [1]). These techniques predict pixel values in the current frame with those from the previous frame and encode the difference between predicted and actual values (prediction error) using intraframe transform coding techniques. The prediction is usually performed by partitioning the current frame into fixed size blocks and estimating a vector identifying a block in the previous frame which provides the best prediction for each block in the current frame. The decoder uses two sources of information to reconstruct each frame. First, the motion vectors are used to construct the prediction of the current frame from previously decoded frames. Then, the prediction error is decoded and added to produce the final decoded frame.

Although hybrid techniques efficiently code blocks of pixels which move with uniform translational motion, they are not efficient at coding blocks which overlap objects moving in different directions because these blocks are not well predicted by any block from the previous frame. Since these blocks are found along discontinuities in the motion field, hybrid coders operating at low bit-rates often give the appearance of blockiness at the edges of moving objects. Puri et al. [2] have proposed adapting the block size of the motion estimates to provide a higher resolution motion field in areas where the larger blocks do not move with uniform motion (transition blocks). Allowing 8x8 blocks to be adapted to 4x4 blocks, this method provided a reduction in the coding cost of the difference signal, tempered by the cost of additional motion vectors.

This paper describes a method for improving the prediction of blocks at edges of moving objects by segmenting the motion field of previous frames and using this segmentation to predict motion field discontinuities in the current frame. As in standard hybrid methods, the decoder receives motion vectors which represent a coarse (block-level) representation of the motion field of each frame. However, after receiving the prediction error, the decoder uses the final reconstructed frame to refine its representation of the motion field to the pixel level. This results in a segmentation of the motion field of the decoded frame. This segmentation is then used to predict motion field discontinuities of new frames and to apply different motion estimates to pixels within a block which are moving in different directions. As a result, blocks on edges of moving object edges are predicted accurately and the prediction error which results is reduced significantly. Ideally, only uncovered areas, or blocks for which the assumption of translational motion (e.g., deformation of object boundary) is violated would result in large prediction errors.

[1] This research was partially supported by an AT&T fellowship. Parts of this work was performed while at AT&T Bell Laboratories, Holmdel, NJ

Let I_n denote frame n in a sequence of images. Let $\hat{v}_n$ denote the motion estimates relating frame I_n to $\hat{I}_{n-1}$, the decoded version of I_{n-1}. Assume that I_n is the next frame to be coded and that both encoder and decoder have access to the last two decoded frames, $\hat{I}_{n-1}$ and $\hat{I}_{n-2}$, along with the motion estimates $\hat{v}_{n-1}$. The outline of the proposed method is:

(a) The encoder and decoder compute a segmentation of the motion field based on $\hat{I}_{n-1}$, $\hat{I}_{n-2}$, and $\hat{v}_{n-1}$. Let $\hat{S}_{n-1}$ denote the computed segmentation. $\hat{S}_{n-1}$ assigns to each pixel in $\hat{I}_{n-1}$ a flag indicating which of 5 candidate motion vectors it matches best. The candidate vectors include the vector from the block to which the pixel belongs and vectors from the 4 nearest neighbor blocks to this block.

(b) The encoder computes motion estimates $\hat{v}_n$ based on I_n, $\hat{I}_{n-1}$, and $\hat{S}_{n-1}$. The encoder uses an iterative procedure to compute each block motion estimate to minimize the energy of the difference signal taken over the largest set of pixels in the block which $\hat{S}_{n-1}$ predicts to be moving with uniform motion.

(c) The decoder uses the motion estimates from (b) and $\hat{S}_{n-1}$ to predict the largest set of pixels in each block which moves with uniform motion. These pixels are motion compensated using the motion vectors received. $\hat{S}_{n-1}$ is used to associate the remaining pixels in each block with an object in some neighboring block and those pixels are motion compensated using the motion vector of that block. The resulting frame serves as the motion compensated prediction of I_n. The encoder mirrors the computations of the decoder to have access to the identical motion compensated prediction.

(d) The encoder forms the difference between the predicted and actual values of I_n and performs transform coding of the difference signal. The decoder updates the predicted values with the coded difference signal, forming $\hat{I}_n$.

Refinement of the motion field is performed within each 8x8 block of the image by testing hypothesis motion vectors applied to each pixel of the block. The hypothesis motion vectors are selected from among those vectors assigned to the present block and its four nearest neighbors. For each hypothesis vector, a small patch of the image centered at the pixel is displaced and a displaced difference is calculated as a weighted sum of the pixels differences within the patch. The pixel is assigned the vector which minimizes the displaced difference. Smoothing of the segmented field is applied to remove isolated segments. Segmentation of frames from the standard test sequence "table-tennis" show that very good sub-block segmentations are produced at the forward edge of moving objects (covered areas). However, uncovered background areas have no match in the previous frame and give a noisy segmentation of the trailing edge of moving objects. This problem can be resolved by segmenting in a reversed time direction when uncovered areas are detected, making the trailing edge become the forward edge of the object and vice-versa.

A number of block motion estimation algorithms have been tested for use with the proposed coding method. These include fast search block-matching algorithms and a gradient based block motion estimation technique [3]. All motion estimation algorithms use 8x8 pixel blocks. The segmentation has been found to be sensitive to the consistency of the motion field estimate. Inconsistent motion fields result in detection of spurious motion field discontinuites during segmentation. However, these discontinuities usually land on block boundaries or produce noisy segmentations in areas where spatial gradients are small.

Techniques have been applied which locate and remove these types of edges by applying smoothness constraints to the motion field.

Tests are being conducted on the proposed method to measure the achievable coding gains and to refine the techniques used for segmentation and motion estimation. The image sequences "salesman" and "table-tennis" have been digitized at a resolution of 288x352 pixels per frame, updated at a 30 Hz rate. Comparisons between the proposed method and standard hybrid based methods applied to these sequences have shown that the proposed method produces significant reductions in prediction error at edges of moving objects.

References

- [1] CCITT SGXV WP XV/4 Specialists Group on Coding for Visual Telephony Doc. #525, "Description of Ref. Model 8," June 9, 1989.
- [2] A. Puri, H.M. Hang and D.L. Schilling, "Interframe Coding with Variable Block-size Motion-compensation," *Proc. IEEE Global Commun. Conf.*, Tokyo, Japan, Nov. 1987, pp. 2.7.1-2.7.5.
- [3] A.N. Netravali and B.G. Haskell *Digital Pictures Representation and Compression* 1988, Plenum Publ. Co., pp. 334-350.