VIDEO CODING USING HYBRID MOTION COMPENSATION

C. Morimoto P. Burlina and R. Chellappa**

Department of Electrical Engineering* and
Center for Automation Research,
University of Maryland,
College Park, MD 20742.
email: {carlos,burlina,rama}@cfar.umd.edu

ABSTRACT

We propose a novel video coding scheme to improve the performance of established block-based motion compensation codecs such as MPEG, H261, and H263. The proposed method is a hybrid scheme which introduces model-based global motion compensation as a pre-processing step to the basic block-based motion compensation technique. Performance evaluation tests show that the new method is capable of achieving higher compression rates with the addition of a very small overhead due to global motion estimation. In this paper we describe a codec based on the hybrid motion compensation technique in the context of H263, and present experimental results, comparing the performance of different codecs using rate-distortion curves for several test sequences.

Keywords: Local and global motion compensation, model and block-based low bit rate video coding.

1. INTRODUCTION

Most current source coding techniques exploit spatio-temporal redundancy present in video sequences by using motion compensated frames (MCF), where blocks of a MCF are modeled as translated blocks of some previous (or future) reference frame, plus a residual image [12]. Therefore, instead of coding each image independently, motion vectors must be coded along with the residual images. The extra bits necessary to code the translation vectors can be significant but do not offset the improvement obtained from the high compression rates which result from coding the residual images instead of the original ones.

Block-based motion compensation (BMC) schemes are simple to implement and perform well under a variety of situations, e.g., when the scene contains several independent moving objects, although, due to the limited translation-only motion model, their performance drops when significant rotation and/or change in scale are present. The use of global motion compensation (GMC) has been proposed in [5, 15, 16].

The advantages of using global motion models are numerous. The BMC set of local translation vectors is substituted by a single set of global motion parameters per frame, and better quality can be achieved at lower bit-rates due to the better motion compensation.

Most global motion techniques are not appropriate though for sequences containing multiple moving objects and large perspective distortions. Methods based on the segmentation of just the dominant or background region using mosaics, or several regions using layered representations are presented in [2, 3, 4, 6, 11, 14]. Other methods which exploit spatio-temporal (3D) redundancy, where the image is segmented based on texture information first, and then grouped according to their motions have also been investigated [7, 12].

In this paper we propose a hybrid motion compensation scheme (HMC) which combines the strengths of both GMC and BMC, and can be implemented using extensions of current standard codecs. Similar schemes have been suggested in [1, 10, 13]. Our GMC module is based on a fast and robust multi-resolution feature-based motion estimation technique presented in [8], which uses a similarity motion model to register consecutive image frames, and can be easily extended to more complex models.

2. HYBRID MOTION COMPENSATION

The baseline algorithm used by most existing video compression standards based on BMC applies transform coding on the reference and residual images; the transform coefficients are then zig-zag scanned, quantized, and the resulting bit stream is entropy coded

*The support of ONR under contract N00014-96-1-0587 is gratefully acknowledged.

(see Figure 1). Inverse quantization and transform coding is also done by the encoder to predict the decoded frame, in order to reduce artifacts. Methods based on GMC can use the same algorithm, with the advantage that only one set of motion parameters (which defines the global image transformation) is necessary for each frame. Unfortunately, a global transformation cannot compensate for all motion in most cases, particularly when multiple independent moving objects are present in the scene. Since only a small set of global motion parameters has to be encoded, the hybrid method can considerably increase the performance of BMC methods without the drawbacks of pure GMC codecs.

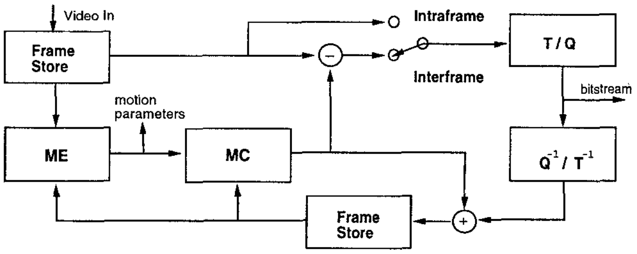

Figure 1: Block diagram of a motion compensation encoder.

Figure 2 shows the block diagram of a hybrid source encoder which uses BMC after the previous frame is compensated using a global transformation. For simplicity, we will only consider forward predicted frames in the following description.

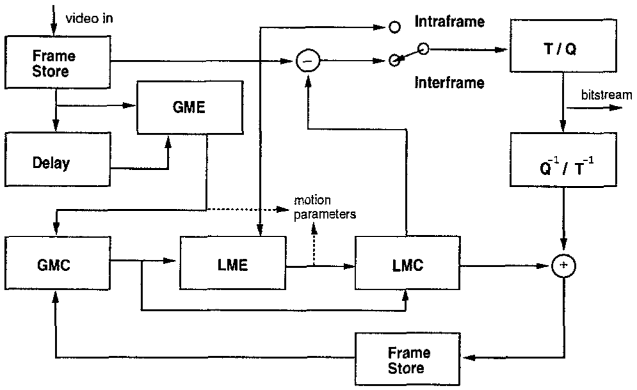

Figure 2: Block diagram of the hybrid motion compensation encoder.

Reference frames (or intraframes) are not affected by the global or block-based motion compensation modules, so that the image is directly encoded, without the need for motion vectors. For compensated frames (or interframes), the *global motion estimation* (GME) module first determines a global transformation which is used by the GMC module to align the previous (decoded) frame with the current one, filling the non-overlapping areas with the corresponding pixels from the previous frame. Then the *local motion estimation* (LME) module computes the translation vectors which minimize the residue (difference) between the global motion compensated frame and the current frame. A residual image is obtained by subtracting the local motion compensated frame from the current frame. The residual image is finally transform coded and quantized, to generate a bitstream which can be further coded using variable length coding (VLC). Inverse quantization and transform are applied to predict the decoded image, which is used for motion compensation in order to reduce artifacts at the decoder, since quantization is lossy. The global transformation is applied only when the PSNR between the global compensated frame and the current frame is higher than the PSNR between the current and previous frame without compensation.

Quantization and VLC are based on the H263 codec. Interframe DCT coefficients are uniform quantized by a single quantization parameter Q, which can be used to dynamically control the rate of the HMC codec. Specifically, when the difference image contains high frequency components such as edges which should be preserved, Q should be increased accordingly. This same scheme is used to quantize the intraframe AC coefficients. The DC intra-frame coefficients though, are DPCM encoded and quantized with quantization step size equal to 8.

3. MOTION ESTIMATION

Global motion is estimated using an algorithm similar to that presented in [9], which is a fast and robust feature-based multi-resolution image stabilization algorithm. Image tokens are automatically detected and tracked between corresponding pyramid levels, and the displacements are used to fit a similarity global motion model. The solution is further refined by the remaining finer pyramid levels, until the highest resolution level is processed. The similarity model is able to compensate for translations parallel to the image plane, rotation around the optical axis, and scaling due to motion along the optical axis. We have also experimented motion estimation algorithms using full affine and pseudo-perspective models, and obtained similar results. We have adopted the simpler similarity model because it is faster to estimate and in general more robust.

Local or block-based motion is estimated using blocks of size 8×8 pixels, and search sizes of ±15 pixels. The displacement vector is obtained by minimizing the sum of absolute differences between the current frame and global motion compensated frame, within the search window. Currently, no subpixel precision is computed for block matching.

4. PERFORMANCE EVALUATION

We evaluate the performance of the BMC, GMC, and HMC codecs based on rate-distortion curves (RDC). The distortion is given by the average PSNR between the original and decoded sequences, and the rate is given in bits per pixel, so that sequences with different image resolutions can be easily compared. A point in the RDC is defined by the average rate-distortion for the whole sequence, obtained with a constant quantization coefficient Q. Better quality and higher bit rates are expected for small values of Q. Large Q values produce low bit rate sequences, but with very low quality. For the construction of the RDC curves, Q was varied from 1 to 31 (the quantization stepsize is actually $2Q$, as in H263).

The BMC and GMC codecs were implemented based on the block diagram of Figure 1, and then combined as described in Section 2 to obtain the HMC codec. Four sequences composed of 10 frames each were used to evaluate these three codecs. The first is a synthetic sequence which simulates dominant rotation around the optical axis of the camera. Figure 3a shows the RDC for the Rotation sequence. Observe that the performances of HMC and GMC are very similar, about 1db better than BMC for the rate of 1 bit per pixel (bpp).

The Harbor sequence contains only zooming motion (changes in scale), which is compensated by the similarity model, but does not fit the block-based motion model. Again, HMC and GMC perform equally well, slightly better than BMC, as shown by Figure 3b. Similar results can be seen from Figure 3c, obtained from the Desert sequence, which starts zooming out, and then pans from right to left.

These results are expected since the global motion estimator is able to compute accurate transformations for the whole sequence, and the scenes are rigid. The Race sequence starts zooming out and then pans from right to left, and contains several racing cars, which do not fit a global motion model. Figure 3d shows the results for this sequence, where it can be seen that the performance of GMC drops considerably, and HMC is almost 2db better than BMC for 1bpp.

5. CONCLUSION

We have presented a hybrid video coding scheme which presents several advantages over standard block-based motion compensation codecs. It is based on a fast and robust motion estimation algorithm which compensates for translations, rotation, and scaling. Global motion compensation schemes offer good image registration with very small overhead due to the motion parameters, and our experimental results indicate that GMC schemes can outperform BMC in several situations. BMC though is more general, and performs better than GMC for sequences with several objects moving independently. HMC combines the strengths of both techniques, and current established motion compensation codecs could be easily extended to include global motion vectors, although this novel scheme would not be compatible with those standards. We are exploring a similar hybrid scheme which operates directly with video coded streams, in order to use the coded translation vectors to estimate the global motion, and create an intermediate, more compact, representation.

6. REFERENCES

[1] D. Adolph and R. Buschmann. 1.15 mbit/s coding of video signals including global motion compensation. *Signal Processing: Image Communications*, 3(2):259–274, 1991.

[2] S. Ayer and H.S. Sawhney. Layered representation of motion video using robust maximum-likelihood estimation of mixture models and MDL encoding. In *Proc. International Conference on Computer Vision*, Cambridge, MA, June 1995.

[3] T. Darrell and A. Pentland. Robust estimation of a multi-layered motion representation. In *Proc. IEEE Workshop on Visual Motion*, pages 173–178, Princeton, NJ, October 1991.

[4] F. Dufaux and F. Moscheni. Background mosaicking for low bit rate video coding. In *Proc. International Conference on Image Processing*, volume I, pages 673–676, Lausanne, Switzerland, September 1996.

[5] M. Hoetter. Differential estimation of the global motion parameters zoom and pan. *Signal Processing*, 16:249–265, 1989.

[6] M. Irani, S. Hsu, and P. Anandan. Mosaic based video compression. In *SPIE Conference on Electronic Imaging, Digital Image Compression: Algorithms and Techniques*, volume 2419, pages 242–253, San Jose, CA, 1995.

[7] F. Marques, P. Salembier, M. Pardas, R. Morros, I. Corset, S. Jeannin, B. Marcotegui, and F. Meyer. A segmentation-based coding system allowing manipulation of objects (SESAME). In *Proc. International Conference on Image Processing*, volume III, pages 679–682, Lausanne, Switzerland, September 1996.

[8] C.H. Morimoto, P. Burlina, R. Chellappa, and Y.S. Yao. Performance analysis of model-based video coding. In *Proc. International Conference on Image Processing*, volume III, pages 279–282, Lausanne, Switzerland, September 1996.

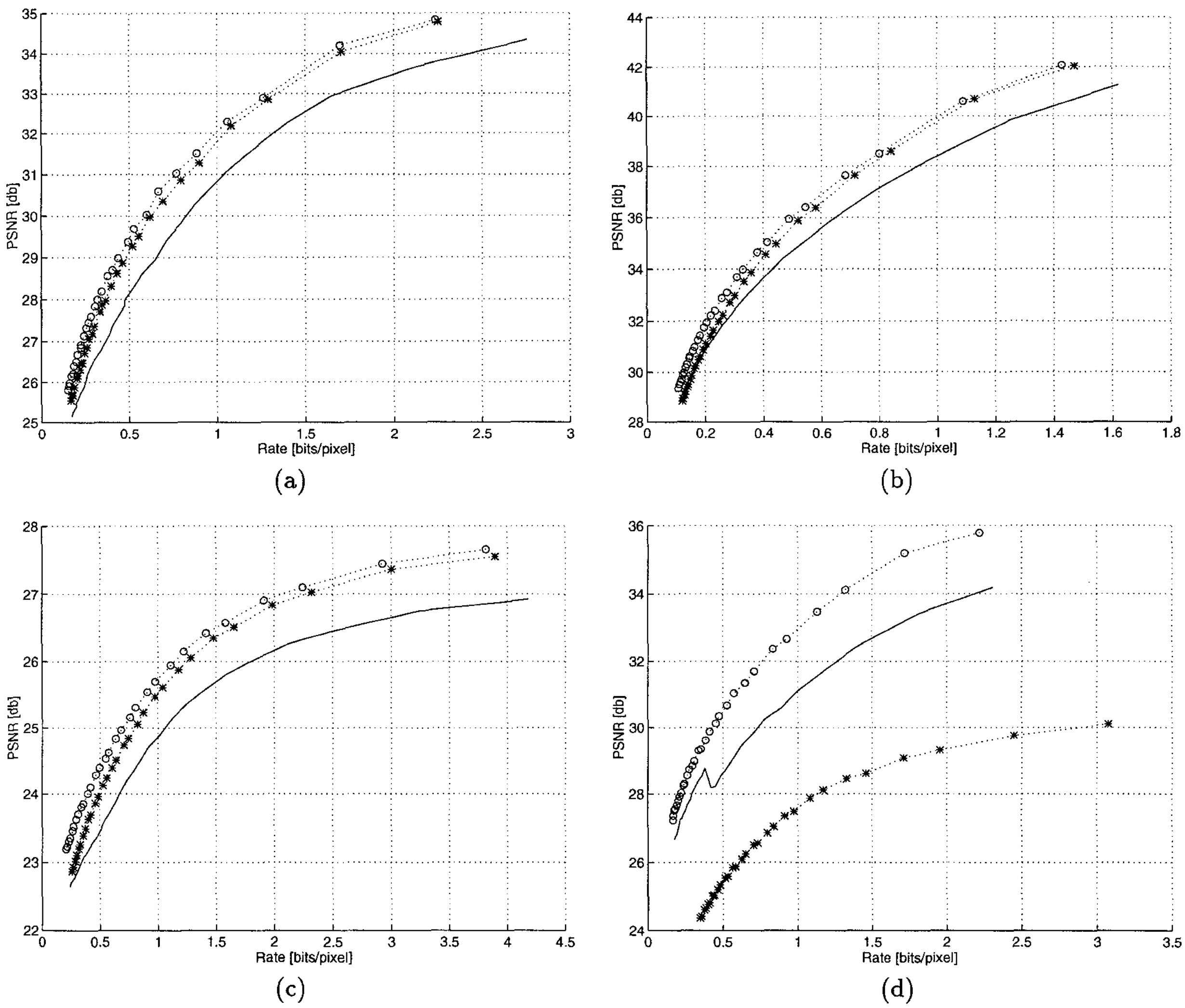

Figure 3: Rate-distortion curves for a) Rotation, b) Harbor, c) Desert, and d) Race sequences. Results from BMC are plotted using the solid line, GMC using the dotted line with stars, and HMC using the dotted line with circles.

[9] C.H. Morimoto and R. Chellappa. Fast electronic digital image stabilization. In *Proc. International Conference on Pattern Recognition*, Vienna, Austria, August 1996.

[10] F. Moscheni, F. Dufaux, and M. Kunt. A new two stage global/local motion estimation based on a background/foreground segmentation. In *Proc. IEEE Conference on Acoustics, Speech and Signal Processing*, pages 2261–2264, Toronto, Canada, May 1995.

[11] H. Sawhney, S. Ayer, and M. Gorkani. Model-based 2d and 3d dominant motion estimation for mosaicing and video representation. In *Proc. International Conference on Computer Vision*, pages 583–590, Cambridge, MA, June 1995.

[12] A.M. Tekalp. *Digital Video Processing*. Prentice Hall PTR, Upper Saddle River, NJ, 1995.

[13] Y.T. Tse and R.L. Baker. Global zoom/pan estimation and compensation for video compression. In *Proc. IEEE Conference on Acoustics, Speech and Signal Processing*, pages 2725–2728, Toronto, Canada, May 1991.

[14] J.Y.A. Wang and E.H. Adelson. Representing moving images with layers. *IEEE Trans. Image Processing*, 3(5):625–638, 1994.

[15] S.F. Wu and J. Kittler. A differential method for simultaneous estimation of rotation, change of scale and translation. *Signal Processing: Image Communications*, 2:69–80, 1990.

[16] A. Zakhor and F. Lari. Edge based 3d camera motion estimation with application to video coding. In M. Ibrahim Sezan and Reginald L. Lagendijk, editors, *Motion Analysis and image sequence processing*, pages 89–123, Boston, MA, 1993. Kluwer.