VIDEO CODING BY MODEL-BASED STABILIZATION

C. Morimoto P. Burlina R. Chellappa and Y. S. Yao

Center for Automation Research,
University of Maryland,
College Park, MD 20742, USA.

ABSTRACT

We study the performance associated with model-based video coding schemes using global motion models for motion compensation. Reference frames and compensated frame differences are coded using a method similar to MPEG, employing transform coding, quantization, and Huffman coding. The traditional block matching motion compensation approach is compared to global motion compensation approaches derived from 3D motion stabilization methods, using similarity and projective transformations. 3D model-based motion compensation is achieved by derotating the input sequence. This yields a projective transformation, which under certain circumstances, is well approximated by a similarity transformation. Experiments are carried out to analyze the performance of each scheme for comparable coding rates.

1. INTRODUCTION

Current source coding techniques applied to video sequences rely on modeling image motion as local 2D translations. In MPEG, the compensated frame differences and reference images are transform coded, and the transform coefficients are quantized and entropy coded (e.g. by Huffman or arithmetic coding). Existing video compression standards are essentially variants of this scheme based on motion compensation and discrete cosine transform (MC-DCT) [1, 4, 7]. These methods yield unacceptable performance at very low bit rates. Alternative approaches to block matching for motion compensation that relies on the use of global motion characteristics have received increased attention [5, 3]. We study a motion compensation scheme derived from 3D model based video stabilization [8]. This method uses derotation which results in compensation through a projective transformation, or it's approximation, a similarity transformation. We present experiments comparing the performance resulting from motion compensation utilizing block matching, similarity and projective transformations, obtained for equivalent coding rates, for various video sequences.

2. COMPENSATION BY STABILIZATION

We summarize below the various steps characterizing our method. We extend the approach reported in [8] and consider motion compensation using 3D model-based stabilization. Distant points are used for estimating the instantaneous rotation of the camera. The sequence is then derotated using a simple projectivity operator whose eight parameters (8P) are computed from the 3 instantaneous rotation parameters and the intrinsic camera parameters [8]. The transformation relating the image plane positions $\mathbf{p}_1$ and $\mathbf{p}_2$ of a given point at t_1 and t_2 are then expressed as

$$\mathbf{p}_2 = (\mathbf{c}^T \mathbf{p}_1 + 1)^{-1} (\mathbf{A}\mathbf{p}_1 + \mathbf{b}) \tag{1}$$

For an arbitrary rigid motion, if all imaged points are coplanar, the above relationship still holds. When the intrinsic parameters are unknown, nominal values are assumed to avoid estimating all the parameters, which most often yields an ill-conditioned system. A projective transformation is well approximated, under certain circumstances, by a similarity transformation (affine),

$$\mathbf{p}_2 = \mathbf{E}\mathbf{p}_1 + \mathbf{f} \tag{2}$$

where E includes only scaling and rotation, resulting in a 4-parameter (4P) transformation. These two approaches (4P and 8P motion compensation) are compared to a traditional block matching (BM) motion compensation scheme.

Local motion estimates are traditionally computed by block matching, by searching for the block that minimizes a given error measure, such as Mean Absolute Difference, or Normalized Cross Correlation [2]. Global motion estimates are computed by tracking a small set of image features using the Sum of Squared Differences. The feature displacements are then used to fit the similarity or projective model to obtain the global motion parameters.

For coding purposes, we use a classic approach used in MC-DCT schemes such as MPEG. The reference frames and compensated frame differences are transform encoded. The transform coefficient is uniform quantized, the resulting diagonally scanned bit stream is run length coded and Huffman encoded. Different quantization matrices are employed for intraframes and compensated frame differences.

3. PERFORMANCE ANALYSIS

Feature tracking was implemented in a Datacube Max Video 200 connected to a SUN SPARCstation 20/612. Motion compensation and frame differences can be processed at about 15 frames.s^{-1} (this includes tracking nine features within displacement of up to ±15 pixels).

We show in this section three experiments, each using sequences composed by 30 frames of resolution $256 \times 256 \times 8$. The first experiment was performed on a sequence exhibiting smooth dominant translation. The first frame is taken as the intraframe and the compensated frame differences are computed and transform coded for each subsequent frame. We show in Figure 1.(a) the resulting PSNR for each decoded frame. Clearly after the 10th frame the 4P compensation yields better results when compared to the BM. This suggests that for large displacements the 4P compensation works better, so that the intraframe does not need to be updated as often (this is further supported by the second experiment). Figure 2 shows the 1st, 10th, 20th and 30th frame of the original sequence, and the corresponding reconstructed frames using BM and 4P motion compensation methods. For a video rate of 10 frames.s^{-1}, the overall coding rate equals to 72Kb.s-1.

The second experiment also compares the BM and 4P based compensations. The sequence was obtained by panning the camera, causing average displacements of 10 pixels per frame. Images from the original and decoded sequences using BM and 4P motion compensation are shown in Figure 3. Subjective analysis shows that the 4P compensation yields superior results. This is confirmed by Figure 1.(b), which shows the performance decay of both schemes in dB. It can be seen that the 4P method always outperforms the BM. This sequence has a coding rate of about 25 Kb.s^{-1}.

The last experiment involves a sequence with dominant rotation, where we compare the BM compensation with the 8P based motion compensation. This sequence has a coding rate of about 40 Kb.s^{-1} The PSNR plot shown in Figure 1.(c) shows that the 8P outperforms the BM after the seventh frame. This can be seen in Figure 4, where it can be observed that the quality of the 8P compensation is better than the BM after 30 frames.

4. CONCLUSION

We have presented extensions of motion stabilization methods using similarity and projective transformations to model based video coding, and compared their performance with a traditional block based matching scheme. The results of our experiments suggest that model based video coding can support higher compression of the difference images and also longer periods between intraframe updates. Future work will study the inclusion of mosaicking and the use of other planes for registration, so as to achieve further motion compensation.

5. REFERENCES

[1] CCITT. CCITT recommendation h.261, video codec for audiovisual services at px64kbits/s. CCITT doc. COM XV-R 37-E, 1990.

[2] M. Ghambari. The cross-search algortihm for motion estimation. *IEEE Trans. Communications*, 38:950–953, 1990.

[3] M. Irani, P. Anandan, and S. Hsu. Mosaic based representations of video sequences and their applications. In *Proc. International Conference on Computer Vision*, 1995.

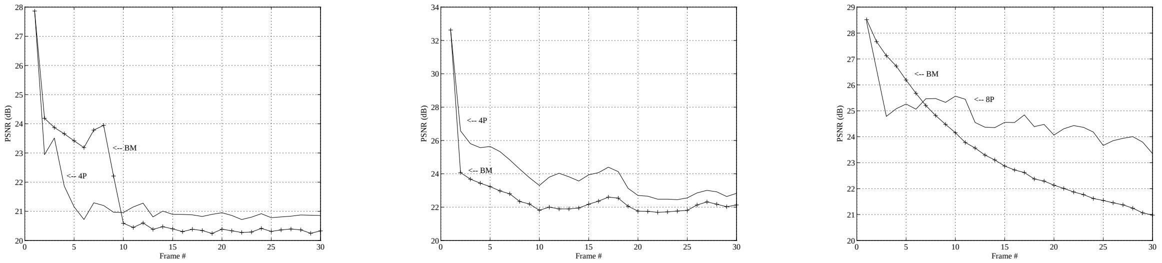
[4] ISO. Information technology - generic coding of moving pictures and associated audio, March 1995.

[5] O.J. Kwon, R. Chellappa, and C. Morimoto. Motion compensated subband coding of video acquired from a moving platform. In *Proc. IEEE Conference on Acoustics, Speech and Signal Processing*, Detroit,MI, 1995.

[6] C. Morimoto, D. DeMenthon, L. Davis, R. Chellappa, and R. Nelson. Detection of independently moving objects in passive video. In *Proc. of the IEEE Intelligent Vehicle Symposium*, pages 270–275, Detroit, MI, 1995.

[7] P. Pancha and M.E. Zarki. Mpeg coding for variable bit rate video transmission. *IEEE Communication Magazine*, 32(5):54–66, 1994.

[8] Y. S. Yao, P. Burlina, and R. Chellappa. Electronic image stabilization using multiple visual cues. In *Proc. International Conference on Image Processing*, pages 191–194, Crystal City, VA, October 1995.



(a) Sequence 1 (b) Sequence 2 (c) Sequence 3

Figure 1: Performace degradation for compensated frame differences for the Sequences 1, 2 and 3.

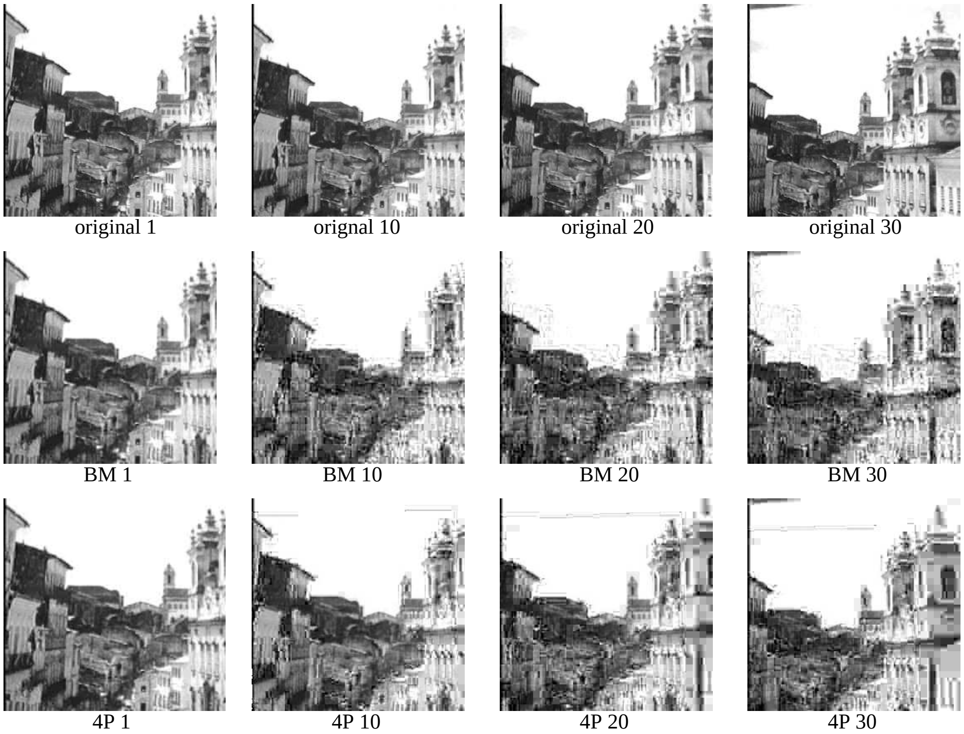
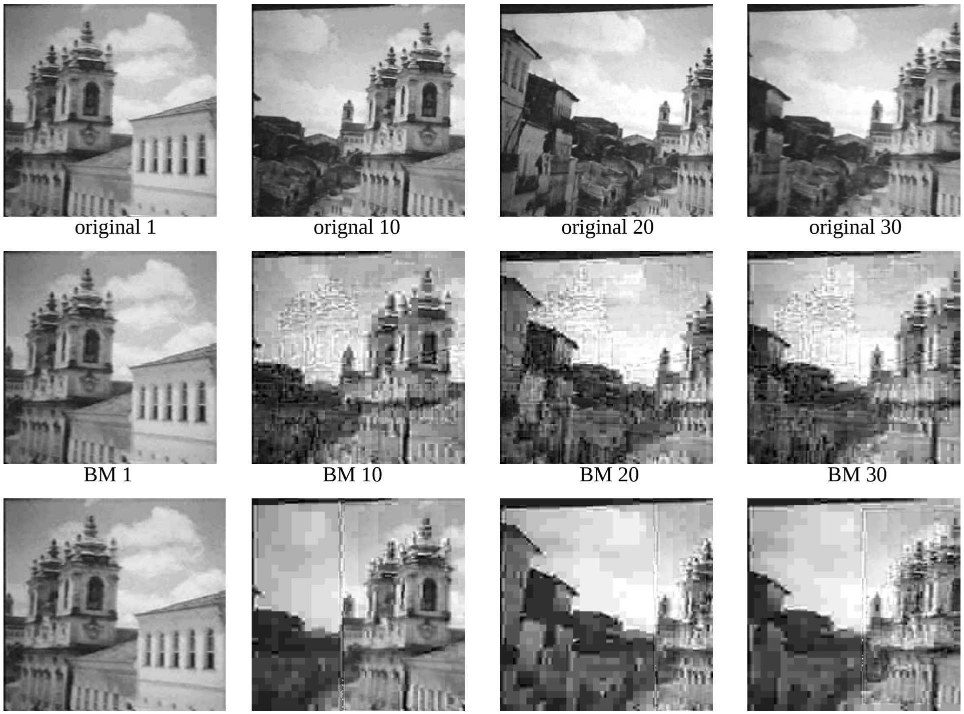


Figure 2: Decoding Results for 1rst Translation Sequence

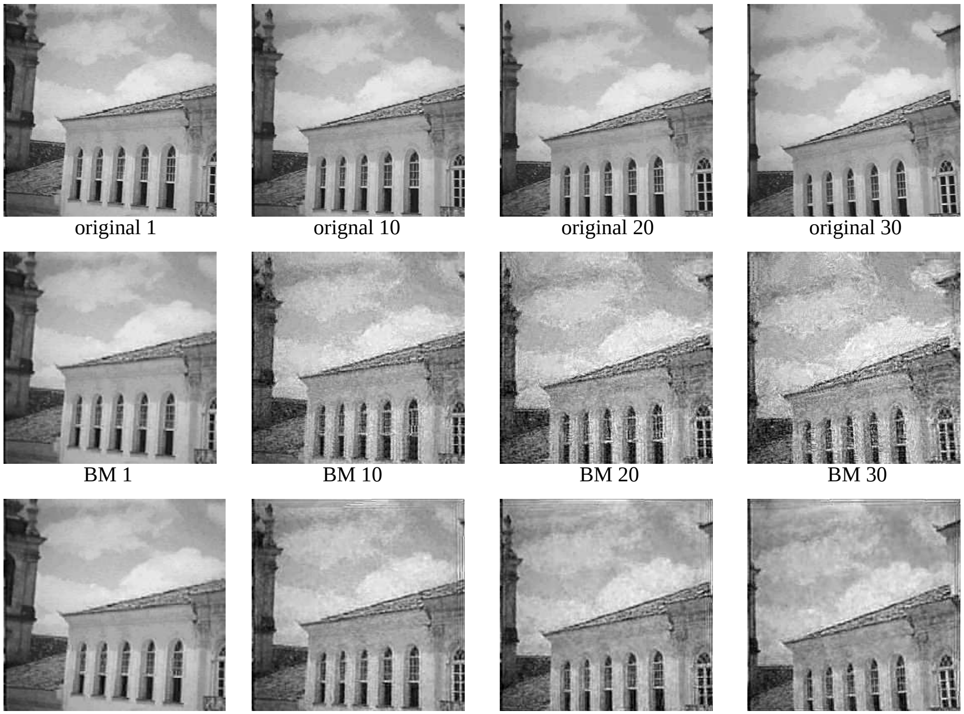


original 1 orignal 10 original 20 original 30

BM 1 BM 10 BM 20 BM 30

4P 1 4P 10 4P 20 4P 30

Figure 3: Decoding Results for 1rst Translation Sequence



original 1 orignal 10 original 20 original 30

BM 1 BM 10 BM 20 BM 30

8P 1 8P 10 8P 20 8P 30

Figure 4: Decoding Results for 1rst Translation Sequence