SEGMENTATION BASED CODING ALGORITHM FOR LOW BIT-RATE VIDEO

Ezzatollah Salari and Sheng Lin

Department of Electrical Engineering
University of Toledo, Toledo, OH 43606
USA

ABSTRACT

In this paper a new motion compensated predictive coding based on object region segmentation is proposed for image sequence coding at low bit-rates. The motion compensated prediction involves segmentation, motion detection, and motion estimation for moving objects. Segmentation is carried out on the reconstructed images in both the encoder and decoder. This will eliminate the need to transmit the region shape information. Also, motion vector prediction is performed in both the encoder and decoder leading to a significant reduction of overhead for motion information. Motion compensated prediction errors are transformed using the Discrete Cosine Transform (DCT) and the coefficients are quantized and entropy coded as recommended by CCITT. Computer simulation shows that the proposed coding algorithm significantly reduces the block artifact which is a dominant distortion associated with the conventional block matching algorithms at low bit-rates.

1. INTRODUCTION

Motion compensation is generally used in an image sequence to decorrelate the data along the temporal direction, and the block matching algorithm [1,2] is most commonly used for motion estimation. Block matching algorithms assume that pixels in a block have the same motion vector. However, if a block contains more than one object moving in different directions, one motion vector for the block will result in inaccurate motion estimation of moving objects, and the prediction error will be large in this block. Consequently, block artifacts will become visible if the algorithm is designed for a low bit-rate application. To avoid the block effect, some new segmentation based coding techniques [3-9] incorporating image structure information have been proposed. In these methods either the still image or the frame difference of two successive frames in an image sequence is first segmented into a number of regions. The region texture and boundary information is then coded for transmission. The segmentation based techniques are considered very promising, however, they require significant amount of codes for encoding the boundary information which is a dominant factor for low bit rate applications.

This paper presents a new segmentation-based motion compensation algorithm for image sequence coding. The proposed scheme provides an alternative method which does not require the region shape information to be transmitted. Also, motion vector prediction is performed in both encoder and decoder, leading to a significant reduction of overhead for motion information.

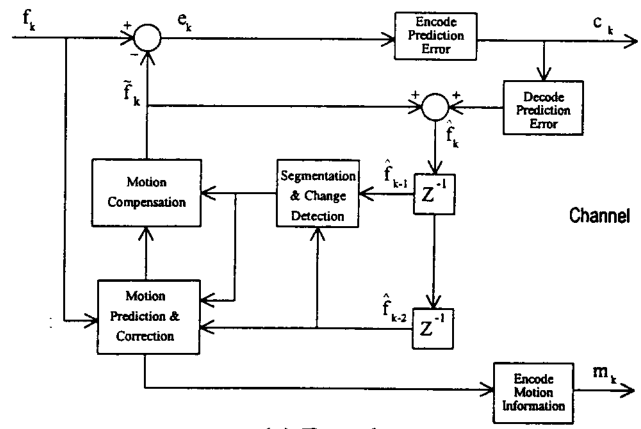
2. CODING ALGORITHM

The proposed algorithm is a combination of motion compensated prediction scheme and transform coding. The motion compensated prediction consists of moving object detection, motion estimation and motion compensation. As shown in Fig. 1, for a current frame f_k, its prediction $\tilde{f}_k$ is obtained by motion compensation of the previously reconstructed frame $\hat{f}_{k-1}$. Suppose two previously reconstructed frames $\hat{f}_{k-1}$ and $\hat{f}_{k-2}$ are available in both encoder and decoder. $\hat{f}_{k-1}$ is first segmented into a number of regions, and the moving regions are detected by change detection. The segmentation is a region growing technique based on human visual system, and was originally developed by Peterson *et al* [10].

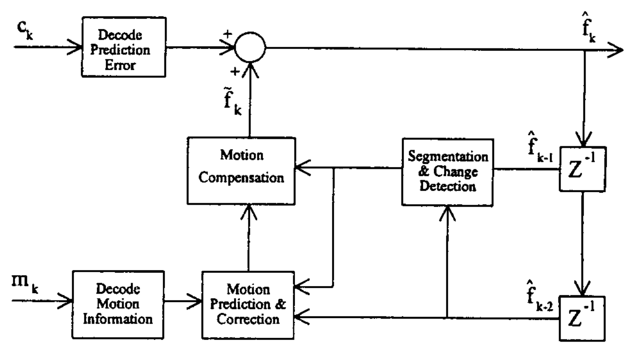
The basic idea of change detection is to find the regions with noticeable motion by thresholding the frame difference of two successive frames $\hat{f}_{k-1}$ and $\hat{f}_{k-2}$. The pixels that have a larger difference than a threshold value are detected as changed region elements. The changed region elements in each segment are counted. If the ratio of the number of the changed pixels to total pixels in a segment exceeds a threshold value, the segment is considered to be a changed segment. After change detection, visually insignificant changed segments are integrated into their surrounding changed segments. Through the change detection, only the changed regions rather than the entire image are motion estimated and compensated, thus reducing the computation time. These segmentation and change detection procedures are carried out on the reconstructed image frames in both encoder and decoder. Therefore, there is no need to transmit the region shape information.

For every changed region of $\hat{f}_{k-1}$, a forward motion vector and a backward motion vector are obtained by searching the corresponding match region in the current frame f_k and the previously reconstructed frame $\hat{f}_{k-2}$, respectively. These two motion vectors are similar because the motion between frames does not change significantly from frame to frame in some applications of image sequence coding, such as video conferencing. If the difference between these two vectors is larger than a threshold value for some regions, the forward motion vectors are coded and transmitted to the decoder. At the decoder, those motion vectors that were not sent, can be obtained from $\hat{f}_{k-1}$ and $\hat{f}_{k-2}$ in the same way as in the encoder side. Thus only a small part of motion vectors are transmitted and hence the overhead for motion information is drastically reduced, which is very important for low bit-rate applications.

The segmented reconstructed previous frame $\hat{f}_{k-1}$, along with the motion vectors of every moving region provide a motion compensated prediction frame. The differential error signals are then calculated by subtracting the predicted frame from the original one and encoded using the discrete cosine transform for spatial redundancy removal. To further reduce the



(a) Encoder



(b) Decoder

Fig. 1 Block diagram of proposed coding algorithm (c_k: coded DCT coefficients; m_k: coded motion correction information)

code amount, transform coding is applied only to the blocks that have large prediction errors. Initially, the pixels in the differential image with values larger than a threshold are detected and counted. If the ratio of the number of pixels with large differences to the total number of pixels in a block exceeds a threshold value, the block is assumed to be a large error block and it is coded using DCT. Otherwise, it will simply be replenished from its predicted values. The DCT coefficients are uniformly quantized (except the dead zone near zero) and entropy coded by the same method as described in CCITT Recommendation H.261[11]. At the receiver, the differential signals will then be added to the predicted values in the reconstruction of the current frame.

3. SIMULATION RESULTS

In this section, we present the effectiveness of the proposed coding algorithm for the low bit-rate applications. The standard video-conference image sequence "Miss America" is used. The image sizes are 360×288 pixels. The simulation is implemented with a bit rate of 64 kbit/s at a frame rate of 30 frame/s. Bit rate is controlled by changing quantization step size for DCT coefficients.

The quality of the reconstructed image is measured using the peak signal-to-noise ratio which is defined by

$$\mathrm{PSNR}(k) = 10 \cdot \log_{10}\left\{ \frac{255^2}{\frac{1}{MN}\sum_{r}\left[f_k(r) - \hat{f}_k(r)\right]^2} \right\}$$

where $M \times N$ is the frame size. Also, when K frames of the reconstructed sequence are considered, the average PSNR is equal to

$$\overline{\mathrm{PSNR}} = 10 \cdot \log_{10}\left\{ \frac{255^2}{\frac{1}{MNK}\sum_{r,k}\left[f_k(r) - \hat{f}_k(r)\right]^2} \right\}.$$

Fig. 2 shows the PSNR of reconstructed images versus frame number. The average PSNR is 32.81 dB with visually good quality at the bit rate of 64 kbit/s, i.e., 0.02 bit/pixel. The subjective quality of the reconstructed images also reveals significant improvements. An original and the reconstructed image frame in the sequence are shown in Fig. 3.

4. CONCLUSION

We presented a new algorithm for image sequence coding based on object region segmentation. The proposed scheme does not require the transmission of the region shape information to the receiver. This is achieved by the segmentation of the reconstructed image both in the encoder and the decoder. Also, motion vector prediction based on the previous two reconstructed frames is implemented both in the encoder and the decoder. The motion vectors are sent for those regions with some significant changes in their motion, otherwise, the motion vectors are estimated at the receiver from the two previously reconstructed frames. The proposed segmentation based coding algorithm will provide a significant improvement in the subjective image quality at low bit-rates.

Fig. 2 PSNR of reconstructed images

REFERENCES

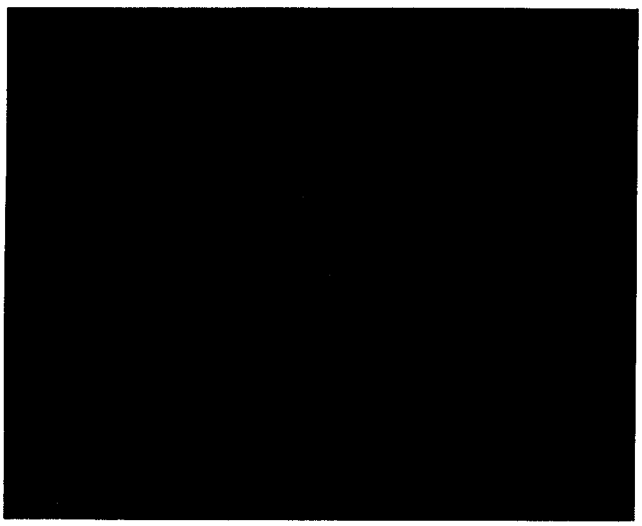
[1] J. R. Jain and A. K. Jain, "Displacement measurement and its application in interframe image coding," IEEE Trans. Commun., vol. COM-29, no. 12, pp. 1799-1808, 1981.

[2] H. G. Musmann, P. Pirsch, and H. J. Grallert, "Advances in picture coding," Proc. IEEE, vol. 73, no. 4, pp. 523-548, 1985.

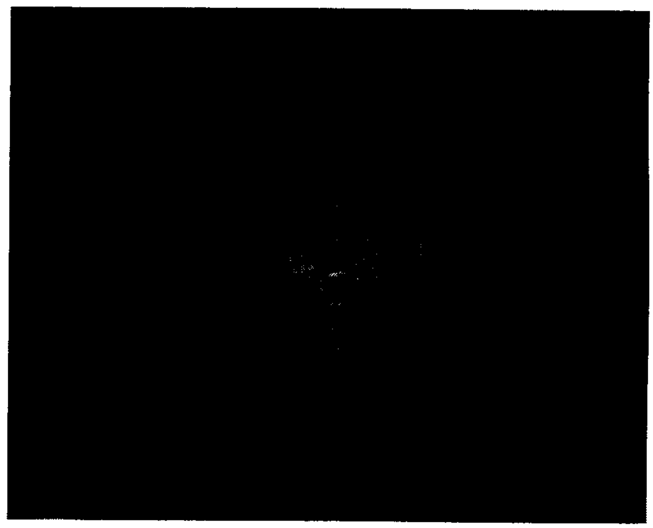
[3] M. Kunt, A. Ikonomopoulos, and M. Kocher, "Second generation image coding techniques," Proc. IEEE, vol. 73, no. 4, pp. 549-574, 1985.

[4] M. J. Biggar, O. J. Morris, and A. G. Constantinides, "Segmented-image coding: performance comparison with the discrete cosine transform," IEE Proc. F, vol. 135, no. 2, pp. 121-132, 1988.

[5] M. J. Biggar and A. G. Constantinides, "Segmented video coding," Proc. Int. IEEE Conf. ASSP'88 (New York), pp. 1108-1111, 1988.
[6] M. Soryani and R. J. Clarke, "Coding moving image sequences by segmentation of difference frames and motion-adaptive frame interpolation," Electron. Lett., vol. 25, no. 12, pp. 794-795, 1989.
[7] F. C. Leou and Y. C. Chen, "A contour-based image coding technique with its texture information reconstructed by polyline representation," Sig. Proc., vol. 25, no. 1, pp. 81-89, 1991.
[8] S. B. Chae, J. S. Kim, and R. H. Park, "Video coding by segmenting motion vectors and frame differences," Opt. Eng., vol. 32, no. 4, pp. 870-876, 1993.
[9] O. J. Kwon and R. Chellappa, "Segmentation-based image compression," Opt. Eng., vol. 32, no. 7, pp. 1581-1587, 1993.
[10] H. A. Peterson, S. A. Rajala, and E. J. Delp, "Human visual system properties applied to image segmentation for image compression," Proc. IEEE GLOBECOM, pp. 91-95, 1991.
[11] CCITT Recommendation H.261, "Video codec for audiovisual services at $p \times 64$ kbit/s," 1990.



(a) Original frame



(b) Reconstructed frame

Fig. 3. An original and the reconstructed frames in Miss America sequence