MODIFIED MOTION COMPENSATED TEMPORAL FRAME INTERPOLATION FOR VERY LOW BIT RATE VIDEO

Chi-Kong Wong and Oscar C. Au

Department of Electrical and Electronic Engineering
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong.
Email: ccou@ee.ust.hk

Abstract

Common techniques such as frame repetition or linear interpolation for reconstructing skipped frames in temporally subsampled video sequence tend to introduce undesirable artifacts. A previously proposed technique, fast motion compensated temporal interpolation (FMCTI), can interpolate video frames in time domain with good interpolated image quality. However, the reconstructed frames tend to be blocky. In this paper, we proposed to incorporate a pixel-based motion classifier into FMCTI to solve the problem. The resulting technique, modified fast motion compensated temporal interpolation (MFMCTI), was found in simulation to give significantly reduced blocking artifacts.

I. INTRODUCTION

Temporal frame subsampling is a simple technique commonly used in conjunction with other techniques such as CCITT H.261 to achieve high compression ratio in very-low-bit-rate (VLBR) video conferencing applications. The skipped frames need to be reconstructed at the receiver. It is well known that simple frame reconstruction techniques such as frame repetition or linear interpolation introduce disturbing artifacts. Frame repetition generates jerky object motions because object motions are not accounted for. Linear interpolation by temporal filtering exhibits blurring in the moving areas because pixels of different object regions are inappropriately used for the interpolation. A postprocessing technique, motion compensated temporal interpolation (MCTI)[1], was proposed to interpolate the skipped frames at receiver from a temporally subsampled videoconferencing sequence. Although MCTI can achieve very good visual quality, it requires a large amount of computation, around 4.4×10^8 additions for each subsampled frame (in CIF format), mainly to perform motion estimation. A fast version of MCTI, FMCTI was also proposed in [1], which can reduce the computation by two order of

magnitude, with reasonably good performance. The FMCTI can potentially be combined with other compression schemes such as CCITT H.261 to achieve very large compression ratio so as to satisfy the very-low-bit-rate requirement of the telephone channels. Since both MCTI and FMCTI use block-based motion estimation, it is inevitable to have blocking artifacts particularly in blocks in which some relatively small objects deform or rotate over time. This includes blinking eyes and fast-rotating hands. In such situations, the assumption of linear translational motion is simply invalid. These blocking artifacts tend to be disturbing to the human visual system because they appear at moving objects which tend to attract attention. In this paper, we propose a modified fast motion compensated temporal interpolation (MFMCTI) scheme by introducing a pixel-based motion classifier based on the information at two instead of one previous frames. Simulation results show that the modification can reduce drastically the blocking artifact resulting in smooth and natural motions in the interpolated sequence. Due to the simple structure of the classifier, the computation requirement is quite moderate in the receiver.

II. MOTION COMPENSATED TEMPORAL INTERPOLATION ALGORITHMS

2.1. Review of FMCTI [1]

For any k , a frame is to be inserted between the $(k-1)^{th}$ and k^{th} subsampled frames so that object motions would appear to be smoother between the adjacent frames. The $(k-1)^{th}$ frame, the k^{th} frame and the inserted frame are divided into blocks of size $N \times N$. Firstly, if the mean absolute difference (MAD) between the block located at (i, j) in the $(k-1)^{th}$ frame and the corresponding block at

(i, j) in the k^{th} frame is less than a threshold T_1 , the block at (i, j) in the inserted frame is classified as stationary block. For each of the non-stationary block B_1 located at (x, y) in $(k-1)^{th}$ frame, a search area of size $(2W+1) \times (2W+1)$ is defined in the k^{th} frame and block-based motion estimation is performed using pixel decimation and search area subsampling. This is called the forward motion search. If the MAD between B_1 in the $(k-1)^{th}$ frame and the best matched block B_2 located at $(x + \Delta x, y + \Delta y)$ in the k^{th} frame is less than a threshold T_2 , backward motion search is not performed. Otherwise, a search area of size $(2W+1) \times (2W+1)$ is defined in the $(k-1)^{th}$ frame and a backward motion estimation is performed for the block at (x, y) in the k^{th} frame. Because of the assumed linear translational motion, the block B_2 should appear at $(x + \frac{\Delta x}{2}, y + \frac{\Delta y}{2})$ in the inserted frame. However, the block B_3 located at $(x + \frac{\Delta x}{2}, y + \frac{\Delta y}{2})$ in the inserted frame usually would not fit into the block grid exactly in the inserted frame. Instead it would usually cover four $N \times N$ blocks. To solve the problem, a list of motion vector candidates is collected for each $N \times N$ block in the inserted frame. The motion vector $(\frac{\Delta x}{2}, \frac{\Delta y}{2})$, either forward or backward, is added to the candidate list of a block in the inserted frame if this block overlaps with the block B_3 . After forward and backward motion estimations, each block in the inserted frame will have a list of candidate motion vectors. The candidate with the maximum associated overlapping area is selected because this motion vector is more reliable than those with smaller overlapping area. A motion field is thus created.

Let the best motion vector for the block B_4 located at (i, j) in the inserted frame be (v_x, v_y) , this block is estimated by averaging the block B_5 located at $(i + v_x, j + v_y)$ in the $(k-1)^{th}$ frame and the block B_6 located at $(i - v_x, j - v_y)$ in the k^{th} frame. One problem of FMCTI is that, if B_5 does not really match B_6 , the estimation by averaging produces undesirable blocking artifacts in the inserted frame.

2.2 Modified Fast Motion Compensated Temporal Interpolation (MFMCCTI)

For the sake of discussion, we assume two frames S_{i-2} and S_i are available. Our task is to construct the skipped frame S_{i-1} . To solve the blocking problem of FMCTI, we propose to incorporate a pixel-based motion classifier, as shown Fig. 1(a), which assigns each pixel in the inserted frame to one of the classes in Table 1, using the information from three successive subsampled frames. Based on the classes, we choose the appropriate coefficients, a_{i-2} and a_i , for the interpolation filter to generate each of the pixels in the inserted frame,

$$\tilde{S}_{i-1}(x, y) = a_{i-2}S_{i-2}(x \mp v_x, y \mp v_y) + a_iS_i(x \pm v_x, y \pm v_y)$$

where if (v_x, v_y) is a forward/backward motion vector, the upper/lower sign is chosen.

Class	Meaning	Criteria	interpolator coefficients
b_1	moving object	$\Delta Q_1 \wedge \Delta Q_2 = 1$	$a_{i-2} = 1/2$ $a_i = 1/2$
b_2	uncovered background	$\Delta Q_1 \wedge \overline{\Delta Q_2} = 1$	$a_{i-2} = 0$ $a_i = 1$
b_3	covered background	$\overline{\Delta Q_1} \wedge \Delta Q_2 = 1$	$a_{i-2} = 1$ $a_i = 0$
b_4	stationary background	$\overline{\Delta Q_1} \wedge \overline{\Delta Q_2} = 1$	$a_{i-2} = 1/2$ $a_i = 1/2$

Table 1: Image classes

The classifier was originally proposed by Bergmann [2] for interpolating skipped fields of television signals. We adopted and modified the classifier so as to tackle the blocking problem of FMCTI. According to Fig. 1(a), the frame differences of the luminance, ΔS_1 and ΔS_2 , are calculated from three successive subsampled frames S_{i-4} , S_{i-2} and S_i where S_i is the current frame. Our classifier is a causal operator, using the information from the two past frames (S_{i-2} and S_{i-4}). This should be contrasted with the non-causal classifier proposed by Bergmann. Then we perform binary thresholding on ΔS_1 and ΔS_2 to

obtain two binary images ΔQ_1 and ΔQ_2 . These two binary images provide masks for the unchanged and changed parts within the three frames. Using the logical combinations stated in Table 1, all pixels in the inserted frame are assigned to one of the four image classes b_1, b_2, b_3 and b_4 , as shown in Fig. 2 by black dots. The four classes can be explained by the example in Fig. 1(b). Some pixels in $\tilde{S}_{i-1}$ are stationary background pixels (b_4) characterized as being "unchanged" in S_{i-4}, S_{i-2} and S_i ($\overline{\Delta Q_1} \wedge \overline{\Delta Q_2} = 1$). These pixels can be generated by taking the average of the corresponding pixels in S_{i-2} and S_i . Some pixels in $\tilde{S}_{i-1}$ belong to moving objects (b_1) and are characterized as being changed between S_{i-4} and S_{i-2} (ΔQ_1), and between S_{i-2} and S_i (ΔQ_2). They can be generated by taking the average of the corresponding pixels in S_{i-2} and S_i . Some pixels in $\tilde{S}_{i-1}$ are uncovered background pixels (b_2) and characterized as "changed" between S_{i-4} and S_{i-2} (ΔQ_1), and "unchanged" between S_{i-2} and S_i ($\overline{\Delta Q_2}$). These pixels should be copied from S_i . Some pixels in $\tilde{S}_{i-1}$ are covered background pixels (b_3) and characterized as "unchanged" between S_{i-4} and S_{i-2} ($\overline{\Delta Q_1}$), and "changed" between S_{i-2} and S_i (ΔQ_2). These pixels should be copied from S_{i-2} .

III. SIMULATION RESULTS AND DISCUSSION

Both FMCTI and MFMCTI are simulated on the luminous component of the "Miss America", "Salesman" and "Claire" sequences which are in CIF (288×352) format. The parameters W and N are both 16. Both FMCTI and MFMCTI work very well on most of sequences.

In Fig. 3, we show the most challenging situations in the "Miss America" sequence: the mouth and eye portions. The three successive subsampled frames are 97th, 99th and 101th of which the 101th frame is the current received frame at receiver. Within the three received frames, there are non-translational motions at the eyes and mouth. The mouth is deformed. The eyes

are fully closed in both 97th and 99th frames and half open in 101th frame. The estimated mouth and eye portions of 100th frame generated by FMCTI are shown in Fig. 3. In this estimated frame, the deformed mouth is approximated well by FMCTI, being very similar to the mouth in the original frame. However, blocking artifact is observed in the left eye in FMCTI due to the averaging of the fully closed eye on the 99th frame and the half-open eye in the 101th frame. In MFMCTI with the pixel-based classifier, both the month and eye portions are interpolated without the blocking artifacts as in FMCTI. Both the eyes and month generated by MFMCTI are very natural. This verifies that MFMCTI is indeed effective in solving the blocking problem of FMCTI.

The Peak-Signal-to-Noise-Ratio (PSNR) of MCTI, FMCTI and MFMCTI for "Miss America" and "Claire" are shown in Fig. 4. The average PSNR of MCTI, FMCTI and MFMCTI are 37.18 dB, 37.45 dB and 37.69 dB respectively in "Miss America" and 39.05dB, 40.18 dB and 41.05 dB in "Claire". These suggest that the proposed MFMCTI is a significant improvement over FMCTI.

IV. CONCLUSION

In this paper, we propose to modify a previously proposed motion compensated temporal interpolator by incorporating a pixel-based motion classifier. Simulation results show that the proposed scheme can reduce significantly the blocking artifacts of the previous scheme.

ACKNOWLEDGEMENT

This work is supported by HKTIIT93/94.EG02.

REFERENCES

- [1] C.K. Wong and O. C. Au, "Fast motion compensated temporal interpolation for video", *Proc. SPIE: Visual Communications and Image Processing '95*, pp.1108-1118, May 1995.
- [2] H. C. Bergmann, "Motion adaptive frame interpolation", *Proc. Int. Zurich Seminar on Digital Communications*, Vol-D, pp. 21-25, Mar 1984.

Fig. 1. (a) Block diagram of the classifier; (b) Classification of pixels.

Figure 2. Image classes of 100th frame of "Miss America" sequence.

Fig. 3. Simulation result on eye and mouth portions of "Miss America" sequence.

Fig. 4. PSNR of the interpolated frames of "Miss America"(left) and "Claire"(right) sequences