A STUDY ON MOTION COMPENSATED PREDICTION IN DCT DOMAIN WITH MULTIPLE MODE SELECTION

Kazuhisa YAMAGISI, Jiro KATTO and Yasuhiko YASUDA

School of Science and Engineering, Waseda University,
3-4-1 Okubo, Shinjuku-ku, Tokyo, Japan.
E-mail:yamagisi@yasuda.comm.waseda.ac.jp

ABSTRACT

In this paper, we propose a motion vector search algorithm that employs the Sum of Absolute Difference (SAD) minimization in DCT domain as a distortion measure. In addition, block based motion estimation is carried out with multiple modes in this frequency domain, and a mode is selected according to the Lagrangian cost function [1, 2, 3]. Four modes are designed. The first two modes are a mode with 1 motion vector (INTER1V) and a mode with 4 motion vectors (INTER4V) per macroblock that are similar to the H.263 [4]. The last two modes use 4 motion vectors per macroblock (INTER1V+) or per block (INTER4V+), where the 8×8 DCT coefficients are firstly separated into 4 groups and a motion vector is independently assigned to each group. The average SNR (Signal to Noise Ratio) gains between 0.3 and 1.0dB are obtained for tested sequences when compared with TMN. Our proposal method provides true motion vectors for low frequency and minimizes distortion for higher frequency. Therefore our proposed method also reduces subjective visual degradation.

1. INTRODUCTION

Today's video coders employ block-based motion compensated prediction. SNR gains achieved by motion compensated prediction are increased by using smaller size blocks. For example, instead of using 16×16 luminance blocks, usage of 8×8 luminance blocks provides better SNR gains. However, the amount of bits are increased when introducing small blocks due to motion vector overheads. Therefore, rate-constrained motion estimation is often employed [1, 2, 3]. In this motion estimation, Lagrangian cost function $J = D + \lambda \cdot R$ is minimized, where D is distortion, R is bits of motion vectors and λ is a Lagrange multiplier, and the macroblock mode decision is carried out based on this Lagrangian cost function. Minimization of this cost function is expected to solve the tradeoff problem on bit assignment between DCT coefficients and motion vector overheads.

With regard to the motion estimation, motion vector search algorithms employ the SAD as a distortion measure in Test Model Near-Term (TMN) [5] for the ITU-T Recommendation H.263, Version 2 [4]. The SAD is calculated between all luminance pixels of the candidate block and the target block. However, this does not necessarily lead to detection of optimal (accurate) motion vectors, and sometimes causes visually remarkable block noises and mosquito noises. A simple solution to alleviate this problem is to differentiate the SAD minimization process according to frequency components.

In this paper, we propose a motion vector search algorithm that employs the SAD minimization in DCT domain as a distortion measure. In addition, block based motion estimation is carried out with multiple modes in frequency domain, and a mode is selected according to the Lagrangian cost function. Finally, we demonstrate effectiveness of our proposed method by computer simulations.

2. MOTION COMPENSATED PREDICTION IN FREQUNCY DOMAIN

Hybrid video coding consists of the motion compensated prediction and the coding strategy for its prediction errors. Different from the ordinary schemes, we try to decrease non-zero DCT coefficients by applying the motion compensated prediction in the DCT domain, where a motion vector search algorithm that employs the SAD minimization in DCT domain as a distortion measure is proposed. As shown in Figure 1, each macroblock is motion estimated and compensated according to one of the four modes that are described below. The first two modes are a mode with 1 motion vector (INTER1V) and a mode with 4 motion vectors (INTER4V) per macroblock that are similar to the H.263, respectively. The last two modes use 4 motion vectors per macroblock (INTER1V+) or per block (INTER4V+), where the 8×8 DCT coefficients are firstly separated into 4 groups and then a motion vector is independently assigned to each group. That is, 16 motion vectors can be used per macroblock at the maximum. In addition, lower frequency com-

ponents and higher frequency components are estimated in an individual manner in order to bring following advantages. For lower frequency, motion estimation is done precisely to detect true motion without being affected by higher frequency noises. Furthermore, it is expectable to reduce block noises that are caused by DC components. For higher frequency, motion estimation is done roughly enough to minimize prediction errors even if motion detection may not be accurate. This combination is promising because subjective visual distortion is expected to be reduced.

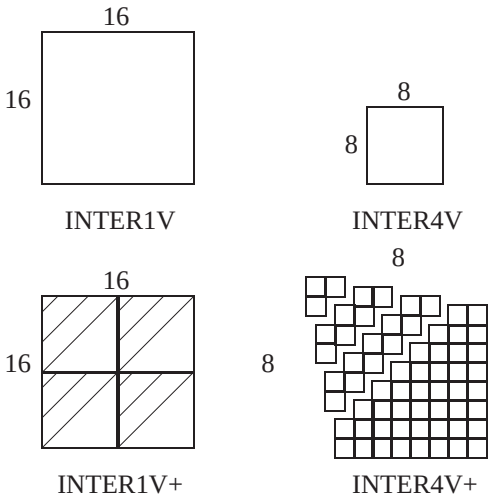

Fig. 1. Proposed Block Division Strategy In Frequency Domain

According to this extension of motion compensation, in our proposal method, additional information is needed to represent a block division pattern in frequency domain; 1bit for a macroblock (INTER1V or INTER1V+) or 4bits for blocks (INTER4V or INTER4V+). Differential encoding of motion vectors among neighboring frequency groups is applied only to motion vectors of the lowest frequency. When a given neighboring block uses INTER1V+ or INTER4V+, a candidate predictor uses motion vectors of the lowest frequency.

Except for the motion compensation strategy discussed above, we had followed conventional encoding methods of the H.263. For example, reference blocks are taken from three surrounding blocks for motion vector coding. After the mode decision, prediction errors are quantized and encoded to produce H.263 compatible bitstreams.

3. MOTION ESTIMATION AND MODE DECISION

In the above, we introduced a new motion compensation method in frequency domain with multiple modes. However the amount of bits may be increased by the motion vector overheads introduced in block division. In order to make this effect as little as possible, block-based motion estimation is carried out by minimizing

$$J_{REF}(B,v) = D_{REF}(B,v) + \lambda_{MOTION}\, R_{REF}(v) \quad (1)$$

where $D_{REF}(B,v)$ is a distortion for the block B between the original (current) frame o and the reference (previous) frame s, that is computed by

$$dif_{REF}(v) = \mathscr{D}(o[x,y] - s[x - v_x, y - v_y]), \quad (2)$$

$$D_{REF}(B,v) = \sum_{x,y \in B} |dif_{REF}(v)|\,, \quad (3)$$

where $\mathscr{D}$ means 8×8 DCT operation, $R_{REF}(v)$ is a bitrate associated with the motion vector including spatial displacement, and λ_{MOTION} represents a Lagrange multilier, respectively.

In our experiment, motion estimation is carried out in a hierarchical manner. Firstly, a window of ± 15 integer pixel units is searched by an INTER1V mode, and a search center is obtained as a median of the predicted motion vectors as defined in [4]. Secondly, the search window is reduced to ± 2 integer pixel units around the search center, and a set of motion vectors is obtained for each mode: INTER1V, INTER1V+, INTER4V and INTER4V+. Finally, half pixel refinement is performed for each mode within ± 1 search window, and the best candidate with half-pel accuracy is determined in each mode.

Given the candidate motion vectors, one of macroblock modes is chosen. Again, we employ a Lagrangian cost function

$$J_{REC}(B,v) = D_{REC}(B,v) + \lambda_{MODE}\, R_{REC}(h,v,c) \quad (4)$$

where $D_{REC}(B,v)$ is a distortion computed as the sum of squared differences (SSD) between pixels of the block B and λ_{MODE} is Lagrange multiplier. $R_{REC}(h,v,c)$ is a bitrate that is needed to transmit and reconstruct a motion vector v and DCT coefficients c, including a macroblock header h that contains additional information about the block division pattern in a frequency domain. The additional mode decision parameter determines a mode to code each macroblock among INTER1V, INTER1V+, INTER4V and INTER4V+. The Lagrange multiplier for motion estimation is chosen as $\lambda_{MOTION} = 0.92 \cdot Q$, and the Lagrange multiplier for mode decision is chosen as $\lambda_{MODE} = 0.85 \cdot Q^2$, where Q is an averaged DCT quantizer value over all macroblocks of the previous coded frame [2, 3].

4. SIMULATION RESULTS

Experiments were conducted using the CIF test sequences. A total of 150 frames of the sequences, Cheerleaders, and a total of 300 frames of the sequences, Foreman and Table Tennis, were used. Their frame rate was fixed at 10Hz. Then, Foreman and Table Tennis sequences were encoded

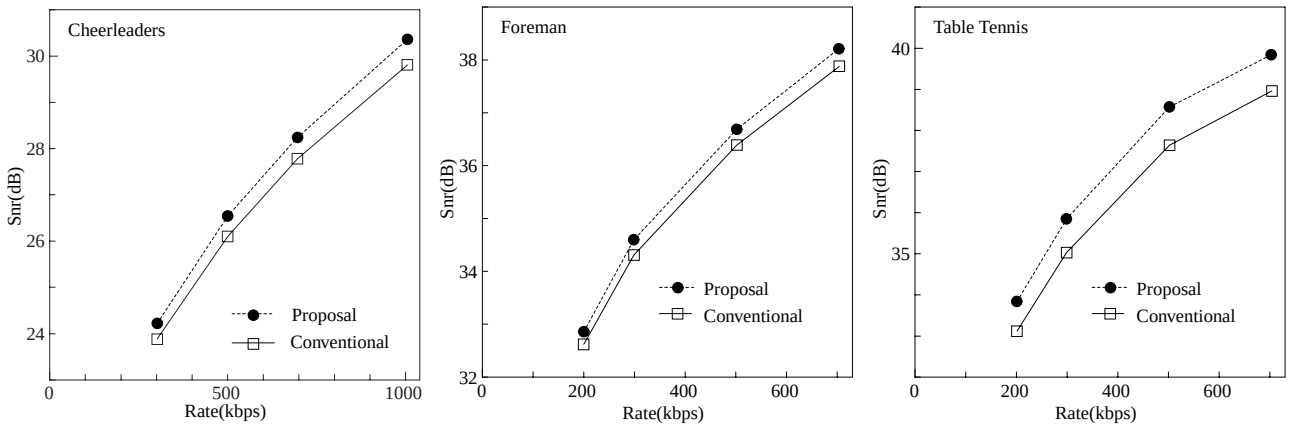

Fig. 2. Snr vs. Rate

at 200, 300, 500 and 700 kbps and Cheerleaders sequences were encoded at 300, 500, 700 and 1000 kbps, respectively. As for rate control, we used the TMN8 rate control method [6].

The average SNR has been improved between 0.3 and 1.0dB (Figure 2), compared with the conventional method. In case of Cheerleaders (500kbps), SNR was improved for all frames (Figure 3). In case of Foreman (500kbps) and Table Tennis (200kbps), there were a few frames on which SNR has degraded due to mode selection overheads, but overall SNR gains were still achieved (Figure 3). In the proposed method, even if SNR is inferior to the conventional method, there were some frames in which visual degradation was improved. This is because the frequency components were effectively differentiated and exploited by introduction of INTER1V+ and INTER4V+.

The selection ratio of INTER1V+ and INTER4V+ became higher and higher as a coding rate became higher. Simulation results also confirmed that the ratio of INTER1V+ and INTER4V+ was large on frames in which target objects moved rapidly (Figure 5). Moreover, the ratio of INTER1V+ and INTER4V+ was effectively controlled by the Lagrangian cost function.

Note that an improvement was large on frames in which target objects moved rapidly. This is because, in the conventional scheme, incorrect motion estimation occurred on such frames and remarkable visual degradation was observed as shown in Figure 5. In contrast, our proposal had drastically reduced the degradation by detecting optimal motion vectors. Furthermore, we want to emphasize following advantages; uncovered areas produced when a person moved were correctly detected, block noises due to discontinuity of DC components between macroblocks were reduced, and mosquito noises due to lack of higher frequency components were also reduced.

5. CONCLUSIONS

In conclusions, we introduced a new motion compensation method in frequency domain with multiple modes. Two new modes with 4 motion vectors per macroblock and per block were introduced, that provides more proper motion vectors for lower frequency and minimizes distortion for higher frequency. Experimental results using actual image sequences show that the proposed method indeed increases SNR gains and reduces visual distortion (block noises, mosquito noises and object disappearance) caused by conventional motion detection limitations.

6. REFERENCES

[1] G. Sullivan and T.Wiegand, “Rate-Distortion Optimization for Video Compression”, IEEE Signal Processing Magazine, vol. 15, no. 6,pp. 74-90, November. 1998.

[2] T.Wiegand and B.Girod, “Lagrangian Multiplier Selection in Hybrid Video Coder Control”, Proc ICIP 2001, Thessaloniki, Greece, October. 2001.

[3] T.Wiegand and B.Andrews, “An Improved H.263 Coder Using Rate-Distortion Optimization”, ITU Study Group 16, Video Coding Experts Group, Document Q15-D-13, Finland, April 1998.

[4] G.Sullivan, “Draft Text of Recommendation H.263 Version 2 (“H.263+”) for Decision”, ITU Study Group 16, Video Coding Experts Group, January 1998.

[5] S.Wenger, “Test model 11”, ITU Study Group 16, Video Coding Experts Group, Monterey, February 1999.

[6] Thomas R. Gardos, “Video Codec Test Model, Near-Term, Version 8 (TMN8)”, ITU Study Group 16, Video Coding Experts Group, Portland, 24-27 June 1997

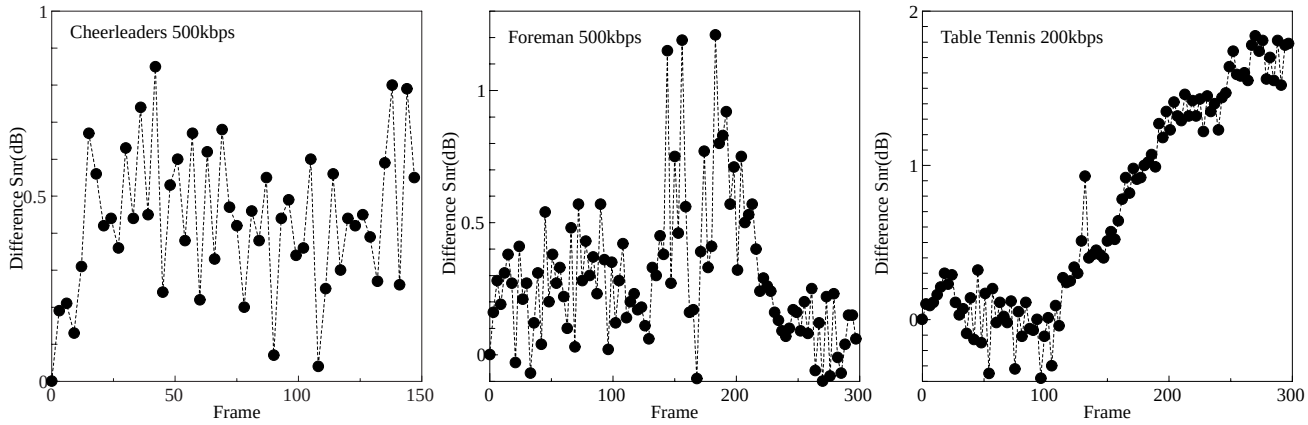

Fig. 3. SNR Gain per Frame agaignst H.263

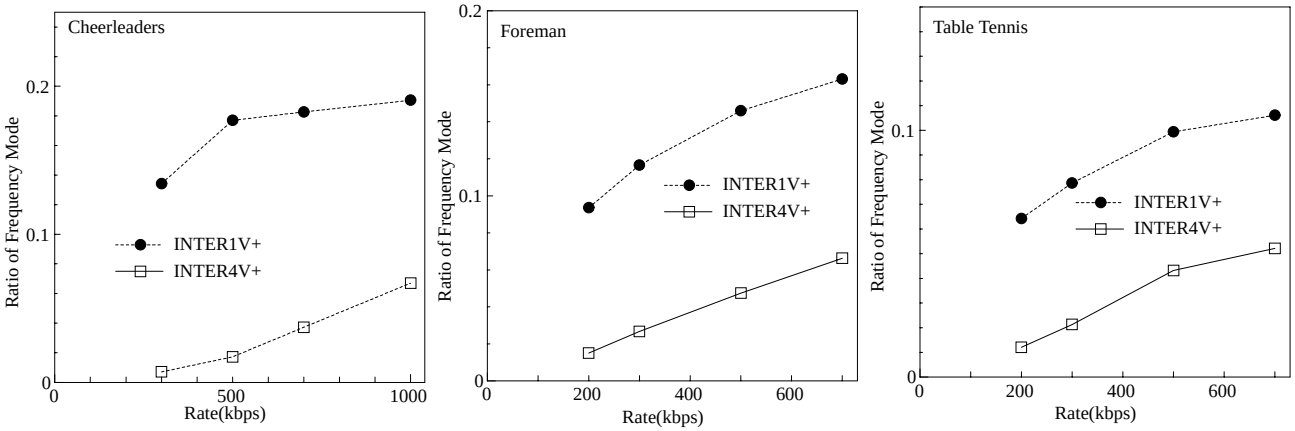

Fig. 4. The Ratio of INTER1V+ and INTER4V+

(a) Proposal method
26.4(dB) 500(kbps)

(b) Conventional method
25.6(dB) 500(kbps)

(a) Proposal method
34.0(dB) 500(kbps)

(b) Conventional method
32.8(dB) 500(kbps)

(a) Proposal method
25.6(dB) 200(kbps)

(b) Conventional method
24.7(dB) 200(kbps)

Fig. 5. Visual Degradation caused by Motion Mismatch in Decoded Frame