VIDEO CODING USING ADAPTIVE GLOBAL MC AND LOCAL AFFINE MC

Hirohisa Jozawa, Kazuto Kamikura, Kazuhisa Yanaka, and Hiroshi Watanabe
NTT Human Interface Laboratories
1-2356 Take, Yokosuka, Kanagawa 238-03, JAPAN
Phone: +81 468 59 2975 FAX: +81 468 59 2829
E-mail: `jozawa@nttvdt.hil.ntt.jp`

ABSTRACT

This paper describes an efficient video coding method using two-stage motion compensation (MC). The proposed MC method employs global MC (GMC) and overlapped block affine MC. GMC is adaptively turned on/off for each macroblock since GMC cannot predict all regions in an image. Simulation results show that the proposed coding method using two-stage MC significantly outperforms H. 263 for sequences with fast motion. Performance improvements in PSNR are about 3-4 dB over H. 263.

1 INTRODUCTION

ITU-T H. 263 [1] is a generic coding standard for low bit-rate video communication. H. 263 significantly outperforms H. 261 owing to its half pel motion compensation (MC), and four efficient option modes: unrestricted motion vector, syntax-based arithmetic coding, advanced prediction, and PB-frames. On the other hand, MPEG-4 is going to develop a low bit-rate video coding standard addressing three major functionalities:

- Content-based interactivity
- Improved coding efficiency
- Error resilience

Thus, MPEG-4 should achieve higher coding efficiency than H. 263 and provide new functionalities not addressed in existing standards. Especially, improved coding efficiency is required for applications such as video transmission over mobile networks or the Internet.

This paper addresses a high compression video coding algorithm based on H. 263. To improve coding efficiency of H. 263, *two-stage motion compensation* (MC) is proposed. The proposed MC consists of global motion compensation (GMC) and local motion compensation (LMC). GMC predicts camera motion such as panning, tilting, and zooming, while LMC predicts local motion of each macroblock. LMC employs an affine motion model in the context of the H. 263's overlapped block MC.

This paper is organized as follows. Section 2 describes the proposed GMC method, and Section 3 presents the LMC method employing overlapped block affine MC. In Section 4, the on/off decision for GMC is described. Finally, in Section 5, performance of the proposed coding method using two-stage MC is compared with that of original H. 263.

2 GLOBAL MOTION COMPENSATION

Global motion caused by camera motion such as panning, tilting, and zooming, is often observed in video sequences. Panning and tilting can be predicted using conventional local MC, however, all macroblocks have the same motion vector in this case. It is redundant to repeat the same motion vector for each macroblock. In addition, conventional MC cannot compensate for zooming. To solve these problems, GMC [2] is employed in the proposed coding method.

2.1 Global Motion Estimation

Local motion vector for each macroblock (16×16) is first estimated using block matching method. The motion search range is ± 127 pels. Then, three global motion parameters (H, V, Z) representing panning, tilting, and zooming are calculated for each macroblock using the following formulae:

$$H = \frac{V_{1x} - V_{2x}}{2} \quad (1)$$

$$V = \frac{V_{1y} - V_{2y}}{2} \quad (2)$$

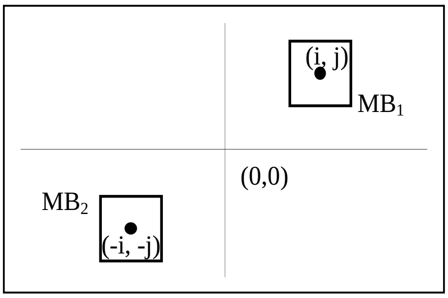

Figure 1 Block pair used for global motion estimation

$$Z = \frac{i(V_{1x} - V_{2x}) + j(V_{1y} - V_{2y})}{2(i^2 + j^2)} \qquad (3)$$

where (i, j) and $(-i, -j)$ are the center coordinates of the two macroblocks MB_1 and MB_2 (Figure 1). The displacement of MB_1 is (V_{1x}, V_{1y}), and that of MB_2 is (V_{2x}, V_{2y}). H and V are quantized to the nearest value of the multiple of two within [-126, 126]. Z is quantized to the nearest value of the multiple of 1/128 within [-31/128, 31/128]. Frequency of each parameter is accumulated over all macroblocks in a frame. Finally, for each parameter, the most frequent value is chosen as the estimated value of the frame.

2.2 Global Motion Compensated Prediction

Global motion compensated image is produced for the use in local MC. GMC is carried out using pel-wise prediction. The displacement vector of pixel (x, y) in the current frame is defined as follows:

$$v_x(x, y) = Zx + H \qquad (4)$$

$$v_y(x, y) = Zy + V \qquad (5)$$

where $v_x(x, y)$ is the horizontal component of the motion vector at (x, y), and $v_y(x, y)$ is the vertical one.

The predicted value for pixel (x, y) is obtained as the value at $(x + v_x(x, y), y + v_y(x, y))$ in the reference frame. When the motion vector components have fractional values, the predicted pixel value is calculated with bi-linear interpolation referring the surrounding four pixels in the reference frame. In addition, when the motion vector points outside the image, the edge pixel is used as prediction for the not-existing pixel as in the unrestricted motion vector mode of H. 263.

3 LOCAL MOTION COMPENSATION

Conventional MC uses translational motion model to describe horizontal and vertical motion of objects. However, moving objects have more complex motions, such as rotation and scaling. To improve prediction performance for sequences containing rotation and scaling of objects, we propose an overlapped block affine motion compensation method.

3.1 Local Motion Estimation Using Affine Motion Model

Horizontal and vertical scaling parameters (C_x, C_y) and rotation parameter θ are estimated in addition to translational motion parameters (t_x, t_y). The prediction unit is a 16×16 macroblock. The five affine parameters for each macroblock are estimated using two-layer search [3] to reduce the number of calculations.

The predicted value at (x, y) in the current frame is obtained as the value at (x', y') in the reference frame:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} C_x & 0 \\ 0 & C_y \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \qquad (6)$$

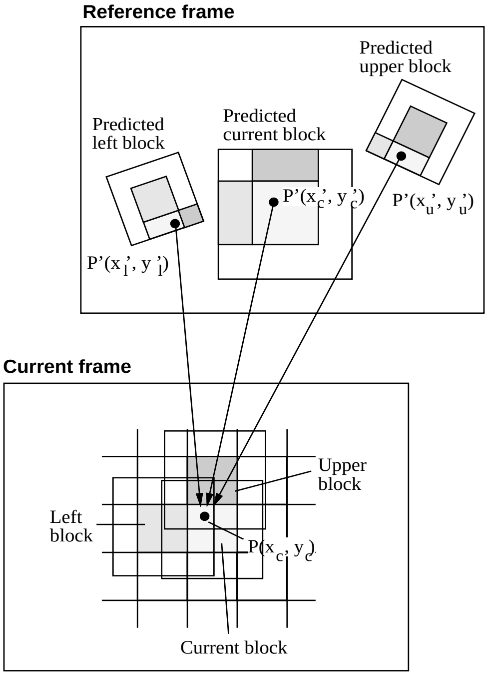

Figure 2 Overlapped Block Affine MC (OBAMC)

Since the x' and y' values can be real numbers, bi-linear interpolation is employed by referring to the four surrounding pixels. When the (x', y') is outside the image, the edge pixel is referred as in the unrestricted motion vector mode of H. 263.

3.2 Overlapped Block MC Prediction

Overlapped block motion compensation using an affine motion model (OBAMC) is employed in the context of the advanced prediction mode in H. 263. Each pixel in a predicted luminance block is a weighted sum of three prediction values, divided by 8. The three prediction values are obtained using three affine parameter sets of:

- the current macroblock,
- the macroblock to the left or right of the current one, and
- the macroblock above or below the current one.

For each pixel, the affine parameter sets of the current and the two nearest blocks are used.

Figure 2 shows an example of OBAMC. The affine parameters of the current block, the block at the left, and the block above are used, since the pixel at (x, y) is located in the upper left part of the block. The predicted value of the pixel at (x, y), $\hat{P}(x, y)$, is obtained using the following equation:

$$\hat{P}(x, y) = \{H_0 P'(x_c', y_c') + H_1 P'(x_l', y_l') + H_2 P'(x_u', y_u')\}/8 \qquad (7)$$

where P' represents the decoded reference image. The values x_c' and y_c' are calculated using Equation (6) and the affine parameters of the current block. The coordinates (x'_l, y'_l) are obtained using the affine parameters of the block to the left of the current one, while (x'_u, y'_u) are obtained using those of the block above the current one. H_0, H_1, and H_2 are the weighting matrices for the current, left, and above blocks defined in H. 263.

4 GMC ON/OFF DECISION

GMC does not perform well for regions where local motion is dominant. In addition, GMC is ineffective when there is no global motion or global motion estimation fails. Therefore, employing GMC for all macroblocks is inefficient in such cases.

To solve the problem above, local MC (OBAMC) is performed both for the reference frame with and without GMC. The Sum of Absolute Difference (SAD) values for GMC on ($GMCON$) and that for GMC off ($GMCOFF$) are calculated as follows:

$$SAD_{GMCON} = \sum_{MB} \left| X - \hat{X}_{GMCON} \right| \quad (8)$$

$$SAD_{GMCOFF} = \sum_{MB} \left| X - \hat{X}_{GMCOFF} \right| \quad (9)$$

where X is the pixel value of the original image to be coded. $\hat{X}_{GMCON}$ is the predicted value of X obtained using both GMC and LMC, while $\hat{X}_{GMCOFF}$ is that obtained using LMC but not using GMC. GMC is selected for the macroblock when $SAD_{GMCON} < SAD_{GMCOFF}$.

5 SIMULATION RESULTS

5.1 Simulation Conditions

To compare the coding performance of the proposed method with that of a reference method, software simulations were carried out. The reference method is the same as H. 263 except that the 8x8 MC was not used. In both methods, unrestricted motion vector mode was used, but syntax-based arithmetic coding and PB-frames mode were turned off. The quantization step-size regulation used in the simulation was the same as that in TMN5 [4]. The frame rate regulation was also the same as that used in TMN5.

A fixed length codeword of 20 bits is inserted into picture header to represent three global motion parameters H (7 bits), V (7 bits), and Z (6 bits). A codeword of 1 bit which signals the use of GMC for a macroblock is also inserted into macroblock layer. Translational parameters (t_x, t_y) for each macroblock are coded differentially using the VLC table defined in H. 263. On the other hand, scaling parameters (C_x, C_y) and rotation parameter Z are coded using variable length codes shown in Tables 1 and 2.

5.2 Results

Four test sequences used in MPEG-4 were examined in the simulations. "Foreman" and "Coast guard" were coded at 112 kb/s in CIF resolution (Y: 352×288, C: 176×144). "Stefan" and "Table tennis" were coded at 320 kb/s in SIF (Y: 352×240, C: 176×120).

Table 1 VLC table for C_x and C_y

C_x, C_y	Bit number	VLC code
0.9	2	10
1.0	1	0
1.1	2	11

Table 2 VLC table for θ

θ	Bit number	VLC code
-0.02π	2	10
0	1	0
0.02π	2	11

Figures 3, 4, 5, and 6 compare the Peak-to-Peak SNR (PSNR) of the decoded images. The results show that the proposed coding method using the two-stage MC significantly outperforms the reference method (H. 263) especially for "Foreman", "Coast guard", and "Stefan" which have fast camera motion. Maximum PSNR improvements are about 3 dB for "Foreman", 2 dB for "Coast guard", and 4 dB for "Stefan" in comparison with H. 263 without two-stage MC. For "Table tennis" sequence, the proposed MC outperforms the reference method by about 0.5 dB at around 50th frame where the camera zooms out.

The proposed method performs slightly worse than H. 263 for stationary scenes, e.g. the first half of "Foreman" and the latter half of "Coast guard" and "Table tennis." This is because GMC does not improve prediction performance for scenes without camera motion. The increased overheads signaling the use of GMC for each macroblock and the additional motion parameters representing rotation and scaling are also the reasons for the degraded performance.

6 CONCLUSION

GMC is effective for coding of video sequences that contain camera motion such as panning, tilting, and zooming. The proposed coding method using two-stage MC significantly outperforms H. 263 by 3-4 dB for sequences with fast camera motion.

The concept of GMC can be extended to the prediction of a certain object in an image. Future works will be focused on GMC for video objects or regions of interests. Improved global motion estimation, efficient representation of GMC on/off flags, and efficient coding of affine parameters will be also studied.

REFERENCES

[1] Draft ITU-T Recommendation H. 263, "Video coding for low bitrate communication," Dec. 1995.

[2] K. Kamikura and H. Watanabe, "Video coding for digital storage media using hierarchical intraframe scheme,"

SPIE Symposium on Visual Communications and Image Processing '90, vol. 1360, pp. 1540-1550, Oct. 1990.

[3] H. Jozawa, "Motion compensated video coding using rotation and scaling models," Picture Coding Symposium '96, Mar. 1996.

[4] ITU-T Study Group 15, Expert's Group on Very Low Bitrate Visual Telephony, "Video codec test model, TMN5," Jan. 1995.

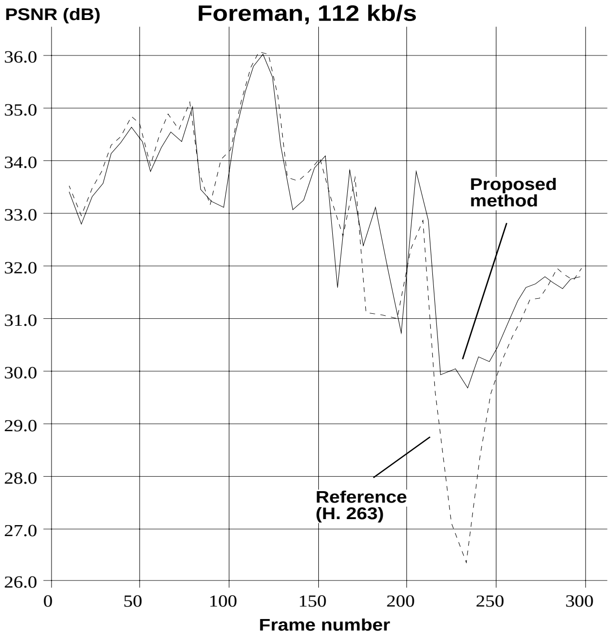

Figure 3 PSNR of decoded images.

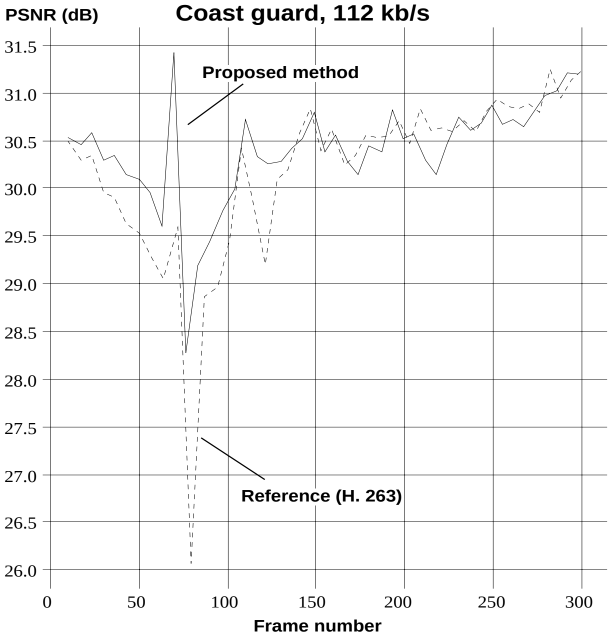

Figure 4 PSNR of decoded images.

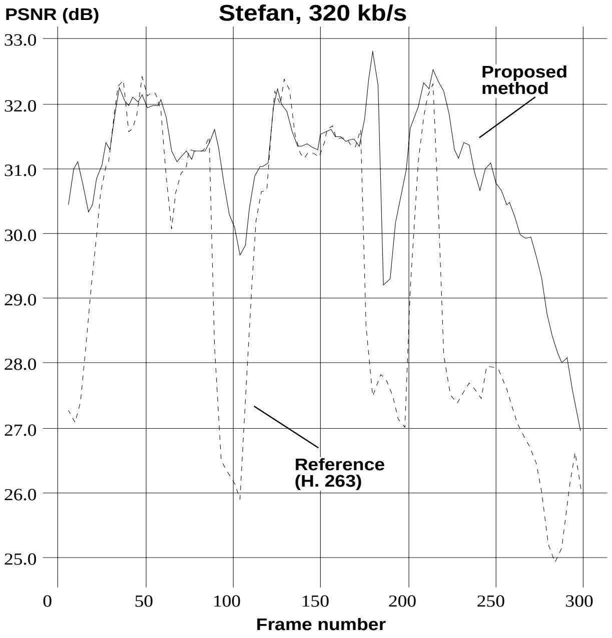

Figure 5 PSNR of decoded images.

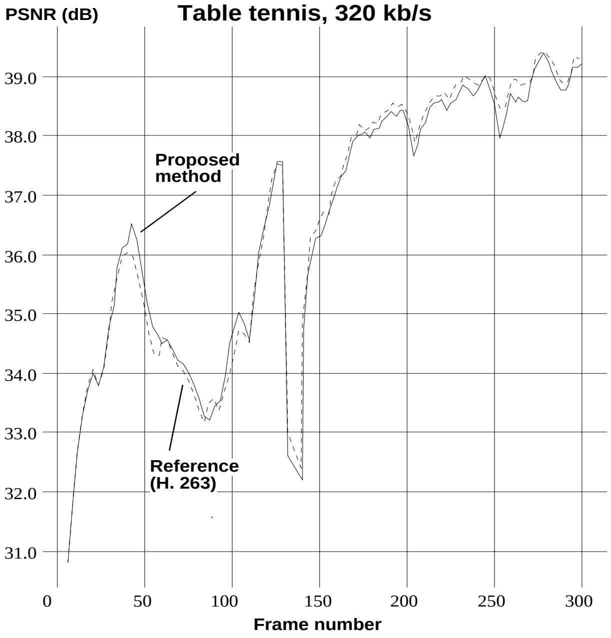

Figure 6 PSNR of decoded images.