USING GLOBAL MOTION COMPENSATION FOR EFFICIENT VIDEO CODING

Z. Ivanovski, Lj. Panovski

Institute of Electronics,
Faculty of Electrical Engineering,
University “SS. Cyril and Methodius” of Skopje,
Karpos II bb, P.O.Box 574,
91000 Skopje, Republic of Macedonia.

ABSTRACT

A way of introducing compensation of camera caused global motion into MPEG-2 is described in this paper. A video-coding method is proposed where global motion estimation of four parameters regarding horizontal and vertical scrolling, zooming and rotation is performed and a global-motion-compensated image is produced for use as a reference frame for MPEG-2 coding. The experimental results have shown an improved performance over MPEG-2 on sequences with pronounced camera zoom and rotation.

1. INTRODUCTION

Predictive coding is a significant aspect of contemporary video coding standards. Motion-compensated type of predictive coding is of utmost importance and is embedded into the MPEG series of video coding standards. Within all of these standards motion compensation is implemented at local level, i.e. it is used to predict of small blocks of size 16×16 pixels within an image, and translational motion is the only motion type that is being estimated and compensated. The motion compensation used in the MPEG standards has a half-pixel accuracy, which provides a 1dB-improvement in SNR over the case when one-pixel accuracy is used.

While MPEG type of coding uses translational motion models to describe the horizontal and vertical motion of objects to implement local motion compensation (LMC), moving objects can exhibit more complex motion, such as rotation, zooming and deformation. In order to treat more complex kinds of object motion non-standard LMC methods using first-order motion models [3], perspective models [4] and second-order motion models [5] have also been studied.

However, besides local motion due to moving objects, video sequences very often contain different kinds of global motion as well, caused mainly by intentional or unintentional camera motion. Possibilities to improve prediction performance through implementation of global motion compensation (GMC) have already been explored [1], and some schemes with adaptive selection of GMC have been proposed [2]. In this paper we combine GMC with standard MPEG-2-type LMC in order to increase the overall coding efficiency. A simple motion model is used to achieve global translation, zooming and rotation.

2. GLOBAL MOTION COMPENSATION

Intentional camera motion (pan, tilt, zoom) or unintentional camera motion (pan, tilt, rotation) introduce global motion effects into video sequences. Under certain conditions, panning and tilting can be approximately treated as horizontal and vertical scrolling, and the latter ones can be predicted using MPEG-2 LMC, i.e. using the translational motion model. However, when a sequence contains only scrolling and no local motion, all macroblocks have the same motion vector and GMC could prevent repeating the same motion vector for each macroblock. On the other hand, sequences containing zooming and/or rotation cannot be compensated with the conventional MPEG-2 LMC.

We employ GMC to either reduce unnecessary overhead in the case of global scrolling or to improve prediction performance in the case of camera zoom and/or rotation. The GMC method proposed here can compensate for global motions such as horizontal and vertical scrolling, zooming and rotation. Four parameters (K_1, K_2, K_X and K_Y) are first calculated from the translational motion vectors obtained between the current and the previous frames. Then, the GMC image is produced by applying the GM parameters to the local decoded version of the reference frame.

A. Global Motion Models

Arbitrary motion from point $\mathbf{P_1} = [X_1,Y_1,Z_1]^T$ to point $\mathbf{P_2} = [X_2,Y_2,Z_2]^T$ in 3D-space can be represented as combination of rotation around an axis through the origin of the coordinate system and translation [6]:

$$\mathbf{P_2} = \mathbf{RP_1} + \mathbf{T} \qquad (1)$$

The rotation matrix **R** is given by:

$$\mathbf{R} = \begin{bmatrix} n_1^2 + (1-n_1^2)c & n_1 n_2(1-c) - n_3 s & n_1 n_3(1-c) + n_2 s \\ n_1 n_2(1-c) + n_3 s & n_2^2 + (1-n_2^2)c & n_2 n_3(1-c) - n_1 s \\ n_1 n_3(1-c) - n_2 s & n_2 n_3(1-c) + n_1 s & n_3^2 + (1-n_3^2)c \end{bmatrix} \qquad (2)$$

where $c = \cos q$, $s = \sin q$, and q is the angle of rotation, while $n_i = \cos a_i$, $i = 1, 2, 3$ define the position of the axis of rotation.

T is the translation vector:

$$\mathbf{T} = [T_X,T_Y,T_Z]^T \qquad (3)$$

Let the camera axis coincide with the *Z*-axis and its image plane be at unit distance from the origin of the coordinate system. The 3D-motion given by (1) will then be projected onto the camera image plane as:

$$\mathbf{p_2} = \frac{Z_1}{Z_1 + T_Z}\mathbf{Rp_1} + \frac{1}{Z_1 + T_Z}\mathbf{T} \tag{4}$$

where $\mathbf{p_1} = [x_1, y_1, 1]^T$ and $\mathbf{p_2} = [x_2, y_2, 1]^T$ are the projections of $\mathbf{P_1}$ and $\mathbf{P_2}$, respectively.

In this paper we work with the following constraints:

$$a_1 = a_2 = 90^o \Rightarrow n_1 = n_2 = 0 \tag{5}$$

$$a_3 = 0^o \Rightarrow n_3 = 1 \tag{6}$$

i.e., we consider that rotation occurs around the camera axis and that zooming is equivalent to object translation along the camera axis.

By introducing constraints (5) and (6) in (2) and (4), we have:

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \frac{1}{Z_1 + T_Z}\left(Z_1 \begin{bmatrix} c & -s \\ s & c \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} T_X \\ T_Y \end{bmatrix} \right) \tag{7}$$

and after rearranging (7) in terms of x_1 and y_1:

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 1 + K_1 & K_2 \\ -K_2 & 1 + K_1 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} + \begin{bmatrix} K_X \\ K_Y \end{bmatrix} \tag{8}$$

where:

$$\begin{aligned} K_1 &= \frac{Z_1 + T_Z}{Z_1}\cos q - 1 \\ K_2 &= \frac{Z_1 + T_Z}{Z_1}\sin q \\ K_X &= -\frac{T_X \cos q + T_Y \sin q}{Z_1} \\ K_Y &= \frac{T_X \sin q - T_Y \cos q}{Z_1} \end{aligned} \tag{9}$$

We consider the four parameters defined in (9) to be global, i.e. applicable to all points of the imaged scene. It should be noted, however, that according to (4), the 2D-projection of a 3D-motion of a point $\mathbf{P_1}$ depends as well on its distance Z_1 from the image plane. So, we expect our results to be strictly valid only if the imaged scenes are planar and parallel to the image plane.

B. Global Motion Estimation

Step One: Local motion estimation using 16×16 block matching is carried out in a range appropriate for the motion extent between the current frame, f_2, and the reference one, f_1. By using the three-step search routine [7], a motion vectors field is obtained with half-pixel accuracy.

Consider two consecutive video frames $f_1(x,y)$ and $f_2(x,y)$. For the estimates of the motion vectors along the *x* and *y*-axis, V_X and V_Y, we have:

$$\begin{aligned} V_X &= i''K_1 + j''K_2 + K_X \\ V_Y &= -i''K_2 + j''K_1 + K_Y \end{aligned} \tag{10}$$

with:

$$\begin{aligned} i' &= i'' + V_X \\ j' &= j'' + V_y \end{aligned} \tag{11}$$

where (i', j') denotes the position of an object point in $f_1(x,y)$ and (i'', j'') denotes the position of the same point in $f_2(x,y)$.

Step Two: Pairs of macroblocks, MB_1 and MB_2, are used to calculate the four global motion parameters, (i_1, j_1) and (i_2, j_2) being the centers of MB_1 and MB_2. By use of (10), (11) and (8), we obtain the following formulas for that purpose:

$$\begin{aligned} K_1 &= \frac{(i_1 - i_2)(V_{1X} - V_{2X}) + (j_1 - j_2)(V_{1Y} - V_{2Y})}{(i_1 - i_2)^2 + (j_1 - j_2)^2} \\ K_2 &= \frac{(j_1 - j_2)(V_{1X} - V_{2X}) - (i_1 - i_2)(V_{1Y} - V_{2Y})}{(i_1 - i_2)^2 + (j_1 - j_2)^2} \\ K_X &= \frac{(V_{1X} + V_{2X}) - (i_1 + i_2)K_1 - (j_1 + j_2)K_2}{2} \\ K_X &= \frac{(V_{1Y} + V_{2Y}) - (j_1 + j_2)K_1 + (i_1 + i_2)K_2}{2} \end{aligned} \tag{12}$$

where, for each pair, V_{1X} and V_{1Y} denote the motion vectors of macroblock MB_1, and V_{2X} and V_{2Y} those of MB_2.

Block pairs are selected according to three patterns. First, calculations of (12) are carried out for all possible block pairs that are positioned symmetrically across the center of the frame, i.e. $i_2 = -i_1$ and $j_2 = -j_1$, and the frequencies of occurrence of the values of each of the four parameters K_1, K_2, K_X and K_Y are determined. Second, calculations are carried out for all possible block pairs satisfying $i_2 = i_1$ and $j_2 = -j_1$, the values of the four motion parameters are calculated and their frequencies of occurrence are determined. Third, all possible block pairs satisfying $i_2 = -i_1$ and $j_2 = j_1$ are taken into account, the values of the four motion parameters are calculated and their frequencies of occurrence are determined. The calculations of (12) are done by rounding the values of parameters K_1 and K_2 to the nearest multiple of 1/512 in the range of 0 - 0.4, while the values of K_X and K_Y are rounded to the nearest multiple of 1/10 in the range of ± 127.

Step Three: The three frequency sets for the values of each of the four motion parameters, obtained using the three patterns for macroblocks matching, are merged into one and the most frequently occurring value for each parameter is chosen as its estimate. In this way the influence of local motion upon the estimated values $\hat{K}_1, \hat{K}_2, \hat{K}_X$ and $\hat{K}_Y$ of parameters K_1, K_2, K_X and K_Y is avoided and good estimates of global motion parameters are obtained.

C. Calculation of Global Motion-Compensated Frames

Using the estimated four global parameters, a global motion-compensated frame, f_{2GMC}, is constructed from the locally decoded reference frame, f_1. The intensity for each pixel of f_{2GMC} is determined from the intensity of the corresponding pixel of f_1, the correspondence being determined through (10) and (11). When necessary, linear interpolation of the intensity of f_1 is used.

By replacing constraint (6) with:

$$a_3 = 90^o \Rightarrow n_3 = 0 \qquad (13)$$

one can readily turn the proposed global motion estimation and compensation procedure into a procedure similar to the one proposed and tested in [2], which provides for horizontal and vertical scrolling and zooming only.

1. LOCAL MOTION COMPENSATION

Local motion compensation is carried out according to MPEG-2, with f_2 as the current frame and f_{2GMC} as the reference frame.

4. EXPERIMENTAL RESULTS

Tests were carried out with gray-scale video sequences containing different kinds of motion. To serve as a reference, the MPEG-2 algorithm was implemented. The stream structure was IPPP…, i.e. B frames were not considered. Frame size was 384×288 pixels. Pixel intensity was 8-bits/pixel.

In situations when the motion is mainly zooming and rotation, the results show an improvement in SNR of approximately 1 dB with respect to MPEG-2, and an improvement of approximately 0.5 dB with respect to the case when the model with constraint (13) is used instead of (6), Figure 1.

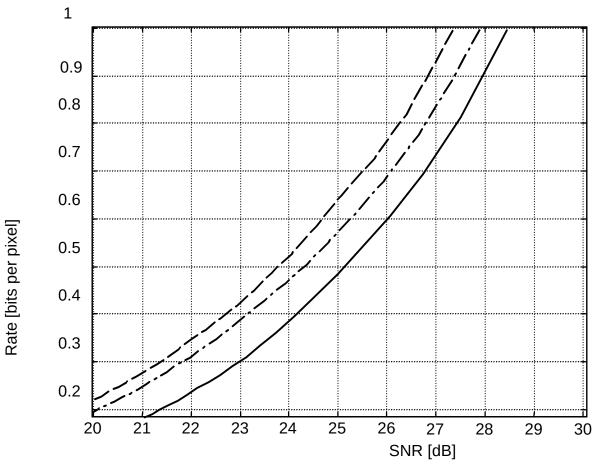

Figure 1 Rate-distortion function

– – – – MPEG-2

–·–·– GMC+MPEG-2, constraint (13) instead of (6)

——— GMC+MPEG-2

As can be seen in Figure 2, better motion compensation achieved as a result of the implementation of GMC, results in smaller prediction errors and consequently in a larger number of non-coded blocks (blocks for which all DCT-coefficients are zero). Consequently, the total number of bits used to code the DCT-coefficients is reduced, as is shown in Figure 3. As one can see comparing the decoded pictures in Figure 4, the visual appearance of the GMC picture is also better.

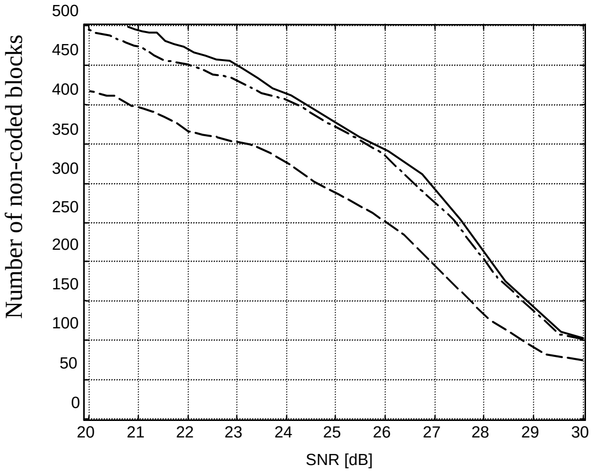

Figure 2 Number of non-coded blocks vs. distorsion

– – – – MPEG-2

–·–·– GMC+MPEG-2, constraint (13) instead of (6)

——— GMC+MPEG-2

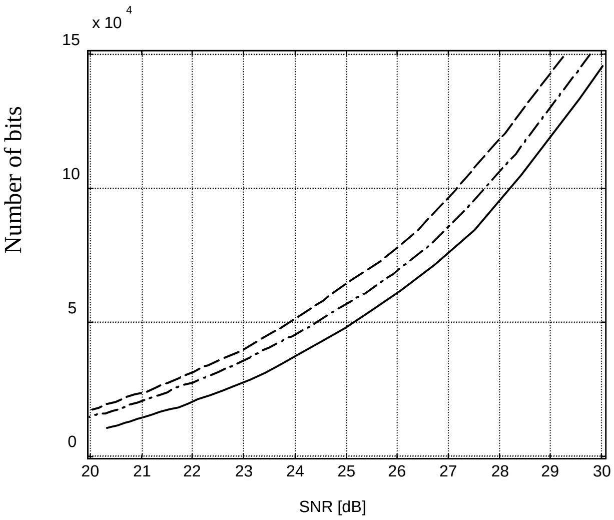

Figure 3 Total number of bits for coding of the DCT-coefficients

– – – – MPEG-2

–·–·– GMC+MPEG-2, constraint (13) instead of (6)

——— GMC+MPEG-2

5. CONCLUSIONS

A video-coding algorithm using GMC in combination with MPEG-2 was developed and tested on a number of video sequences. MPEG-2 was used as reference. The results have shown an improved performance over MPEG-2 on sequences with pronounced camera zoom and rotation. Thus, despite the fact that GMC was based on a simple four-parameter model, describing horizontal and vertical scrolling, zooming and rotation, the potentials of including GMC in MPEG-2 have been clearly demonstrated.

a) Original frame

b) Decoded MPEG-2 frame

c) Decoded GMC frame

Figure 4 Original and decoded frames

REFERENCES:

[1] M. Hoetter, "Differential Estimation of the Global Motion Parameters Zoom and Pan", *Signal Processing*, vol. 16, No.3, pp. 249-265, 1989.

[2] H. Jozawa, K. Kamikura, A. Sagata, H. Kotera, J. Watanabe, "Two-Stage Motion Compensation Using Adaptive Global MC and Local Affine MC", *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 7, No.1, Feb. 1997.

[3] H. Jozawa, "Motion Compensated Video Coding Using Rotation and Scaling Models", *Picture Coding Symposium '96*, pp. 309-312, March 1996.

[4] R.Y.Tsai and T.S. Huang, "Estimation Three-Dimensional Motion Parameters of a Rigid Planar Patch", *IEEE Trans.Acoust., Speech, Signal Processing*, Vol. ASSP-29, pp. 1152-1157, Dec. 1981.

[5] Y. Nakaya, H. Harashima, "Motion Compensation Based on Spatial Transformations", *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 4, pp. 339-356, June 1994.

[6] R.Y. Tsai, T.S. Huang, "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 6, No. 1, Jan. 1984.

[7] V. Bhaskaran, K. Konstantinides, "Image and Video Compression Standards", *Kluwer Academic Publishers,* Boston, 1997.