BLOCK-MATCHING TRANSLATION AND ZOOM MOTION-COMPENSATED PREDICTION

Ka-Man Wong, Lai-Man Po, and Kwok-Wai Cheung[#]

kmwong@ee.cityu.edu.hk, eelmpo@cityu.edu.hk, kwcheung@chuhai.edu.hk
Department of Electronic Engineering, City University of Hong Kong, Hong Kong
Department of Computer Science, Chu Hai College of Higher Education, Hong Kong [#]

Abstract - In modern video coding standards, motion compensated prediction (MCP) plays a key role to achieve video compression efficiency. Most of them make use of block matching techniques and assume the motions are pure translational. Attempts toward a more general motion model are usually too complex to be practical in near future. In this paper, a new Block-Matching Translation and Zoom Motion-Compensated Prediction (BTZMP) is proposed to extend the pure translational model to a more general model with zooming. It adopts the camera zooming and object motions that becomes zooming while projected on video frames. Experimental results show that BTZMP can give prediction gain up to 2.25dB for various sequences compared to conventional block-matching MCP. BTZMP can also be incorporated with multiple reference frames technique to give extra improvement, evidentially by the prediction gain ranging from 2.03 to 3.68dB in the empirical simulations.

Index Terms – Video Coding, Motion Compensated Prediction, Translation and Zoom Motion.

I. INTRODUCTION

Block-based motion-compensated prediction (MCP) is the core concept contributing to the high coding efficiency of the modern video coding schemes. To apply this, a frame is divided into non-overlapping blocks. Then, motion estimation is carried out to find a prediction for each block based on the data in previously encoded frame. A residue block is created by subtracting the prediction from the current block. Only the residue block and the data (motion vector) required to reproducing the prediction are encoded. The compression performance highly depends on the prediction accuracy. In all video standards like H.26X and MPEG-X, MCP is based on a translation motion model as shown in Figure 1(a) in which the video frames consists of rigid objects. The object motion is limited to translation only from frame to frame. Deviation from this translation motion model is encoded in the residue block. Thus, the accuracy of this model in representing the real motion greatly influences the coding efficiency. The problem with this 2D frame based motion model is that it has the following assumptions on the objects in real life scene from which we capture the video frame.

1. Objects are rigid without rotation.
2. Objects are moving in a 2D plane perpendicular to the video camera.

Practically, only the first assumption is close to reality (at least when small block size is used in MCP). However, objects may move in any direction in the 3D world. There may have some objects moving towards the camera while other objects moving away from the camera. In this case, the motion is better modeled by a more general translation and zoom motion model as illustrated in Figure 1(b).

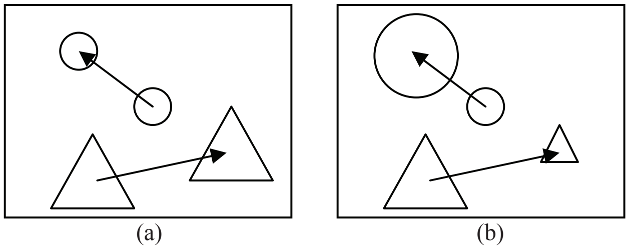

Fig. 1 2D Frame based motion model (a) Translation motion model (b) Translation and zoom motion model

To improve the coding efficiency, one approach is to increase the prediction accuracy assuming the same translation motion model. For instance, the latest standard H.264 (aka MPEG-4 part 10) [1-4] employs multiple reference frames (MRF), variable block size (VBS) to improve the motion compensated prediction accuracy. Rate-distortion optimization (RDO) [5] is used to optimize the tradeoff between the increased accuracy and the associated side bits introduced by these techniques.

Another approach is to use a more general affine model [9-11] involving translation, rotation and zoom motion to increase the prediction accuracy. In [9], affine parameter sets are estimated and multiple “wrapped” frames are generated based on the parameter sets as references and the affine parameters are transmitted if the block in a wrapped frame is selected as the best prediction candidate. In [10], motion vectors of other blocks are used to make an affine transformed block as a searching candidate. It focuses on local complex general motion. In [11], the global camera motion is estimated to form a parametric model to assist the motion compensated prediction. As these methods require affine parameter estimation, the complexity limits the deployment of these motion models in practical application.

In this paper, a motion-compensated prediction model based on the translation and zoom motion model as shown in Figure 1(b) is studied. By combining translation and zoom motion components, the proposed MCP can better suit the real motion. In addition, by using block-based implementation, the new MCP model can be easily deployed in the existing video coding framework. This paper presents the promising results of our preliminary study on applying this MCP model to video coding. It should be noted that the major difference between this work and the methods in [9-11] is that our method does not require affine parameter estimation for generating additional reference frames. Zoomed reference frames are indicated as the index of selected zoomed frame similar to MRF, thus no side information about zoom scale is needed to be transmitted. The rest of this paper is organized as follows. In section 2, block-matching model for MCP will be explained and the proposed model will be introduced in a way to utilize the zooming property. Section 3 provides experimental results for assessing the achievable gains with the proposed technique. Finally, conclusions will be drawn in section 4.

978-1-4244-4291-1/09/$25.00 ©2009 IEEE

II. Block-Matching Translation and Zoom Motion-Compensated Prediction

2.1. Translation motion based motion Estimation

In block matching motion estimation with multiple reference frames(MRF)

$$\{F_{t-1}, F_{t-2}, F_{t-3}, \ldots , F_{t-P}\} , \quad (1)$$

F_{t-d} is the frame with delay d and $F(a,b)$ denotes the intensity of a pixel located at (a,b). The motion for a NxN block in the current frame F_t is represented by the motion vector

$$MV(u,v,d) \quad (2)$$

using the minimum sum of absolute difference (SAD) criterion

$$SAD(x,y,u,v,d) = \sum_{i=1}^{N}\sum_{j=1}^{N}\left|F_t(x+i,y+j) - F_{t-d}(x+i+u,y+j+v)\right| \quad (3)$$

between the block at (x,y) in the current frame F_t and the matched block at (u,v) in the reference frame F_{t-d} .

Since MRF use frames with different time delays, the weakness of single reference frame in case of temporary occlusions and periodic deformations can be resolved by selecting frames at other time that does not have such problems. Fig. 2 shows an example of MRF in case of temporary occlusion and periodic deformation.

2.2. Translation and Zoom motion estimation

To handle a more realistic motion model shown in Fig. 1(b), an additional zoom axis for zoomed frames is introduced and the set of reference frames becomes

$$\{F_{t-1,0} , F_{t-2,0} , F_{t-3,0} , \ldots , F_{t-P,0}\};$$

$$\{F_{t-1,-Z} , \ldots, F_{t-1,-1} , F_{t-1,0} , F_{t-1,+1} , \ldots , F_{t-1,+Z}\} \quad (4)$$

where Z is the maximum number of zooming levels for zoom-in and zoom-out, the range ($-Z, +Z$) represents the zooming range used. The block matching criterion becomes:

$$SAD(x,y,u,v,d,s) = \sum_{i=1}^{N}\sum_{j=1}^{N}\left|F_{t,0}(x+i,y+j) - F_{t-d,s}(x+i+u,y+j+v)\right| \quad (5)$$

where $F_{t-d,\, s}$ is the frame with delay d and with zoom level s that gives minimum SAD. The motion vector of a block is written as $MV(u, v, d, s)$. Fig. 3 shows an example for BTZMP. The time axis shows the motions of two objects. The triangular object moves right and zooms in; and the circular object move left and zooms out. The zoom axis shows the zoomed frames. As the zooming is not periodic, MRF does not have the most appropriate frames for motion estimation. However, BTZMP zoom in frames and zoom out frames also provide suitable frames for motion estimation. In this case BTZMP gives a better predicted block for motion compensated prediction. Thus the prediction accuracy increased. Although providing zoomed frames for every zoomed object will make the motion prediction gain higher, it is not realistic to provide too many zoomed frames for the motion estimation. It will increase the memory requirement, and the computational requirement will also be higher. So, the number of zoomed frames and the zoom scales provided should also be considered to make BTZMP robust and efficient.

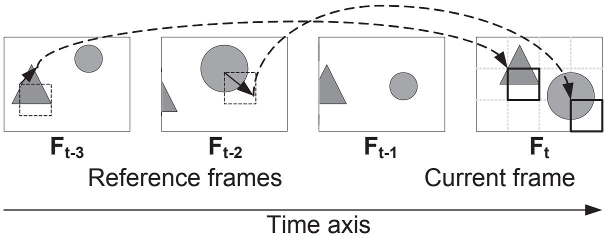

Fig. 2 An example of MRF where the triangular object moving out of the frame for two frames and back; and the circular object moving downward and resizes periodically.

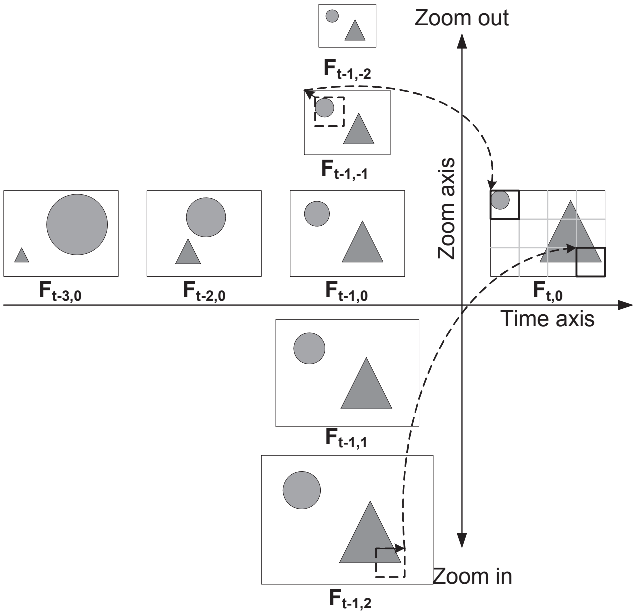

Fig. 3 An example of block motion estimation with MRF and BTZMP where the triangular object moving right and zooming in with time, and the circular object moving left and zooming out with time.

2.3 The zooming scheme

BTZMP uses several zoomed decoded frames which are obtained by a variable zoom factor. It should be noted that the proposed method does not have the procedure to detect the zoom motion before the generation of zoomed frames like [9-11]. In such design, number of zoomed frames and the zooming factors for these frames should be pre-determined. Fixed zoom factors with regular step size for the zoomed frames are used due to the simplicity of encoding and decoding. Thus the factors to be determined become the zooming step size and the number of zoom levels.

Small zooming step size is used. Small zooming step size with fractional zooming scales adapts to the zooming property of real video sequence. Similar to the center biased property found in Fast Block Matching Algorithm (FBMA) [12], MVs are concentrated within a small region. It is also found that zoom motion in video sequence also concentrate within a small zoomed region. Since the biased property found, the suitable number of zoom levels based on the zoom distribution can be determined.

2.4 Zoomed frames generation

Interpolation techniques are used to generate the zoomed frames for motion compensated prediction. The encoder and the decoder should use the same interpolation algorithm for generating the zoomed frames because different interpolation algorithms will give different residue blocks. Bilinear interpolation is selected since arbitrary scale zooming can be incorporated directly and the interpolation can be implemented efficiently. In bilinear interpolation, the weighted average of four surrounding points in the original resolution will be taken. Fig. 4 illustrates the arbitrary scale bi-linear interpolation.

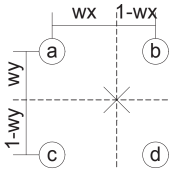

Fig. 4 Bi-linear interpolation

The bi-linear interpolated point X can be calculated by the following equation, suppose *a*, *b*, *c* and *d* are the pixels in the original image,

$$p(X) = (1-w_x)(1-w_y)a + (1-w_y)w_x b + (1-w_x)w_y c + w_x w_y d \cdot \quad (6)$$

The interpolated points can be calculated directly.

III. EXPERIMENTAL RESULTS AND ANALYSIS

The motion compensated prediction gain of BTZMP will be presented via several experiments. Firstly, prediction gain of various video sequences will be shown to provide a picture of the improvement of BTZMP and a suitable zooming step size will also be selected. Secondly, distribution of zoom motion vectors will be investigated and the effective number of zoom levels will be selected. Finally, additional prediction gain of combining BTZMP and MRF will also be given by experimental results.

4.1. Experiment setup

In the experiments, sequences *akiyo*, *foreman*, *mobile* and *stefan* are used, they are at CIF resolution and 300 frames in 30fps frame rate. The macroblock size is fixed at 16x16 and the search windows size is set at ±16 and exhaustive search is used within the search window. Sum of absolute difference (SAD) is used for the block distortion analysis. Since these tests are used to explain the potential of BTZMP improves the motion compensated prediction, original frames and zoomed original frames are used for the BTZMP. It is different from a real video codec that use reconstructed frames for motion compensated prediction, so the bit rate is not available at this stage.

TABLE 1 List of zooming scales in zoom step analysis

Step size	Zoom scales of zoomed frames
1/16	{5/16, …, 15/16, 16/16, 17/16, …, 26/16}
1/32	{21/32, …, 31/32, 32/32, 33/32, …, 42/32}
1/64	{53/64, …, 63/64, 64/64, 65/64, …, 74/64}
1/128	{117/128, …, 127/128, 128/128, 129/128, …, 138/128}

4.2. Prediction accuracy improvement of BTZMP

To show the improvement of BTZMP, several zoom steps from 1/16 to 1/128 are tested and we use ±11 zoom level in this test thus 23 frames are used for each motion prediction, the detailed zoom scales used are tabulated in TABLE 1.

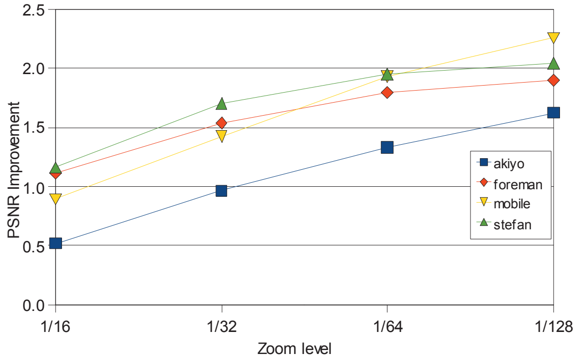

Fig. 5 PSNR improvement of BTZMP with various zooming step sizes with bi-linear interpolation

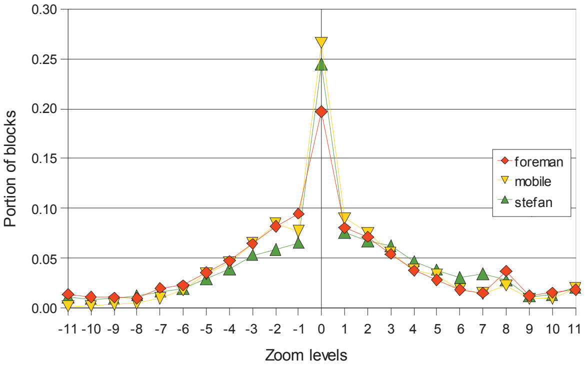

Fig. 6 Zoom motion distribution with zooming step size of 1/128

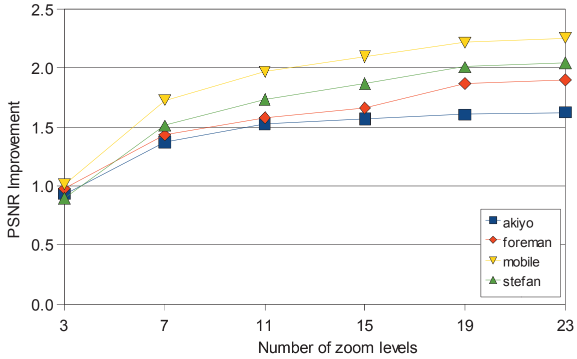

Fig. 7 PSNR improvement of BTZMP with various number of zooming steps

Fig. 5 shows the PSNR improvement of BTZMP. The labels 1/16, 1/32, 1/64, 1/128 relate to the various zooming step size and the improvement is relative to the ordinary block motion compensated prediction without zooming. It can observed that BTZMP significantly improved the estimation accuracy, and also the improvement keep growing with smaller zooming steps are used.

We can also see that improvement is up to 1.62-2.25dB for these sequences. The improvement for *akiyo* sequence is significantly lower because the sequence has a static background that has no zooming. Among the zooming step sizes, 1/128 step size gives the highest PSNR improvement that indicates there may have many slight zooming in these sequences. For CIF resolution, 1/128 step size can be selected as the BTZMP zooming step size, since there is no difference in computational complexity and memory requirement between different step sizes.

4.3. Number of zoom levels

Although the estimation can be further improved by adding more zoom levels and using smaller zoom step size, it enlarges the memory requirement and increases the computation complexity, an effective number of levels should be found to obtain the most significant improvement. The analysis first find the major range of zoom motions and then various number of zoom levels will be tested. Fig. 6 shows the distribution of zoom motion of *foreman*, *mobile* and *stefan* with zooming step size of 1/128. It can be seen that more than 75% of blocks are selected within ±5 zoom levels. In *akiyo* sequence most blocks are static and more than 90% of blocks selected the frame without zooming, it is not shown in Fig. 6.

To investigate the impact on using fewer zoom levels, an experiment for using various number of zoom level is also provided. Fig. 7 shows an analysis with 3-23 zoom levels, 11 levels can give improvement close to the improvement with 23 zoom levels. From fig. 6 and 7, it can be suggested that 11 levels can make BTZMP to be robust and efficient.

4.4. Combine with multiple reference frames

As mentioned in introduction, BTZMP is to provide a new aspect to existing techniques for improving prediction accuracy. In this part, analysis on BTZMP compared to and combined with MRF is provided. TABLE 2 tabulates the PSNR results of these techniques. It is clear that BTZMP is very competitive with multiple reference frames. It can also be seen that BTZMP+MRF always give the highest prediction improvement. BTZMP+MRF with 11 zoom levels and 11 reference frames can give a prediction gain up to 3.22dB compared to conventional method, and it further go up to 3.68dB while 23 zoom levels and 23 reference frame are used. The combination of BTZMP and MRF gives about additional improvement about 0.23 - 1.34dB in sequences compare to either MRF or BTZMP. From these observations it can be seen that these techniques can be combined very well and make an extra gain in prediction accuracy improvement.

TABLE 2 PSNR comparison of BTZMP and multiple reference frames

	akiyo	foreman	mobile	stefan
Ordinary full pixel MCP	43.1343	32.3451	24.8408	24.9212
MRF, 11 references	44.1688	34.3512	27.0406	26.1204
BTZMP, 11 zoom steps	44.6603	33.9251	26.8091	26.6548
BTZMP + MRF, 11 zoom steps, 11 references	**45.0234**	**35.1572**	**28.0584**	**27.1619**
MRF, 23 references	44.2976	34.7183	27.1735	26.0930
BTZMP, 23 zoom steps	44.7591	34.2465	27.0971	26.9669
BTZMP + MRF, 23 zoom steps, 23 references	**45.1652**	**35.7395**	**28.5162**	**27.5493**

IV. Conclusions

In this study, a new technique for motion estimation that exploits the zooming effect in video sequences is presented. The conventional pure translational motion model is extended by a zooming axis that covers zooming effect from real world motions projected on the video frame.

BTZMP is proposed for this translation and zoom motion model by using multiple zoomed reference frames. The proposed method does not require detection of zooming effect between current frame and reference frame. It generates the zoomed frame within a small zoom range and fixed zooming scales.

Analysis found that using step size of 1/128 gives prediction improvement up to 2.25dB with CIF video sequences. Analysis also found that using 11 zooming step with 1/128 step size already cover more than 75% of zooming blocks and it also give prediction gain close to the maximum improvement. These figures show that the prediction gain achievable with BTZMP is significant.

BTZMP is also found very competitive with MRF. It improves motion prediction in a different aspect, combination of BTZMP and MRF take the prediction improvement to another stage. It can be concluded that BTZMP will be a very promising and also achievable technique for future multimedia standards.

ACKNOWLEDGMENT

The work described in this paper was substantially supported by a grant from Applied R & D Funding (ARD) of City University of Hong Kong [Project No. 9667014] and also partially supported by a contrast research project for "Future Multimedia Standard" from ASTRI (Hong Kong Applied Science and Technology Research Institute Company Limited).

References

[1] A. Luthra, G. J. Sullivan, and T. Wiegand, "Introduction to the special issue on the H.264/AVC video coding standard," IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 7, pp. 557–559, July 2003.

[2] I. E. G. Richardson, H.264 and MPEG-4. New York: Wiley, 2003.

[3] T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 7, pp. 560–576, Jul. 2003.

[4] "Coding of Audio-Visual Objects – Part 2: Visual," MPEG-4 Visual Version 1, ISO/IEC 14 496–2, 1999.

[5] Gary J. Sullivan and Thomas Wiegand, "Rate-Distortion Optimization for Video Compression," IEEE Signal Processing Magazine, Vol. 15, No. 6, pp. 74-90, November 1998.

[6] "Video Coding for Low Bitrate Communication,", ITU-T Recommendation H.263 Version 3 (H.263++), 2000.

[7] T. Wiegand, X. Zang, and B. Girod, "Long-term memory motion-compensated prediction," IEEE Trans. Circuits Syst. Video Technol., vol. 9, no. 1, pp. 70–84, Feb. 1999.

[8] G. J. Sullivan and R. L. Baker, "Rate-distortion optimized motion compensation for video compression using fixed or variable size blocks," in Proc. GLOBECOM'91, Phoenix, AZ, Dec. 1991, pp. 85–90.

[9] Thomas Wiegand, Eckehard Steinbach, and Bernd Girod, "Affine Multi-Picture Motion-Compensated Prediction," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 15, No. 2, pp. 197-209, February 2005.

[10] R. C. Kordasiewicz, M. D. Gallant, S. Shirani, "Affine Motion Prediction Based on Translational Motion Vectors," IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 10, pp. 1388 - 1394, Oct. 2007.

[11] F. Lari, A. Zakhor, "Video compression based on camera motion," In Conference Record of The Twenty-Sixth Asilomar Conference on Signals, Systems and Computers, 1992. vol. 2, pp. 1004-1010. 26-28 Oct. 1992.

[12] C. H. Cheung and L. M. Po, "A Novel Cross-Diamond Search Algorithm for Fast Block Motion Estimation," IEEE Trans. on Circuits and Systems for Video Technology, Vol.12, No. 12, pp. 1168-1177, Dec. 2002.

[13] W. K. Pratt, Digital Image Processing, New York: Wiley, 1991.

[14] B. D. Lathi, Signal Processing and Linear Systems, Berkeley Cambridge Press, 1998.