G-013

Incorporating Zoom Motion Estimation in Video Compression: A More General Approach for Motion Compensated Predictive Coding

Morsalin Uz Zoha and **Jun Ohya**
Waseda University, 1-3-10 Nishi-Waseda Shinjuku-ku, Tokyo, Japan
E_mail: zoha@fuji.waseda.jp, ohya@waseda.jp

Abstract: In order to remove temporal redundancy between successive image frames, most of the video coding standards (e.g. ITU-T H.261, ISO/IEC standard MPEG-2, the latest video coding standard H.264/AVC through the joint efforts of ITU/T and ISO/IEC or proposed HEVC) use block based motion estimation technique. But conventional block based motion estimation techniques have a major drawback that these are limited to detect only the parallel translation of objects' motion, i.e. horizontal and vertical motions. Although many attempts have been found in literature to resolve the existences of other geometric transformations such that zoom, rotation, tilting motion of objects, these are not accepted by industry for real-time video CODEC because of their computational complexities. However, in video sequences for example digital movie videos, zoom motion often occurs to get images on 2D image surface from 3D data. To address this issue, this paper demonstrates the feasibility of incorporating a newly proposed zoom motion estimation algorithm in the latest video coding standard AVC/H.264 for improving digital video compression. Executing implementation using H.264 standard, the experimental results will justify the claimed performance of the proposed zoom motion estimation technique.

Keywords: Zoom Motion Estimation, H.264, Block-based Motion Estimation, Video Coding.

1. Introduction:

In inter frame coding, only residual block data i.e. a difference value set between the macroblock and its best correlated block, is encoded and transferred with overhead information (motion vector and, block size mode). Therefore, we know that motion estimation and compensation play a key role in transmission and storage of video signals at reduced bit rates because, the main subject of the encoding is the residual data generated by motion estimation. For this reason, many approaches have been proposed for efficient motion estimation [1]-[5]. For example, compensation for an object with complex texture can be achieved by dividing a large block into small square or rectangular blocks from 16x16 to 4x4 using the Variable Block Size (VBS) structure coding [1], [2] adopted in H.264/AVC. The sub-pixel motion estimation technique [3], [4] in H.264/AVC greatly increases motion vector accuracy (up to 4 times) by interpolation of the reference frame.

Block-based conventional motion estimation can efficiently remove temporal redundancy, but it has limitation that it is only able to detect horizontal and vertical object motion. However, an object with geometric transformations produces more complex motion in the image plane, so block-based conventional motion estimation cannot achieve effective compensation using the parallel translation motion model. Actually, real image sequences represent a three dimensional real world on a 2D image surface and, thus, zoom motion commonly occurs in image sequences. Usually, zoom motion occurs for two major reasons: First, objects can move toward or away from a camera and, second, the camera can zoom in or out on objects. Using zoom motion, the size of objects within an image plane is changed. If these characteristics are reflected in the ME procedure, the compression efficiency of a coded video image will be greatly increased.

In literature, there have been many attempts to improve the motion estimation efficiency by using algorithms for detecting zoom motion. But these have some sort of limitations to be used in real-time video CODEC. In this paper the feasibility of incorporating a newly proposed zoom motion estimation technique termed as BTZMP (Block-Matching Translation and Zoom Motion-Compensated Prediction) proposed by Wong et. al. [6] is examined to be used in the latest video coding standard H.264 to get a compression efficient real-time encoder. The rest of the paper is organized as follows: BTZMP has been overviewed in section 2 and section 3 provides the experimental results. This paper concludes in Section 4 with concluding remarks.

2. Overview of BTZMP:

In block matching motion estimation with multiple reference frames, the motion for a N x N block in the current frame F_t is represented by the motion vector MV (u,v,d) that can be calculated using the minimum SAD (Sum of Absolute Difference) criterion.

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

Here, F_{t-d} is the frame with delay d and F(x, y) denotes the intensity of a pixel located at (x, y).

Since multiple reference frames 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.

To get a more realistic motion model, an additional zoom axis for zoomed frames is introduced in BTZMP. So, the block matching criterion becomes

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

Here, $F_{t-d,z}$ is the frame with delay d with zoom level z for the range (-Z, +Z) that provides minimum SAD.

3. Experimental Results:

BTZMP technique's claimed accomplishment was that BTZMP would be a very promising and achievable technique for future multimedia standards. In this section, by incorporating this zoom motion estimation algorithm in H.264, its claimed performance is justified. In the experiment YUV format CIF (352 x 288) sized video sequences are used and the features enabled in the Main Profile of H.264 are only considered where the frame rate of H.264/AVC is 30 fps, slice mode is OFF, rate control is OFF, RDO is OFF, hadamard is OFF, search range is 32, search is not restricted, symbol mode is CABAC, no partition mode is used and out file mode is Annex B. H.264 is implemented with H.264 reference model version JM6.1e [7].

Table I lists the PSNR comparison among the translation-based motion compensated prediction (MCP), MCP with multiple reference frames and BTZMP for a bit rate of 512kbps.

It is obvious from the table that BTZMP always makes an extra gain in prediction accuracy in terms of PSNR, which is in accordance with BTZMP's claim [6].

TABLE I
PSNR COMPARISON

Test Sequence	MCP	MCP with multiple ref. frames	BTZMP
BRIDGE (CLOSE)	36.75	39.21	41.29
BRIDGE (FAR)	34.93	36.84	39.40
MISS AMERICA	40.21	42.09	43.61
WATERFALL	28.43	31.65	33.98
NEWS	41.25	43.78	44.94
HIGHWAY	30.11	32.29	35.79

Conclusion: It is a fact that video compression technology has changed our life in the past by introducing DVD, Digital TV, iPod, Blu-Ray, YouTube etc. In future we can expect by popularizing 3D TV, Super Hi-Vision, Super Reality, video coding will play a major role. Not only bringing change in entertainment of our life, video coding can be used in a more serious issue- the field of healthcare- to develop a real-time telemedicine system. However, what else the purposes, most of the real-time video CODECs limit their practical usage of affine transformation of objects rather real-time CODECs only deal with composition of translation and rotation of moving rigid-body objects. Here in the paper, the feasibility is examined for incorporation of zoom motion in a real-time encoder to get a more general affine model for motion compensated predictive coding.

References:

[1] Ahmad, N. Khan, S. Masud and M.A. Maud, "Selection of variable block sizes in H.264", Proc. of IEEE. ICASSP., vol. 3, pp. 173 - 176, May. 2004.

[2] G. J. Sullivan and R. L. Baker, "Rate-distortion optimized motion compensation for video compression using fixed or variable size blocks", Global Telecommunications Conference, vol. 6, pp. 85 - 90, Dec. 1991.

[3] Z. B. Chen, P. Zhou and Y. He., "Fast Integer Pel and Fractional Pel Motion Estimation for JVT", Joint Video Team Meeting, Awaji Island, Japan, JVT-F017, Dec. 2002.

[4] T. Wedi, "A Fast Sub-Pixel Motion Estimation Algorithm for H.264/AVC Video Coding", IEEE Trans. Circuit Syst. Video Technol., vol. 21, no. 2, pp. 237 - 242, Feb. 2011.

[5] A. Liu, W. Lin, M. Paul, F. Zhang, C. Deng, "Optimal Compression Plane for Efficient Video Coding", IEEE Trans. Image Proc., vol. 12, no. 99, Apr. 2011.

[6] K. Wong, L. Po, K. Cheung, "Block-Matching Translation and Zoom Motion-Compensated Prediction", Proc. of 16th IEEE International Conference on Image Processing (ICIP), pp. 1597 – 1600, 7-10 Nov. 2009.

[7] Reference Model JM6.1e, JVT. [Online]. Available: ftp://ftp.imtc-files.org.