WAVELET VIDEO COMPRESSION USING REGION BASED MOTION ESTIMATION AND COMPENSATION

Danny Lazar   Amir Averbuch

School of Computer Science
Tel Aviv University, Tel Aviv 69978, Israel

ABSTRACT

We present wavelet based video compression algorithms. The motion estimation and compensation, which is an essential part in the compression, is based on segment movements. The proposed based codec sequentially employs the following steps: We first divide each frame in the sequence into blocks resulted from the block matching algorithm. The, we appply Hough transform, in order to group it into segments the blocks that have similar motion parameters. We estimate the motion parameters for each segment using several motion models and least-squares algorithm and apply motion compensation upon the frames in each group. In the wavelet domain an optimal vector bit allocation is being used to distribute the bit budget among consecutive frames. The manipulated frames in each group are quantized and coded using techniques from still image compression. The residual error of the compression is reduced by applying error correction in the wavelet domain to enhance the quality of the reconstructed sequence.

1. INTRODUCTION

Data compresson of video squences has recently received considerable attention due to its applicability to communication, broadcasting and storage. Recently, consistent efforts have been invested to standardize the video coding scheme. The latest MPEG-4 standard [7, 11] recommends to utilize some form of object extraction and segmentation to increase the compression rate without degradation of the quality.

Wavelet based implementations of hierarchical motion compensation ([10]) and 3-D subband coding for multiresolution compression ([9]) have been reported in [5, 9, 8]. The wavelet transform has been shown to be an efficient coding method for still images and video. Unlike block transform-based coding, it does not suffer from blocking artifacts and hence it is able to produce better subjective quality especially at low bit rates.

In our preliminary work ([1]), we suggest a coding scheme which, for scenes where the camera holds still or pans doesn't demand any motion computations but uses the wavelet compression and bit allocation to compress groups of several frames, in image sequence. We also inroduced an improved compression that exploits the motion compensation resulting from motion vectors computed from fixed-sized block matching.

In order to improve the quality of the resulted frames for more complicated scenes which are three dimensional or contain camera zooming we propose an improved video scheme which exploits the appropriate motion estimation and compensation, based on defining segments in the sequence and compute their motion parameters. This paper applies the wavelet theory to compress full-motion color video signals using a motion estimation and compensation, using segmentation, while images are predistored by spatial reorientation prior to 2-D subband coding. In this work we propose a video compression scheme which utilizes motion compensation and exploits the wavelet compression characteristic which keeps, for a given bit budget, the most important features that should be used for high quality reconstruction.

We split the video sequence into n groups of M_n consecutive frames. The proposed baseline coder sequentially combines wavelet still compression, bit allocation, segmentation of the motion estimation and compensation. The following is an outline of the main steps of the combined algorithm:

Motion detection -

1. The video sequence is splitted into groups of consecutive frames where the number of frames in each group depends on the scene's characteristics.
2. Block matching is applied between the central frame and the rest of the frames in the group. We compute the motion vectors for each block relatively to the central frame. The sizes of the blocks can be either fixed (say 16×16) or varied (using quatree structure).

3. The motion vectors of the central frame are segmented into moving regions. Each segment contains blocks that have similar motion.
4. The motion is detected in the following way: the motion parameters, which describe each segment's motion from the central frame into every other frame in the group, are computed. The motion parameters for each segment are determined using least-squares on the motion vectors of the blocks in this segment. Different motion models are examined such as using 3,4,6 or 8 model parameters according to the scene's charateristics (2D or 3D).
5. When we have the segments of the central frame and their motion parameters, we predict from them the other frames in the group. The first half of the frames in the group are derived from the central frame in the group using backward motion prediction of the segments, while the other half is derived from the central frame using forward segments motion prediction.

Compression -

1. Application of motion compensation. For each frame in the group, we compute the difference between the frame and its motion predicted frame.
2. A fast multiscale wavelet based expansion is applied on each frame and its predicated frame in the group as was described in [6].
3. An optimal bit allocation on the differencing between frames in each group and their predictions is applied.
4. The central frame and the other frames' differences in the group are quantized and coded using the techniques of still image compression [4, 5]. The distribution of the bit budget among the M_n frames in the nth group is determined by the bit allocation.
5. In order to reduce the error resulted from the motion prediction and the wavelet compression, we apply error correction method. We estimate the error of the quantization through decompression and add this approximation to the main quantized frames.

2. MOTION MODEL ANALYSIS

Motion in image sequence is mostly contained in the movements of objects in a 3D scene and by camera motion. Thus, motion modeling is based on 3-D motion parameters that describe rotation, zoom and translation. The better accurate parameters we receive, a higher object quality is reconstructed. The 3D motion of objects and cameras become a 2D motion on the image plane via a projection is also called apparent motion or optical flow, which finds diverse applications in video processing, compression and computer vision. The motion parameters helps to remove temporal data redundancy and attains high compression ratios. Motion estimation became a fundamental component of standards such as H.261, H.263 and the MPEG family. Motion models which are used by these standards are very simple (one 2D motion vector per block which describes the translation in the x and y axis. The emerged MPEG-4 and MPEG-7 standards offer a region-based model that allows increased efficiency and flexibility. In our work we follow the region based approach.

Our approach is based on the split of the frame into blocks (using either a quadtree structure for variable-size blocks or fixed size blocks - both depend on the video sequence characteristics). Then we compute the 2D motion vectors per block. In section 3 we explain how to perform motion estimation and define new segments from these motion vectors. These segments-regions have similar motion parameters.

In our paper we describe the geometric changes in video images. We discuss the influence of these changes on the resulted image and we investigate the changes of the projection of a point in the world on the image plane as a result of those changes. Also, several parametric models for the representation of the motion vector fields are discussed. The appendix explains how to describe the motion changes (zoom, pan and rotation) and motion models which describe these changes.

The goal of this work is to derive high rate video compression which utilizes also motion analysis which is based on segmented motion vectors. When we demonstrate how the proposed compression method operates on several video sequences, the different motion model will be taken into consideration: in one sequence called "ping-pong" we have two-dimensional scenes. Therefore, 3 or 4 parameters are sufficient to describe it. In the other sequence called "garden" the scenes are three dimensional. Therefore, we need 6 or 8 parameters to describe the motion model.

3. MOTION SEGMENTATION

Most real image sequences contain multiple moving objects or multiple motions. Motion segmentation means labeling the pixels that are associated with each independently moving 3-D object in a sequence featuring multiple motions. The motion vectors that were derived from block matching, which are associated with the same motion, are grouped together. Proper feature selection facilitates effictive motion segmentation. In our work, a parametric model-based approach has been adopted for motion-based video segmentation where the model parameters constitute the features. The accuracy of the segmentation results clearly depends on the accuracy of the estimated segments and the parameters. Although we may not always be able to determine the actual 3-D motion and surface structure parameters from the mapping parameters, for image coding application this does not pose a serious problem since we are mainly interested in predicting the next frame in the group from the current frame.

We split the central frame in the group into moving regions (segments) that consist of areas that have similar motion. Then, the prediction of the frames which are adjacent to the central frame is based on these segments. Our segmentation method describes the motion in 2 or 3-dimensional scenes by dividing the frame into different regions (segments). These regions are contained in different planes.

Practically, the image is partitioned into nonoverlapping segments of rectangular regions of fixed size (or variable size if quadtree structure is applied) referred to as blocks, whose union covers the whole image. In sequences with clear camera pan/zoom, substantial savings can be achieved compared to standard methods which are based solely on local block motion estimation. In particular cases such as simple motion fields where all or most of the points are displaced in a uniform manner, such as in videophone or simple 2D sequences, we define the whole image to be one region. Then the computed motion is global. The saving in this case is enormous since all (or most of) the image points can be described by a small set of parameters. The global motion model has been used in computer vision, but has only recently been adopted in phase II of the MPEG-4 standard.

In figure 1 we observed the partition of demo frames from "ping-pong" and "garden" video sequence into four segments using 4 parameters motion model.

4. REFERENCES

[1] Averbuch, A., Lazar, D., *Wavelet Based Video Coder via Bit Allocation*, submitted to IEEE trans Circuits Systems for Video Technology.

[2] Lazar, D., Averbuch, A., M. Israeli, *Image Compression using Vector Quantization on Wavelet coefficients*, IEEE Trans. on Image Processing, No.1, pp. 4-15, Jan. 1996; Also in IEEE Data Compression Conference 1993 (DCC'93), March 30-April 2, pp. 459, 1993, USA.

[3] Lazar, D., *Image Compression using Vector Quantization on Wavelet Coefficients*, M.Sc Thesis, School of Mathematical Sciences, Tel Aviv University, August 1992.

[4] Averbuch A., Nir R., *Still Image Compression using Coded Multiresolution Tree*, Preprint 1995.

[5] A. Averbuch, R. Coifman, F. Meyer,*Motion compensation of wavelet coefficients for very low bit rate video coding*, IEEE 1997 International Conference on Image Processing (ICIP97), October 26-29, Santa Barbara, CA, USA, 1997.

[6] A. Averbuch, F. Meyer, J-O Strömberg, *Fast Adaptive Wavelet Packet Image Compression*, IEEE Trans. on Image Processing, Vol. 9, No. 5, pp. 792-800, May 2000.

[7] R. Koenen, *MPEG-4 - Multimedia for our time*, IEEE Spectrum, 36(2):26-33, 1999

[8] Uz KM, Vetterli M, Le Gall D,*Interpolative multiresolution coding of advanced television with compatible subchannels*. IEEE trans Circuits Systems for Video Technology, 1(1):86-99, 1991.

[9] Lewis AS, Knowles G, *Video compression using 3D wavelet transforms*. Electronics Letters, 26(6):396-397, 1990.

[10] LeGall, D, *MPEG: A video compression standard for multimedia applications*, Communication ACM, 34(4), pp. 46-58, 1991.

[11] MPEG-4: The whole issue of *Signal Processing: Image Communication*, Vol. 15, pp. 269-478, 2000.

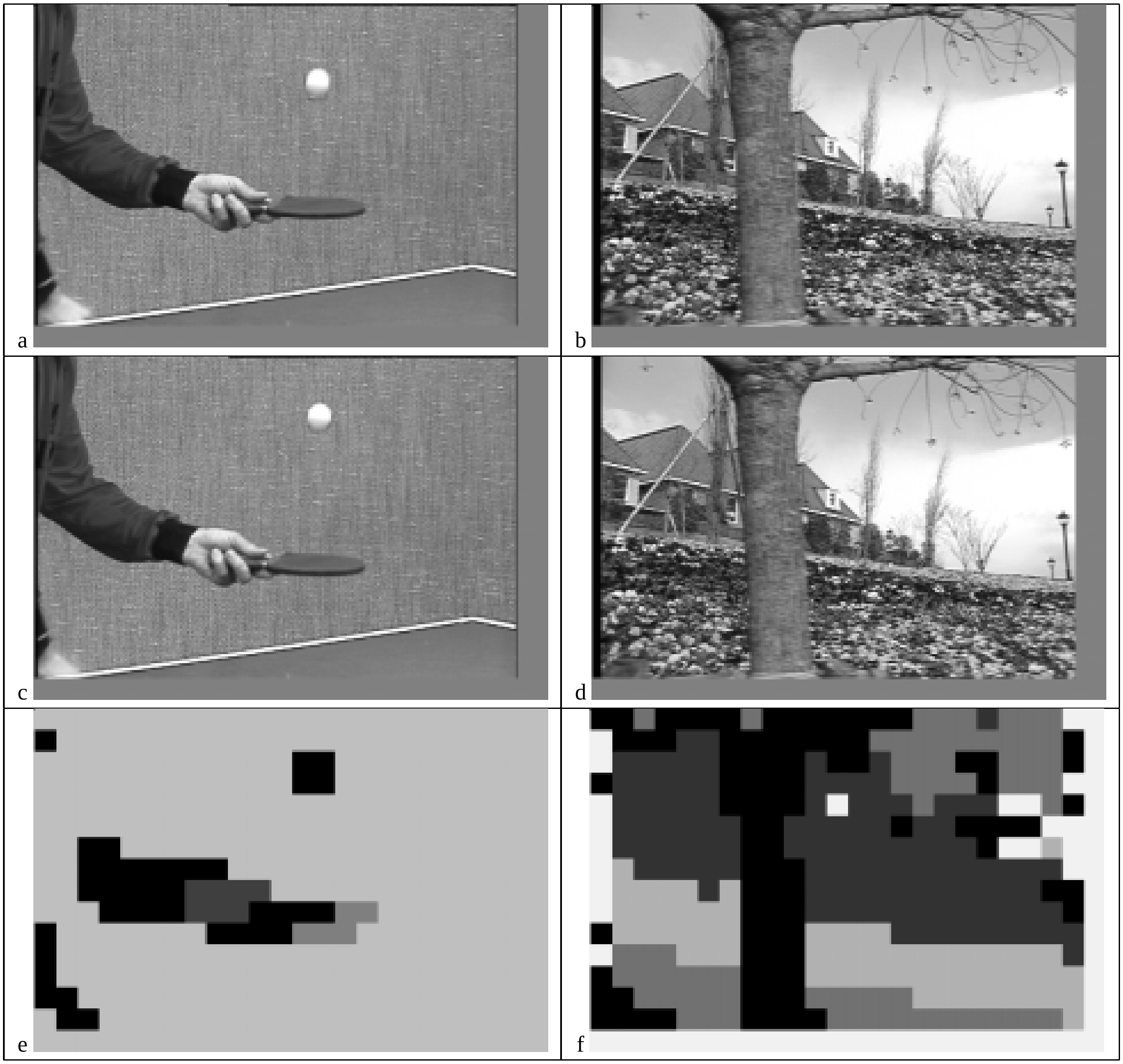

Fig. 1. Frames from the ping-pong and garden sequences and their segments. 4 segments are detected using 4 parameters motion model. Different colours for different regions. (a) frame ping-pong 000, (b)frame garden 000, (c) frame ping-pong 001, (d) frame garden 001, (e) ping-pong segments (f) garden segments.