MOTION COMPENSATED SUBBAND CODING OF VIDEO ACQUIRED FROM A MOVING PLATFORM

Oh-Jin Kwon, Rama Chellappa and Carlos Morimoto

Center for Automation Research
University of Maryland
College Park, MD20742-3275

ABSTRACT

We improve the performance of conventional motion compensated Discrete Cosine Transform video coding. For motion compensation, we employ a two step algorithm in which the camera motion is compensated first and then the motion of moving objects is estimated. We use a feature matching algorithm for camera motion compensation. Motion compensated frame differences are divided into three regions called stationary background, moving objects, and newly emerging area. A region adaptive subband image coding scheme is used for spatial coding of these regions.

1. INTRODUCTION

Due to the importance of digital video communications, numerous video coding techniques have been developed in conjunction with image coding techniques. Extensive research in the past two decades has made this field sufficiently mature so that several standards are now available. They are the ITU-T H.261 (formerly CCITT Recommendation H.261) [1] for teleconferencing, the ISO/IEC MPEG-1 [2, 3] for digital storage media, like CD-ROM's, and the MPEG-2 [4, 5] for high-quality coding of possibly interlaced video, including HDTV. All of these standards are based on the same hybrid coding structure, namely Motion Compensated (MC) temporal coding combined with Discrete Cosine Transform (DCT) spatial coding.

The basic structure of well studied MC-DCT video coding [1]-[5] is shown in Fig. 1. The current frame of input video is first predicted based on the encoded version of the previous frame available in the receiver. The Motion Compensated Frame Difference (MCFD), which is a difference between the current frame and the predicted frame, is divided into nonoverlapping blocks, mostly 8×8 in size, and transformed by a two dimensional (2D) DCT. The coefficients obtained from this transformation are then encoded using a quantizer normally coupled with a variable length coder and designed to suit the statistical characteristics of the coefficients. For motion compensation, Block-Matching Algorithm (BMA) is extensively used. In BMA, the prediction of motion is also performed on a block by block basis.

The support of the Advanced Research Projects Agency (ARPA order no. 8459) and the U.S. Army Topographic Engineering Center under contract DACA 76-92-C-0079 is greatfully acknowledged.

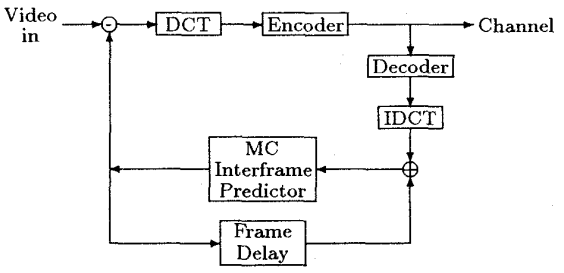


Figure 1: Basic structure of MC-DCT video coding.

The current frame is partitioned into nonoverlapping rectangular blocks and, for each block, a search is carried out for the displacement which produces the best match among the neighboring blocks in the previous frame.

Although this MC-DCT coding has been widely used for video coding, it has some disadvantages. First, the BMA has been developed under the assumptions of rigid and translational motion and constant illumination condition between the frames so that it suffers from the obvious problem that the true motion is not piecewise constant. This is especially true if the scenes are acquired from a moving camera whose motion includes rotation, zoom, and pan as well as translation. Second, the underlying block structure for the DCT spatial coding does not adequately represent the arbitrarily shaped objects and backgrounds in a scene resulting in annoying blocking effects at low bit rates.

In this paper, we propose two methods for improving the performance of conventional MC-DCT video coding in the following ways.

1. To improve BMA, we substitute the BMA by a two step motion compensation algorithm [6] in which the camera motion (global motion) is compensated first and then the motion of moving objects (local motion) in a scene is estimated.
2. To compensate for the disadvantage of block based DCT spatial coding, we employ the Region Based Subband Image Coding (RB-SBIC) method described in [7].

We use the Feature Matching Algorithm (FMA) for Global Motion Compensation (GMC). In FMA, several fea-

0-7803-2431-5/95 $4.00

ture points are extracted using a Sobel operator [8] and global motion parameters (translation, rotation, and zoom) are computed by matching these feature points. For completion of the GMC, the illumination change between frames is also considered. The GMC is followed by the Local Motion Compensation (LMC) in which the moving objects are detected and their velocities are also calculated.

When we temporally predict the current frame from the MC version of the previous frame, we assume that pixels in the current frame belong to one of the following three different regions.

1. Stationary background: the region is unchanged with respect to the previous frame and compensated well by GMC.
2. Moving objects: the moving parts of the current frame. LMC is needed for the prediction of this region.
3. Newly emerging areas: the parts of the current frame which are not present in the previous frame. This region can be subdivided into two parts, 1) the newly emerging background and objects on the boundary of the frame due to global motion and 2) the background occluded by moving objects in the previous frame and uncovered due to local motion. Newly emerging areas are not temporally predictable with respect to the previous frame. Spatial prediction based on the boundary pixels available in the previous frame is instead used for the temporal prediction of this region.

For spatial coding of MCFD, we employ an RB-SBIC scheme in which we decompose the MCFD into several image subbands, study the energy distribution of the decomposed MCFD , and use different quantizers for each region in each subband.

This paper is organized as follows. Section 2 describes the GMC and LMC methods employed for temporal prediction. In Section 3, we design the RB-SBIC scheme for spatial coding of MCFD. Section 4 compares the experimental results of our proposed video coding system to that of conventional MC-DCT scheme.

2. MOTION COMPENSATION

2.1. Global Motion Compensation (GMC)

The GMC used in our video coding scheme is based on a 2D image registration algorithm [9] that estimates the 2D camera translation and rotation around the axis perpendicular to the center of image plane and the zoom factor.

If we assume that the distance between the camera and the scene is large enough that camera panning motion can be approximated by translation, the relationship between two image frame coordinates can be approximated by [9]

$$\begin{pmatrix} X_2 \\ Y_2 \end{pmatrix} = z \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} + \begin{pmatrix} \Delta X_2 \\ \Delta Y_2 \end{pmatrix} \quad (1)$$

where, (X_i, Y_i) is the frame coordinate at time t_i, for i = {1,2}, $(\Delta X_2, \Delta Y_2)$ is the translation measured in the image coordinate system of frame t_2 , θ is the rotation angle between the two frames, and z is the zoom factor.

Global motion parameters, $\{\Delta X_2, \Delta Y_2, \theta, z\}$, are estimated using FMA. We first extract feature points of frame t_1 using a Sobel operator. We convolve the frame with a 5×5 Sobel operator and select the points whose value is greater than a threshold value. To avoid all the feature points getting concentrated in a small area, we divide the entire frame into 16 equally spaced blocks and include the point showing the maximum value in a block as a feature point. After extracting the feature points of frame t_1, we find the corresponding feature points in frame t_2 using a correlation matching process. For the neighborhood of each feature point extracted from frame t_1, we search for the best matching point in frame t_2 using the cross correlation coefficient as the matching criterion. This correlation matching process can only provide the best pixel to pixel match. To improve the accuracy to subpixel level, we also apply a subpixel matching algorithm using a differential method described in [6].

If the feature point pairs of frame t_1 and t_2 are available, the zoom factor z can be easily estimated prior to other global motion parameters because the Euclidean distances of the feature points in a frame are only dependent on the zoom factor and are invariant to rotation and translation. We calculate the Euclidean distances of the feature points for each frame and use Least-Square Estimation (LSE) for the estimation of zoom factor [9]. With the zoom factor determined, the rotation and translation parameters can be estimated using (1). We assume that the rotation angle θ is very small so that we can approximate $\cos\theta$ and $\sin\theta$ up to the first linear terms, end up with linear equations, and can estimate the parameters, θ, ΔX_2, and ΔY_2, using LSE [9]. If the estimated global motion parameters are not accurate enough, we can also apply the multiscale recursive matching and estimation refinement procedures described in [9].

Using the estimated global motion parameters, we can transform the two image frames into a common coordinate system and perform GMC by taking the difference between two successive frames. In the stationary background, the Global Motion Compensated Frame Difference (GMCFD) normally has small values. Image noise and illumination change between two frames may cause the pixel values to be non-zero. To compensate for the effect of illumination change, the mean of GMCFD is adjusted to be zero in the stationary background.

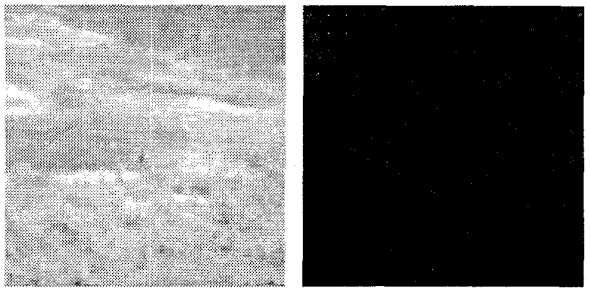
2.2. Local Motion Compensation (LMC)

Moving objects in a scene are detected based on the absolute value of GMCFD. When we detect moving objects from the GMCFD, the following factors are considered.

1. Large values of GMCFD are usually due to two factors, image noise and moving objects.
2. Large values of GMCFD due to moving objects occur as blocks while those due to image noise are isolated.
3. The values of GMCFD are dependent on the local contrast in the scene.

We employ an adaptive moving object detector [6] which suppresses the image noise by checking the local average of GMCFD and uses different thresholding values based on the local contrast.

When more than two successive GMCFD's are available, we can estimate the velocity of the moving object by



(a) (b)

Figure 2: Sample image sequence "helicopter" (a) and its MCFD energy distribution in the 16 × 16 STFT tree decomposition (b).

calculating the displacement of the centroid of the moving object [6].

3. RB-SBIC SPATIAL CODING

In this section, we describe the RB-SBIC for the spatial coding of MCFD. In RB-SBIC, the input MCFD is first decomposed into several subbands using a bank of analysis filters and then each subband is down-sampled, encoded, and transmitted through a channel. Therefore, the problem of designing an RB-SBIC can be subdivided into the following three problems: 1) design of a bank of analysis filters, 2) design of an encoder characterized by quantizers and bit allocation methods, and 3) design of a specific decomposition tree structure.

For our choice of the analysis filter bank, we use a perfect reconstruction filter bank that recursively employs the ideal two band Lowpass (LP)-filter and Highpass (HP)-filter in which LP and HP filterings are done by 1) extending the discrete-time input signal of length N to the length $(2N-2)$ symmetric signal, 2) performing the Discrete Fourier Transform (DFT), 3) doing the filterings in the frequency domain, and 4) performing the inverse DFT.

We first study the energy distribution of MCFD in subband decompositions and propose an RB-SBIC encoder. If both the analysis filter bank and the encoder are given, the optimal subband decomposition tree structure can be found using the *bottom-up* search method coupled with the *principle of separate minimization* [10]. The resulting optimal decomposition structures for MCFD of sample image sequences are presented in Section 4.

Fig. 2(a) shows the first frame of sample image sequence "helicopter" taken here for the observation of MCFD energy distribution. The image sequence is composed of 15 frames whose size is 464 × 464, represented using 8 bits/pel, and obtained from a moving platform.

Fig. 2(b) illustrates the MCFD energy distribution of sample image sequence in the 16 × 16 Short-Time-Fourier-Transform (STFT) tree decomposition [7]. For display purposes, we calculated the logarithm of the squared subband decomposed MCFD values and normalized them between 0 and 255. It may be seen that the lower frequency subbands have higher energy than the higher frequency subbands. The corresponding *energy ratios* [7] are calculated for this

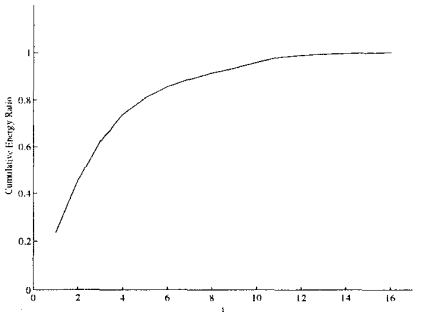


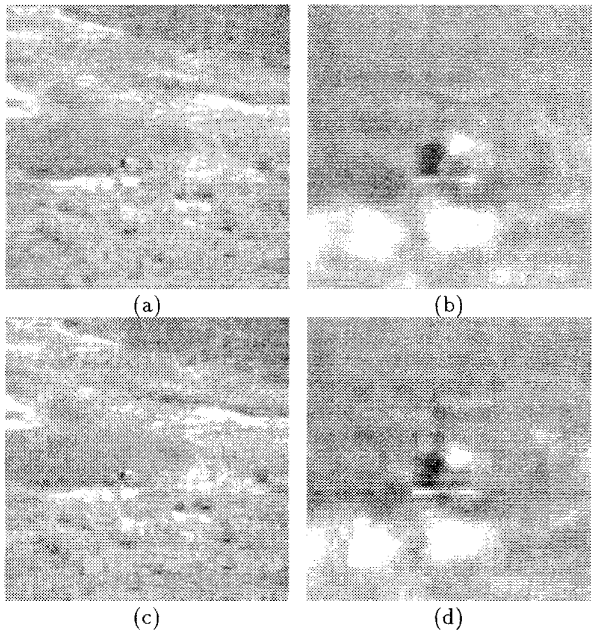
Figure 3: Example of cumulative *energy ratio* of MCFD.

decomposition. This time, we include the Lowest Frequency Subband (LFS) for the calculation of *energy ratio* since the LFS and the Higher Frequency Subbands (HFS) of MCFD have similar statistical characteristics. Cumulative values of *energy ratio* are shown in Fig. 3. It is observed that MCFD preserves the *Energy Packing Property towards the Lower Frequency Subbands (EPPLFS)* [7] very well and has more than 90% of its total energy in the lower half of subbands.

For spatial coding of MCFD, the RB-SBIC proposed in [7] can be modified as follows.

1. Assume that a decomposition tree structure is given and global and local motion parameters are available.
2. Using global and local motion parameters, divide each subband of MCFD into three region - stationary background, moving objects, and newly emerging area.
3. For each region of each subband (including the LFS), calculate the variance and transmit it.
4. For each region of each subband, design an Entropy Constrained Quantizer (ECQ) by modeling the shape of distribution using the Generalized Gaussian Distribution (GGD).
5. Based on the variance and the rate-distorsion performance of the designed quantizer, allocate the number of bits for each region of each subband.
6. Employ PCM using the above mentioned quantizer and bit allocation.

As done in [7], we have performed the KS-test to find the best fitting GGD parameter α for 4×4, 8×8, and 16×16 STFT decomposed MCFD subbands. The distributions of MCFD subbands have been found to be similar to those of DPCM residuals of the LFS in [7] so that they can be modeled by the Laplacian distribution. The (UTQ,HC) pair for the Laplacian distribution, designed in [11], is employed as a method of ECQ and the same bit allocation scheme of [7] can be used here. Due to the EPPLFS, variances of each region are decreasing as the distance of the subband becomes larger. For the variances of the LFS regions, we have assigned 12 bits. For the variances of the regions in other higher frequency subbands, we have transmitted the difference between the variance of the current region and that of the corresponding region in the nearest lower frequency subband using 8 bits. If no bits were assigned to the corresponding region in the nearest lower frequency subband, the variance of the current region is not transmitted.



(a) (b) (c) (d)

Figure 4: Simulation results for "helicopter" image sequence at 0.05 bits/pel: proposed video coding (a) and enlarged version (b) and MC-DCT video coding (c) and enlarged version (d).

4. EXPERIMENTAL RESULTS

Simulation results have been obtained for the "helicopter" image sequence. The 16 × 16 STFT tree decomposition structure has been selected as an optimal decomposition structure because this structure has shown the best result among all the possible tree structures up to 16 × 16 STFT decomposition and decomposing the MCFD into the smaller sized subbands resulted in negligible improvements at low bit rates.

Fig. 4(a) shows the reconstructed images of the last frame at 0.05 bits/pel. The moving object area of Fig. 4(a) is enlarged and displayed in Fig. 4(b). The results of conventional MC-DCT video coding are also included for subjective comparisons. While MC-DCT coding suffers from blocking effects, the proposed scheme shows blurring effects at low bit rates.

Using the PSNR measure, objective tests were performed for all frames (Fig. 5). It is seen that more than 1-dB improvement on average has been achieved.

GMC and LMC algorithms and design of analysis filters to reduce blurring effects at low bit rates can be considered as future research topics.

5. REFERENCES

[1] *CCITT Recommendation H.261, Video Codec for Audiovisual Services at p×64 Kbits/s*, CCITT doc. COM XV-R 37-E, 1990.

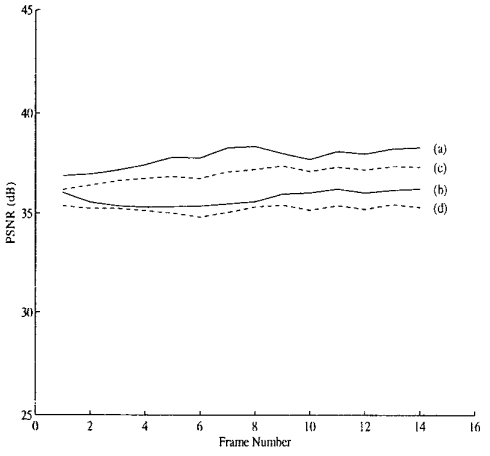


Figure 5: PSNR results for "helicopter" image sequence: proposed video coding at 0.1 bits/pel (a) and 0.05 bits/pel (b) and MC-DCT video coding at 0.1 bits/pel (c) and 0.05 bits/pel (d).

[2] D. J. LeGall, "MPEG: A video compression standard for multimedia applications", *Commn. of the ACM*, vol.34, no.4, pp.46-58, Apr. 1991.

[3] P. Pancha and M. E. Zarki, "MPEG coding for variable bit rate video transmission", *IEEE Commn. Mag.*, vol.32, no.5, pp.54-66, May 1994.

[4] *Information Technology-Generic Coding of Moving Pictures and Associated Audio Recommendation H.262*, ISO/IEC 13818-2 Committee Draft, Nov. 1993, Seoul.

[5] T. Chiang and D. Anastassiou, "Hierarchical coding of digital television", *IEEE Commn. Mag.*, vol.32, no.5, pp.38-45, May 1994.

[6] Q. Zheng and R. Chellappa, "Motion detection in image sequences acquired from a moving platform", in *Proc. IEEE ICASSP-93*, vol. 5, pp.201-204, April 1993.

[7] O. Kwon and R. Chellappa, "Region based subband image coding scheme", *Proc. IEEE ICIP-94*, vol. 2, pp.859-863, Nov. 1994.

[8] A. K. Jain, *Fundamentals of Digital Image Processing*, Englewood Cliff, NJ: Prentice-Hall Inc., 1989.

[9] Q. Zheng and R. Chellappa, "A computational vision approach to image registration", *IEEE Trans. on IP*, vol. IP-2, pp.311-326, July 1993.

[10] G. J. Sullivan and R. L. Baker, "Efficient quadtree coding of images and video", *IEEE Trans. on IP*, vol.IP-3, pp.327-331, May 1994.

[11] N. Farvardin and J. M. Modestino, "Optimum quantizer performance for a class of non-Gaussian memoryless sources", *IEEE Trans. on IT*, vol.IT-30, pp.485-497, May 1984.