Motion Compensated Subband Video Coding with Arbitrarily Shaped Region Adaptivity

Oh-Jin Kwon and Seokrim Choi

The performance of Motion Compensated Discrete Cosine Transform (MC-DCT) video coding is improved by using the region adaptive subband image coding [18]. On the assumption that the video is acquired from the camera on a moving platform and the distance between the camera and the scene is large enough, both the motion of camera and the motion of moving objects in a frame are compensated. For the compensation of camera motion, a feature matching algorithm is employed. Several feature points extracted using a Sobel operator are used to compensate the camera motion of translation, rotation, and zoom. The illumination change between frames is also compensated. Motion compensated frame differences are divided into three regions called stationary background, moving objects, and newly emerging areas each of which is arbitrarily shaped. Different quantizers are used for different regions. Compared to the conventional MC-DCT video coding using block matching algorithm, our video coding scheme shows about 1.0-dB improvements on average for the experimental video samples.

Manuscript received June 6, 2000; revised July 4, 2001 and Oct. 15, 2001.

Oh-Jin Kwon (phone: +82 3408 3295, e-mail: ojkwon@sejong.ac.kr) and Seokrim Choi (e-mail: schoi@kunja.sejong.ac.kr) are with International Research Institute for Information and Telecommunications (IRIIT) and the Department of Electronics Engineering, Sejong University, Seoul, Korea.

I. INTRODUCTION

Due to the importance of digital video communications, numerous video coding techniques have been developed in conjunction with image coding techniques. Most of these techniques have attempted to remove spatial and temporal redundancies in the input video to improve the coding efficiency. As for removing spatial redundancies, transform coding methods for still image compression have been employed. The Discrete Cosine Transform (DCT) based image coding techniques [1], the subband decomposition based techniques [2], and vector quantization based techniques [3] are known as the most popular techniques for removing spatial redundancies. Concerning the removal of temporal redundancies, motion compensation techniques [4], [5] relying on Block Matching Algorithms (BMA) for temporal estimations and predictions are popular and have appeared in most of the video coding methods.

Extensive research of the past two decades has made this video coding research field sufficiently mature so that several standards are now available. They are the ITU-T H.261 (formerly CCITT Recommendation H.261) [6] for teleconferencing, the ISO/IEC MPEG-1 [7], [8] for digital storage media, like CD-ROM's, the MPEG-2 [9] for high-quality coding of possibly interlaced video, including HDTV, the MPEG-4 [10] for interactive multimedia. Most of these standards are based on the hybrid coding structure, namely Motion Compensated (MC) temporal coding combined with DCT spatial coding.

In MC-DCT video coding, 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 non-overlapping blocks and transformed by a two-dimensional 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, BMA is extensively used. In BMA, the prediction of motion is also performed on a block by block basis. The current frame is partitioned into non-overlapping 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 works quite well for typical low data rate scenes, it has some disadvantages. First, the BMA has been developed on the assumptions of rigid and translational motion and constant illumination condition between the frames, so it suffers from the obvious problem of the true motion being not piecewise constant. This is especially true if the scenes are acquired from the camera on a moving platform and the motion includes the camera's zoom, rotation, tilt and pan as well as the platform's rotation and translation. Second, the underlying block structure for the DCT spatial coding does not fit in arbitrarily shaped objects and backgrounds in a scene, causing annoying blocky effects at low coding rates.

Recently, there has also been active research in trying to apply the subband image coding techniques to video coding [11]-[17]. These techniques are attracting much attention for their capability to support scalability. In subband video coding, input video frames are decomposed into multiple subbands and coded in a coarse-to-fine fashion so that pyramid decoding of the entire sequence at various quality levels is possible [16], [17]. Another motivation for subband video coding is that it compacts signal energy more efficiently than the MC-DCT coding and can be tuned effectively to the spatial characteristics of the human visual system [2]. Most of the experimental results of previous works [11]-[17] have shown that subband coding produces less blockier picture quality than the MC-DCT based coding.

In this paper, two methods to improve the performance of conventional MC-DCT video coding are proposed. 1) To compensate for the disadvantage of block based DCT spatial coding, the Region Adaptive Subband Image Coding (RA-SBIC) [18] is employed. 2) To improve BMA, it is substituted by a two-step motion compensation algorithm [19] in which the camera motion (global motion) is compensated first and then the motion of moving objects (local motion) in a scene is compensated.

We use the Feature Matching Algorithm (FMA) for Global Motion Compensation (GMC). In FMA, several feature points are extracted using a Sobel operator [20] 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:

- Stationary background: the region is unchanged with respect to the previous frame and compensated well by GMC.
- Moving objects: the moving parts of the current frame. LMC is needed for the prediction of this region.
- Newly emerging areas: the parts of the current frame that 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 simple bilinear interpolation [21] using 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 RA-SBIC scheme [18] in which we decompose the MCFD into several image subbands, study the energy distribution of the decomposed MCFD, and use different quantizers and different number of bits for each region of each subband.

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

II. MOTION COMPENSATION

1. Global Motion Compensation

The GMC used in our video coding scheme is based on a two-dimensional (2D) image registration algorithm [19] that estimates the zoom factor and the 2D camera translation and rotation around the axis perpendicular to the center of image plane.

If we assume that the distance between the camera and the scene is large enough, the relationship between two image frame coordinates can be approximated by [19]

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

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 the local peak and greater than a threshold value T_g. To prevent all the feature points locating 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 be formulated as follows:

- Denote pixel values of the frame t_1 and t_2 by $f_1(m,n)$ and $f_2(k,l)$, respectively.
- The cross correlation coefficient between $f_1(m,n)$ and $f_2(k,l)$ is calculated by

$$\rho_{f_1 f_2}(m,n;k,l) = \frac{1}{\sigma_1 \sigma_2 (2w_m+1)^2} \sum_{i,j=-w_m}^{i,j=w_m} [f_1(m+i,n+j) - \mu_1] \cdot [f_2(k+i,l+j) - \mu_2] \quad (2)$$

where $(2w_m+1)\times(2w_m+1)$ is the size of matching window and μ_i and σ_i^2, $i = 1,2$, are the local means and variances for each frame t_i, respectively.

- Assume that there exist N feature points extracted by the Sobel operator in frame t_1 and denote them by $\{f_1(m_i, n_i),\ i = 1,2,...N\}$.
- The corresponding best matching feature points in frame t_2, $\{f_2(k_i, l_i),\ i = 1,2,...N\}$, are searched for the $(2w_s+1)\times(2w_s+1)$ neighborhood of each $f_1(m_i, n_i)$ such that $\rho_{f_1 f_2}(m_i, n_i; k_i, l_i)$ is maximized.

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 [19]. The relationship between the zoom factor z from frame t_1 to t_2 and the Euclidean distances of the N feature points for each frame can be represented by

$$\begin{pmatrix} d_1^1 \\ d_2^1 \\ \cdot \\ \cdot \\ \cdot \\ d_N^1 \end{pmatrix} = z \begin{pmatrix} d_1^2 \\ d_2^2 \\ \cdot \\ \cdot \\ \cdot \\ d_N^2 \end{pmatrix}, \quad (3)$$

where

$$d_i^1 = \sqrt{(m_i - \bar{m})^2 + (n_i - \bar{n})^2} \quad (4)$$

$$d_i^2 = \sqrt{(k_i - \bar{k})^2 + (l_i - \bar{l})^2} \quad (5)$$

$$\bar{m} = \frac{1}{N}\sum_{i=1}^{N} m_i \quad (6)$$

$$\bar{n} = \frac{1}{N}\sum_{i=1}^{N} n_i \quad (7)$$

$$\bar{k} = \frac{1}{N}\sum_{i=1}^{N} k_i \quad (8)$$

$$\bar{l} = \frac{1}{N}\sum_{i=1}^{N} l_i, \quad (9)$$

and the LSE of z is computed as

$$z = \left[\begin{pmatrix} d_1^2 \\ d_2^2 \\ \cdot \\ \cdot \\ \cdot \\ d_N^2 \end{pmatrix}^T \begin{pmatrix} d_1^2 \\ d_2^2 \\ \cdot \\ \cdot \\ \cdot \\ d_N^2 \end{pmatrix} \right]^{-1} \begin{pmatrix} d_1^2 \\ d_2^2 \\ \cdot \\ \cdot \\ \cdot \\ d_N^2 \end{pmatrix}^T \begin{pmatrix} d_1^1 \\ d_2^1 \\ \cdot \\ \cdot \\ \cdot \\ d_N^1 \end{pmatrix} = \frac{\sum_{i=1}^{N} d_i^2 d_i^1}{\sum_{i=1}^{N} d_i^2 d_i^2}. \quad (10)$$

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. If we substitute $\cos\theta$ and $\sin\theta$ in (1) using 1 and θ, respectively, we end up with the following equation

$$\begin{pmatrix} X_2 \\ Y_2 \end{pmatrix} = z\begin{pmatrix} 1 & \theta \\ -\theta & 1 \end{pmatrix}\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} + \begin{pmatrix} \Delta X_2 \\ \Delta Y_2 \end{pmatrix}. \quad (11)$$

If we represent (11) using the feature point pairs, it becomes

$$\begin{pmatrix} k_i \\ l_i \end{pmatrix} = z\begin{pmatrix} 1 & \theta \\ -\theta & 1 \end{pmatrix}\begin{pmatrix} m_i \\ n_i \end{pmatrix} + \begin{pmatrix} \Delta X_2 \\ \Delta Y_2 \end{pmatrix}, \quad i = \{1,2,...,N\} \quad (12)$$

or

$$\begin{pmatrix} k_1 - zm_1 \\ l_1 - zn_1 \\ k_2 - zm_2 \\ l_2 - zn_2 \\ \cdot \\ \cdot \\ \cdot \\ k_N - zm_N \\ l_N - zn_N \end{pmatrix} = \begin{pmatrix} zn_1 & 1 & 0 \\ -zm_1 & 0 & 1 \\ zn_2 & 1 & 0 \\ -zm_2 & 0 & 1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ zn_N & 1 & 0 \\ -zm_N & 0 & 1 \end{pmatrix} \begin{pmatrix} \theta \\ \Delta X_2 \\ \Delta Y_2 \end{pmatrix}. \quad (13)$$

The rotation and translation parameters can then be computed as

$$\begin{pmatrix} \theta \\ \Delta X_2 \\ \Delta Y_2 \end{pmatrix} = \left[\begin{pmatrix} zn_1 & 1 & 0 \\ -zm_1 & 0 & 1 \\ zn_2 & 1 & 0 \\ -zm_2 & 0 & 1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ zn_N & 1 & 0 \\ -zm_N & 0 & 1 \end{pmatrix}^T \begin{pmatrix} zn_1 & 1 & 0 \\ -zm_1 & 0 & 1 \\ zn_2 & 1 & 0 \\ -zm_2 & 0 & 1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ zn_N & 1 & 0 \\ -zm_N & 0 & 1 \end{pmatrix} \right]^{-1} \times \begin{pmatrix} zn_1 & 1 & 0 \\ -zm_1 & 0 & 1 \\ zn_2 & 1 & 0 \\ -zm_2 & 0 & 1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ zn_N & 1 & 0 \\ -zm_N & 0 & 1 \end{pmatrix}^T \begin{pmatrix} k_1 - zm_1 \\ l_1 - zn_1 \\ k_2 - zm_2 \\ l_2 - zn_2 \\ \cdot \\ \cdot \\ \cdot \\ k_N - zm_N \\ l_N - zn_N \end{pmatrix}. \quad (14)$$

Using the obtained global motion parameters, we can transform frame t_1 into the coordinate system of frame t_2 and perform GMC by taking the difference. For the pixel to pixel matching of the two frames, bilinear interpolation has been employed.

In the stationary background, the Global Motion Compensated Frame Difference (GMCFD) normally has small values. Imaging noise and illumination change between two frames may cause the pixel values to be non-zero. To compensate the effect of illumination change, the mean of GMCFD is adjusted to be zero in the stationary background.

2. Local Motion Compensation

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.

- Large values of GMCFD are usually due to two factors, imaging noise and moving objects.
- The large values of GMCFD due to moving objects occur with nearby large values while those due to imaging noise are isolated.
- The values of GMCFD are dependent on the local contrast of the scene.

We employ an adaptive moving object detector [19] which suppresses the imaging noise by checking the local average of GMCFD and uses different thresholding values based on the local contrast. Denoting the GMCFD by $d(k,l)$, we classify a pixel (k,l) as belonging to a moving object in frame t_2 if

$$\min\{\overline{d(k,l)}, d(k,l)\} > \max\{\gamma D(k,l), T\} \quad (15)$$

where

$$\overline{d(k,l)} = \frac{1}{(2w_n+1)^2} \sum_{i,j=-w_n}^{i,j=w_n} d(k+i,l+j), \quad (16)$$

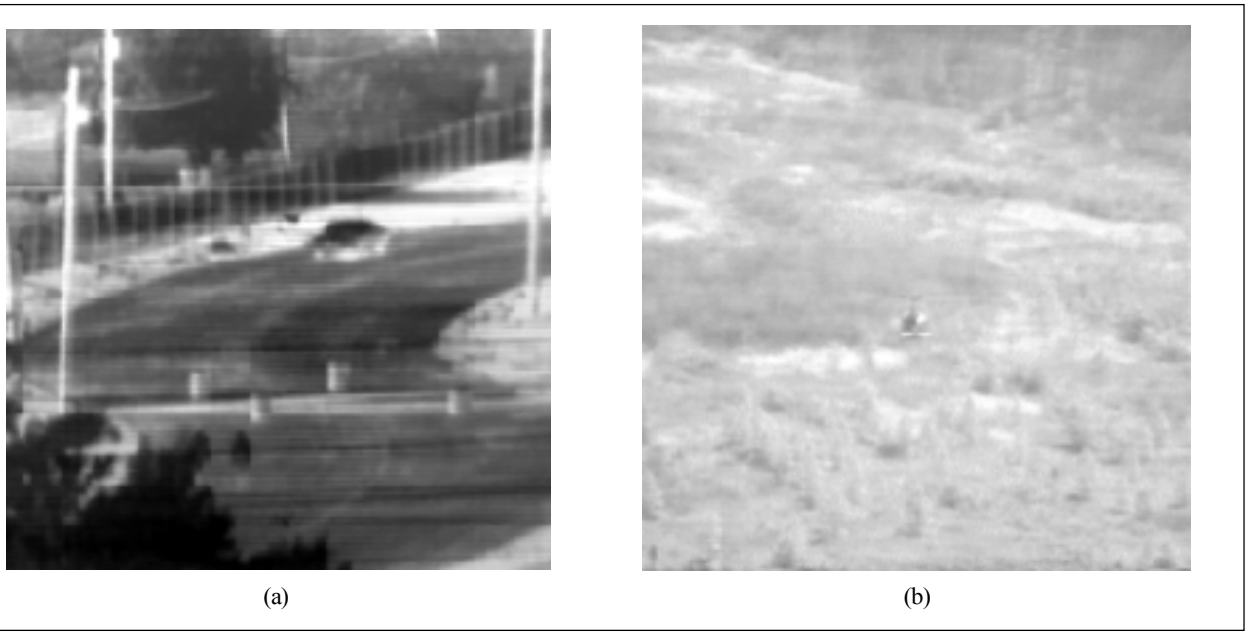
$$D(k,l) = \sup_{(i,j)\in(\pm w_n,\pm w_n)} \{f_2(k+i,l+j)\} - \inf_{(i,j)\in(\pm w_n,\pm w_n)} \{f_2(k+i,l+j)\} \quad (17)$$

$(2w_n+1)\times(2w_n+1)$: Size of neighborhood window

γ : Threshold for relative difference

T : Threshold for absolute difference.

When more than two successive GMCFD's are available, the velocity of a moving object is estimated by calculating the



(a) (b)

Fig. 1. Sample image sequences: (a) "helicopter" and (b) "TI-car".

displacement of the centroid of the moving object [19] and used for the Local Motion Compensation (LMC) of this area.

3. Refinement of the GMC and the LMC

In our motion compensation scheme described in Sections II-1 and II-2, the global motion is compensated prior to the local motion. If any of the feature points used to estimate the global motion parameters in Section II-1 is included in the moving objects detected in Section II-2, those feature points should be excluded for the estimation of global motion, since they are not globally moving points in a scene. In this case, we refine the aforementioned global and local motion compensation scheme by reapplying the scheme excluding those feature points for the estimation of global motion.

III. RA-SBIC SPATIAL CODING

In this section, we describe the RA-SBIC for the spatial coding of MCFD. In RA-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 RA-SBIC can be subdivided into the following three problems: 1) design of a bank of analysis filters, 2) design of an encoder characterized by employed 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 Lowpass (LP)-filter and Highpass (HP)-filter (described in [18]), 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. Bearing in mind that, when the DFT representations are used, all the sequences in both spatial domain and frequency domain behave as if they are periodic, we have chosen the modulo $(2N-2)$ symmetric extension, since this extension ends up with both the real valued and the modulo $(2N-2)$ symmetric DFT coefficients for the real input signal [18]. It is also possible to use the length $(2N-1)$ or the length $2N$ symmetric extension. However, the length $(2N-1)$ extension gives us the odd number of DFT coefficients, so that the exact 2-band LP and HP filterings cannot be done. The length $2N$ extension does not preserve the modulo $2N$ symmetric, so that we end up with the complex values at the end of filterings [18].

We first study the energy distribution of MCFD in subband decompositions and propose an RA-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* [22] described in [18]. The resulting optimal decomposition structures for MCFD of sample image sequences are presented in Section IV.

Figure 1 shows the first frames of two sample image sequences taken here for the observation of MCFD energy distribution. The "helicopter" image sequence is composed of 15 frames whose size is 464×464 and the "TI-car" sequence is composed of 48 infrared image frames whose size is 208×208.

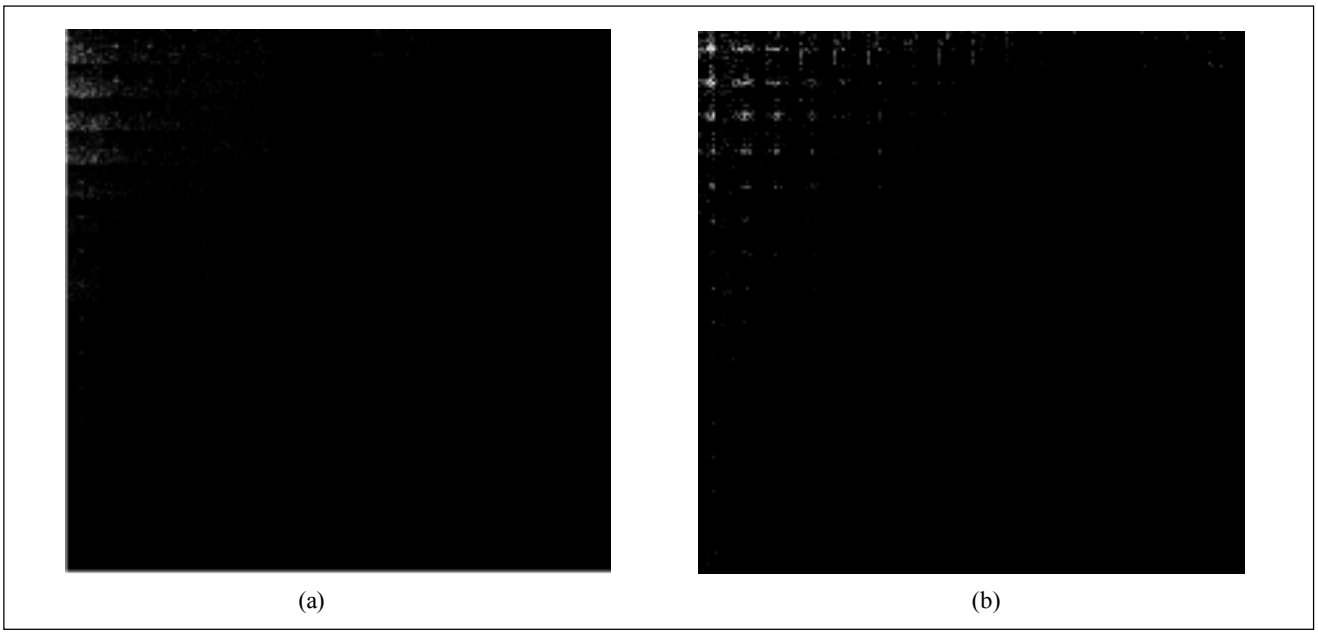


Fig. 2. Examples of MCFD energy distribution in the 16 × 16 STFT tree decomposition: (a) "helicopter" and (b) "TI-car".

Both sequences are represented using 8 bits/pel and obtained from a moving platform.

Figure 2 illustrates the MCFD energy distribution of sample image sequences in the 16 × 16 Short Time Fourier Transform (STFT) tree decomposition [ch.2 of [2]], [18] which analyzes the MCFD into 16 × 16 equally spaced subbands and subsamples them by 16 × 16. 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* [18] are calculated and 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* [18] very well and has more than 90% of its total energy in the lower half of subbands.

For spatial coding of MCFD, the RA-SBIC proposed in [18] can be modified as follows:

- Assume that a decomposition tree structure is given and global and local motion parameters are available.
- Using global and local motion parameters, divide each subband of MCFD into three regions – stationary background, moving objects, and newly emerging area.
- For each region of each subband (including the lowest frequency subband), calculate the variance and transmit it.
- For each region of each subband, design an Entropy Constrained Quantizer (ECQ) by modeling the shape of distribution using the Generalized Gaussian Distribution (GGD).

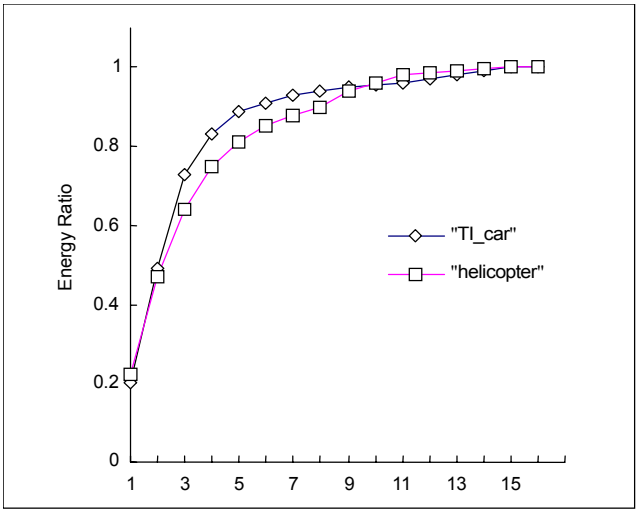


Fig. 3. Examples of cumulative values of *energy ratio* of MCFD.

- Based on the variance and the rate-distortion performance of the designed quantizer, allocate the number of bits for each region of each subband.
- Employ Pulse Coded Modulation (PCM) using the above mentioned quantizer and bit allocation.

As done in [18], we have performed the Kolmogorov-Smirnov(KS)-test to find the best fitting GGD parameter α for 4 × 4, 8 × 8, and 16 × 16 STFT decomposed MCFD subbands and found that they can be modeled by the Laplacian distribution.

The uniform threshold quantizer (UTQ) combined with the Huffman coder (HC) for the Laplacian distribution, designed in

[23], is employed as a method of ECQ and the same bit allocation scheme of [18] can be used here. The variances of regions in each subband are also transmitted using the same method described in [18].

IV. EXPERIMENTAL RESULTS

Simulation results have been obtained for the two sample image sequences. The 16×16 STFT tree decomposition structure has been selected for both image sequences because this structure has shown the best result among all possible tree structures up to 16×16 STFT decomposition and decomposing the MCFD into the smaller sized subbands has gained almost negligible improvements.

For the GMC, the matching window parameter w_m and the searching window parameter w_s are set to 7 and 5, respectively. The threshold value T_g controlling the number of feature points is set to be three times as large as the average of the Sobel operated pixel values in a frame. In our experiments, the number of feature points detected in each frame has been shown to be between 16 and 30.

For our choice of the LMC parameters, the neighborhood window parameter w_n and the threshold for relative difference γ are set to 1 and 0.6, respectively, and the threshold for absolute difference T is chosen as to be three times as large as the average of GMCFD values.

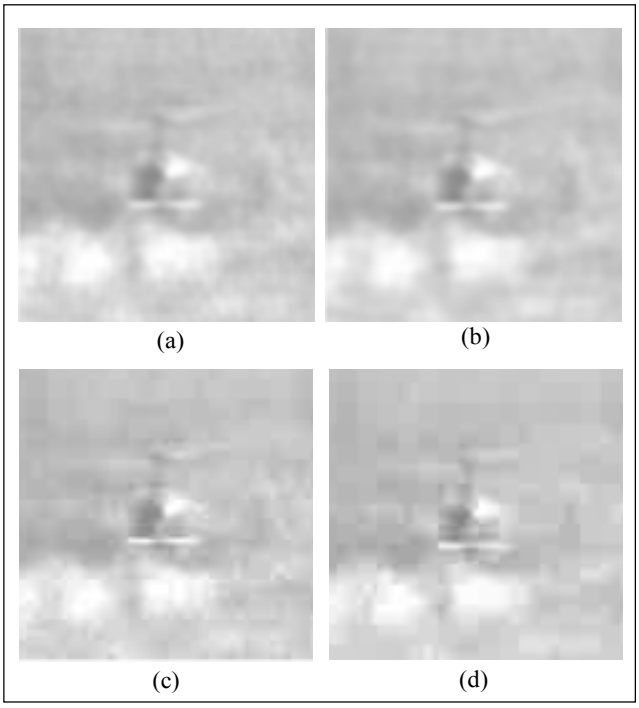


Fig. 5. Enlarged version of simulation results for "helicopter" image sequence: RA-SBIC video coding at (a) 0.1 bits/pel, (b) 0.05 bits/pel; MC-DCT video coding at (c) 0.1 bits/pel, and (d) 0.05 bits/pel.

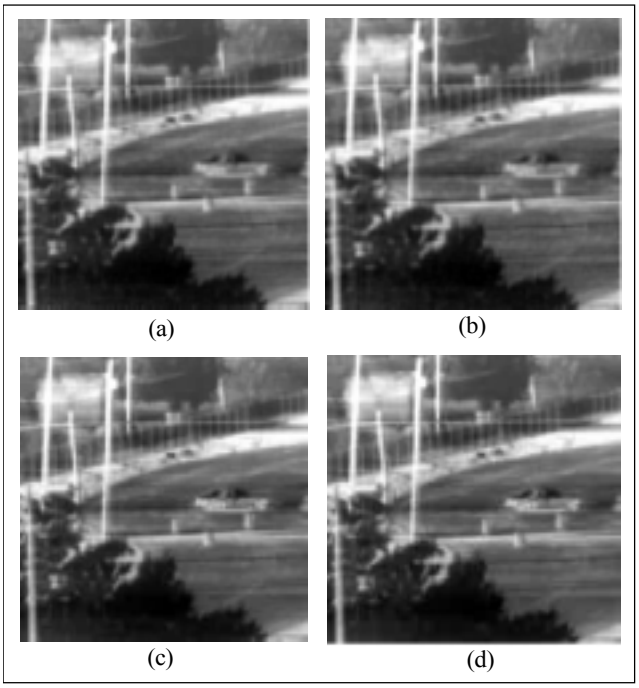


Fig. 6. Simulation results for "TI-car" image sequence: RA-SBIC video coding at (a) 0.5 bits/pel, (b) 0.3 bits/pel; MC-DCT video coding at (c) 0.5 bits/pel, and (d) 0.3 bits/pel.

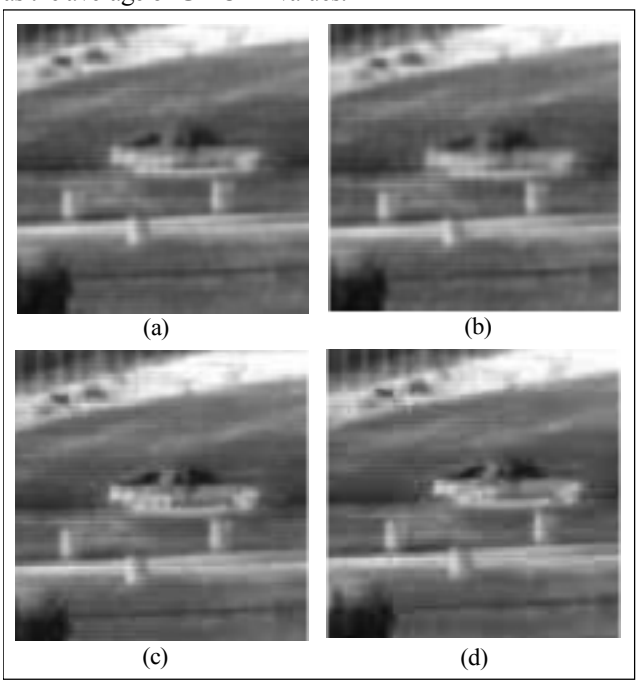


Fig. 7. Enlarged version of simulation results for "TI-car" image sequence: RA-SBIC video coding at (a) 0.5 bits/pel, (b) 0.3 bits/pel; MC-DCT video coding at (c) 0.5 bits/pel, and (d) 0.3 bits/pel.

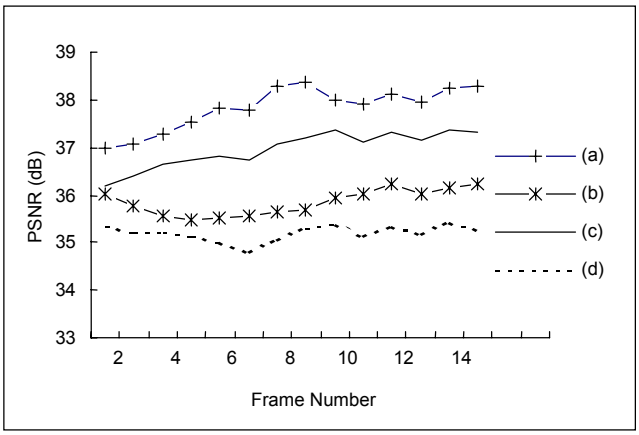


Fig. 8. PSNR results for "helicopter" image sequence: RA-SBIC video coding at (a) 0.1 bits/pel, (b) 0.05 bits/pel; MC-DCT video coding at (c) 0.1 bits/pel, and (d) 0.05 bits/pel.

Figure 4 and 6 show the reconstructed images of the last frames at different rates. The results of conventional MC-DCT video coding are also included for subjective comparisons. Parts of Fig. 4 and 6 including the moving objects are enlarged and displayed in Fig. 5 and 7, respectively. As for the conventional MC-DCT video coding method, the MPEG2 Test Model 5 [24] which uses 8×8 block DCT and generates only the P pictures has been employed. While MC-DCT cod-from blocking effects, the proposed scheme shows blurring effects at low bit rates.

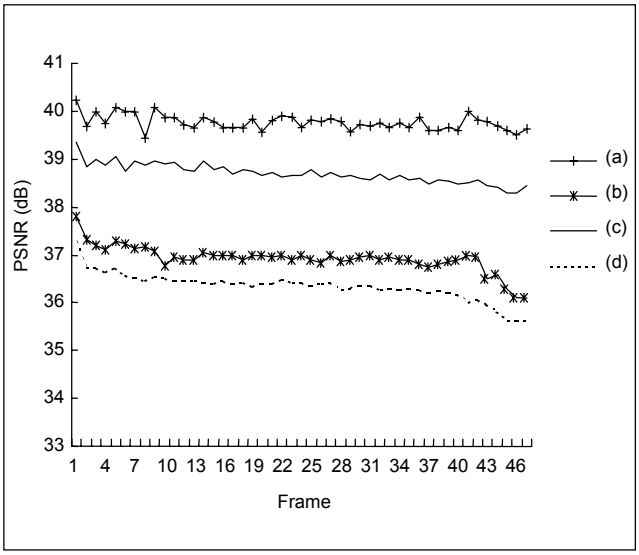


Fig. 9. PSNR results for "TI-car" image sequence: RA-SBIC video coding at (a) 0.5 bits/pel, (b) 0.3 bits/pel; MC-DCT video coding at (c) 0.5 bits/pel, and (d) 0.3 bits/pel.

Using the Peak Signal-to-Noise Ratio (PSNR) measure, objective tests are performed for all frames (Fig. 8 and 9) at various rates. It is seen that more than 1.0 dB improvements on average have been achieved. Compared to the conventional MC-DCT video coding using BMA, our video coding scheme shows better performance.

V. CONCLUSIONS AND FUTURE RESEARCH

The main contribution of this work is the application of region adaptive subband image coding to the video coding. In order to reduce temporal redundancies, both the global and the local motion compensation algorithms are employed. Global motion is compensated by using a feature matching algorithm, and the illumination change between frames is also compensated. Motion compensated frame differences are subdivided into three regions (stationary background, moving objects, and newly emerging areas) each of which has arbitrarily shapes. Local motion of moving objects is also compensated. A novel region adaptive subband image coding is designed for spatial coding of these arbitrarily shaped regions. Experimental results showed up to 1 dB improvement in PSNR over the MC-DCT video coding such as the H.261 and MPEG-2 standards, with also much better subjective quality.

Although the proposed video coding scheme removes blocky effects mostly appearing in the MC-DCT video coding, it suffers from blurring effects at low bit rates. Global and local motion compensation algorithms and methods of subband filterings to reduce blurring effects at low bit rates can be considered as future research topics.

ACKNOWLEDGEMENT

The authors are grateful to Dr. Q. Zheng and Professor R. Chellappa for letting us use their global and local motion compensation program.

REFERENCES

[1] W. Pennebaker and J. Mitchell, *JPEG Still Image Compression Standard,* New York, Van Nostrand Reinhold, 1993.

[2] J.W. Woods, Ed., *Subband Image Coding,* Boston, Kluwer, 1991.

[3] A. Gersho and R.M. Gray, *Vector Quantization and Signal Compression,* Boston, Kluwer, 1992.

[4] M.I. Sezan and R.L. Lagendik, *Motion Analysis and Image Sequence Processing,* Boston, Kluwer, 1992.

[5] B. Girod *et al.,* "Guest editorial: Introduction to Special Issue on Image Sequence Compression," *IEEE Trans. on Image Processing*, vol. 3, Sept. 1994, pp.465-468.

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

[7] D.J. LeGall, "MPEG: A video Compression Standard for Multimedia Applications," *Comm. of the ACM*, vol. 34, no. 4, Apr.

1991, pp.46-58.
[8] P. Pancha and M.E. Zarki, "MPEG Coding for Variable Bit Rate Video Transmission," *IEEE Comm. Magazine*, vol. 32, no. 5, May 1994, pp.54-66.
[9] J. Watkinson, *MPEG-2*, Butterworth-Heinemann, 1999.
[10] R. Koenen, *Overview of the MPEG-4 Standard*, ISO/IEC JTC1/SC29/WG11 N3156, Dec. 1999.
[11] G. Karlsson and M. Vetterli, "Three Dimensional Subband Coding of Video," *Proc. IEEE ICASSP-88*, Apr. 1988, pp.1100-1103.
[12] T. Ebrahimi and M. Kunt, "Image Sequence Coding Using a Three Dimensional Wavelet Packet and Adaptive Selection," *Proc. SPIE Conf. Vis. Comm. and Image Processing*, Nov. 1992, pp.222-232.
[13] Y.Q. Zhang and S. Zafar, "Motion Compensated Wavelet Transform Coding for Color Video Compression," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 2, Sept. 1992, pp.285-296.
[14] P.J. Toutier *et al.,* "Motion Compensated Subband Coding Schemes for Compatible High Definition TV Coding," *Signal Processing: Image Comm*., vol. 4, Aug. 1992, pp.325-344.
[15] J.R. Ohm, "Temporal Domain Sub-band Video Coding with Motion Compensation," *Proc. IEEE ICASSP-92*, Apr. 1992, pp.229-232.
[16] J. Lee and B.W. Dickinson, "Subband Video Coding with Scene-Adaptive Hierarchical Motion Estimation," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 9, no. 3, Apr. 1999, pp.459-466.
[17] F. Kossentini *et al.,* "Rate-Distortion-Constrained Subband Video Coding," *IEEE Trans. on Image Processing*, vol. 8, no. 2, Feb. 1999, pp.145-153.
[18] O. Kwon and R. Chellappa, "Region Adaptive Subband Image Coding," *IEEE Trans. on Image Processing*, vol. 7, May 1998, pp.632-648.
[19] Q. Zheng and R. Chellappa, "Motion Detection in Image Sequences Acquired from a Moving Platform," *Proc. IEEE ICASSP-93*, vol. 5, Apr. 1993, pp.201-204.
[20] A.K. Jain, *Fundamentals of Digital Image Processing*, Englewood Cliff, NJ., Prentice-Hall, 1989.
[21] R. Crane, *A Simplified Approach to Image Processing – Classical and Modern Techniques in C,* Hewlett-Packard Company, Upper Saddle River, NJ., Prentice-Hall, 1997.
[22] G.J. Sullivan and R.L. Baker, "Efficient Quadtree Coding of Images and Video," *IEEE Trans. on IP*, vol. IP-3, May 1994, pp.327-331.
[23] N. Farvardin and J.M. Modestino, "Optimum Quantizer Performance for a Class of Non-Gaussian Memoryless Sources," *IEEE Trans. on IT*, vol. IT-30, May 1984, pp.485-497.
[24] *Coded Representation of Pictures and Audio Information, Test Model 5,* ISO/IEC JTC1/SC29/WG11, 1993.



Oh-Jin Kwon received the B.S. degree from Hanyang University, Seoul, Korea, in 1984, the M.S. degree from the University of Southern California, Los Angeles, in 1991, and the Ph.D. degree from the University of Maryland, College Park, in 1994, all in electrical engineering. From 1984 to 1989, he was a research staff member at the Agency for Defense Development, Korea, and from 1995 to 1999, he was a head of Media lab in Samsung SDS Co., Ltd., Seoul. Since 1999, he has been a faculty member with Sejong University, Seoul, Korea, and is currently Assistant Professor. His research interests are image and video coding, watermarking, and processing.



Seokrim Choi received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, Korea in 1981 and 1983, respectively, and the Ph.D. degree in electrical engineering from Syracuse University, Syracuse, NY. in 1992. In 1983, he was a Researcher at ETRI. From 1983 to 1985, he was with Korea Broadcasting System, Seoul, Korea, and from 1993 to 1996, he was a head of HDTV lab in Hyundai Electronics Industries Co., Ltd. Since 1996, he has been a faculty member with Sejong University, Seoul, Korea, and is currently Professor and Chairman of the Department of Electronics Engineering. His current research interests include image communication, object segmentation, and MPEG.