Multiresolution Motion Compensation in The Wavelet Domain for Video Coding

Chang-Mo Yang, Dong-Keun Lim, Jong-Tae Kim and Yo-Sung Ho

Kwang-ju Institute of Science and Technology (K-JIST)
1 Oryong-dong, Puk-gu, Kwangju, 500-712, Korea
E-mail: {changmo, dklim, jtkim, hoyo}@kjist.ac.kr

ABSTRACT

In this paper, we propose new algorithms for multiresolution motion estimation and motion compensation in the wavelet domain for video coding. We also propose a wavelet-based residual quantization method which exploits the interband relationship among wavelet coefficients and block partitioning. The proposed video coding scheme includes multi-level wavelet decomposition, motion estimation and motion compensation, raster scanning within each subband, formation of block trees, partitioning of block trees, and adaptive arithmetic coding. Although the proposed video coding scheme is simple, it produces bitstreams having SNR scalability from the embedded nature of the wavelet transform. Experimental results demonstrate that the proposed video coding schemes are quite competitive to other wavelet-based video coders in the literature.

1. INTRODUCTION

In most video coding schemes, some type of inter-frame prediction is employed to remove the temporal redundancy existing in video sequences. Motion estimation and motion compensation has been used as an efficient tool for removing temporal redundancy. This is usually achieved by estimating motion vectors of macroblocks between successive frames to realign each frame with respect to the previous one. The video frame can be represented efficiently as a combination of a previous reference frame and the difference between the original and the reference images. Existing video compression standards use two-dimensional discrete cosine transform (2-D DCT) to encode motion-compensated residual errors.

In recent years, the discrete wavelet transform (DWT) has been popularly employed in image and video coding applications. Using the discrete wavelet transform, we can divide the input signal into a number of segments, each corresponding to a different frequency subband. Therefore, processing of each subband can be much easier than processing the whole image frame. In addition, wavelet-based coders overcome blocking artifacts and mosquito noises, which are main problems of DCT-based coding algorithms especially at low bit rates.

Several algorithms have been proposed to predict wavelet coefficients by motion compensation in the wavelet domain. Zhang and Zafar proposed an algorithm for multiresolution motion estimation and motion compensation (MRME/MRMC) [1]. In their scheme, each video frame is decomposed into multiple layers of different resolutions. Lower resolution layers are smoothed versions of higher resolution layers. Motion activities at different layers are highly correlated since they actually characterize the same motion structure at different resolution scales and different frequency ranges. From this property, MRME/MRMC schemes significantly reduce the searching and matching time for motion vectors and produce a smooth motion vector field. Their motion compensation is based on a block-matching algorithm (BMA) in each layer.

In three-dimensional (3-D) subband coding schemes [2,3], 2-D image coding techniques are extended to 3-D video coding. They utilized an embedded zerotree wavelet (EZW) [4] coding scheme to encode error frames. Tham and Ranganath proposed a 3-D subband video coding scheme for very low bit-rate applications [2]. After they perform 3-D wavelet decomposition of a group of video frames, they encode important wavelet coefficients using a data structure, called tri-zerotrees (TRI-ZTR).

In a 3-D set partitioning in hierarchical trees (3-D SPIHT) [3] algorithm, they extend the 2-D SPIHT algorithm [5] into 3-D video coding. In 3-D SPIHT, 3-D spatial-temporal orientation trees are coupled with SPIHT sorting and refinement. They also extend the scheme to color-embedded video coding without explicit bit allocation, which can be used for color plane representation.

Although 3-D subband coding schemes have reduced blocking effects and mosquito noises significantly, they also have some drawbacks. They generally have blurred moving objects in temporally low-resolution pictures due to the averaging effect of low-pass filtering.

Motion-compensated 3-D subband video coding techniques have been proposed to alleviate this problem. Besides, they improve compression efficiency by concentrating signal energy in the temporal low subband. However, they require large frame memory for 3-D subband decomposition because these techniques operate on several consecutive frames. As the target number of temporal layers increases, the number of required frames increases exponentially and unavoidable excessive encoding and decoding delay is introduced.

In this paper, we propose a new wavelet-based video coding scheme which includes wavelet decomposition, motion estimation and motion compensation, raster scanning within each subband, formation of block trees, partitioning of block trees, and adaptive arithmetic coding. After applying an MRME/MRMC operation, we encode residual errors by a new quantization method. For entropy coding, we use an adaptive arithmetic coder that is employed in JPEG. In order to control the bit rate, we exploit the embedded property of the wavelet transform during the residual quantization.

2. SYSTEM OVERVIEW

Motion estimation and motion compensation play a very important role in video coding and generally determines the performance and complexity of the coder. In this section, we propose new algorithms for multiresolution motion compensation and residual quantization.

Fig. 1 shows a video coding scheme using MRME. Once motion information in the current frame is estimated in the discrete wavelet transform domain, we do not need to perform the inverse discrete wavelet transform to obtain motion information for the next frame. Therefore, its computational complexity is relatively lower than conventional DCT-based video coding schemes where motion information is performed in the spatial domain and the inverse discrete cosine transform is required.

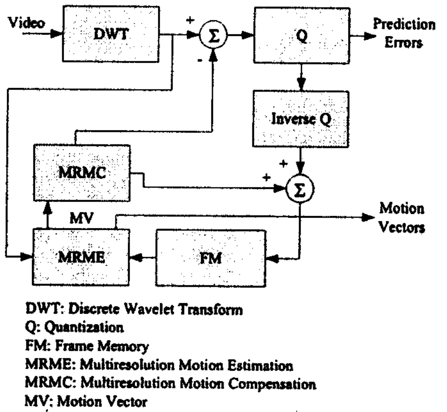

Figure 1. MRME-based Video Coder

As shown in Fig. 1, after the input frame is decomposed into subbands by the discrete wavelet transform, MRME/MRMC is performed in the wavelet domain and then prediction errors are quantized and entropy coded. Finally, the motion compensated frame is added back to the reconstructed error frame to make the reference frame for the next frame.

3. MOTION ESTIMATION

3.1. Concept of Multiresolution Representation

Block transform coding schemes suffer from blocking artifacts at low bit-rates. Subband decomposition provides an alternative approach to represent nonstationary video signals and residual signals after motion compensation. Subband representation is flexible and can be easily adapted to the nature of the human visual system. It is also free from blocking artifacts due to the nature of its global operation.

Using the wavelet transform, we can decompose a video frame into a number of subbands with different resolutions, each corresponding to a different frequency subband. These multiresolution subbands also provide a representation of motion structure at different scales.

In MRME schemes, motion vectors in higher resolutions are predicted by the motion vectors in the lower resolution and refined at each step. Although we can use a different coding scheme in each layer, we apply a general block-matching motion estimation algorithm in the wavelet domain, instead in the spatial domain [6-8].

In the block-matching motion estimation algorithm, each block in the current frame is matched to a block within the search area of the previous frame to find the displacement of the block. In order to find the best matching position, we need to calculate matching measures for all candidate blocks. Then, we choose the best candidate as the motion vector of the block. Generally, MSE (Mean Squared Error) and MAD (Mean Absolute Difference) are employed as matching criteria. Because of its simplicity, we use MAD as a matching criterion in this work.

Let $I_t(k,l)$ be the luminance pixel values in the current frame, and $I_{t-1}(k,l)$ be the luminance pixel values in the previous frame. MAD can be defined as

$$MAD(dx,dy)=\frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\left|I_t(k,l)-I_{t-1}(k+dx,l+dy)\right| \quad (1)$$

where (dx,dy) is the motion vector and N is the block size.

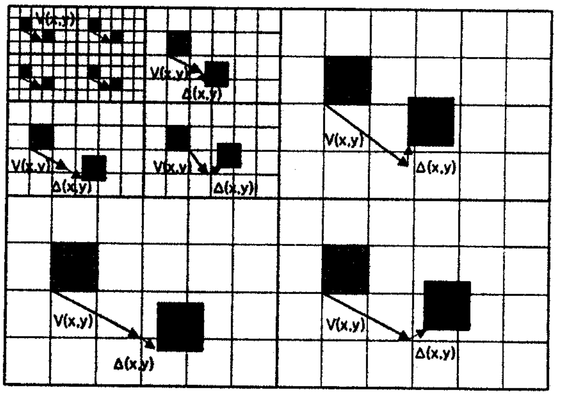

Figure 2. Multiresolution Motion Estimation

In the multiresolution representation, a video frame can be decomposed into multiple layers of different resolutions and different frequency subbands by the wavelet decomposition, as shown in Fig. 2. Motion activities at different layers and subbands are highly correlated since they actually characterize the same motion structure at different scales and frequency ranges. In multiresolution motion estimation, motion vectors are initially estimated in the lowest subband. Then, the motion vector is used as a basis for motion vectors of higher subbands. The motion vector in the lowest subband is scaled appropriately and used as the initial motion estimate in higher subbands.

The lowest subband has only the 1/64 size of the original image in the three-level wavelet decomposition. Although its size is small, it has a large portion of the signal energy of the original image. Therefore, detailed motion estimation is needed in the lowest subband. In addition, errors in motion vectors at the lowest subband will be propagated and expanded to all subsequent higher subbands. In order to increase the accuracy of motion estimation, we can use a sub-pixel motion search in the lowest band. However, it does not improve compression efficiency due to increased overheads for motion vectors [6].

3.2 Structure of Multiresolution ME/MC

In our work, a video frame is decomposed by the wavelet transform, as shown in Fig. 3. It is well known that human vision is more perceptible to errors in lower subbands than those incurred in higher subbands. As shown in Fig. 3, the subband S_8 has only 1/64 of the original size, but contains a large percentage of the total signal energy. In multiresolution motion estimation algorithm, motion vectors are initially estimated for the lower subbands, S_8, W_8^V, W_8^H, and W_8^D, where V, H and D indicate the band for vertical, horizontal and diagonal locations, respectively.

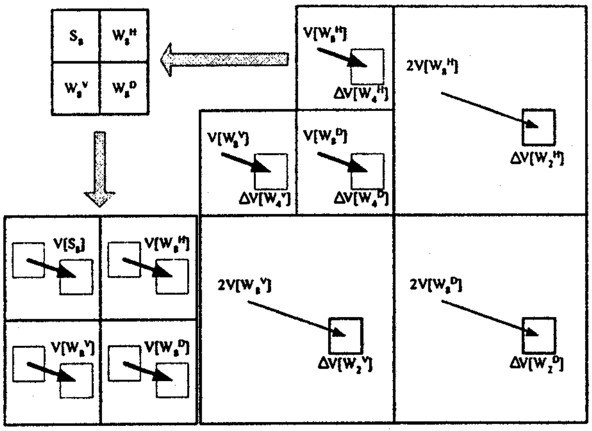

Figure 3. MRME Procedure

Motion vectors in higher subbands are refined by motion information obtained in lower subbands. During this operation, errors in lower subbands can be propagated and expanded to all subsequent higher subbands.

In conventional wavelet coding algorithms, they used variable block-size MRME/MRMC, with consideration of human vision characteristics. Smaller block size is used in lower subbands, while bigger block size is used in higher subbands. Therefore, numbers of motion vectors in each subband are the same.

In the proposed algorithm, block sizes of S_8 and W_8 subbands are 2×2, while those of W_4 and W_2 subbands are 4×4, and 8×8, respectively. The number of blocks are (width/16)× (height/16) for each subband. One pixel shift in S_8 is corresponding to 8-pixel shift in the spatial domain. In the lowest subband, motion vectors are within [-2, 2] pixels. Using the sub-pixel motion estimation method, we can increase the accuracy of motion compensation for sub-images. We use 1/2-pixel motion estimation in subbands S_8, W_8^V, W_8^H, and W_8^D, respectively.

After we decompose the previous and the current frames to ten subbands by the three-level wavelet decomposition, S_8 and three W_8 subbands of the current and the reference frames are interpolated by two. While conventional methods interpolate only the previous frame, the propose method interpolates both the previous and the current frames. We adopt the general block matching algorithm, where the search area is [-4,4] pixels. We can find motion vectors using the MAD matching criterion. Motion vectors in S_8 subband are estimated by

$$V(W_8{}^j)_{j=V,H,D} = ME[f_n(W_8{}^j), f_{n-1}(W_8{}^j)]_{i=V,H,D} \quad (2)$$

For other subbands, we predict the initial motion vector by scaling the motion vector in the lower subband. We exploit the parent-child relationship [4,5] of wavelet coefficients because motion activities in lower subbands are highly correlated to those of higher subbands. We can represent this operation by

$$E\left[V_{2^i}^j(W_{2^i}^j)\right]_{\substack{i=1,2\\ j=V,H,D}} = 2^{M-i-1} V_M^j(W_8^j) \quad (3)$$

where i and j means the decomposition level and the position of subbands, respectively.

The next step is motion vector refinement. For neighboring pixels of the estimated location, we find motion vectors in the search area Ω by

$$\Delta V_{2^i}^j(W_{2^i}^j) = \left\{(dx, dy) \mid \min\left[\sum_{p\in\Omega}\sum_{q\in\Omega} | I_t(x,y) - I_{t-1}(x+p, y+q)\right]\right\}_{\substack{i=1,2\\ j=V,H,D}} \quad (4)$$

Finally, we can get $V_{2^i}^j(x,y)$ by

$$V_{2^i}^j(W_{2^i}^j)_{\substack{i=1,2\\ j=V,H,D}} = E\left[V_{2^i}^j(W_{2^i}^j)\right] + \Delta V_{2^i}^j(W_{2^i}^j) \quad (5)$$

The multiresolution motion estimation algorithm can be summarized as follows:

① Apply wavelet decomposition to the reference and the current frames.
② Interpolate the lowest subband by 2, and then estimate motion vectors using block-matching motion search.
③ Set initial motion vectors for other subbands by Eq. (3).
④ Refine the motion vectors $V(W_8^j)$ by Eq. (4).
⑤ Set motion vectors in each subband by adding the initial and refined motion vectors by Eq. (5).

Fig. 4 shows the overall procedure for multiresolution motion compensation. MRMC is performed in the similar way as MRME.

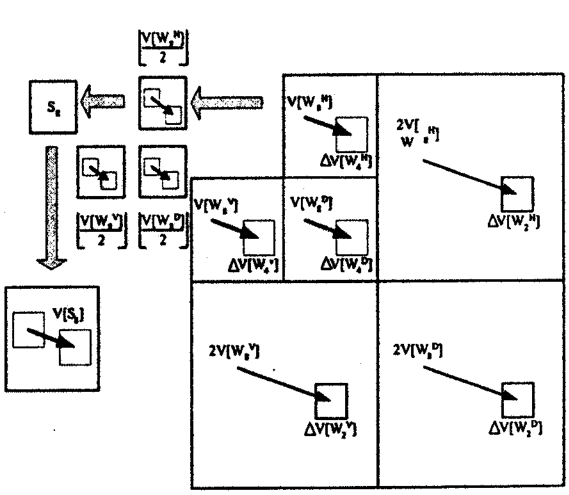

Figure 4. MRMC Procedure

4. RESIDUAL QUANTIZATION

4.1. Motivation

In this section, we describe a new quantization method. After the residual image frame is decomposed into several frequency subbands, we divide each subband into small blocks and construct block trees using interband magnitude correlation among wavelet coefficients. Significant coefficients in block trees are encoded by a set partitioning algorithm and an adaptive arithmetic coder.

Once the input image is decomposed into subbands, one wavelet coefficient at a given scale is related to other coefficients corresponding to the same spatial location at the same scale but different orientations. We call the other coefficients as cousins of the given coefficient. Therefore, except for the lowest frequency subband, every coefficient has two cousins. For the lowest frequency subband, no coefficient has cousins. We define this relationship as the coefficient-cousin relationship. Fig. 5 illustrates this interband magnitude relationship.

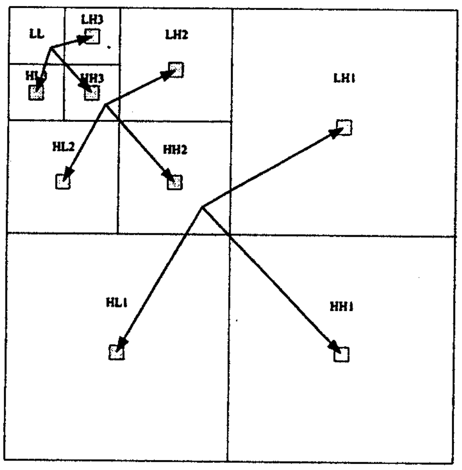

Figure 5. Relationship among Cousins

In order to construct block trees using the coefficient-cousin relationship, we divide each subband into small blocks, nominally with the dimension of 64×64 pixels. After each subband is divided into small blocks, we can construct block trees using the relationship among cousins, as shown in Fig. 6.

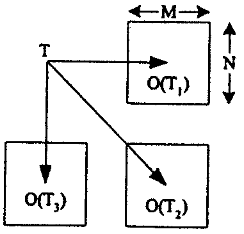

Figure 6. Formation of Block Trees

Once we construct block trees, we identify the significance of a tree T using $c_{i,j}$, the coefficient in position (i,j), and quantization step n by

$$S_n(T)=\begin{cases}1, & if\ 2^n \le \max\limits_{(i,j)\in T} |c_{i,j}| < 2^{n+1}\\ 0, & else\end{cases} \tag{6}$$

where "1" means significant and "0" means insignificant with respect to n. Similarly, we identify the significance of a coefficient $c_{i,j}$ by

$$S_n(c_{i,j})=\begin{cases}1, & if\ 2^n \le |c_{i,j}| < 2^{n+1}\\ 0, & else\end{cases} \tag{7}$$

When a tree T is identified to be significant with respect to n, it is partitioned into four small trees of the same size by the partitioning operation, as shown in Fig. 7.

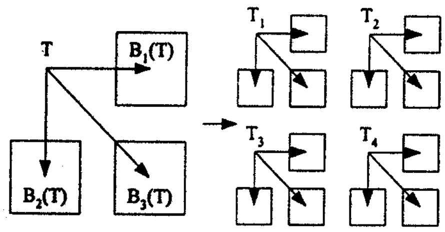

Figure 7. Set Partitioning in Block Trees

The significant tree is recursively split until the size of blocks $B_i(T)$, $i\in\{1,2,3\}$, is 4×4. When the size of blocks $B_i(T)$, $i\in\{1,2,3\}$, is 4×4, all coefficients in T are encoded. The main motivation for the proposed partitioning is to find high-energy regions quickly and encode them first.

4.2. The Proposed Quantization Algorithm

In the proposed quantization algorithm, the significance information is stored in three ordered lists: list of insignificant pixels (LIP), list of insignificant blocks (LIB), and list of significant pixels (LSP). For all the lists, each entry is identified by a coordinate (i,j) that represents an individual pixel in LIP and LSP, and represents the starting position of the tree in LIB. The proposed algorithm consists of four coding passes: initialization, sorting pass, refinement pass, and update of the quantization step.

In the initialization pass, we divide each subband into smaller blocks and construct block trees. Then, we calculate the initial quantization step n in the transform image X as follows:

$$n=\left\lfloor \log_2\left(\max_{\forall(i,j)\in X} |c_{i,j}|\right)\right\rfloor \tag{8}$$

After we determine entries of LIP, LIB, and LSP, all coefficients in the lowest subband are used as initial LIP entries, and all block trees are used as initial LIB entries. LSP are set to null in this initialization pass.

In the sorting pass, we identify the significance of pixels and trees in LIP and LIB. When a pixel in LIP is identified to be significant, the sign of the pixel is also identified, and the pixel is moved to LSP. Similarly, we evaluate the significance of trees in LIB sequentially. When a tree T in LIB is identified to be significant, it is partitioned according to the partitioning rule, and removed from LIB. All partitioned small trees T_i, $i \in \{1,2,3,4\}$, are added back to LIB. When the size of blocks $B_i(T)$, $i \in \{1,2,3\}$, is 4×4, significance of all pixels in the tree T is identified and the tree is removed from LIB. If a pixel is identified to be significant, the sign of the pixel is identified and the pixel is added to LSP. Otherwise, the pixel is added to LIP.

In the refinement pass, the n-th most significant bit (MSB) of entries in LSP is identified with respect to n. In this pass, we do not consider entries included in the last sorting pass with the same n.

In the update of quantization step, we decrease the quantization step n by one, and repeat the coding step from the sorting pass.

5. EXPERIMENTAL RESULTS

Experiments are performed with a Pentium 800 MHz personal computer with 128MBytes RAM. Each frame of test sequences is decomposed by dyadic 9/7-tap wavelet filters. For the entropy coding, we use an adaptive arithmetic coder with seven different contexts for encoding pixels and four contexts for encoding signs. We encode the significance identification of each tree, the refinement bit and motion vectors with one fixed context. For unavailable neighboring coefficients, the corresponding context bits are set to zero. As a performance measure, we use the peak signal-to-noise ratio (PSNR).

5.1. Performance of Motion Estimation

Fig. 8 shows motion-compensated images of FOOTBALL. We compare image quality of the proposed motion compensation to that of the spatial domain motion compensation. The motion estimation is performed by block matching with 16×16 pixel blocks and search area within [-16,16] pixels. As shown in Fig. 8, the motion compensated image in spatial domain has some blocking artifacts. On the other hand, there are some smoothing effects in the motion compensated image in the wavelet domain.

5.2. Performance of Intra-Frame Coding

Fig. 9 compares performances of intra-frame coding by the proposed and SPIHT schemes. For performance comparison, two popular monochrome images, BOATS and BARBARA, of size 576×720 pixels are used. Fig. 9 demonstrates that the proposed scheme provides good results and often outperforms SPIHT. For BOAT, the proposed scheme provides slightly better objective quality than SPIHT. However, the proposed scheme significantly outperforms SPIHT for BARBARA mainly because it has a lot of high-frequency components.

5.3. Performance of Video Coding

We apply video coding algorithms to two popular 4:2:0, 30-fps color sequences, MOTHER AND DAUGHTER and HALL MONITOR, of size 176×144 pixels. We encode the video sequence at 30 and 60 kbps (kbits per second) with 10 fps (frame per second). Since the original test sequences have 30 fps, every third frame is encoded and reconstructed. Fig. 10 and Fig. 11 show performances of the proposed video coding schemes at 30 and 60 kbps, respectively. Fig. 10 and Fig. 11 demonstrate that the proposed coding scheme provide good performance. Table 1 lists average PSNR values of the proposed and 3-D SPIHT [3] video coding schemes at 30 kbps and 60 kbps, respectively. As shown in Table 1, the proposed video coding algorithm provides slightly higher PSNR values than the 3-D SPIHT algorithm.

(a) Motion-Compensated Image in the Spatial Domain

(b) Motion-Compensated Image in the Wavelet Domain

Figure 8. Motion-Compensated Images

6. CONCLUSIONS

In this paper, we have proposed a new video coding scheme based on a new motion estimation and quantization methods. In order to exploit temporal redundancies in video sequences, we use multiresolution motion estimation. The prediction error is quantized by a new method exploiting the relationship among wavelet coefficients. The embedded property of the wavelet decomposition enables an accurate rate control. The proposed video coding algorithm has a low computational complexity. Although it is simple, the proposed scheme performs competitively to other wavelet-based video coding schemes. In addition, the proposed algorithm is basically free from blocking artifacts, and suitable for scalable video coding.

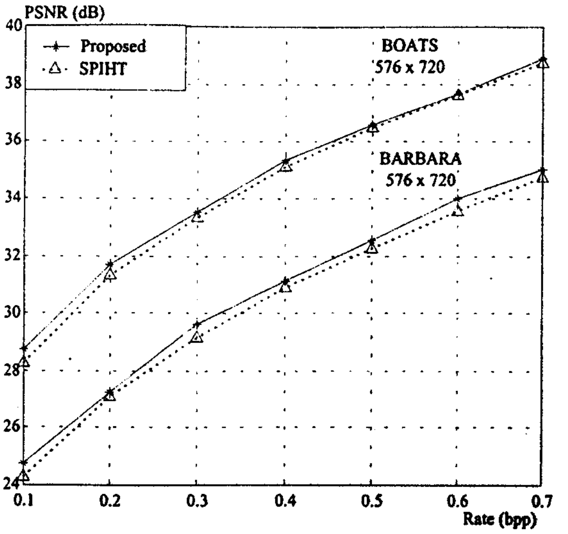

Figure 9. Performance for Intra-Frame

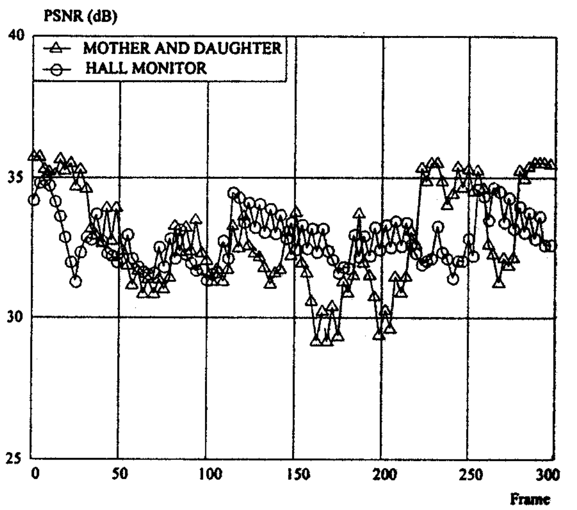

Figure 10. Performance at 30 kbps

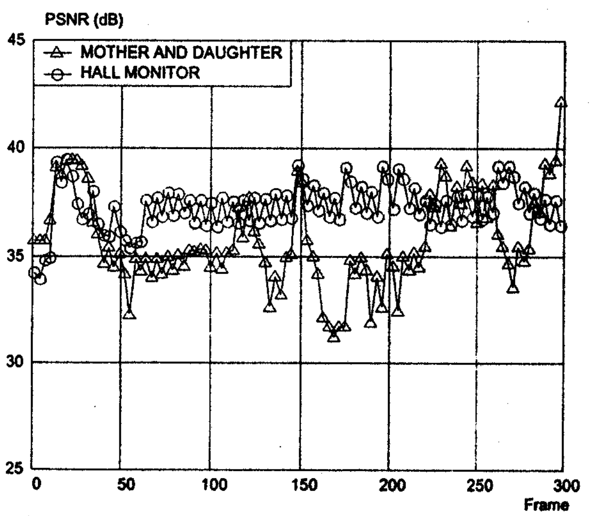

Figure 11. Performance at 60 kbps

Table 1. Average PSNR Values

Coding Scheme	Bit Rate	MOTHER AND DAUGHTER	HALL MONITOR
MRMC-RQ (Proposed)	30 kbps	32.85 dB	32.86 dB
	60 kbps	35.70 dB	37.31 dB
3-D SPIHT	30 kbps	32.71 dB	32.95 dB
	60 kbps	35.57 dB	37.95 dB
MC 3-D SPIHT	30 kbps	32.78 dB	32.30 dB
	60 kbps	35.69 dB	37.36 dB

7. ACKNOLEDGEMENT

This work was supported in part by the Korea Science and Engineering Foundation (KOSEF) through the Ultra-Fast Fiber-Optic Networks (UFON) Research Center at Kwangju Institute of Science and Technology (K-JIST), and in part by the Ministry of Education (MOE) through the Brain Korea 21 (BK21) project.

REFERENCES

[1] Y. Q. Zhang and S. Zafar, "Motion Compensated Wavelet Transcoding for Color Video Compression", *IEEE Trans. on Video Tech.*, vol.2, no.3, pp.285-296, Sept. 1992.

[2] J. Y. Tham, S. Ranganath and A. A. Kassim, "Highly Scalable Wavelet-based Video Codec for Very Low Bit-Rate Environment", *IEEE Journal on Selected Areas in Comm.*, vol. 16, no. 1, pp. 12-27, Jan. 1998.

[3] B. J. Kim, Z. Xiong and W. A. Pearlman, "Low Bit-Rate Scalable Video Coding with 3-D Set Partitioning in Hierarchical Trees", *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 10, no. 8, pp. 1374-1387, Dec. 2000.

[4] J. Shapiro, "Embedded Image Coding using Zerotrees of Wavelet Coefficients", *IEEE Trans. on Signal Proc.*, vol. 41, no. 12, pp. 3445-3462, Dec. 1993.

[5] A. Said and W. Pearlman, "A New, Fast and Efficient Image Codec based on Set Partitioning in Hierarchical Trees", *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 6, no. 3, pp. 243-250, June 1996.

[6] S. Zafar, Y. Q. Zhang and B. Jabbri, "Multiscale Vide Representation using Multiresolution Motion Compensation and Wavelet Decomposition", *IEEE Trans. on Selected Areas in Comm.*, vol.11, no.1 pp.24-35, Jan. 1993.

[7] F. G. Meyer, A. Averbuch and R. R. Coifman, "Motion Compensation of Wavelet Coefficient for Very Low Bit Rate Picture Coding," *Int'l Conf. on Image Proc.*, pp.606-609, Oct. 1995.

[8] H. W. Park and H. S. Kim, "Motion Estimation using Low-Band Shift Method for Wavelet-Based Moving Picture Coding", *IEEE Trans. on Image Proc.*, vol.9, no.4, pp.577-587, April 2000.