Hindawi Publishing Corporation
EURASIP Journal on Applied Signal Processing
Volume 2006, Article ID 71643, Pages 1–8
DOI 10.1155/ASP/2006/71643

Fast Motion Estimation and Intermode Selection for H.264

Byeong-Doo Choi, Ju-Hun Nam, Min-Cheol Hwang, and Sung-Jea Ko

Department of Electronics Engineering, Korea University, Anam-Dong, Sungbuk-Ku, Seoul 136-701, South Korea

Received 1 August 2005; Revised 5 June 2006; Accepted 11 June 2006

H.264/AVC provides various useful features such as improved coding efficiency and error robustness. These features enable mobile devices to adopt H.264 standard to achieve effective video communications. However, the encoder complexity is greatly increased mainly due to motion estimation (ME) and mode decision. In this paper, we propose a new scheme to jointly optimize intermode selection and ME using the multiresolution analysis. Experimental results show that the proposed method is over 3 times faster than other existing methods while maintaining the coding efficiency.

1. INTRODUCTION

Recent advances in wireless communication technology have introduced various mobile services such as multimedia message services, video on demand, and mobile video communications. Especially, there are increasing demands on mobile video communications with prevailing demands of the mobile devices equipped with camera module. To realize this service, video sequences have to be compressed with high coding efficiency and error robustness. The H.264/AVC is the state-of-the-art video compression standard recently developed by the ITU-T/ISO/IEC Joint Video Team [1].

The H.264/AVC provides various useful features such as improved coding efficiency, error robust data partitioning, and network friendliness with the network abstraction layer (NAL). These features enable mobile devices to adopt the H.264/AVC standard to achieve effective video communications [2]. H.264/AVC supports multiple reference frames and various block sizes for ME. It uses tree-structured hierarchical macroblock (MB) partitions. There are 7 different block sizes (16×16, 16×8, 8×16, 8×8, 8×4, 4×8, and 4×4 blocks) that are used in a macroblock. The current H.264/AVC reference software is based on a rate-distortion optimization (RDO) framework for both ME and mode decision.

Among all modules in the H.264/AVC encoder, ME and mode decision require a heavy computation, especially when RDO is used. ME has to be performed for every MB coding mode to find the best matching block. For mode decision, all possible combinations of coding modes are considered to obtain the MB with minimum cost. Moreover, since these operations are performed in multiple reference frames, the computational load significantly increases at the encoder.

The computational burden of ME can be reduced by applying fast ME methods, such as the three-step search [3], the four-step search [4], the diamond search [5], and the hexagon search [6]. Recently, uneven multihexagon search (UMHexagonS) has been adopted for the fast ME in the H.264/AVC encoder reference software (JM 84) [7]. It reduces significantly the processing time of ME while maintaining the coding efficiency, but is designed without considering the multiple reference frames.

To optimize the mode decision process, the H.264/AVC software adopts the full mode decision algorithm (FMD) inducing an exhaustive computation [8]. For fast mode decision, the early termination technique [9] reduces the number of potential prediction modes. In [10, 11], the classification methods are proposed to reduce the average number of block types while maintaining the coding performance, but require the additional processing of edge detection.

In this paper, we propose new fast ME and intermode selection techniques using the multiresolution analysis for the H.264/AVC encoder. The proposed method is based on the multiframe/multiresolution ME using hexagon searching (MFMRME-HS). For the fast intermode selection (FIMS), we introduce a bottom-up merge method using a hypothesis testing. The proposed method first splits all MBs into 4×4 sub-MBs and then merges the sub-MBs when they are classified as the same class by using a new hypothesis-and-test based method.

The organization of the paper is as follows. The proposed fast ME method is introduced in Section 2. In Section 3, the

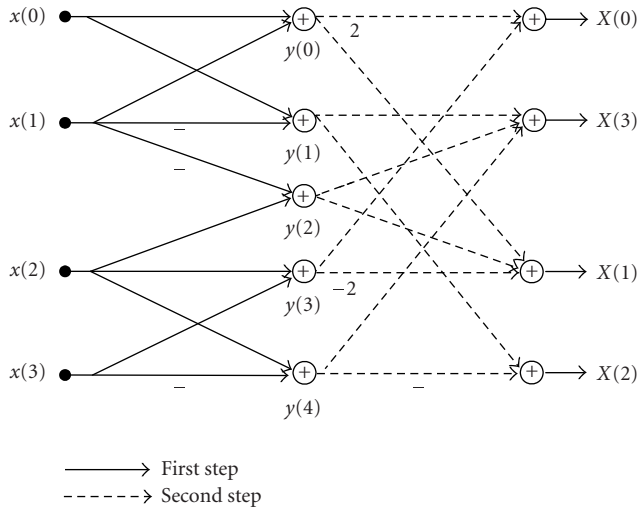


Figure 1: Modified implementation of fast integer transform.

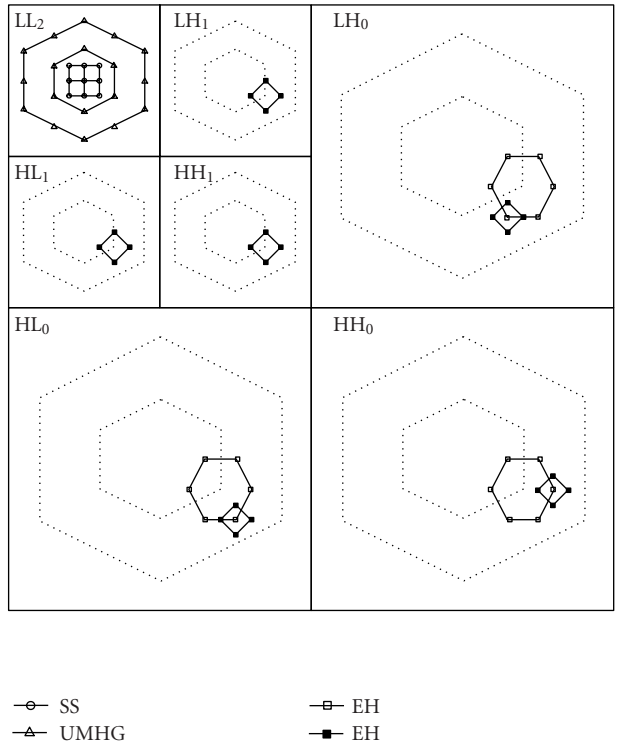


Figure 2: Search pattern for multiresolution.

proposed fast intermode decision method is given. Finally, the simulation results are shown in Section 4 and the conclusion is described in Section 5.

2. THE PROPOSED MOTION ESTIMATION METHOD

The multiresolution ME (MRME) technique is a fast ME method that is an alternative to the conventional block-matching algorithm. In the conventional MRME, motion vector (MV) is estimated at the lowest resolution (LL band) and then the estimated MV is appropriately scaled to be used as an initial bias and refined on the remained subbands of wavelet transform. By using the MRME scheme, we can reduce the number of search points while maintaining the accuracy of the estimated MV. However, adopting MRME for the H.264/AVC encoder requires discrete wavelet transform (DWT) as well as fast integer transform (FIT), which results in additional computations. The encoding architecture consisting of two different transforms is not efficient in terms of system optimization due to its additional computations and increased memory size.

In order to overcome this problem, we propose a modified FIT. Figure 1 shows the proposed four-tap modified FIT. Compared with Malvar's optimal FIT approach [12], it requires a little additional computations and memory spaces. However, its resultant coefficients include the coefficients from both FIT and three-level Haar wavelet transform. For MRME, reference frames and a current frame should be decomposed into multilayers. After processing the modified FIT, the resultant coefficients are saved separately in the memory space of each corresponding layer. These coefficients are reused when we perform MRME for following frames. Thus, by adopting the proposed modified FIT, we can decompose the frames into three layers without additional DWT.

The detailed procedure of the modified FIT is as follows. The first step of the modified FIT decomposes 4 pixels into 5 intermediate coefficients. The two intermediate coefficients $y(1)$ and $y(4)$ become the coefficients for the H_0 layer. In the second step, $X(0)$ and $X(1)$ become the coefficients for the L_2 and H_1 layers, respectively. For MRME, the resultant coefficients are reallocated for each layer in Figure 2. In addition, the high frequency coefficients in multilayers are utilized in the proposed mode decision method. The four coefficients from the modified FIT, $X(0)$, $X(1)$, $X(2)$, and $X(3)$, are exactly the same as those from the original FIT.

With the modified fast integer transform, we propose a fast ME algorithm using MRME, hexagon searching, and multiframe reference. The proposed method exploits the cross-correlation among the multilayers of the wavelet transform on multiframes, to reduce the computational complexity. It can achieve a smaller number of search points over other fast methods and can maintain similar or even smaller distortion error.

Figure 2 shows the proposed searching patterns: the small square (SS), the uneven multihexagon grid (UMHG), and the extended hexagon (EH). The SS and UMHG searching patterns are applied to find the coarse MV at layer 2. The EH is used to refine the MV at lower layers.

Figure 3 does show the concept of the proposed MFMRME-HS (multiframe/multiresolution ME using hexago nal search) method consisting of four steps. The detailed procedure is followed.

Step 1. Search the MV in the object region with small motion by using the SS searching pattern at layer 2 of the reference

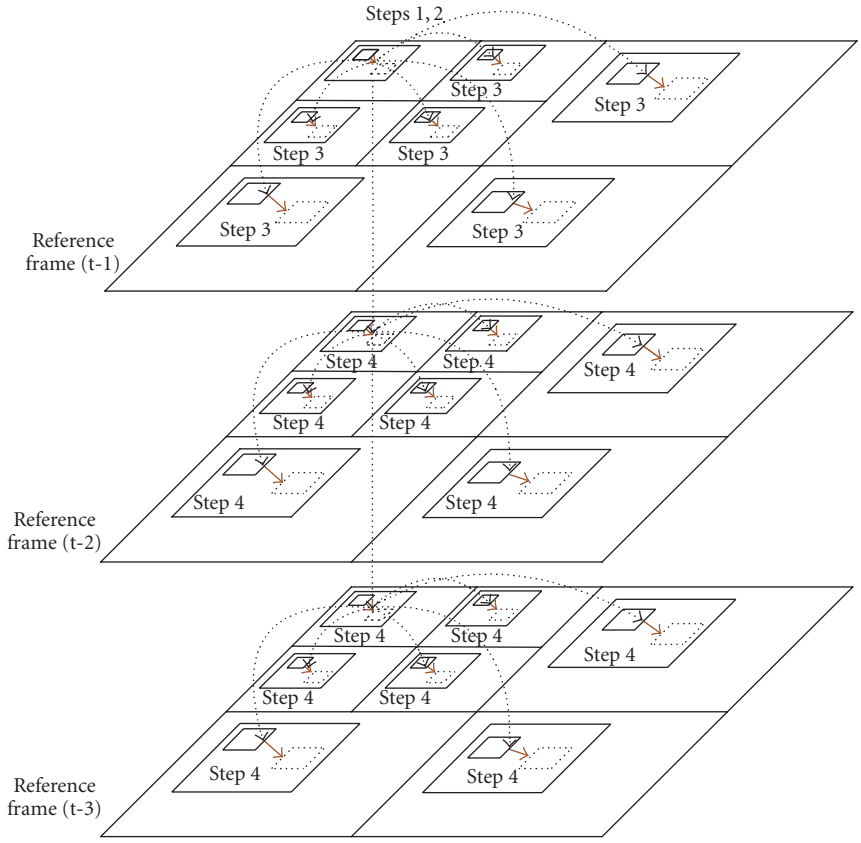


Figure 3: Multiframes/multiresolution motion estimation scheme.

frame (t-1) shown in Figure 3. If the minimum sum of absolute difference (SAD) is smaller than an initial threshold, go to Step 3 to perform MV refinement at lower layers. Otherwise, go to Step 2 to keep searching the MV at layer 2. The MV obtained from Step 1 becomes the initial search center for the next step.

Step 2. Find the MV by using the UMHG pattern at layer 2 of the reference frame (t-1) shown in Figure 3. The MV obtained from Step 2 becomes the search center for the next step, and its corresponding minimum SAD becomes the threshold for the fast reference frame selection (FRFS) in Step 4.

Step 3. Use the EH to refine the MV obtained from Steps 1 and 2 around the search center at lower layers. The MV refined in this step becomes a candidate of the best MV.

Step 4. Steps 2 and 3 are iterated at all the remained reference frames. For the FRFS, if the minimum SAD at layer 2 is over a threshold, MV refinement (Step 3) is not performed for its corresponding reference frame. As a result, this FRFS improves the performance of the proposed MFMRME-HS. Finally, select the best MV with the minimum SAD among all candidate MVs obtained from Step 3.

The proposed MFMRME-HS algorithm utilizes the correlation between multiframes and subbands of the wavelet transform to reduce the complexity of ME.

3. THE PROPOSED FAST INTERMODE SELECTION METHOD

In the H.264/AVC, there are totally 7 different block sizes (16×16, 16×8, 8×16, 8×8, 8×4, 4×8, and 4×4) that are utilized in the variable size. The reason for adopting seven different block sizes in H.264/AVC is to represent more accurate motion field of moving objects to reduce the residual error. In general, the macroblocks in a moving object have the same MVs and homogeneous regions.

In the intermode RDO implementation of H.264/AVC, ME is performed for all the possible block sizes to find the one with the least rate-distortion cost. The intermode decision process using Lagrange multiplier requires an extremely large time consumption, minimizing the following cost function:

$$J(s,c,\mathrm{MODE}\mid \mathrm{QP},\lambda_{\mathrm{MODE}}) = \mathrm{SSD}\,(s,c,\mathrm{MODE}\mid \mathrm{QP}) + \lambda_{\mathrm{MODE}}\cdot R(s,c,\mathrm{MODE}\mid \mathrm{QP}), \quad (1)$$

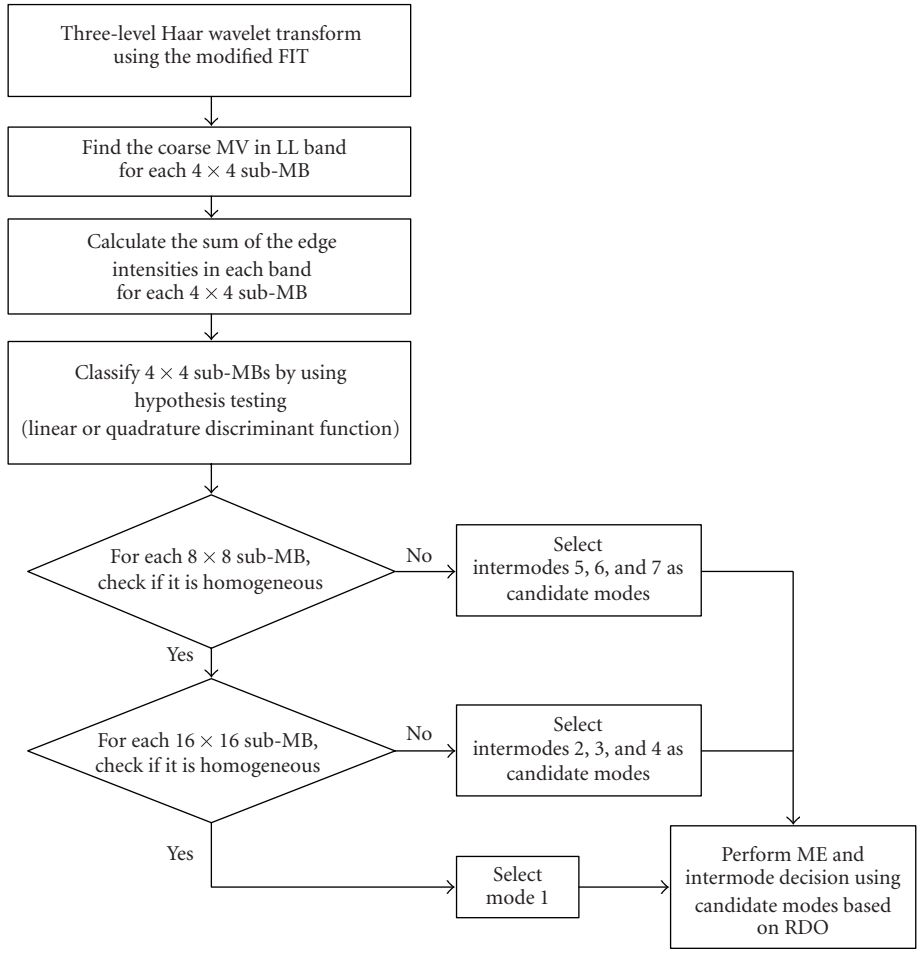


Figure 4: Flowchart of fast intermode decision by classifying sub-MBs.

where s and c are the source video signal and the reconstructed video signal, respectively, QP is the quantization parameter, λ_{MODE} is the Lagrange multiplier, SSD is the sum of the squared differences between s and c, MODE indicates an MB mode.

In this section, we propose a bottom-up merge method for fast intermode selection (FIMS). Figure 4 shows the proposed bottom-up merge method. The proposed FIMS splits all 16×16 MBs into 4×4 sub-MBs and determines the class of each 4×4 sub-MB by using both MVs in the LL band and the edge information. The 4×4 sub-MBs with the same class can be merged into three ways such as modes 4, 5, or 6 shown in Figure 5, and 8×8 sub-MBs merged as mode 4 can be further grouped into one of three block modes; modes 1, 2, and 3.

In order to classify 4×4 sub-MBs, we propose a new classification method based on a statistical hypothesis testing. A region tends to be homogeneous if the textures in the region have very similar spatial property. It was observed in natural video sequences that there are a lot of homogeneous regions belonging to the same video objects. When the objects move, the various parts of the objects move in a similar manner. Homogeneous blocks in the picture would have similar motion and are very seldom split into smaller blocks [11]. An effective way of determining the homogeneous region is to use the edge information, since sub-MBs in a homogeneous region have similar edge patterns.

In the proposed method, the vertical, horizontal, and diagonal edge information as well as the motion vectors in the LL band is used to determine the homogeneous region. The vertical, horizontal, and diagonal edge information can be obtained from the absolute values of the wavelet coefficients in LH, HL, and HH bands, respectively. The test of homogeneity hypothesis is as follows. Let $X = \{x_0, x_1, x_2, x_3, x_4\}$ be the feature vector for hypothesis testing, where x_0 and x_1 are the horizontal and vertical elements of the motion vector in the LL band, x_2 is the sum of the amplitude of the coefficients in the LH band (representing the horizontal edge information), x_3 is the sum of the amplitude of the coefficients in the HL band (representing the vertical edge information), and x_4 is the sum of the amplitude of the coefficients in the HH band (representing the diagonal edge information). Assume that the elements of the feature vector are mutually independent. Since Haar transform is an orthogonal transform, the

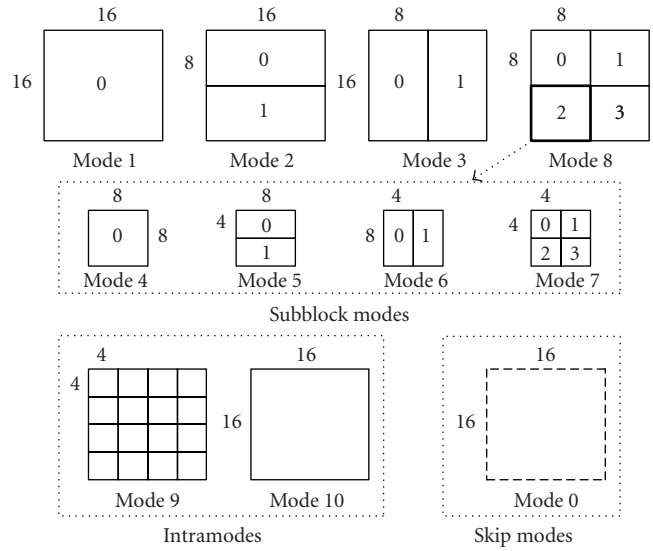


Figure 5: Block modes of H.264.

amplitude of each edge is independent. Using the model that is the most popular statistical model for DWT coefficients and MVs in [13, 14], we assume that the distributions for the elements of the feature vector are Laplacian given by

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x|}. \tag{2}$$

When the above assumptions are adopted, the proposed classifier can be approximated to the minimum distance classifier from the Bayes decision rule. The approximated discriminant function for Laplacian distribution is given by

$$g_i(x) = \sum_{j=0}^{3} \lambda_j \left| x_j - \mu_{i,j} \right|, \tag{3}$$

where i is the index of classes and j is the index of the feature vector. λ_j denotes a weight for the jth element. $\mu_{i,j}$ indicates a predetermined mean value of the jth element of the ith class.

We should classify each 4×4 sub-MB by calculating $g_i(x)$ for all classes and obtaining the minimum $g_i(x)$. In the proposed FIMS, four different classes are defined, such as smooth region with small motion, smooth region with large motion, rough region with small motion, and rough region with large motion. If the number of 4×4 subblocks having the same class in an 8×8 block is three or four, the 8×8 block is recognized as the homogeneous region. For 16×16 blocks, the same processing is used to classify the four 8×8 blocks. Then, RDO using the Lagrangian cost function is performed to select the best mode among the candidate modes selected by the hypothesis testing.

The proposed FIMS does not require the optimization process minimizing the Lagrange function of (1) for all intermodes. It can reduce the number of potential intermodes. The minimum distance classifier in the FIMS requires smaller computations than the Lagrange optimization method used in the current intermode RDO implementation of the H.264/AVC. Therefore, we can reduce the computational complexity of the whole ME and intermode decision by using the proposed method.

4. EXPERIMENTAL RESULTS

For mobile video, good coding efficiency and low complexity are required. To implement a real-time H.264 encoding and transmission system, the profile and level are constrained due to the limited memory size, low computing power, and narrow bandwidth of the mobile network, In general, H.264 baseline profile (BP) and level 3.0 or lower are adopted for mobile video applications. The BP supports intra- and intercoding (using I-slices and P-slices) as well as entropy coding with context-adaptive variable length codes (CAVLC) [2].

In this section, to demonstrate the effectiveness of the proposed MFMRME-HS and FIMS, simulations using test sequences including "table tennis" and "foreman" have been conducted under the following profile constraints and experimental conditions using JM 95 in Intel Pentium IV 2.8 GHz PC with 512 MB RAM:

(i) baseline profile,
(ii) level 3.0,
(iii) QCIF sequence: 30 frames,
(iv) reference frames: 10,
(v) adaptation of RD optimization,
(vi) quantization parameter (QP): 28 and 32,
(vii) search range: ± 16,
(viii) GOP structure: IPPP.

We first demonstrate the improvement of processing time. The MFMRME-HS method utilizes the modified FIT which requires the additional computations compared to the optimal FIT in [12]. Moreover, it requires the extra memory to store the DWT coefficients. The extra memory access can slow down the processing speed. In spite of the above weak

Table 1: Features of motion estimation algorithms (assuming M is the number of reference frames size and N is the size of image).

	Computations of transforms (for 4×4)	Number of search points	Required memory size
Full search	Only FIT (4 multiplies + 16 additions)	$M \times N^2$	$M \times N^2$
UMHexagonS	Only FIT (4 multiplies + 16 additions)	$M \times N \log N$	$M \times N^2$
Conventional MRME	FIT + DWT (16 multiplies + 28 additions)	$M \times N \log N$	$M \times N^2 \log N$
Proposed MFMRME-HS	Modified FIT (4 multiplies + 22 additions)	$\frac{M}{2} \times N \log(\log N)$	$M \times N^2 \log N$

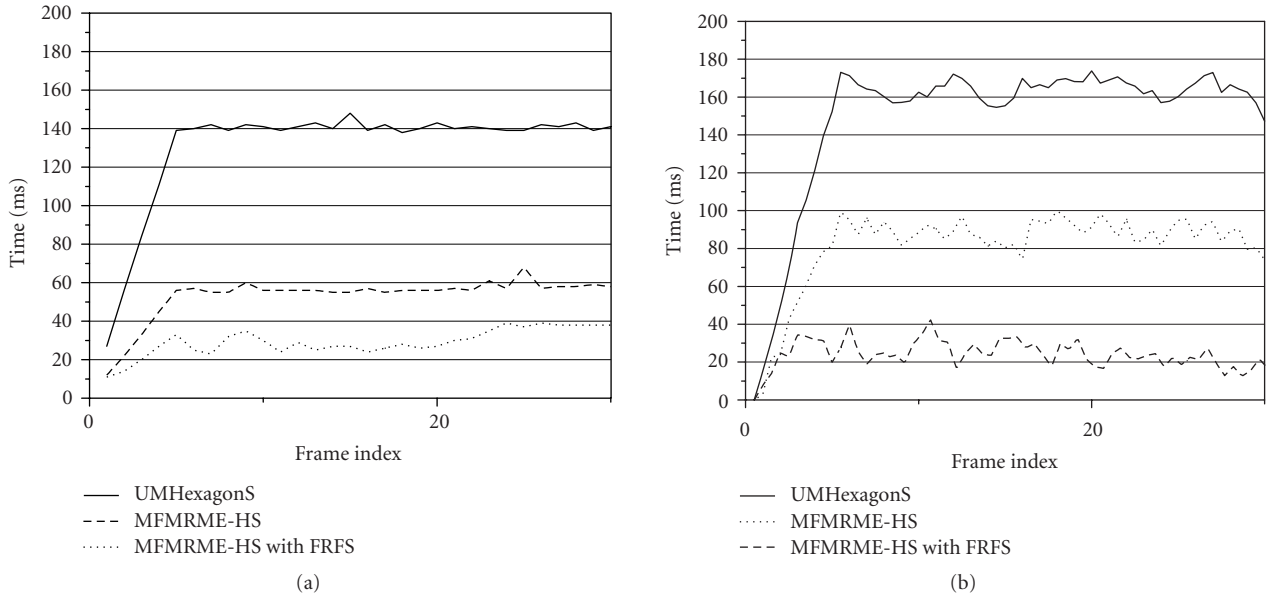


Figure 6: Comparison of processing times. (a) "Table tennis," (b) "foreman."

points, the proposed method adopting the MRME scheme can reduce the overall processing time due to a smaller number of search points. Moreover, by using the proposed FRFS, we can fast select the reference frame containing the best matching block. Table 1 summarizes the additional computations and advantages of the proposed MFMRME-HS. Figure 6 shows the processing time of ME for multiple references. To demonstrate the performance of the proposed ME, the processing time of the MFMRME-HS with the modified FIT is compared to that of the UMHexagonS with the conventional FIT. The proposed method is faster than UMHexagonS approach over 3 times.

Figure 7 shows the value of peak-to-peak signal-to-noise ratio (PSNR) of reconstructed images for the H.264/AVC encoder reference software with the full search, UMHexagonS, and the proposed MFMRME-HS algorithm. The PSNR curves obtained by these methods are almost the same. As far as the processing time is concerned, the proposed method is quite efficient.

Figures 8(a) and 8(b) show simulation results of intermode decision with FMD and the proposed FIMS. The modes selected by the proposed methods are close to those obtained by FMD. A group of experiments were carried out on the test sequence with 2 quantization parameters (QP = 28, 32). Tables 2 and 3 show the results according to quantization parameters. When measuring the processing time of the proposed FIMS, we did not consider the processing time of the integer transform or any other part in H.264. The experimental results show that the proposed method reduces the encoding time by 60% on average. The PSNR loss is negligible with the highest loss at 0.15 dB.

5. CONCLUSION

In this paper, we have proposed a new scheme to jointly optimize intermode selection and ME using the multiresolution analysis. Experimental results show that the proposed MFMRME-HS method is over 3 times faster than existing

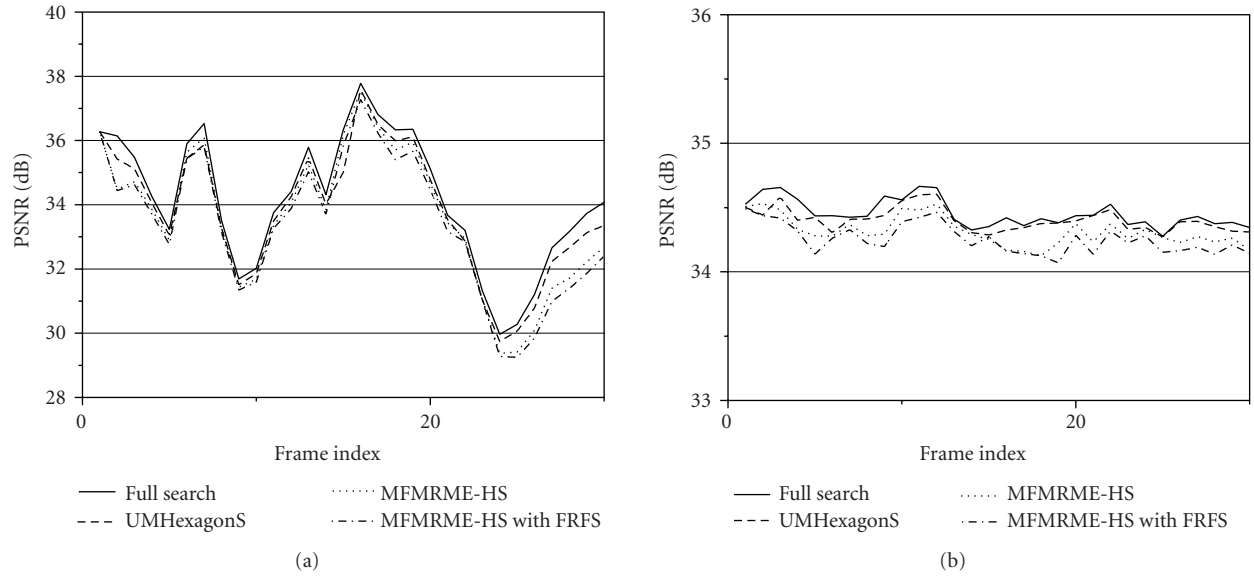
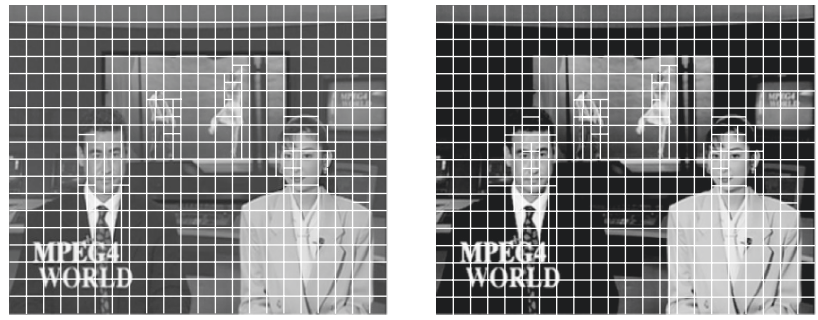


Figure 7: Comparison of PSNR: (a) "table tennis," (b) "foreman."



(a) (b)

Figure 8: Simulation results of intermode selection: (a) FMD, (b) proposed FIMS.

Table 2: Comparison of the intermode decision (QP = 28).

	GOP structure	Change of PSNR (dB)	Saving of processing time (%)
News	IPPP	−0.03	62.50
Table tennis	IPPP	−0.06	51.45
Foreman	IPPP	−0.10	50.30

Table 3: Comparison of the intermode decision (QP = 32).

	GOP structure	Change of PSNR (dB)	Saving of processing time (%)
News	IPPP	−0.04	62.50
Table tennis	IPPP	−0.08	42.25
Foreman	IPPP	−0.15	41.30

ME methods while maintaining the visual quality. Moreover, the proposed mode decision method has reduced the encoding time by 60% on average. The PSNR loss is negligible with the highest loss at 0.15 dB.

Experimental results indicate that the proposed fast motion estimation and interprediction method enable the H.264/AVC coder to be effectively adopted for the mobile video communication.

REFERENCES

[1] T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[2] A. N. Netravali and B. G. Haskell, *Digital Pictures*, Plenum Press, New York, NY, USA, 2nd edition, 1995.

[3] "Draft ITU-T recommendation and final draft international standard of joint video specification ITU-T recommendation, H.264/ISO/IEC 14496-10 AVC," Joint Video Team (JVT) of IEC MPEG and ITU-T VCEG, JVT-G050, March 2003.

[4] R. Li, B. Zeng, and M. L. Liou, "A new three-step search algorithm for block motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 4, no. 4, pp. 438–442, 1994.

[5] L.-M. Po and W.-C. Ma, "A novel four step search algorithm for fast block motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 313–317, 1996.

[6] J. Y. Tham, S. Ranganath, M. Ranganath, and A. A. Kassim, "A novel unrestricted center-biased diamond search algorithm for block motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 4, pp. 369–377, 1998.

[7] C. Zhu, X. Lin, and L.-P. Chau, "Hexagon-based search pattern for fast block motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 5, pp. 349–355, 2002.

[8] Z. Chen and Y. He, "Fast Integer and Fractional Pel Motion estimation," Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-F017, December 2002.

[9] G. Sullivan, T. Wiegand, and K. P. Lim, "Joint model reference encoding methods and decoding concealment methods," Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-J049, December 2003.

[10] P. Yin, H.-Y. C. Tourapis, A. M. Tourapis, and J. Boyce, "Fast mode decision and motion estimation for JVT/H.264," in *Proceedings of IEEE International Conference on Image Processing (ICIP '03)*, vol. 3, pp. 853–856, Barcelona, Spain, September 2003.

[11] C.-H. Cheung and L.-M. Po, "A novel cross-diamond search algorithm for fast block motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 12, pp. 1168–1177, 2002.

[12] H. S. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, "Low-complexity transform and quantization in H.264/AVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 598–603, 2003.

[13] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 1, pp. 246–250, 1997.

[14] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice Hall, Englewood Cliffs, NJ, USA, 1984.

Byeong-Doo Choi received the B.S. and M.S. degrees in electronics engineering from the Department of Electronics Engineering at Korea University, Seoul, South Korea, in 2001 and 2003, respectively. He is currently working towards the Ph.D. degree in multimedia processing and communication at Korea University. His research interests include video compression, transmission, and reconstruction.

Ju-Hun Nam received the B.S. degree in electronics engineering from Dong-A University in 1995 and the M.S. degree in 1997. He is now a Ph.D. candidate in electronics engineering with the Department of Electronics Engineering at Korea University. His research interests include JPEG2000, MPEG4, source and channel codings, and multimedia communication based on DSP/FPGA.

Min-Cheol Hwang received B.S. degree in electronics engineering from the Department of Electronics Engineering at Korea University, Seoul, South Korea, in 2003. He is currently working towards the Ph.D. degree in multimedia signal processing and communication at Korea University. His research interests are in the areas of image compression, such as JPEG2000 and H.264, and multimedia communications.

Sung-Jea Ko received the Ph.D. degree in 1988 and the M.S. degree in 1986, both in electrical and computer engineering, from State University of New York at Buffalo, and the B.S. degree in electronics engineering at Korea University in 1980. In 1992, he joined the Department of Electronics Engineering at Korea University where he is currently a Professor. From 1988 to 1992, he was an Assistant Professor of the Department of Electrical and Computer Engineering at the University of Michigan-Dearborn. From 1986 to 1988, he was a Research Assistant at State University of New York at Buffalo. He has published more than 200 papers in journals and conference proceedings. He also holds over 10 patents on data communication and video signal processing. He is currently a Senior Member in the IEEE, a Fellow in the IEE and a Chairman of the Consumer Electronics Chapter of IEEE Seoul Section. He is the 1999 Recipient of the LG Research Award given to the Outstanding Information and Communication Researcher. He received the Hae-Dong Best Paper Award from the IEEK (1997) and the Best Paper Award from the IEEE Asia Pacific Conference on Circuits and Systems (1996).