Realtime H.264 Encoding System using Fast Motion Estimation and Mode Decision

Byeong-Doo Choi, Min-Cheol Hwang, Jun-Ki Cho, Jin-Sam Kim, Jin-Hyung Kim, and Sung-Jea Ko

Department of Electronics Engineering, Korea University
5-1 Anam-Dong, Sungbuk-ku, Seoul 136-701, Korea
sjko@dali.korea.ac.kr

Abstract. H.264 provides various useful features such as improved coding efficiency and error robustness. These features enable mobile device to adopt H.264/AVC standard to achieve effective video communications. However, the encoder complexity is greatly increased mainly due to motion estimation (ME) and mode decision. We implemented a realtime H.264 encoding system based on DSP and FPGA chip using the propose scheme to jointly optimize inter mode selection and ME using the multi-resolution analysis. In our paper, we focus on reducing the complexity while maintaining the visual quality output. Experimental results show that the proposed algorithm is over 3 times faster than the existing method while maintaining the coding efficiency and the implemented system. The implemented H.264 encoder meets the common requirement of the real-time video coding [30 frames/sec].
Keyword: H.264, inter mode decision, motion estimation, MRME, mobile video communication.

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 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 [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 H.264/AVC standard to achieve effective video communications [2]. H.264 supports multiple reference frames and various block sizes for ME. It uses tree-structured hierarchical macroblock (MB) partitions. There are 7 different block sizes (16x16, 16x8, 8x16, 8x8, 4x8 and 4x4 blocks) that are used in a macroblock.

The current H.264 reference software is based on a rate- distortion optimization framework for both ME and mode decision.

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

The computational burden 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, the UMHexagonS has been adopted for fast ME in H.264 encoder reference software (JM 84) [7]. To optimize the mode decision process, the H.264 software adopts the full mode decision algorithm (FMD) [8] inducing an exhaustive computation. For fast mode decision, the early termination technique [9] reduces the number of potential prediction modes. In [10], a classification method is proposed to reduce the average number of block types while maintaining the coding performance.

In this paper, we propose new fast ME and inter mode decision techniques using the multi-resolution analysis for the H.264/AVC encoder. For fast ME, the proposed method is based on the multi-frame/multi-resolution ME using Hexagon searching. For the fast decision of inter coding mode, we introduce a down-top merge method. The proposed method splits all MBs into 4x4 sub-MBs and 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 II. In Section III, the proposed fast inter mode decision method is given. Finally, the implemented H.264 embedded system and the simulation results are shown in Section IV and the conclusions are described in Section V.

2 THE PROPOSED MOTION ESTIMATION METHOD

The multi-resolution ME (MRME) technique is an alternative ME method to the conventional block-matching algorithm. In the conventional MRME, MVs are estimated at the lowest resolution and the estimated MV is scaled appropriately to be used as an initial bias and refined for remained sub-bands on wavelet transform. However, adopting the MRME for the H.264 requires wavelet transform as well as the integer transform, which results in additional computations.

To solve this problem, we propose a modified fast integer transform that can obtain wavelet coefficients by factorizing integer transform into two steps properly. Fig. 1 shows the modified implementation of a 4x4 fast integer transform. Unlike Malvar's fast integer transform approach [11], its first step is equal to the Haar wavelet transform. The first step of the modified fast integer transform

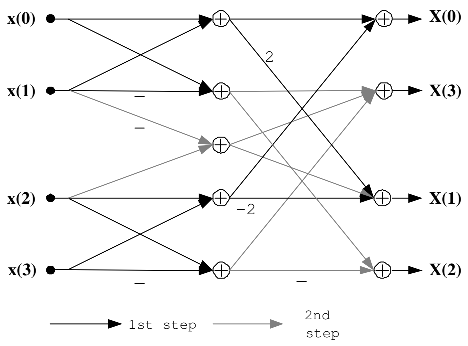


Fig. 1. Modified implementation of fast integer transform.

decomposes a whole image into three layers. The coefficients in multi-layers are utilized in the proposed ME and mode decision methods.

With the modified fast integer transform, we propose a fast ME algorithm using MRME, Hexagon Searching, and Multi-frame reference. The proposed method exploits the cross correlation among multi-frames and layers of wavelet transform 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.

Fig. 2 shows the proposed searching patterns; the small square (SS), the uneven multi-hexagon-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.

Fig. 3 shows the concept of the proposed MFMRME-HS method consisting of four steps. The detailed procedure is followed:

1. Search the MV in the object region with small motion by using the SS searching pattern at layer 2 of the reference frame (t-1) shown in Fig. 4. 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.
2. Find the MV by using the UMHG pattern at layer 2 of the reference frame (t-1) shown in Fig. 4. 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.
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.
4. Step 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

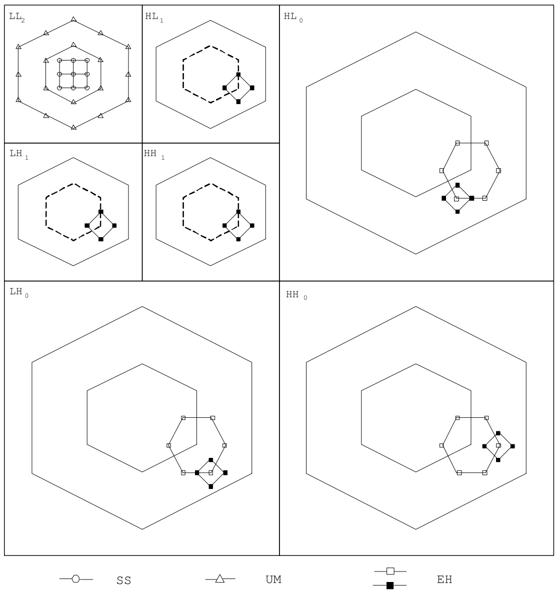


Fig. 2. Search pattern for multi-resolution.

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 multi-frames and sub-bands of wavelet transform to reduce the time complexity of ME.

3 THE PROPOSED FAST INTER MODE SELECTION METHOD

In this section, we propose a down-top merge method for fast inter mode decision. Fig. 4 shows the proposed down-top merge method. The proposed method splits all 16x16 MBs into 4x4 sub-MBs and determines the class of each 4x4 sub-MBs by using both MVs in the LL band and the edge information. The 4x4 sub-MBs with the same class can be merged into three ways such as modes 4, 5, or 6 shown in Fig. 5, and 8x8 sub-MBs merged as mode 4 can be further grouped into one of the three block modes; modes 1, 2, and 3. After inter mode decision, ME is performed using MFMRME-HS only for the selected mode.

In order to classify 4x4 sub-MBs, we propose a new classification method based on statistical hypothesis testing. A region tends to be homogeneous if the textures in the region have very similar spatial property. Since blocks in the homogeneous region of the image can have similar MVs, they are not further split

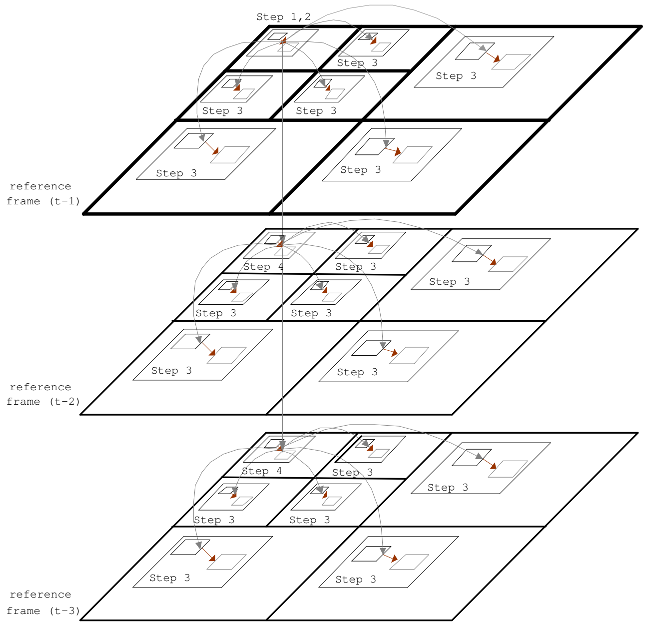


Fig. 3. Multi-frames/multi-resolution motion estimation scheme.

into smaller blocks. 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 edges in each sub-band of wavelet transform are used to determine the homogeneous region. The test of homogeneity hypothesis is as follows: Let $X = \{ x_0, x_1, x_2, x_3, x_4 \}$ be the feature vectors for hypothesis testing where x_0 is the absolute of motion vector in the LL band, x_1 is the average of pixel intensities in the LL band, x_2 is the sum of the amplitude of the horizontal edge in the LH band, x_3 is the sum of the amplitude of the vertical edge in the HL band, and x_4 is the sum of the amplitude of the diagonal edge in the HH band. Assume that the elements of the feature vector are independent each other. Since wavelet transform is an orthogonal transform, the amplitude of each edge is independent. Moreover, it is assumed that the distributions for the elements of the feature vector are Laplacian given by

$$f(x) = \frac{\lambda}{2} e^{-\lambda |x-\mu|} \tag{1}$$

This model is the most popular statistical model for coefficients in frequency domain and MVs [12] $\sim$ [14]. When the above assumptions are adopted, the proposed classifier can be approximated to the minimum distance classifier from

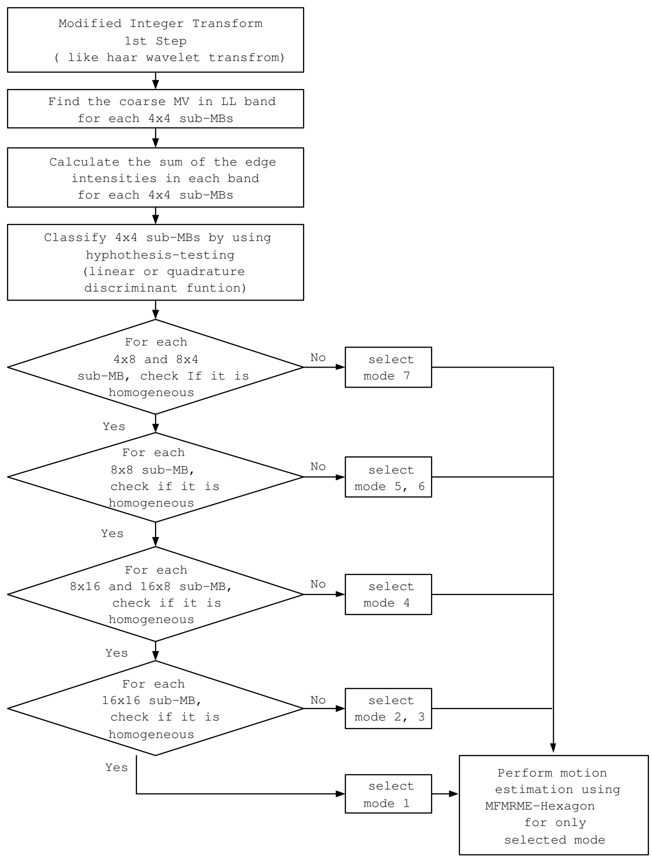


Fig. 4. Flow chart of fast inter mode decision by classifying sub-MBs.

Bayes decision rule. The approximated discriminant functions for Laplacian distribution is given by

$$g_i(x) = \sum_{j=0}^{n-1} \lambda |x_j - \mu_{ij}| \tag{2}$$

where i is the index of classes and j is the index of the feature vector. We can classify each 4x4 sub-MB by calculating $g_i(x)$ for all classes and obtaining the minimum $g_i(x)$. Unlike conventional methods, the proposed inter mode decision method does not require ME and rate distortion optimization for every inter modes. However, the minimum distance classifier with a small computation is needed. Therefore, the proposed down-top merge method achieves fast inter mode decision.

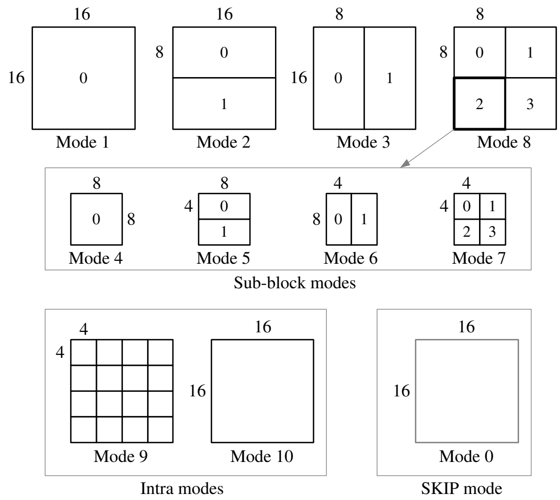


Fig. 5. Block modes of H.264/AVC.

4 Experimental Results

To demonstrate the performance of the proposed system, we implemented the H.264 system which includes a 32-bit DSP processor and a FPGA, as shown in Fig. 6. The implemented H.264 encoder system targets a cost effective and high performance micro controller solution for system on chip. This system is under development with ALTERA Stratix (25,660 LEs) and TMS320C6416 (600 Mhz, 4800 MIPS, 128kb cache). Simulations have been conducted using JM 84 and the computational complexity is measured by the number of DSP clock cycles.

Fig. 7 shows the processing time of ME for multiple references. The proposed method is faster than UMHexagonS approach over 3 times since the redundant process among reference frames is reduced.

Fig. 8 shows the change of peak-to-peak signal-to-noise ratio (PSNR) and bitrate of reconstructed images for H.264 encoder reference software with the full search, UMHexagonS search, 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.

Fig. 9(a) and (b) shows simulation results of the inter mode decision with FMD and the proposed method. 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. Table 1 and 2 show the results according to quantization parameters. 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 0.05 dB.

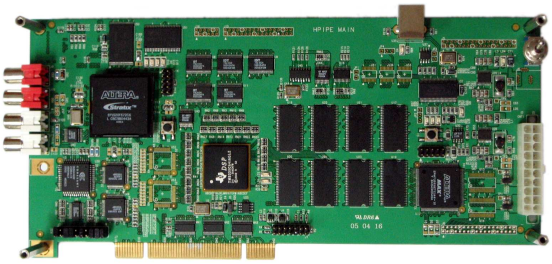


Fig. 6. The implemented H.264 system using DSP and FPGA.

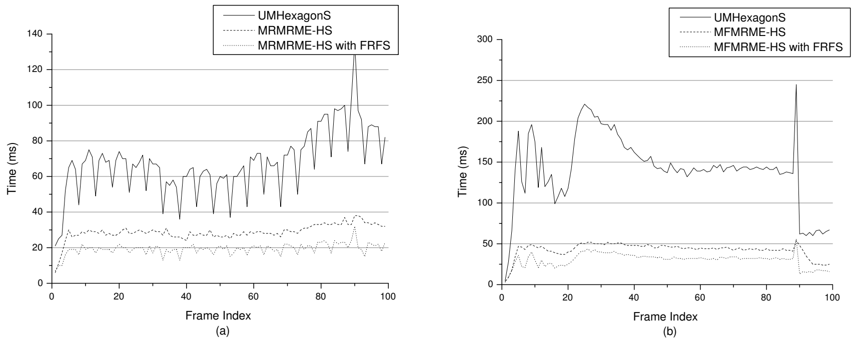


Fig. 7. Comparison of processing times. (a) "News", (b) "Table tennis"

5 Conclusions

In this paper, we have proposed a new scheme to jointly optimize inter mode selection and motion estimation using the multi-resolution analysis and show the implemented H.264 embedded system. Experimental results show the proposed MFMRME-HS method is over 3 times faster than the existing ME method 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 0.05 dB. Experimental results indicate that the proposed fast motion estimation and inter-prediction method enable the H.264/AVC coder to be effectively adopted for the mobile video communication.

References

[1] Wiegand, T., Sullivan, G.J.: Overview of the H.264/AVC video coding standard. IEEE Trans. Circuits Syst. Video Technol. 13 (2003) 560-576 Netravali, A.N.,

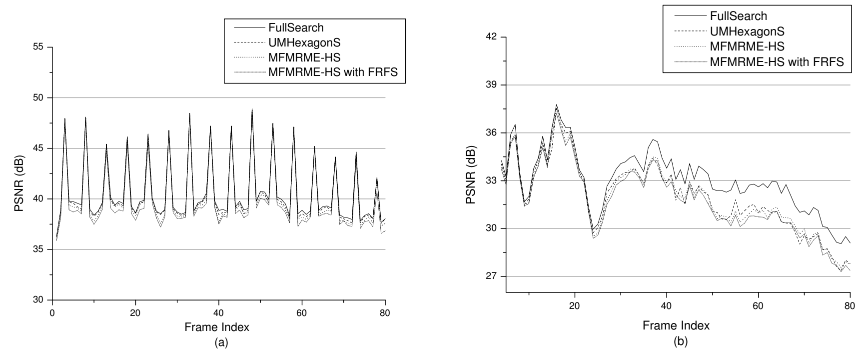


Fig. 8. Comparison of PSNR's. (a) "News", (b) "Table tennis"

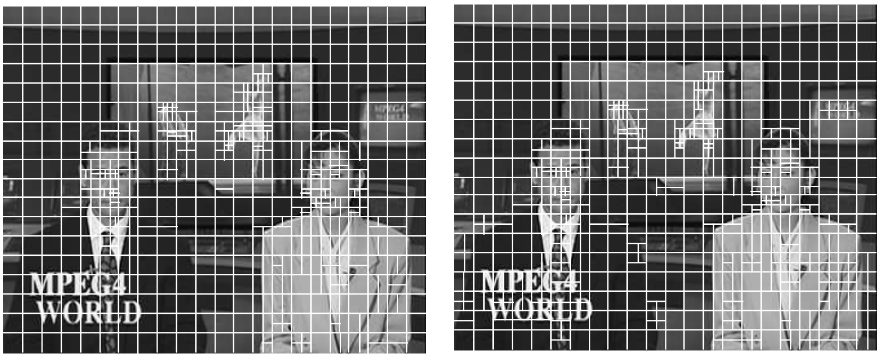
Table 1. TEST CONDITION VIDEO SEQUENCE COMPARISON (QP=28)

Sequence	GOP Structure	Change of PSNR(dB)	Processing Time Saving(%)
News	IPPP	-0.04	67.75
Table Tennis	IPPP	-0.02	55.25

Haskell, B.G. : Digital Pictures, 2nd ed., Plenum Press (1995) 613-651
[2] ITU-T recommendation, H.264/ISO/IEC 14496-10 AVC : Draft ITU-T recommendation and final draft international standard of joint video specification. Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-G050. (2003)
[3] Li, R., Zeng, B., Liou, M.L.: A new three-step search algorithm for block motion estimation. IEEE Trans. Circuits Syst. Video Technol. 4 (1994) 438-443
[4] Po, L.M., Ma, W.C.: A novel four-step search algorithm for fast block motion estimation. IEEE Trans. Circuits Syst. Video Technol. 6 (1996) 313-317
[5] Tham, J.Y., Ranganath, S., Ranganath, M., Kassim, A.A.: A novel unrestricted center-biased diamond search algorithm for block motion estimation. IEEE Trans. Circuits Syst. Video Technol. 8 (1998) 369-377

Table 2. TEST CONDITION VIDEO SEQUENCE COMPARISON (QP=32)

Sequence	GOP Structure	Change of PSNR(dB)	Processing Time Saving(%)
News	IPPP	-0.05	67.50
Table Tennis	IPPP	-0.03	47.75



(a) (b)

Fig. 9. Simulation results of inter mode selection. (a)FMD, (b)The proposed method

[6] Zhu, C., Lin, X., Chau, L.: Hexagon-based search pattern for fast block motion estimation. IEEE Trans. Circuit Syst. Video Technol. 12 (2002) 349-355

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

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

[9] Yin, P., Tourapis, H.C., Tourapis, A.M., Boyce, J.: Fast mode decision and motion estimation for JVT/H.264. Proc. IEEE Int. Conf. Image Processing, Barcelona, Spain. 3 (2003) 14-17

[10] Lim, K.P., Wu, S., Wu, D.J., Rahardja, S., Lin, X., Pan, F., Li, Z.G.: Fast inter mode selection. Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-I020 (2003)

[11] Malvar, H.S., Hallapuro, A., Karczewicz, M., Kerofsky, L.: Low-complexity transform and quantization in H.264/AVC. IEEE Trans. Circuits Syst. Video Technol. 13 (2003) 598-603

[12] Chiang, T., Zhang, Y.Q.: A new rate control scheme using quadrature distortion mode. IEEE Trans. Circuit Syst. Video Technol. 7 (1997) 246-250

[13] Jayant, N., Noll,P.: Digital coding of waveforms. Englewood Cliffs, NJ: Prentice Hall (1984)

[14] Cheung, C.H., Po, L.M.: A novel cross-diamond search algorithm for fast block motion estimation. IEEE Trans. Circuit Syst. Video Technol. 12 (2002) 1168-1177