

# COMPLEXITY ADAPTATION IN H.264/AVC VIDEO CODER FOR STATIC CAMERAS

*Chanyul Kim and Noel E.O'Connor*

CLARITY:Centre for Sensor Web Technologies, Dublin City University, Ireland

## ABSTRACT

H.264/AVC uses variable block size motion estimation (VBS-ME) to improve coding gain. However, its complexity is significant and fixed regardless of the required quality or of the scene characteristics. In this paper, we propose an adaptive complexity algorithm based on using the Walsh Hadamard Transform (WHT). VBS automatic partition and skip mode detection algorithms also are proposed. Experimental results show that 70% - 5% of the computation of H.264/AVC is required to achieve the same PSNR.

**Index Terms**— H.264/AVC, skip mode, WHT

## 1. INTRODUCTION

The superior performance of the H.264/AVC mainly comes from advanced technologies, such as 1/4-pixel accurate variable block size motion estimation (VBS-ME) with multiple reference frames, intra prediction and advanced entropy coding. However, VBS-ME is very computationally demanding due to its usage of Rate-Distortion Optimization (RDO) to decide the best block partition [1]. There are two main approaches to reduce the computational cost. The first is to use fast algorithms for ME such as enhanced predictive zonal search (EPZS) and hybrid unsymmetrical cross multi-hexagon grid search, which are implemented in the Joint Model (JM) reference software. The second approach is early termination algorithms for both block size [2] and skip mode [3] decisions.

For most static camera applications such as surveillance and video telephony, fixed block sizes (16x16 or 8x8) are used. Object-based coding could be a candidate solution to avoid block based motion search. However, it requires additional segmentation schemes, and is still an unsolved problem, particularly for partial motion in a large object area with small motion [4]. Moreover, the assumption that most static camera applications require low bit-rate (low quality) is not valid due to the emergency of mega-pixel IP camera technology based on H.264/AVC [5]. For this reason we target VBS-ME for enhancing quality. Our experiments reveal that the performance of VBS-ME worsens with increasing quantization parameter (QP). For this reason, we target computational

complexity savings at low bit rates. The basic idea presented in this paper is to adapt the complexity according to the required quality and scene complexity. In order to achieve this, we propose simple detection of an block containing an edge using the Walsh-Hadamard Transform (WHT) which is very a simple transform widely used in the image coding [6] and pattern recognition [7]. Then an automatic allocation algorithm for VBS based on detected edge blocks and skip mode detection is presented reusing WHT coefficients obtained when detecting edge blocks.

## 2. RELATED WORK AND MOTIVATION

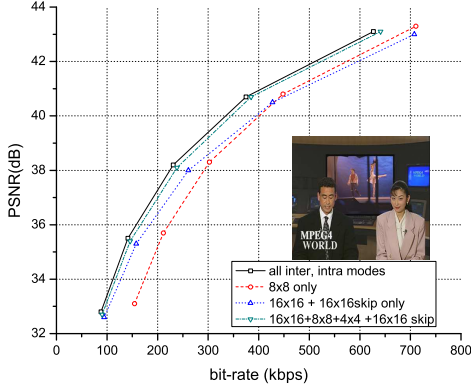
The proposed method consists of two main function; VBS partitioning without Rate-Distortion calculation and deciding the skip mode according to QP. Lauri *et al.* [8] presented VBS partitioning and skip mode detection based on lightweight segmentation where they used edge detection. Their approach intuitionally finds the skip mode, but causes a lot of degradation at low QP. Other methods related to VBS partitioning depend on a threshold based on Rate-Distortion optimization. For deciding skip mode, the common approach is to find a SAD (Sum of Absolute Difference) lower than a pre-defined threshold, and this block is considered as a skip block. However, SAD does not reflect the characteristics of the DCT fully and the overall quality depends on the pre-defined threshold obtained empirically. To avoid this problem, Zhenyu *et al.* [9] presented Hadamard Transform (HT) based skip block detection where all blocks are divided into 4x4 sub-blocks and HT coefficients found. Their approach is similar to our proposed method. The main difference is that in their approach all blocks need to be transformed even though complexity is reduced by using HT instead of DCT and thus this still represents computational complexity.

The proposed method uses three block types (16x16, 8x8 and 4x4). Fig. 1(a) shows the RD degradation of selected modes compared to all modes in H.264/AVC, where negligible difference is obtained between all modes and three modes. Moreover, more changing to variable complexity function is finding skip modes shown in Fig. 1(b). As QP increase, required modes are dominant to skip modes and 16x16 modes because mode bits for large block require less bits than those of the small block, where mode bits are more important part in total encoded bits than data bits. Thus there is much room

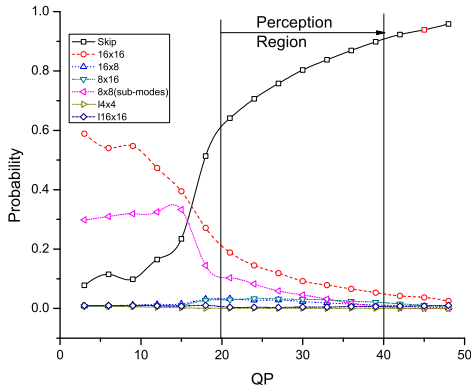
---

Authors would like to acknowledge the support of Samsung Electronics and Science Foundation Ireland under grant 07/CE/I1147.

for saving computational complexity at high QP. In this paper, we present the WHT based skip mode decision and VBS partitioning via edge block detection for differencing image. Our approach reuses WHT coefficients of small block used in edge block detection for finding skip modes.



(a) Rate Distortion according to block size; There is negligible degradation between full modes and the three modes (16x16,8x8 and 4x4)



(b) Variable Block decision probability; skip mode, 16x16 and 8x8 sub-block should all be considered for low values of QP, but only skip mode and 16x16 are enough for high values of QP

**Fig. 1.** Effect of VBS in the News test sequence

### 3. THE PROPOSED ALGORITHM

The WHT is an orthogonal transform whose basis functions consist of a set of rectangular discontinuous waveforms that take the value -1 and 1 [10]. The main advantage over other orthogonal transforms is its low computational complexity since a small number of additions and subtractions are required. The transformation kernels of the WHT are defined as:

$$f(x, y, u, v) = \frac{1}{N} \prod_{i=0}^{n-1} [(-1)^{p_i(x)p_{n-1-i}(u)} (-1)^{p_i(y)p_{n-1-i}(v)}] \quad (1)$$

and its inverse kernel is

$$i(x, y, u, v) = f(x, y, u, v)$$

where  $n = \log_2(N)$ ,  $p_i(m)$  represents the  $i$ th bit in the binary representation of  $m$ . If a image block is projected on a sequency ordered WH kernel (Eq. (1)), the projection on the first few WHT coefficients can capture a considerable portion of the information in a block, named energy compaction [6]. We use the characteristics of energy compaction and the orthogonal property in this paper. The proposed method proceeds as follows:

1. The difference image between the current and previous frame is divided into non-overlapped 2x2 blocks and WHT coefficients are calculated.
2. If the distribution of sequency coefficients is greater than a pre-defined threshold ( $\tau$ ), this block is considered as a block containing an edge.
3. If the current macro-block (MB) does not have an edge block, the current MB is treated as a 16x16 partition. If not, the current MB is partitioned into 8x8 and follow the same procedure.
4. In 16x16 partitioned blocks in the image, the first four WHT coefficients are obtained from the 2x2 sub-block coefficients.
5. If the sum of WHT coefficients after quantization is zero, the current MB is considered as a skip mode.

#### 3.1. VBS Partitioning based on an Edge Block

Various edge detection algorithms can be used for detecting a homogeneous block. The Sobel operator is widely used to perform a 2-D spatial gradient measurement on an image due to its simplicity. However, its usage for image coding still presents a burden. For example, in order to decide if a pixel is an edge or not, it performs a convolution operation with a 3x3 mask which requires ten additions and four multipliers per an edge pixel. It also performs thresholding for each direction. The proposed method uses a 2x2 WHT and only needs twelve additions per four pixels. The WHT coefficients bear pseudo frequency information.

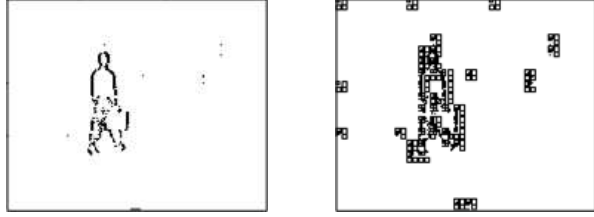
Let coefficients of the 2x2 WHT be defined as  $H(u, v)$ , where  $(u, v)$  represents the position in a block. If a block has a homogeneous region, most energy is packed into  $H(0, 0)$ . Relatively, the variance of the rest of coefficients would be a zero. If Eq. (2) is larger than pre-defined threshold (we set this to 25 in this paper), the block is considered as homogeneous.

$$\sigma = E(H(u, v)^2) - E(H(u, v))^2 \quad (2)$$

$$(u, v) \in \{(1, 0), (0, 1), (1, 1)\}$$

where  $E(\cdot)$  represents expectation. Fig. 2(a) shows example results of the proposed edge detection.

The VBS partitioning has a simple procedure. In a homogeneous region indicated as white in Fig. 2(b), the current MB is set as a 16x16 block which can be a candidate for skip mode. If the current block contains edge blocks, it is divided into four 8x8 blocks and it is decided which 8x8 block contains edge blocks. When a 8x8 block contains edge blocks, this block is partitioned into four 4x4 mode, otherwise, it is considered as a 8x8 mode block.



(a) Detected edge blocks in the differenced frame  
(b) VBS partitioning; white, black and elsewhere represent 16x16, 4x4 and 8x8 respectively

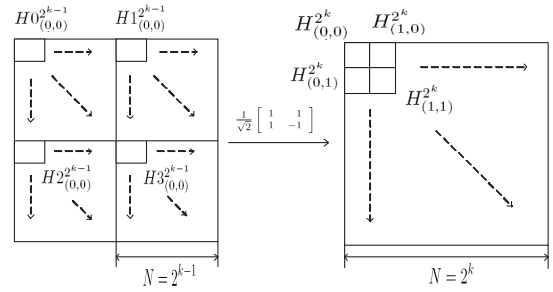
**Fig. 2.** Edge and VBS partitioning of 45th frame of Hall Monitor

### 3.2. Skip Mode Decision

The most common skip mode detection for H.264/AVC is performed after deciding VBS partition based on mode competition. It has no computational advantage when an image is encoded with a high value of QP. Moreover, the characteristic of an image does not affect the computational efficiency. After VBS partitioning, skip mode detection is performed at 16x16 block size. The procedure for choosing skip mode in H.264/AVC focuses on 4x4 sub-blocks. If the DCT coefficients of all 4x4 blocks included in a MB after ME is zero, the MB is selected as a skip MB. However, this sub-block based zero coefficient searching algorithm sets a strict condition for determining a skip mode. For example, if a block has only few coefficients in the furthest area from (0,0) (high frequency alive considering a 16x16 whole block), this could be treated as a skip mode. However, some coefficients of a sub-block in this MB might not zero, so this block is not decided as a skip mode in H.264/AVC.

The proposed skip mode algorithm focuses on finding the important coefficients (positioned at (0,0),(1,0),(0,1),(1,1)) driven by the previously calculated 2x2 WHT coefficients obtained when detecting an edge block. The mathematical relationship between a  $2^k \times 2^k$  block and its sub-partitioned blocks can be driven as 2-D WHT shown in Eq. (3). This is only valid for low sequency ordered coefficients located at (0,0), (1,0), (0,1), (1,1).

$$\begin{bmatrix} H_{(0,0)}^{2^k} & H_{(1,0)}^{2^k} \\ H_{(0,1)}^{2^k} & H_{(1,1)}^{2^k} \end{bmatrix} = WHT_2 \times \begin{bmatrix} H_{(0,0)}^{2^{k-1}} & H_{(0,0)}^{1^{2^{k-1}}} \\ H_{(0,0)}^{2^{k-1}} & H_{(0,0)}^{3^{2^{k-1}}} \end{bmatrix} \times WHT_2^T \quad (3)$$



**Fig. 3.** WHT coefficients relationship with its quadrant sub-blocks

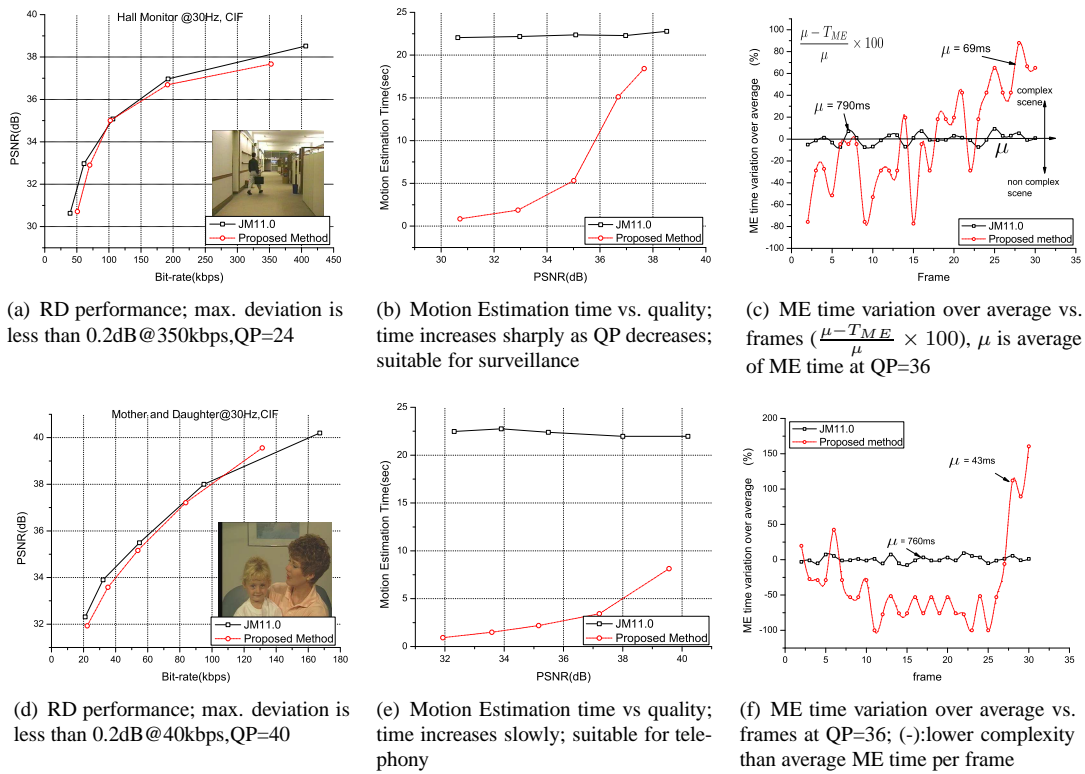
where all variables are visually defined in Fig. 3,  $WHT_2 = WHT_2^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ . Important coefficients of a MB could be identified by taking the WHT using its quadrant sub-blocks' zero sequency term ( $H_{(0,0)}^8$ ).  $H_{(0,0)}^8$  of 8x8 sub block can be calculated as weighted ( $\frac{1}{2^2}$ ) sum of  $H_{(0,0)}^2$  of all 2x2 sub blocks in a 8x8 block. The  $H_{(0,0)}^2$  of a 2x2 block is already calculated when detecting an edge block, so the important coefficients of an MB can be obtained very efficiently using a 2x2 WHT. After quantization with Eq. (4) which is used in H.264/AVC, if the sum ( $\sum_{u,v \in \{0,1\}} \frac{|H_{(u,v)}^{16}|}{Th(QP)}$ ) is zero, this MB is taken as a skip mode.

$$Th(QP) = \omega \times \frac{2^{qp\_bits} - qp\_const}{quant\_coef[qp\_rem][0][0]} \quad (4)$$

where QP is the quantization parameter,  $qp\_rem = QP \% 6$ ,  $qp\_bits = QP / 6 + 15$ ,  $qp\_const = (1 \ll qp\_bits) / 6$ ,  $quant\_coef$  is the scaling matrix defined in H.264, and  $\omega$  is a compensation constant between the DCT and the WHT based on empirical measurements (set as 10).

## 4. EXPERIMENTAL RESULTS

The proposed method is embedded into the reference software JM11.0 provided by JVT and tested on the ‘‘Hall monitor’’ and ‘‘Mother and daughter’’ sequences with CIF resolution. These sequences are representative of static camera surveillance and telephony applications. The test conditions are as follows; (1) low complexity mode enabled, (2) the number of reference frames is one, (3) GOP is IPPP with an I frame inserted every 30 frames, (4) the Hadamard transform is used and (5) QP=24,28,32,36,40. The computational complexity of the proposed method tends to be in proportion to the QP selected. In Hall monitor, only 5% computational complexity is needed compared to H.264/AVC at QP=40 as shown in Fig. 4(b). As QP increases, the required complexity also increases (almost 70% of H.264). The degradation in PSNR however is less than 0.2dB within the whole range as shown in Fig. 4(a). In the Mother and daughter sequence which has small motion and a large moving object area, a similar result



**Fig. 4.** RD performance and Encoding time (a),(b)&(c) Hall monitor (d),(e)&(f) Mother and Daughter sequence.

to Hall monitor is obtained – Fig. 4(d). However, even less complexity can be achieved compared to Hall monitor due to the less complex nature of the scene –Fig. 4(e)&(f). Therefore, complexity is clearly being adapted depending on the required quality (QP) (Fig. 4(b)&(e)) and the complexity of the scene (Fig. 4(c)&(f)).

## 5. CONCLUSIONS

In this paper, we have proposed an adaptive complexity algorithm according to required quality by user (QP) and scene characteristics. The algorithm consists of VBS partitioning based on edge and skip mode detection which are performed in Walsh-Hadamard Domain. The result shows that no degradation quality compared to H.264 and much saving complexity is achieved. We focus on immobile camera which gives us a true motion which reduces our edge detection accuracy. Edge blurring is occurred in the fast moving and camera movement, which generate error. In that case, we have to consider inside the edges, which means some level of object segmentation should be required. Future work includes not only study of light segmentation using WHT but also require various test for lots of databases which include different resolution and scene characteristics.

## 6. REFERENCES

- [1] Yibo FAN, Takeshi IKENAGA2, and Satoshi GOTO, "Reconfigurable variable block size motion estimation architecture for search range reduction algorithm," *IEICE Transactions on Electronics*, vol. E91-c, pp. 440–448, 2008.
- [2] Libo Yang, K. Yu, Jiang Li, and Shipeng Li, "An effective variable block-size early termination algorithm for h.264 video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 6, pp. 784–788, June 2005.
- [3] D. Wu, K. P. Lim, W. Yao, T. K. Chiew, and J. Y. Tham, "A novel approach to skip mode decision for H.264," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Jan. 2008, vol. 6822 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*.
- [4] Y. Shireen, M. Elhabian Khaled, El-Sayed, and H. A. Sumaya, "Moving object detection in spatial domain using background removal techniques - state-of-art," *Recent Patents on Computer Science*, vol. 1, pp. 32–54, 2008.
- [5] Michael J, "The impact of video surveillance technology on future video coding standards," Tech. Rep., ITU-T, 2005.
- [6] C. M. Mak, N. Li, and W. K. Cham, "Fast motion estimation in walsh hadamard domain," in *Intelligent Signal Processing and Communication Systems, 2005. ISPACS 2005. Proceedings of 2005 International Symposium on*, Dec. 2005, pp. 349–352.
- [7] M. Hassan, I. Osman, and M. Yahia. F. "Walsh-hadamard transform for facial feature extraction in face recognition," *ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY*, vol. 23, 2007.
- [8] Lauri Koskinen, Ari Paasio, and Kari Halonen, "Cnn-type algorithms for h.264 variable block-size partitioning," *Image Commun.*, vol. 22, no. 9, pp. 797–808, 2007.
- [9] Zhenyu Liu, Yang Song, Yiqing Huang, Lingfeng Li, S. Goto, and T. Ikenaga, "Edge gradient based fast inter motion estimation for h.264/AVC," in *Image and Signal Processing and Analysis, 2007. ISPA 2007. 5th International Symposium on*, Istanbul, Sept. 2007, pp. 52–57.
- [10] J.L. Walsh, "A closed set of normal orthogonal functions," *AM.J. Math*, vol. 45, pp. 5–24, 1923.