# Edge-Based Coding Tree Unit Partitioning Strategy in Inter Prediction

Maria Santamaria and Maria Trujillo

Multimedia and Vision Lab., Universidad del Valle,
Ciudadela Universitaria Meléndez, Cali, Colombia
{maria.santamaria,maria.trujillo}@correounivalle.edu.co

**Abstract.** The High Efficiency Video Coding standard increases in the range of $[30, 40]\%$ data compression ratio compared to H.264/MPEG-4 (AVC), but it requires bigger number of operations. HEVC uses a quadtree coding structure. The quadtree partitioning process is a high complexity operation since it employs an exhaustive process, called rate distortion optimisation, which involves all possible combinations of quadtree partitions. In this paper, a frame partitioning strategy is addressed using motion features. Motion features are edges extracted using Gaussian smoothing, the Sobel operators, and the Otsu's method. The strategy achieves CU partitions according to the amount of motion in content, and reduces the number of operations in the inter prediction mode.

**Keywords:** edge detection, high efficiency video coding, coding tree unit.

## 1 Introduction

Motion estimation (ME) consists in estimating the displacement of image content from one frame to another. It is commonly used to remove temporal redundancy – inter-frame prediction – between consecutive frames, and it is perhaps the most time consuming part in video coding, being adopted by standards such as H.263, H.264/MPEG-4 (AVC), and HEVC/H.265 [3]. The most widely used technique to estimate motion is the block-matching algorithm (BMA). In a BMA the current frame is split into non-overlapping blocks of size $n \times n$ and for each one, the algorithm searches for the block of the same dimensions that matches most. The search is made within a search window of size $(n + 2p) \times (n + 2p)$ in the reference frame by minimising a block distortion measure (BDM). The parameter $p$ is called the search parameter and represents the maximum motion displacement.

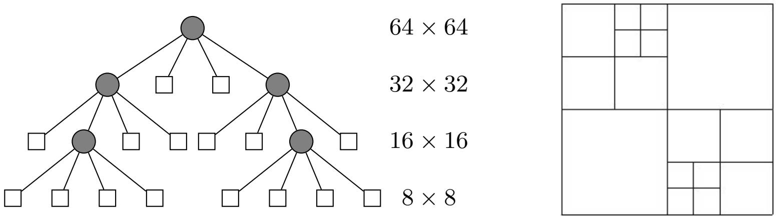
The High Efficiency Video Coding (HEVC/H.265) standard improves compression performance compared to existing standards, in the range of $[30, 40]\%$ bit rate reduction. The video coding layer employs an hybrid approach, which combines prediction and transformation for reducing redundancy in a video signal, using: motion analysis, temporal prediction, motion compensation, and

E. Bayro-Corrochano and E. Hancock (Eds.): CIARP 2014, LNCS 8827, pp. 46–53, 2014.

space-time transformations. Moreover, HEVC adopts a quadtree coding structure for estimating the coding tree units (CTUs). The CTU is the basic unit of prediction in HEVC (analogous structure to a macroblock in prior standards), and consists of a luma CTB and the corresponding chroma CTBs. Each CTU is split recursively into multiple coding units (CUs) (see Fig. 1). Each CU has an associated partitioning into prediction units (PUs) and a tree of transform units (TUs). The best combination of HEVC quadtree partitioning is obtained using rate-distortion optimisation (RDO) [7]:

$$\min J = D + \lambda R, \tag{1}$$

where $J$ is the rate-distortion (RD) cost and it is minimised for a particular value of $\lambda$. RD represents the number of bits, measured at a rate $R$, to transmit a reconstructed signal without exceeding a given distortion $D$ [8]. Since RDO is performed exhaustively – testing all possible combinations –, the minimisation involves high complexity, which may not be suitable for real-time applications [7].



(a) Quadtree structure (b) Block structure

**Fig. 1.** Division of a CTU into CUs

In this paper, an approach for frame partitioning, in order to reduce the computational cost of RDO is presented. The strategy is top-down and it uses motion features to decide whether to split a CU. An initial motion estimation is calculated using the difference between two frames. The Gaussian mask and the Sobel operators are used on the resulting image for indicating where motion regions are located. A threshold is used on the image-gradient magnitude for obtaining motion features. Results show that CTU partitioning based on motion features achieves CUs according to the amount of motion of the content, and therefore, homogeneous prediction blocks (PBs) are obtained for performing motion estimation.

The remaining sections are organised as follows: Section 2 presents some related works. Section 3 contais the partitioning strategy; Section 4 is focused on the experimental evaluation; and Section 5 includes final remarks.

## 2 Related Works

Different strategies for frame partitioning can be found in the literature. Liu et al. [4] proposed a method based on block edge information, which consists in determining when a block is suitable for subsampling. For this purpose, a block is classified as a flat block or edge block. A flat block is one whose elements have high likelihood, while an edge block is one whose elements have low likelihood. Edge blocks have high amount of high frequency signal. If these blocks are subsampled, the prediction error will be higher. For this reason, a block is subsampled when it is classified as a flat block. Kim and O'Connor [2] proposed a strategy based on edge detection using Walsh-Hadamard transform (WHT) and a skip mode detection according to quantisation parameter. If a block has an edge, it is split into four blocks of same size. If the sum of WHT coefficients after quantisation is zero, the current block is considered as a skip mode. Wang et al. [10] proposed a block size selection method which estimates the initial motion vector and the edge direction of a block, which are used to perform merging and splitting decisions.

Gohokar and Gohokar [1] strategy selects block sizes taking into account texture information which is based on the energy in the AC coefficients of the discrete cosine transform. The algorithm stops block size reduction for visually irrelevant regions. Mera and Trujillo [5] proposed a strategy that assesses variability in each block in order to determine homogeneity using different measures. The intensity variation of a block is compared with the parent for deciding whether to split the block. Zhang et al. [11] strategy is based on entropy. If the entropy of a CU is extremelly small or is smaller than the average entropy, the CU is considered as optimal. If the entropy of a CU is extremely large, the CU is partitioned.

## 3 Partitioning of Coding Tree Units Based on Edges

The partitioning of CTUs is determined by the homogeneity of the content, which is assessed using motion features. Motion features are edges extracted on an initial motion estimation image. The use of edge detectors – in the ideal case –, may produce a set of curves that indicate the boundaries of objects and may reduce irrelevant information, whilst preserving the important structural properties of an image. The partitioning strategy is presented along this section and can be summarised in three main steps: initial motion estimation, gradient based edge detection, and partitioning decision. Fig. 2 shows an output of these steps.

### 3.1 Initial Motion Estimation

The first step consists in calculating a difference between a current frame and a reference frame. The difference image is a simple method for estimating motion between two frames at a low computational cost, which will be used to perform frame partitioning only in regions with motion.

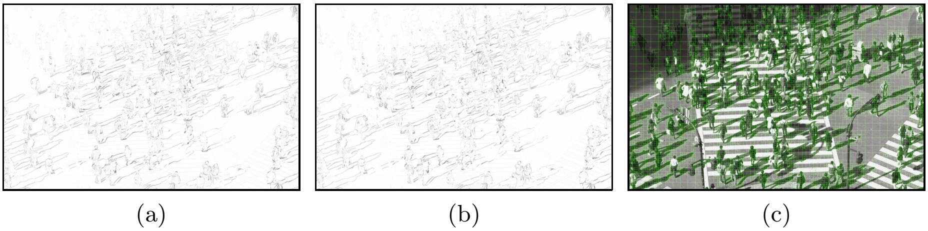


**Fig. 2.** Output of the main steps of the proposed approach using two frames from PeopleOnStreet, (a) is the difference image, (b) the motion region map, and (c) the partitioned frame

### 3.2 Gradient Based Edge Detection

Edges are pixels at, or around which, the local intensities undergo a sharp variation along a particular orientation. Edge detectors produce a binary map – also called edge map – that highlights edges in an image. The simplest method of edge detection consists in thresholding the image-gradient magnitude [9], which can be estimated by smoothing an image and differentiating it:

$$J_x = \frac{\partial I}{\partial x} = ((I * G) * H_x)(x) \quad \text{and} \quad J_y = \frac{\partial I}{\partial y} = ((I * G) * H_y)(y), \tag{2}$$

where $*$ is the convolution operator, $I$ is the input image, $G$ is a Gaussian kernel, and $H_x$ and $H_y$ are the Sobel operators [9]. The image-gradient magnitude is calculated as:

$$||\nabla I|| = \sqrt{J_x^2 + J_y^2}. \tag{3}$$

The edge map is obtained by thresholding the image-gradient magnitude. The threshold $T$ is determined using the Otsu's method, which assumes that an image contains pixels from two classes, whose intensities are unknown. The algorithm finds a threshold $T$ such that background and foreground distributions are maximally separated, which implies minimising the intra-class variance or maximising the between-class variance [6]. Resulting edge map contains motion features between the current and the reference frame. In this context, edge maps are called motion region maps.

### 3.3 Partitioning Decision

A current frame is split into CTUs of 64 × 64 size. At each node, a test is performed to decide whether the content of the CU has motion. The test consists in determining if the CU of the current node contains at least one motion region – based on the corresponding motion region map. If the test is true, the CU is split into four CUs of same size. Otherwise, the node becomes a leaf and motion estimation is applied. This process is repeated while a CU contains a motion region and the minimum CU size has not been reached. Figure 3 contains three examples of CTU partitioning.

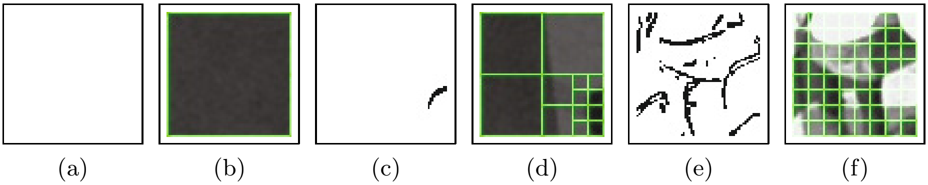


**Fig. 3.** Illustration of CTUs of 64 × 64 partitioning using three different settings: (a), (c), and (d) are the motion region maps, whilst (b), (d), and (f) show obtained partitions on frame contents

# 4 Experimental Evaluation

The proposed approach was implemented in C++ using the OpenCV (Open Source Computer Vision) library – version 2.4.8.

The CTU partitioning and the motion estimation are calculated using the luma channel. CUs size varies from $64 \times 64$ to $8 \times 8$. The search parameter $p$ is set to 7 and the sum of absolute differences (SAD) is used as BDM. The search is performed using the full-search BMA, which compares all $(2p + 1)^2$ possible PBs in the search window to find the best match. Furthermore, for a current frame, a reference frame is the one preceding it.

Eight benchmark videos are used for the experimental evaluation, available at the ftp server of the Leibniz Universität Hannover (ftp.tnt.uni-hannover.de). Characteristics of the videos are presented in Table 1, and illustration of the content is in Fig. 4.

**Table 1.** Characteristics of benchmark videos

| Class | Sequence name | Spatial resolution | Frame rate | Frames | Camera motion |
|---|---|---|---|---|---|
| A | PeopleOnStreet | 2560 × 1600 | 30 | 150 | No |
| B | Kimono | 1920 × 1080 | 24 | 240 | Yes |
| B | ParkScene | 1920 × 1080 | 24 | 240 | Yes |
| C | BQMall | 832 × 480 | 60 | 600 | Yes |
| C | PartyScene | 832 × 480 | 50 | 500 | Yes |
| D | RaceHorses | 416 × 240 | 30 | 300 | Yes |
| D | BasketballPass | 416 × 240 | 50 | 500 | Yes |
| E | MobileCalendar | 1280 × 720 | 60 | 600 | Yes |

The proposed approach is compared with the strategy proposed by Kim and O'Connor [2], which works based on edge detection using the WHT. This strategy uses a $2 \times 2$ matrix, and requires twelve operations per four samples. On the other hand, the proposed approach performs three convolutions and a thresholding

(a) PeopleOnStreet (b) Kimono (c) ParkScene (d) BQMall

(e) PartyScene (f) RaceHorses (g) BasketballPass (h) MobileCalendar

**Fig. 4.** Benchmark videos

for calculating the motion region map. Thus, the algorithm requires at least ten operations per sample, what makes it more computationally expensive than edge detection based on WHT.

The WHT approach generates coarse edges, whilst the Sobel operators provide more information about intensity changes. Thus, it is expected that CTUs have been highly partitioned in different ways using the proposed partitioning, as shown in Table 2.

**Table 2.** Comparison of CTUs partitions between the proposed approach and the WHT based approach

| Sequence name | # of CTUs | CTUs equally split | # of CUs | |
|---|---|---|---|---|
| | | | Proposed approach | WHT approach |
| PeopleOnStreet | 149000 | 0.18 | 5225042 | 2304095 |
| Kimono | 504529 | 0.84 | 3723988 | 874306 |
| ParkScene | 504529 | 0.83 | 3931420 | 2241841 |
| BQMall | 69484 | 0.26 | 2182936 | 1460839 |
| PartyScene | 57884 | 0.23 | 1880819 | 1456049 |
| RaceHorses | 35880 | 0.62 | 311232 | 301269 |
| BasketballPass | 59880 | 0.71 | 427461 | 301269 |
| MobileCalendar | 149391 | 0.42 | 5154387 | 3878661 |

Table 2 contains the number of CTUs of 64 × 64 multiply by the number of frames in the video sequences, in the second column; the ration between the number of CTUs equally split by the two approaches and the total of CTUs to split, in the third column; the number of CUs obtained using the proposed approach, in the fourth column; and the number of CUs obtained using the WHT based approach, in the fifth column.

The PeopleOnStreet video captures an scene with moving objects (people) using an static camera. Thereby, the motion is focused mainly on the people crossing the street. For this video it is expected to have a low percentage of CTUs equally partitioned, due to an edge map represents such moving objects, and the edges obtained by the approaches compared are highly different.

The proposed approach achieves good results due to regions with higher motion produce larger CUs, while regions with small motion produce smaller CUs. Furthermore, the proposed approach avoids performing all possible partitions performed by RDO, providing an efficient method for partitioning a frame.

The evaluation of motion estimation results is based on two criteria: efficiency and prediction quality. The efficiency is determined by the mean number of search points (# of sp), and the prediction quality is given by the peak signal-to-noise ratio (PSNR). The proposed approach presents a higher prediction quality than [2], in Table 3. However, the evaluation is only on the inter prediction component. An evaluation with the whole video coding standard is required to draw a final conclusion on the estimation quality.

**Table 3.** Motion estimation performance using full-search algorithm

| Sequence name | PSNR | | # of sp | |
|---|---|---|---|---|
| | Proposed approach | WHT approach | Proposed approach | WHT approach |
| PeopleOnStreet | 31 | 28 | 223 | 223 |
| Kimono | 34 | 32 | 221 | 215 |
| ParkScene | 31 | 30 | 221 | 220 |
| BQMall | 31 | 30 | 220 | 219 |
| PartyScene | 28 | 28 | 218 | 217 |
| RaceHorses | 30 | 30 | 213 | 213 |
| BasketballPass | 31 | 30 | 213 | 211 |
| MobileCalendar | 31 | 31 | 221 | 220 |

## 5 Final Remarks

In this paper, an approach for CTUs partitioning is presented. The proposed approach is based on motion features and achieves good results on noticeable regions with no motion and reduces the number of operations in the inter-prediction mode.

The proposed approach calculates more motion features than the WHT based approach. However, both approaches have to be implemented and tested in the HEVC Test Model (HM) Reference Software in order to determine which one achieves CUs partitions equals to the optimal CUs calculated by the Reference Software.

If the content, between the current frame and the reference frame, of a PB is unchanged, then that PB may be coded as skip mode. In this way, the proposed approach can be used for deciding whether to mark a PB as skip mode.

Since new objects in a scene may produce large amount of motion features, the proposed approach can be used for deciding whether to use the intra-prediction mode.

## References

1. Gohokar, V.V., Gohokar, V.N.: Adaptive selection of motion estimation block size for rate-distortion optimization. International Journal of Computer Applications 17(4), 44–48 (2011)
2. Kim, C., O'Connor, N.E.: Complexity adaptation in H.264/AVC video coder for static cameras. In: Picture Coding Symposium (PCS), pp. 1–4 (2009)
3. Li, L., Liu, S., Chen, Y., Chen, T., Luo, T.: Motion estimation without integer-pel search. IEEE Transactions on Image Processing 22(4), 1340–1353 (2013)
4. Liu, Q., Chen, Z., Goto, S., Ikenaga, T.: Fast motion estimation algorithm based on edge block detection and motion vector information. In: International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), pp. 590–593 (2007)
5. Mera, C., Trujillo, M.: Using dispersion measures for determining block-size in motion estimation. DYNA 79(171), 97–104 (2012)
6. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics 9(1), 62–66 (1979)
7. Sullivan, G.J., Ohm, J.R., Han, W.J., Wiegand, T.: Overview of the high efficiency video coding (HEVC) standard. IEEE Transactions on Circuits and Systems for Video Technology 22(12), 1649–1668 (2012)
8. Sullivan, G.J., Wiegand, T.: Rate-distortion optimization for video compression. IEEE Signal Processing Magazine 15(6), 74–90 (1998)
9. Trucco, E., Verri, A.: Introductory Techniques for 3-D Computer Vision. Prentice Hall (1998)
10. Wang, X., Sun, J., Xie, R., Yu, S., Zhang, W.: An improved block size selection method based on macroblock movement characteristic. Multimedia Tools and Applications 43(2), 131–143 (2009)
11. Zhang, M., Qu, J., Bai, H.: Entropy-based fast largest coding unit partition algorithm in high-efficiency video coding. Entropy 15(6), 2277–2287 (2013)