A FAST BACKGROUND MODEL BASED SURVEILLANCE VIDEO CODING IN HEVC

Li Ma[1], *Honggang Qi*[1], *Siyu Zhu*[1], *Siwei Ma*[2]

[1]University of Chinese academy of sciences, China
[2]Institute of Digital Media, Peking University, China

ABSTRACT

High Efficiency Video Coding (HEVC) is the most up to dated video coding standard which reduces the bit rate by almost 40% with the same objective quality compared to H.264/AVC but at about 40% extra encoding complexity overhead. The main reason for HEVC complexity increase is inter prediction that accounts for 60-70% of the whole encoding time. Especially for surveillance video with huge amount of data, the reducing of coding complexity is a very pressing work. So in this paper, a novel fast method is proposed based on background model to speed up the searching procedure of surveillance video coding. This paper analyzes the proportion of different block partitions and reference frames for multiple types blocks, and puts the analysis results into practice by developing optimized mode decision and reference frame selection schemes for the HEVC encoder according to spatial-temporal corresponding CUs (coding units) or PUs (prediction units). Simulation results show that the proposed algorithm can achieve more than 41% reduction of coding time with comparable rate distortion (RD) performance.

***Index Terms*— HEVC, surveillance video coding, background modeling, reference frame selection, mode decision**

I. INTRODUCTION

HEVC is currently being prepared as the newest video coding standard of the ITU-T Video Coding Experts Group and the ISO/IEC Moving Picture Experts Group. And with the development of the society and the improvement of living standards, surveillance video coding has been much emphasized over the past years. Unlike common test sequences, surveillance video is a typical video with less active motion and less texture. Taking advantage of the special characteristics of surveillance many kinds of technologies such as object-based, background model based and nonparametric background generation in [1–3] are used for the surveillance video. That can get great performance but high complexity, so an adaptive fast algorithm is needed to reduce the complexity without negligible performance loss.

Many fast coding algorithms are proposed to reduce the complexity of the encoder. In [4]- [5] *Shen* proposed a fast multiframe selection method to deduct unnecessary reference frame searches and modes using the information of the neighboring blocks on common sequences. In [6] An Efficient Inter Prediction Mode Decision Method is employed in ME (motion estimation), which has a limited reducing computational complexity of the ME up to 51.76% instead of the whole coding procedure while ME only accounts for less than 70% in whole. A couple of other works has focused on speeding up the mode decision through early termination mechanisms in [7–9], but the techniques proposed in this paper are shown to outperform them in the majority of the test cases. In this paper, three types involve FCUs (foreground CUs), BCUs (background CUs) and MCUs (mixed CUs) are classified, and then some prediction modes are directly removed and the reference frames will be reduced for different types of CUs or PUs respectively. The proposed method can deduct the unnecessary coding time with the almost similar coding performance. Our novel fast arithmetic proposed work is implemented using *Zhang's* background model method in [10] on HEVC platform.

The remainder of this paper is organized as follows. In Section II, The comprehensive analysis and fast mode decision algorithm based on the background model are presented. Overall are described in Section III. Simulation results are presented in Section IV. Finally Section V concludes the paper.

II. PROPOSED FAST ALGORITHM BASED ON BACKGROUND MODELING

A. Background Generation

In Zhang's method [10], the background frame is built using a simple background modeling procedure and periodically updated while encoding without object detection, tracking or segmentation. It uses a non-parametric background modeling method based on mean shift algorithm to generate a background frame which will be encoded into bit stream and then is employed as the long-term reference frame for prediction as depicted in Fig.1. This background modeling method achieves more significant complexity reduction and better coding performance compared to other object-oriented or background-prediction based methods.

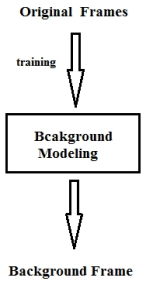

Fig. 1: The framework of background model

978-1-4799-9/14/$31.00

B. CU Classification

In the proposed algorithm, CUs are classified into three categories: FCUs, BCUs and MCUs according to the difference between the original CU and the co-located CU in the reconstructed background modeling frame, which is generated and updated through the method in [8] as depicted in Fig. 2.

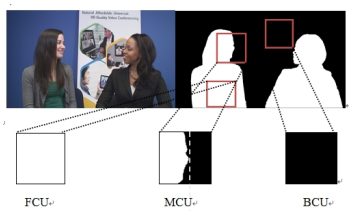

Fig. 2: The different types of CUs

Before classifying a CU, the pixels in the CU are first decided to be foreground or background pixels:

$$P_{x,y}(\boldsymbol{C}) = \begin{cases} P_F, \left| P^c_{x,y} - P^b_{x,y} \right| > \tau_0 \\ P_B, \left| P^c_{x,y} - P^b_{x,y} \right| <= \tau_0 \end{cases} \quad (1)$$

where $P_{x,y}(\boldsymbol{C})$ is the type of the pixel, $P^c_{x,y}$ is the pixel of current block, and $P^b_{x,y}$ is the collocated pixel in the background frame. P_F is the foreground pixel, and P_B is the background pixel. Threshold τ_0 is set to 5 for current pixel, which may be optimal in our experiments. The result of pixels classification with the $\tau_0 = 5$ which is the most stable through experiments from 1 to 10 is shown in Fig.3.

Fig. 3: (a) Original video of crossroad (cif) (b)Mask of result of subtraction of crossroad (cif), the black area is the background, the white area is the foreground

After the classification of pixels, the category of the current coding unit $CU(\boldsymbol{C})$ will be obtained through calculating and comparing the proportion of P_F. The classification method is expressed as follows:

$$CU(\boldsymbol{C}) = \begin{cases} FCU, \sum_{x=0}^{N-1}\sum_{y=0}^{N-1}(P_{x,y}(\boldsymbol{C}) == P_F)/N^2 > 1-\theta \\ BCU, \sum_{x=0}^{N-1}\sum_{y=0}^{N-1}(P_{x,y}(\boldsymbol{C}) == P_F)/N^2 < \theta \\ MCU, \textbf{\textit{others}} \end{cases} \quad (2)$$

where x and y are the horizontal and vertical location of every pixel in the current CU. N is the width or height of the current CU. The threshold $\theta = 1/16$ is the empirically value which accords with the background partitioning by human eyes the most.

However, the inter mode decision process in HEVC is performed using all the possible prediction modes and reference frames to find the one with the least RD cost. This achieves the highest coding efficiency but requires a very high computational complexity.

C. Adaptive Early Termination of CU Mode Decision

In HEVC [11], a picture is divided into multiple largest coding units (LCUs). LCU can be further split into CUs, and their sizes can vary from 64x64 to 8x8, determined by recursively splitting according to the RD cost. The prediction unit (PU) is the basic block to process inter prediction, which the size is set with the corresponding CU size and then the partition mode is chosen. There are 8 partition modes for inter PU as shown in Fig.4, and the second line is AMP (asymmetric motion partition).

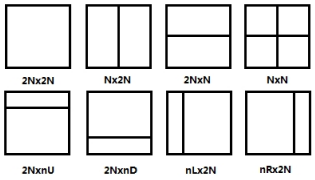

Fig. 4: 8 partition modes for inter PU

It is easy to understand that the large PU size is more suitable for CU with homogeneous texture. Experimental results show that small partition sizes like intra mode and AMP mode cover only a small fraction of the possible mode decision schemes for BCU or FCU inter coding. Table 1 tabulates the ratios of selection of INTRA and AMP for different types of CU:

Table 1: Ratios of intra and AMP for both BCU and FCU

CIF	Average accurate rate		SD	Average accurate rate	
	INTRA	AMP		INTRA	AMP
bank	0.0107	0.0263	crossroad	0.0274	0.0262
campus	0.0090	0.0289	classover	0.0188	0.0264
overbridge	0.0201	0.0355	overbridge	0.0177	0.0247
snowroad	0.0079	0.0282	campus	0.0159	0.0280
average	0.0119	0.0297	average	0.0199	0.0263

Hence, the two modes should be removed for reducing the computation load. What's more, if the CU is FCU or BCU, CU splitting is terminated because we can get the picture region from the reference frame easily. The CU coding candidate modes set Ω can be mathematically expressed as follow:

$$\Omega = \begin{cases} \Lambda - \{M_{intra}, M_{AMP}, CU_{split}\}, \ CU(\boldsymbol{C}) = BCU or FCU \\ \Lambda, \ CU(\boldsymbol{C}) = MCU \end{cases} \quad (3)$$

where Λ denotes the set including all CU modes defined in HEVC in the equation.

In general, through the mode maps of surveillance video, the motion and texture characteristics are almost not in the background area. So there is still room for reducing the complexity by using the mode relations among the spatial-temporal CUs as shown in Fig.5 for BCU. If the temporally collocated CU in the nearest forward frame and the neighboring CUs in this frame are all skip mode, only skip mode and 2Nx2N mode can be used for the current BCU.

The validity of the prior method is verified by experimental results on a variety of video sequences in different resolutions in Table 2. The accurate ratio is over 96% on average which is high enough to ensure the performance without loss. The QPs (quantization parameters) of every sequence are 27, 32, 38, and 45.

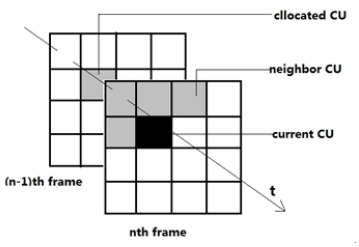

Fig. 5: Temporal-spatial correlations of CUs for BCU (left CU, above CU, above left CU, above right CU, the collocated CU in the preciously frame)

Table 2: CU mode correlations among the contiguous CUs for BCU

CIF	ratio	SD	ratio
bank	0.9780	bank	0.9618
campus	0.9674	classover	0.9913
snowroad	0.9736	campus	0.9674
average	0.9730	average	0.9735

D. Adaptive Reference Frame Selection

A picture can be coded through temporal predicting from multiple reference pictures which are organized in two reference picture lists and the reference frames are ordered in advance in HEVC. The reference index identifies which of the reference pictures in the list should be used for creating the prediction signal in [12]. However, each reference frame are of different importance in the list. It should be noticed that the bitrate will increase with the enlargement of the reference list. Although the maximum number of reference frame is allowed in HEVC, in most cases the FCU and BCU will select only very few reference frames. According to our experimental results, the ratio of the usage of the BG frame which is used as the last reference frame for FPU (foreground partition unit) is 0.05 on average, and the ratio of choosing the first reference frame and BG frame as the optimal frame for BPU (background partition unit) is 0.87 on average. Thus, only the first reference frame and BG frame are selected for BPU, and the other common reference frames in the list are chosen rather than the BG frame for FPU. The reference selection can be expressed as:

$$\phi = \begin{cases} \phi_B = \{ref_0, ref_{bg}\}, & BPU \\ \phi_F = \{ref_0, ref_1, \cdots, ref_n\}, & FPU \end{cases} \quad (4)$$

where ref_{bg} represents the background frame with the last reference index and $ref_0, ref_1, \cdots, ref_n$ are the frames in the reference list without ref_{bg}.

The proportion of FCU is much less in the surveillance video coding, thus the complexity can be reduced efficiently. Moreover, images tend to have a large successively area of background, so the correlation among adjacent CUs is high under most conditions. The neighboring PUs' modes are used to determine the current PU's mode if the PU is BPU. If all of the neighboring PUs, including the left PU, the above PU, the above left PU, the above right PU and the left bottom PU, have the same reference frame index, the reference list only includes the reference frame for the current PU. The spatial relationship of the current PU and it's neighboring PUs are illustrated in Fig 6.

Table 3 demonstrates the ratio of the reference frames being identical for both current PU and all its neighboring PUs. It can be seen that the ratio is over 85% on average. We get the average statistical data of each sequence with the QPs are 27, 32, 38, and 45.

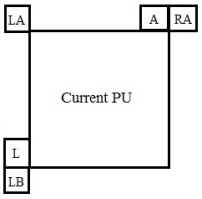

Fig. 6: Spatial correlations of PUs for BPU (LA: left above PU, A: above PU, RA: right above PU, L: left PU, LB: left bottom PU)

Table 3: Reference frame correlations among the neighboring PUs

CIF	ratio	SD	ratio
campus	0.8625	classover	0.8001
snowgate	0.9033	overbridge	0.8635
snowroad	0.8507	campus	0.8428
average	0.8722	average	0.8355

III. OVERALL ALGORITHM

The proposed fast algorithm is enabled after the background frame is generated and reconstructed as a reference frame. The framework is depicted in Fig.7 and the detailed overall algorithm of a CU is described as follows:

1) Performing the procedure of the Eq.1 and Eq.2.
2) If the current CU is BCU and the modes of all the corresponding CU as shown in Fig.6 are skip mode, only skip mode and 2Nx2N mode can be used; then performing the procedure of the Eq.3.
3) If it is BPU and the reference frames of all the neighboring PUs as shown in Fig.7 is the same, only this reference frame is selected; then preforming the procedure of Eq.4.
4) Performing the remaining inter and intra prediction procedure and get the minimal RD cost.
5) Determining the best prediction mode according to the minimal RD cost.

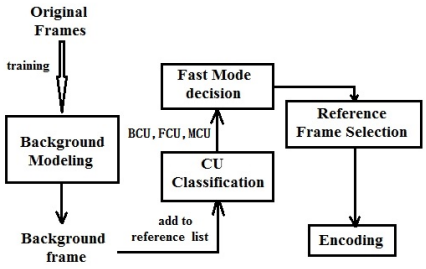

Fig. 7: Coding framework of our proposed method

IV. EXPERIMENTAL RESULTS

To verify the performance of the proposed algorithm, the proposed algorithm is implemented on HEVC reference software HM12.0. The anchor is *Zhang's* background model without the proposed fast algorithm. The tested sequences are surveillance videos including *bank*, *campus*, *snowgate*, *snowroad* in cif format and *bank*, *campus*, *classover*, *overbridge* in sd format.Here, for the modeling background frame, the generation period is 120 fames and the updating modeling time is 600 frames. The QP of background frame is 10 and the set of QPs are 22, 27, 32, and 37. The experimental results are demonstrated in Table 4.

Table 4: BD-rate and time reduction compared with anchor

Sequence		BD-rate			Time saving
		Y	U	V	
CIF	bank	0.5%	-0.4%	-3.0%	22.87%
	campus	1.5%	-0.9%	-0.1%	47.82%
	snowgate	0.6%	-1.9%	-4.9%	49.70%
	snowroad	0.7%	-2.7%	-1.1%	37.44%
SD	bank	1.3%	-0.6%	-1.5%	41.66%
	campus	1.0%	-0.6%	-1.2%	44.97%
	classover	2.6%	-0.5%	-1.4%	52.59%
	overbridge	1.5%	-0.5%	-1.3%	38.16%
Average		1.2%	-1.0%	-1.8%	41.91%

Table 4 tabulates that the proposed fast mode decision method can achieve almost 42% time saving on average under low delay P configuration, which the time savings is defined as equation 5. On the other hand, the loss of the proposed for all test sequences is no more than 1.2% loss in luma but increasing performance on chroma 1.0% and 1.8% respectively, which indicates that the proposed method get good results for most surveillance videos. This method can also be used in other platform with a long-term reference frame.

$$\Delta T = (T_{Anchor} - T_{Proposed})/T_{Anchor} \quad (5)$$

Moreover, we give a comparison of coding complexity results of original background modeling method and the proposed fast method. Two sequences of QP 22 in different resolutions are chosen visualized in Fig.8, the proposed shows much complexity reducing against anchor.

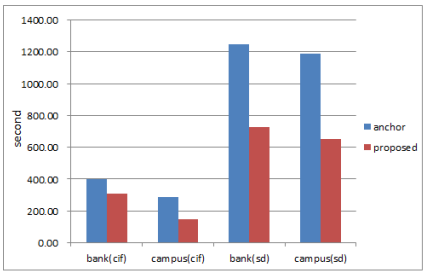

Fig. 8: Coding complexity comparison of original background modeling method and the proposed fast method

V. CONCLUSIONS

This paper focuses on a novel background model based fast mode decision schemes for surveillance in HEVC. In the proposed algorithm, variable-sized CUs are classified into three kinds: FCUs,BCUs and MCUs, then the rarely used prediction modes and reference frames are removed for different CUs and PUs adaptively using both the characteristics of the surveillance video and the information of the corresponding CUs or PUs in spatial-temporal direction. Experimental results show that the proposed method can achieve significant coding time saving while still maintaining almost the same coding performance.

VI. ACKNOWLEDGMENT

This research is supported by the National Hightech R&D Program of China (863 Program, 2012AA010805), the National Science Foundation of China (61322106, 61379100, 61472388 and 61103088), which are gratefully acknowledged.

VII. REFERENCES

[1] Isabel Martins and Luis Corte-Real, "A video coder using 3-d model based background for video surveillance applications," in *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*. IEEE, 1998, vol. 2, pp. 919–923.

[2] Divya Venkatraman and Anamitra Makur, "A compressive sensing approach to object-based surveillance video coding," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*. IEEE, 2009, pp. 3513–3516.

[3] Yazhou Liu, Hongxun Yao, Wen Gao, Xilin Chen, and Debin Zhao, "Nonparametric background generation," *Journal of Visual Communication and Image Representation*, vol. 18, no. 3, pp. 253–263, 2007.

[4] Liquan Shen, Zhi Liu, and Zhaoyang Zhang, "Fast multiframe selection algorithm based on spatial-temporal characteristics of motion field," *Journal of Electronic Imaging*, vol. 17, no. 4, pp. 043004–043004, 2008.

[5] Liquan Shen, Zhi Liu, Ping An, Ran Ma, and Zhaoyang Zhang, "Low-complexity mode decision for mvc," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 21, no. 6, pp. 837–843, 2011.

[6] Alex Lee, DongSan Jun, Jongho Kim, Jinwuk Seok, Younhee Kim, Soonheung Jung, and Soo Choi Jin, "An efficient inter prediction mode decision method for fast motion estimation in hevc," in *ICT Convergence (ICTC),2013 IEEE International Conference on*. IEEE, 2013, pp. 502 – 505.

[7] Qin Yu, Xinfeng Zhang, and Siwei Ma, "Early termination of coding unit splitting for hevc," in *APSIPA ASC, 2012 IEEE International Conference on*. IEEE, 2012, pp. 1 – 4.

[8] Shen X, Yu L, and Chen J, "Fast coding unit size selection for hevc based on bayesian decision rule," in *Picture Coding Symp(PCS), 2012 IEEE International Conference on*. IEEE, 2012, pp. 453–456.

[9] Kim Jaehwan, Yang Jungyoup, Won Kwanghyun, and Jeon Byeungwoo, "Early termination of coding unit splitting for hevc," in *Picture Coding Symposium(PCS), 2012 IEEE International Conference on*. IEEE, 2012, pp. 449 – 452.

[10] Xianguo Zhang, Luhong Liang, Qian Huang, Yazhou Liu, Tiejun Huang, and Wen Gao, "An efficient coding scheme for surveillance videos captured by stationary cameras," in *Visual Communications and Image Processing 2010*. International Society for Optics and Photonics, 2010, pp. 77442A–77442A.

[11] Rickard Sjoberg, Ying Chen, Akira Fujibayashi, Miska M Hannuksela, Jonatan Samuelsson, Thiow Keng Tan, Ye-Kui Wang, and Stephan Wenger, "Overview of hevc high-level syntax and reference picture management," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1858–1870, 2012.