Fast ROI-Based HEVC Coding for Surveillance Videos*

Huaying Xue, Yuan Zhang, Yunong Wei
Information Engineering School
Communication University of China
Beijing 100024, China
aimmeexue@outlook.com, yzhang@cuc.edu.cn

***Abstract*—In recent years, surveillance videos are playing an important role in modern security systems. As a state-of-the-art video coding standard, High Efficiency Video Coding (HEVC) has been deployed in many surveillance cameras. Nevertheless, the request of real-time and high subjective quality coding still poses the great challenging problem of efficient representation and compression for high complexity HEVC. In this paper, we present a fast Region-of-Interest (ROI) based HEVC coding system for surveillance videos. First of all, the automatic segmentation is conducted to generate the ROI mask. The mask is then used in the quadtree-based block partitioning algorithm and the rate control algorithm so as to reduce the computational costs on non-ROI regions. In particular, we compress the non-ROI regions with a two-step method: fast prediction mode selection and fast coding unit(CU) depth level (size) selection. The ROI-based rate control method is performed as well. The bit allocation between different regions depends on the proportion of ROI region in the whole picture. The proposed system can significantly reduce the computational costs with negligible loss in terms of PSNR at a given bitrate, while the visual quality of ROI can be much improved.**

Keywords—HEVC; Region of Interest(ROI); fast decision; Surveillance Video

I. INTRODUCTION

During the last decade, video surveillance system has been flooding into lives of humans. Limited by the storage capacities and transmitted bandwidth, a huge amount of raw video data must be compressed. Fortunately, the emerging High Efficiency Video Coding (HEVC) standard [1][2] is conducted under the efforts of Joint Collaborative Team on Video Coding (JCT-VC) jointly established by ISO/IEC MPEG and ITU-T VCEG, which provides a doubling in coding efficiency with respect to the H.264/AVC high profile.

HEVC adopts a quadtree coding structure, in which the largest coding unit (LCU) can be a maximum of $2N_{MAX}\times2N_{MAX}$ (N_{MAX}=32) pixels and the smallest coding unit (SCU) can be a minimum of $2N_{MIN}\times2N_{MIN}$ (N_{MIN}=4) pixels. Each CU (depth level=k) can be recursive split into four equal sub-CUs (depth level=k+1) and each leaf sub-CU can be split into multiple prediction units (PU) as well. Fig.1 illustrates CU and PU coding structure of HEVC.

Eight partition patterns (or size) are used to define the PUs. Inter-coded CUs can apply all modes, while Intra-coded CUs

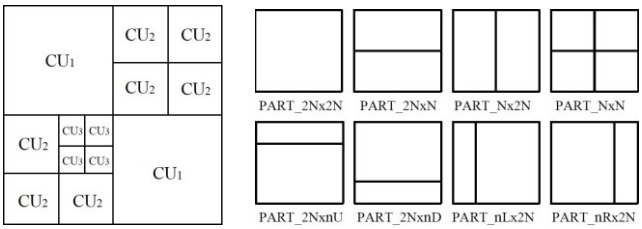

Fig. 1. CU and PU coding structure of HEVC, where k indicates the depth level of CU.

can apply only PART_2Nx2N and PART_NxN. Note that PART_NxN is adopted only when the current CU is the SCU. A Skip-coded CU is considered to be a special inter prediction mode without coding of motion vector differences and residual information, which only has PART_2Nx2N. Motion estimation (ME) and motion compensation (MC) are performed with all prediction modes and depths to find the least rate-distortion cost (RDcost). The newly introduced quadtree-based block partitioning can significantly improve the coding efficiency, yet causes a dramatic increase in coding complexity, due to the necessity of finding the best combinations of 11 prediction modes and 4 depth levels in quadtree.lu

To solve the above problem, a number of algorithms [3]-[7] have been proposed to speed up its selection. Early skip detection (ESD) [3] and coded block flag (CBF) fast mode decision (CFM) [4] are mainly based on the terminate detection algorithms by analysis of coding information, including Skip mode and CBF. Leng et al. [5] combined the fast mode decision algorithm on frame level and CU level, predicting the current CU mode by referring co-located CU in the previous frame and the neighboring CUs. Early CU(ECU) termination algorithm terminate the splitting process if mode of current CU is decided as Skip mode[6]. Shen et al. [7] propose a Bayesian-based fast CU size decision algorithm, which employs important and computational-friendly features to assist decisions on CU splitting by minimizing the Bayesian risk.

On the other hand, the surveillance videos are captured by stationary cameras. The region of moving object in the surveillance video is the main region that people pay attention to, which is called Region-of-Interest (ROI) area. The aforesaid HEVC coding algorithms haven't taken the characteristic and requirements of surveillance system into consideration. Intuitively, the encoder should allocate more resources including computational resource and bitrates to the ROI area

Project supported by National Nature Science Foundation of China (No.61472389).

so as to improve the overall visual quality. Therefore, it is reasonable to compress ROI and non-ROI separately, naturally leading to the ROI-based algorithms.

There are three main problems that ROI-based algorithm needs to deal with: ROI detection, ROI-based rate control and ROI-based computational resource allocation. Deng et al. [8] generated a pixel-wise weight map according to the extraction of face regions and proposed a weight-based unified rate-quantization (URQ) scheme. Next, they proposed another method [9] to determine the maximum depth of LCU on the basis of average weight of visual importance and the target complexity. Some other scheme designed a fix prediction mode candidate list and CU partitioning termination conditions for different regions [10], which is based on the improved coding structure. These methods suffer from the main drawback that it fails in dealing with all the three problems with a common HEVC model.

In this paper, we propose a fast ROI-based HEVC coding system for surveillance videos. The main idea is to reduce the bitrate and computational costs on non-ROI regions. At the frame level, a fast ROI detection is performed before coding, which outputs a binary picture as a mask. Subsequently, bitrates are allocated to different region with a fix factor K, aiming at enhancing the quality of ROI. At the CU level, we propose a two-step method: fast prediction mode selection and fast CU depth level selection for non-ROI region, on the basis of the mode decision statistics and the correlation among consecutive frames. It should be noted that, we keep the exhaustive coding decision procedure for ROI, because ROI is essentially important for surveillance videos.

The rest of this paper is organized as follows. In Section II, an overview of the proposed coding system is presented. Section III introduces ROI detection of surveillance videos. Section IV describes the proposed ROI-based fast algorithm and rate control algorithm in detail. Experimental results are presented in Section V, followed by the conclusion in Section VI.

II. Overview of the Proposed System

Fig. 2 gives the flowchart of the proposed fast ROI-based HEVC coding system. This system can be divided into three modules: ROI detection, ROI-based fast decision and ROI-based rate control. The algorithms are performed at two levels: frame level and CU level.

At the frame level, we perform a background subtraction algorithm to generate a mask, precisely identifying the ROI regions. Some modifications are introduced to meet the requirements of further process. Note that the first frame is coded in the traditional way so as to initialize the background model. The detailed algorithm will be described in Section III. Then, the bitrates are assigned to the ROI region K times more than non-ROI region according to the rate control model established in [11].

At the CU level, we firstly classify each CU into two categories: CU of ROI region and CU of non-ROI region, represented by CU_R and CU_N respectively. The independent bit allocation and quantization parameter(QP) computing of

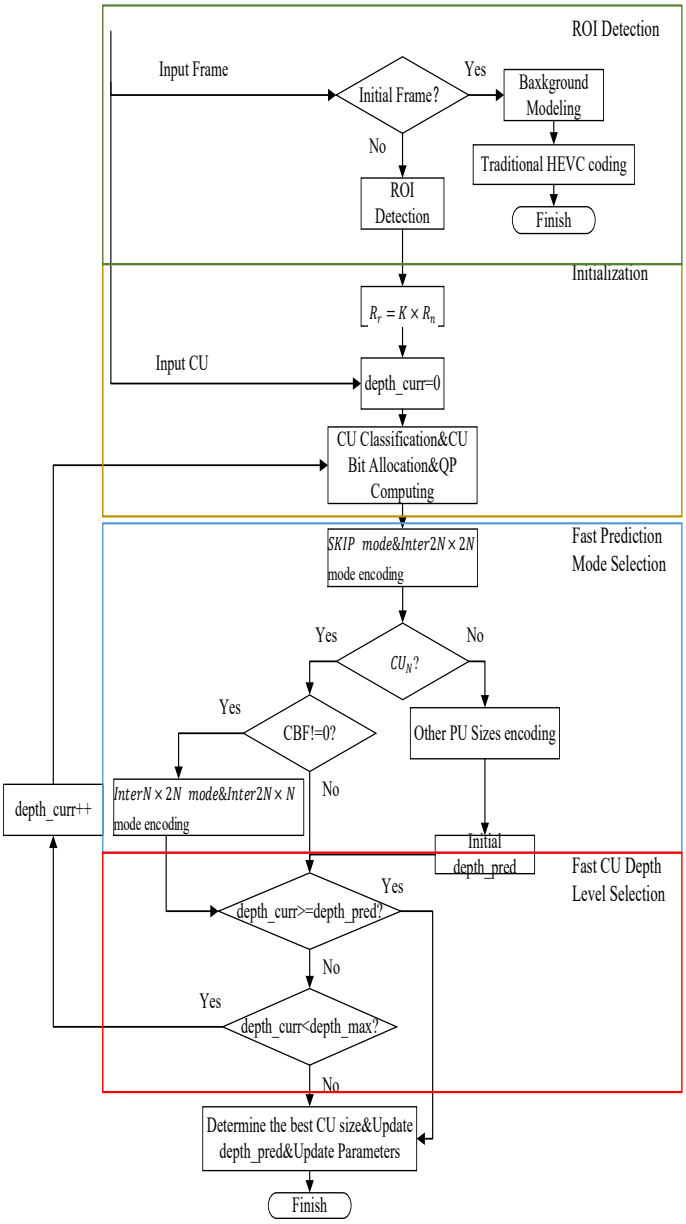

Fig. 2. Flowchart of the proposed system.

CUs are made for different regions. Furthermore, we propose a two-step method: fast prediction mode selection and fast CU depth level selection only for CU_N. Meanwhile, we still use rate-distortion- optimization (RDO) based exhaustive mode search to select the best coding sizes for CU_R. With the analysis of prediction mode distribution and depth-information correlation between temporal adjacent CU_N, some CU sizes and its modes can be adaptively skipped to terminate the exhaustive RDO search process, which will be described in Section IV.

III. ROI Detection

Many background subtraction techniques have been proposed, such as Gaussian Mixture Model (GMM) [12], the W4 model [13] and the Σ-Δ motion detection filter [14].

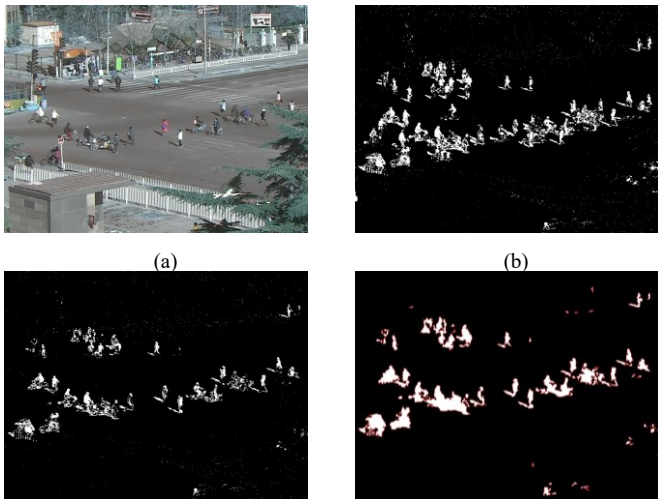

Fig. 3. Example of ROI detection. (a) The 44th original frame of the Crossroad sequence (b) detected mask obtained with ViBe (c) default subsampling factor=1 (d) final detected ROI mask.

Although these methods could get accurate results, they are hard to be implemented due to the high computational cost. Therefore, we utilize a fast yet efficient background subtraction technique called visual background extractor(ViBe) [15]. It is a software module for extracting background information from moving images. The background model is initialized with a single frame and built with a set of samples for each pixel. By comparing pixel value with the set of samples in a two dimensional Euclidean color space, each frame can generate a binary mask, indicating the foreground(ROI) and background (non-ROI) regions. Inevitably, some tricky problems would affect the performance of ViBe, including dynamic background, sudden illumination changes etc. As shown in Fig. 3(b), ghost is the most annoying artefact caused by two main reasons: the presence of a moving object in the first frame and the inclusion of foreground objects in the background model if they remain static for too long.

To address the above problem, we alter the default subsampling factor from 16 to 5, and even 1 to speed up the inclusion of ghosts in the background model. And the spatial update mechanism of ViBe ensures that the process is faster than the inclusion of real static foreground objects (see Fig. 3(c)).

In addition, we perform some post-processing operations for preserving the integrity of region segmentation. Firstly, dilation operation is applied to enlarge the boundaries of foreground regions. The areas of foreground pixels grow in size while holes within those regions become smaller. Secondly, we remove the object from foreground if its contour parameter is smaller than L which is calculated as:

$$L = \frac{image->height + image->width}{perimscale} \quad (1)$$

where *perimscale* is a control parameter (set to be 60 in our system).The final detected ROI mask is shown in Fig. 3(d).

IV. Fast ROI-Based HEVC Algorithm

Compared with the fixed 16×16 macroblocks in H.264/AVC, the flexible size of CU is regarded as one of the most significant contributions to HEVC. However, the exhaustive searching strategy to compare all possible CU sizes and prediction modes is time-consuming.

Generally speaking, surveillance videos are captured by cameras at a fixed location for a long period, during which the background is relatively motionless. As shown in Fig. 4, there are some features of CU_N that can be utilized for possible early termination.

In this section, we take several experiments to study the relationship between CU_N and the best coding choices. Then we propose fast prediction mode selection and fast CU depth level selection for CU_N, which will be discussed in Section A&B. In Section C, we take the ROI-Based rate control method [11] to improve the subjective quality of ROI region.

Note that we classify CU with different depth (size) into CU_R and CU_N according to the binary segmentation mask. If the pixel number of ROI region in a CU is zero, then this CU will be categorized as CU_N; otherwise, it will be a CU_R.

A. Fast Prediction Mode Selection

To discover the rule of optimal selection, we conduct an experiment implemented on the HEVC reference software (HM 14.0) to get the distribution of best prediction mode at each depth.

Based on the percentage of different modes, we classify all possible prediction mode into three major categories: SKIP& $Inter_{2Nx2N}$, $Inter_{2NxN}$& $Inter_{Nx2N}$ and others. As Table I shown, large sizes like Skip mode and Inter mode of 2Nx2N, 2NxN and Nx2N account for more than 95% on average, while the total average percentage of CUs with other modes is about 3.7%. There are a lot of "flat" or "homogeneous" blocks in non-ROI region with similar motion vector (e.g., roads, plants and buildings), which are more suitable to be coded with larger prediction size. Choosing a large size may lead to less side information and will not result in much larger residual, and it's considered as very efficient for CU_N.

In order to avoid the exhaustive search on those most impossible prediction modes, we redefine the candidate prediction mode list only with large sizes: Skip, $Inter_{2Nx2N}$, $Inter_{2NxN}$ and $Inter_{Nx2N}$. On the other side, the proportion of Skip and $Inter_{2Nx2N}$ exceed 90%. Zero transformed coefficient levels are exploited to perform early termination of Skip and $Inter_{2Nx2N}$ for further speed up. To be specific, we will get CBF

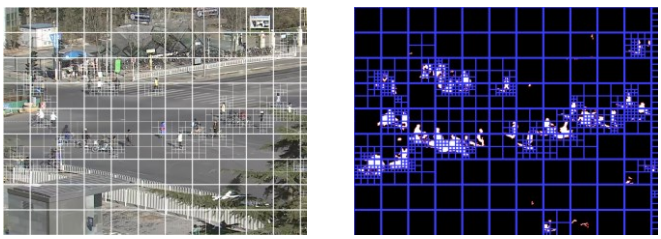

Fig. 4. Illustration of prediction mode and CU depth level distribution in the 44th frame of the Crossroad sequence, QP=32.

TABLE I. DISTRIBUTION OF PREDICTION MODES OF CU_N (%) ,IPPP, 100 FRAMES

PU / Sequence	SKIP&Inter$_{2Nx2N}$	Inter$_{2NxN}$&Inter$_{Nx2N}$	Others
Campus	91.49	5.19	3.32
Classover	91.50	5.52	2.98
Crossroad	83.41	9.44	7.15
Bank	89.69	6.02	4.29
Intersection	96.76	2.35	0.90
Average	90.57	5.70	3.73

after checking each prediction mode, indicating whether the Transform Unit (TU) contains any non-zero coefficients. If CBF is zero after encoding one luma and two chromas, check of remaining prediction modes will be skipped.

B. Fast CU Depth Level Selection

In HM, CU depth has a fixed range from 0 to 3 for every QP, which means 85 ($4^0+4^1+4^2+4^3$) more calculations for different CU are required for each LCU. The computational complexity increases significantly as depth grows. Similar to the choice of prediction mode, small depth levels are always selected for static background or slow motion region. Moreover, for the same sequences, Table II shows that the percentage of small depth level is larger at higher QPs (e.g., QP=37), which indicates that larger block size are more suitable for CUs at higher QPs. This can be explained by the RD cost function. When QP is large, the bitrate dominates the RD cost. Therefore, large size is preferred.

Obviously, the exhaustive selection within fixed depth range is inefficient. As we can see from Table II, only 9.7% of the CU quadtree blocks choose the depth level "3" on average. In low bitrate (high QP) situation, the possibility of selecting depth level "3" is very low with 0.5%. Thus, ME and MC on large depth level could be skipped in most cases with negligible loss of coding performance. In addition, the CUs in the non-ROI region are highly content-dependent, and most of them don't show a wide variation of depth from the co-located CU in the previous frames. Therefore, we can make use of the temporal correlations of depth information to predict the best max depth for current CU_N.

In the proposed method, we set the depth level with higher appearance frequency with prior search order. And the predicted depth level(*depth_pred*) will be set to the depth value with top priority. The remainder depth search process will be skipped after *depth_pred.* However, the main drawback of this mechanism is that it's impossible for CU to get large *depth_pred* again, which has great influence on the CUs where sudden change occurs. Because the frequency number of depth level, which is larger than *depth_pred*, is invariable, and can't be set to *depth_pred* again. To address the above problem, we use a control variable CBF to evaluate the coding performance of current depth (*depth_curr*). Only when *depth_pred* is less than or equal to *depth_curr* and the CBF of *depth_curr* is equal to zero can we stop CU splitting.

We have made some experiments to verify the performance of the proposed two fast algorithms. By exploiting the exhaustive search of prediction modes and CU depth levels in HM under the condition mentioned above, the results, demonstrated in Table III, show the effectiveness of our method with a high average accuracy (more than 90%).

TABLE II. DISTRIBUTION OF CU DEPTH LEVEL OF CU_N (%) ,IPPP, 100 FRAMES

Level / QP	0	1	2	3
22	18.79	22.09	29.45	29.71
27	58.17	22.65	13.36	5.82
32	78.75	14.34	5.47	1.45
37	86.09	10.39	3.05	0.50
Average	60.45	17.37	12.83	9.37

TABLE III. ACCURACY OF PROPOSED FAST ALGORITHM (%) ,IPPP, 100 FRAMES

Method / Sequence	Fast Prediction Mode Selection	Fast CU Depth Level Selection
Campus	96.71	94.77
Classover	96.59	92.71
Crossroad	95.39	91.35
Bank	95.38	93.15
Intersection	99.12	95.84
Average	96.64	93.56

C. ROI-Based Rate Control

Although many algorithms have been proposed for ROI-based rate control, the related research on HEVC platform is pretty limited, especially for the newest R-λ model, which is initially introduced in HM.10 and has been improved in a recent version of the reference software. In our system, we take the scheme [11], proposed by Marwa Meddeb, to improve the subjective quality of ROI region.

After ROI detection, the frame level bitrate budget is divided into two parts according to the number of pixels in per region and a positive constant K. K is a selected factor to control the ratio of allocated bits per pixel between the ROI and non-ROI region. And at the CU level, the bit allocation depends on the total bits allocated per region and on the weight of each CU of the same region. The parameters of corresponding region will be updated once the CU is coded.

V. EXPERIMENTAL RESULTS

The proposed algorithm has been implemented on the HEVC reference software HM 14.0 under the HEVC low-delay common test conditions as indicated in JCT-VC[16].Fast Encoder Decision (FED) and Fast Decision for Merge (FDM) are enabled. Only the first frame is encoded as

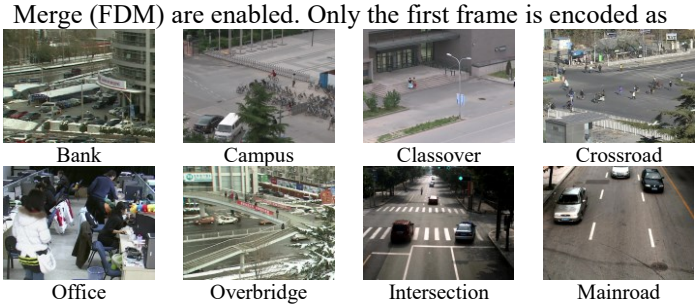

Fig. 5. The tested surveillance videos

TABLE IV. PERFORMANCE EVALUATION FOR EACH SCHEME COMPARED TO HM

Sizes	Sequence	CU_N (%)	BDPSNR (dB)		BDBR (%)		$\Delta PSNR_R$ (dB)		TS (%)	
			Ref	Prop	Ref	Prop	Ref	Prop	Ref	Prop
720×576	Bank	74.68	-0.003	-0.012	1.408	1.185	-0.325	0.010	65.349	63.008
	Campus	74.05	-0.024	-0.038	1.215	1.636	-0.181	0.001	62.620	61.917
	Classover	94.08	-0.001	0.007	0.351	-0.391	-0.172	-0.010	73.962	75.770
	Crossroad	45.46	-0.055	-0.043	1.915	1.519	-0.164	0.004	44.327	43.886
	Overbridge	48.97	-0.024	-0.026	1.251	1.033	-0.172	-0.008	56.553	49.998
	Office	38.11	-0.035	-0.041	1.351	1.505	-0.096	-0.004	39.905	33.950
1600×1200	Intersection	83.31	-0.005	-0.029	0.208	1.048	-0.118	0.000	64.736	70.619
	Mainroad	81.56	-0.034	-0.050	7.669	0.943	-0.046	-0.012	42.460	45.329
	Average	67.53	-0.023	-0.029	1.921	1.060	-0.159	-0.002	56.239	56.060

TABLE V. PERFORMANCE EVALUATION FOR THE PROPOSED SCHEME (K=20)

Sizes	Sequence	ΔPSNR (dB)	ΔBitrate (%)	$\Delta PSNR_R$ (dB)	$\Delta PSNR_N$ (dB)	TS (%)
720×576	Bank	-0.11	-0.29	0.33	-0.13	64.91
	Campus	-0.11	0.00	0.25	-0.12	63.68
	Classover	-0.06	-0.51	0.73	-0.08	76.29
	Crossroad	-0.12	0.00	0.33	-0.13	45.47
	Overbridge	-0.05	-0.12	0.15	-0.06	52.51
	Office	-0.08	0.01	0.12	-0.10	36.44
1600×1200	Intersection	-0.18	-0.85	0.49	-0.26	59.30
	Mainroad	-0.42	-0.61	2.08	-0.70	46.41
	Average	-0.14	-0.30	0.56	-0.19	55.63

an I-frame, and the remaining frames are encoded as P-frames.

As shown in Fig. 5, totally eight SD&HD surveillance videos (each with 100 frames at 30 frames per second) are tested, which is obtained from the PKU-SVD-A dataset [17][18]. The scenes of the videos include dark and bright (DA/BR), fast motion and slow motion (FM/SM), indoor and outdoor (ID/OD), large foreground and small foreground (LF/SF). BDPSNR(dB) and BDBR(%) [19] are utilized as metrics for coding performance. Moreover, we use ΔPSNR, $\Delta PSNR_R$, $\Delta PSNR_N$ to denote the gains in PSNR of whole picture, ROI region and non-ROI region separately. ΔBitrate indicates the percentage increase of bit rate, and TS(%) is used for representing the coding time saving.

To evaluate the performance of our system, we compare it with the fast HM (with CFM, ECU, and ESD fast options enabled), indicated as Ref, which still goes through all CUs without considering the subjective quality. As shown in Table IV, the proposed and Ref methods can save almost the same encoding time, both by around 56%, with a marginal BDPSNR loss compared with the original HM. Different from the Ref, the proposed method saves encoding time only on the quality of the ROI region. As a result, we could notice that, for every sequence, $PSNR_R$ of the proposed method is higher than that of the Ref under the similar coding time. Meanwhile, the termination only determined by coding information may fail when coding a sequence with a combination of high large PU size and small CU size (e.g., Mainroad). Figs. 6 and 7 also show the R-D performances of all methods.

In the further testing, we open the rate control option, and set the target bitrates as we got from the different QP settings. The detailed coding results with comparison between the proposed scheme and HM are shown in Table V. We notice that the overall quality of the ROI region is improved with an average gain of 0.56 dB in terms of PSNR. On the other hand, the overall PSNR and $PSNR_N$ slightly decrease. Fig. 8 shows the R-D performance of the proposed method and the HM, in which both PSNR and $PSNR_R$ are marked. Compared with the HM, the gaps between our method and HM are modest at low bitrates. This is because the bits are automatically allocated to the high-complexity regions while the left CUs are coded with Skip mode. Nevertheless, the gains of our method in terms of $PSNR_R$ are increased as the bitrate increases.

We have also evaluated the subjective visual quality. Obviously, more details in the ROI region (e.g., licence plate number) can be well preserved with our method, as shown in Fig. 9. It should be noticed that the quality gain is obtained with the encoding even much faster than the HM, as indicated in Table V.

VI. CONCLUSION

In this paper, we have proposed a fast ROI-based HEVC coding system for the surveillance videos, targeting at reducing both the bits and the computational costs on the non-ROI regions. We have adopted a classic segmentation scheme to generate the mask. Based on the statistics, we have proposed a two-step method to early estimate the prediction

modes and depths with the low possibility, so that the complicated selection process could be partially saved. Different from the existing early termination methods, the proposed one has fully taken the characteristic and requirements of a surveillance system into account. We have also presented the experimental results in terms of time saving, R-D performance, and visual quality, which show that the proposed method can significantly reduce the encoding time while improving the quality of ROI region at a given bitrate.

ACKNOWLEDGMENT

This work was supported in part by National Natural Science Foundation of China(61472389).

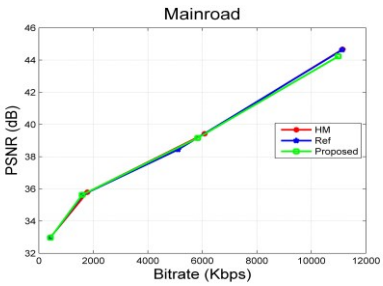

Fig. 6. R-D performance of "Mainroad" under different QP settings .

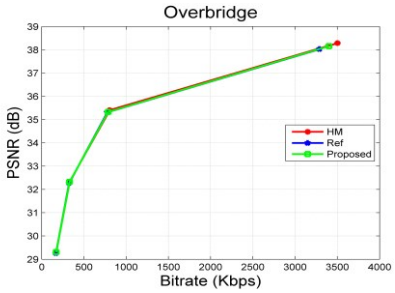

Fig. 7. R-D performance of "Overbridge" under different QP settings .

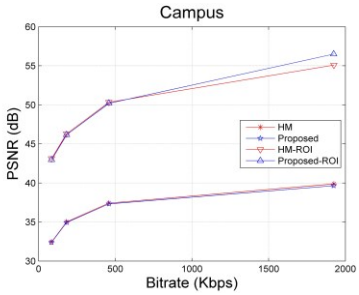

Fig. 8. R-D performance of "Campus" under rate control .

REFERENCES

[1] ITU-T,Recommendation H.265(04/15), Series H: Audiovisual and Multimedia Systems, Infrastructure of audiovisual services- Coding of Moving Video Coding, Online: http://www.itu.int/rec/T-REC-H.265-201504-I.

Fig. 9. Visual quality comparison – the 52th encoded frame of "Mainroad", 425 kbps (*left:HM right:Proposed*).

[2] Sullivan G J, Ohm J R, Han W J, et al. Overview of the high efficiency video coding (HEVC) standard[J]. IEEE Transactions on circuits and systems for video technology, 2012, 22(12): 1649-1668.

[3] J. Yang, J. Kim, K. Won, H. Lee, and B. Jeon, "Early SKIP Detection for HEVC, document JCTVC-G543," *JCT-VC*, 2011.

[4] R. H. Gweon and Y.-L. Lee, "Early Termination of CU Encoding to Reduce HEVC Complexity, document JCTVC-F045," *JCT-VC*, 2011.

[5] Leng, Jie, et al, "Content based hierarchical fast coding unit decision algorithm for HEVC," in *Multimedia and Signal Processing (CMSP)*, 2011, p. 56-59.

[6] K. Choi, S.-H. Park, and E. S. Jang," Coding Tree Pruning Based CU Early Termination, document JCTVC-F092," JCT-VC, 2011.

[7] Shen, Xiaolin, Lu Yu, and Jie Chen, "Fast coding unit size selection for HEVC based on Bayesian decision rule," in *Picture Coding Symposium (PCS)*, 2012, p. 453-456.

[8] Deng, Xin, et al. "Complexity control of HEVC based on region-of-interest attention model," in *Visual Communications and Image Processing Conference*, 2014, p. 225-228.

[9] Deng X, Xu M, Wang Z, "A ROI-based bit allocation scheme for HEVC towards perceptual conversational video coding," in *Advanced Computational Intelligence (ICACI)*, 2013, p. 206-211.

[10] Zhang Xianguo, et al, "Optimizing the hierarchical prediction and coding in HEVC for surveillance and conference videos with background modeling," *IEEE transactions on image processing*, vol. 23,pp. 4511-4526, Oct. 2014.

[11] Meddeb, Marwa, Marco Cagnazzo, and Béatrice Pesquet-Popescu, "Region-of-interest-based rate control scheme for high-efficiency video coding," APSIPA Transactions on Signal and Information Processing., vol. 3,pp. 16-34, Dec. 2014.

[12] Stauffer, Chris, and W. Eric L. Grimson, "Adaptive background mixture models for real-time tracking," in *Computer Vision and Pattern Recognition*, 1999, vol. 2, p. 246–252.

[13] Haritaoglu, Ismail, David Harwood, and Larry S. Davis, "W 4: Real-time surveillance of people and their activities." *Pattern Analysis and Machine Intelligence.*, vol. 22, pp. 809-830, Aug. 2000.

[14] Manzanera, Antoine, and Julien C. Richefeu, "A new motion detection algorithm based on Σ–Δ background estimation," *Pattern Recognition Letters.*, vol. 28, pp. 320-328, Feb. 2007.

[15] Barnich, Olivier, and Marc Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *Image Processing.*, vol. 20, pp. 1709-1724, Dec. 2011.

[16] Bossen, Frank, "Common test conditions and software reference configurations, document JCTVC- H1100," *JCT-VC*, 2012.

[17] X. Zhang, T. Huang, Y. Tian, W. Gao, "Background Modeling Based Adaptive Prediction for Surveillance Video Coding", *Image Processing*, vol. 23, pp. 769-784, Nov. 2014.

[18] W. Gao, Y. Tian, T. Huang, S. Ma, X. Zhang, "IEEE 1857 Standard Empowering Smart Video Surveillance Systems", *IEEE Intelligent Systems.*, vol. 29, pp. 30-39, Sep. 2013.

[19] Bjontegaard, Gisle, "Calcuation of average PSNR differences between RD-curves, document VCEG-M33," ITU-T Q6/16, Apr. 2001.