A BACKGROUND PROPORTION ADAPTIVE LAGRANGE MULTIPLIER SELECTION METHOD FOR SURVEILLANCE VIDEO ON HEVC

Long Zhao, Xianguo Zhang, Yonghong Tian, Ronggang Wang, Tiejun Huang

National Engineering Laboratory for Video Technology, Peking University, Beijing 100871, China
{yhtian, tjhuang}@pku.edu.cn

ABSTRACT

In the recent video coding standards, the selection of Lagrange multiplier is crucial to achieve trade-off between the choices of low-distortion and low-bitrate prediction modes. For surveillance video coding, the rate-distortion analysis shows that, a larger Lagrange multiplier should be used if the background in a coding unit took a larger proportion. Therefore, a modified Lagrange multiplier might be better for rate-distortion optimization. To address this problem, we perform an in-depth analysis on the relationship between the optimal Lagrange multiplier and the background proportion, and then propose a Lagrange multiplier selection model to obtain the optimal coding performance for surveillance videos. Following this, we further develop a Lagrange multiplier optimized video coding method. Experimental results show that our coding method can averagely achieve 18.07% bitrate saving on CIF sequences and 11.88% on SD sequences against the background-irrelevant Lagrange multiplier selection method.

Index Terms— Lagrange multiplier selection, surveillance video, background proportion, Lambda Factor, HEVC

1. INTRODUCTION

In current hybrid video coding standards, multifarious coding modes are adopted to achieve high coding efficiency, including kinds of intra prediction modes with different prediction directions and the various-block-pattern inter prediction modes. The choice of what modes to adopt for the current coding unit is determined through a Lagrange multiplier based mode decision process. The Lagrange multiplier used in the mode decision process balances between the choices of low-distortion modes and low-bitrate modes. Therefore, the optimization of Lagrange multiplier selection is an important research topic and it is very significant for the improvement of video coding efficiency.

To investigate an optimal Lagrange multiplier, some rules and methods have been proposed in some famous pioneer works. In [1] and [2], Thomas Wiegand et al. derived the relationship between the Lagrange multiplier, the distortion and the rate in the following equation,

$$\lambda = -\frac{dD}{dR} \tag{1}$$

where λ represented the Lagrange multiplier, D denoted the distortion and R was the coding rate. They finally derived the computation method of the Lagrange multiplier in the following equation,

$$\lambda = \mathrm{c}Q^2 \tag{2}$$

where Q represented the quant value and c represented a constant determined by their experimental results.

In the recent coding standard HEVC [3] which has achieved 50% bit-saving than H.264/AVC, the largest coding unit (LCU) is in the size of 64×64. The LCU can be further divided into coding units (CUs) from 32×32 to 8×8 by a quad tree partition. Correspondingly, a large number of inter prediction patterns are introduced for the prediction unit (PU). Besides, the number of intra prediction modes is also increased to 35. As is listed, there are much more coding modes to be chosen from the Lagrange multiplier based mode decision process. As a result, a forward step of Lagrange-multiplier selection is employed in HEVC. In the reference software HM8.0 [4], the computation method of the Lagrange multiplier is in the following equation,

$$\lambda = \alpha \times W_k \times 2^{((QP-12)/3.0)} \tag{3}$$

where α is a factor dependent on pictures according to whether they are referenced, and W_k represents weighting factor dependent on encoding configuration and QP offset.

Besides the recent optimization in HEVC, J. Zhang et al. [5] proposed a method of selecting Lagrange multiplier based on the context of the video, taking motion vector of the scene into consideration. With different Lagrange multipliers in different coding layers, a Lagrange multiplier selection method for scalable video coding was proposed in [6]. P. Sangi et al. [7] proposed a Lagrange multiplier selection method for block-based motion estimation criteria.

However, these methods above were not specially designed for surveillance video, which usually has its own properties. The surveillance cameras are always deployed on a fixed position, thus a large proportion of background region exists. Intuitively, these background regions can be predicted with low distortion by each existing mode.

Therefore, it is not difficult to suspect that the low-bitrate coding modes should be more appreciate for these regions since they can achieve large bit-saving without large distortion increase. In further, as the Lagrange multiplier trades off between the choices of low-distortion modes and low-bitrate modes, a modified Lagrange multiplier might be better for rate-distortion optimization. In summary of our conjecture above, a better Lagrange multiplier selection method should be exploited to improve the coding efficiency of surveillance video.

To validate the conjecture for the Lagrange multiplier modification, this paper firstly analyzes the rate-distortion results which are performed on frames and LCUs with different background proportions. Results show that, a larger Lagrange multiplier should be chosen for frames and LCUs with larger background proportions. To build up the Lagrange multiplier selection model for each LCU, we further conduct experiments to figure out the optimal Lagrange multipliers for some fixed background proportions Based on the results, a mapping from the background proportion of LCU to Lagrange multiplier is built up.

Moreover, a Lagrange Multiplier Optimized (LMO) encoder is proposed to achieve better coding efficiency for surveillance video. The LMO encodes each LCU by the following steps: It firstly trains the parameters of the Lagrange multiplier model using a fixed number of frames; Then it calculates the background proportion for the input LCU according to its relationship with the reference frames and classifies the LCU into a category according to the background proportion; Thirdly, an optimal Lagrange multiplier is selected for this LCU category utilizing the Lagrange multiplier selection model; Afterwards, the LCU is encoded with kinds of exiting intra-and-inter prediction modes in various CU partitions, when the rates and distortions are also obtained; Finally, with each mode's coding rate and distortion, the selected optimal Lagrange multiplier is employed to select the best prediction mode in the best CU partition and the corresponding code stream.

With the proposed LMO encoder, we conduct experiments on the HEVC reference software HM8.0 to evaluate its performance. The test surveillance sequences are in the resolution from CIF to SD. Experiment results show that our method can averagely achieve 18.07% bitrate saving on CIF sequences and 11.88% bitrate saving on SD sequences.

The rest of this paper is organized as follows. Sec.2 analyzes the reason for adopting a larger Lagrange multiplier for background regions. Sec.3 derives the proposed Lagrange multiplier selection model. Sec.4 describes our LMO encoder. Experiments and conclusion are given respectively in Sec.5 and Sec.6.

2. PROBLEM ANALYSIS

A conjecture has been made in Sec. 1 about surveillance video coding: low-bitrate coding modes should be selected more to achieve large bit-saving and a modified Lagrange multiplier might be better for rate-distortion optimization. To validate this conjecture, this section gives a detailed experimental analysis on the optimal Lagrange multiplier, which will be derived from the rate-distortion result of coding surveillance video. As referred in Eq.1, an approximation of the computation method of Lagrange multiplier can be made as $-dD/dR$. This shows the Lagrange multiplier represents the relationship between the distortion change and its corresponding coding rate change at a fixed QP point. Consequently, $-dD/dR$ and the value of Lagrange multiplier can be approximately calculated from the $-\Delta D/\Delta R$ at each two adjacent points in the MSE(mean square error)-bitrate curve. Therefore, we can find tendency of appreciating the Lagrange multiplier value for each sequence, frame or LCU from its coding MSE-bitrate curve.

With such prior knowledge, to verify the specialized Lagrange multiplier selection property of surveillance video, as in the referred conjecture, experiments for the statistics of surveillance and non-surveillance video coding results are conducted. Further on, to verify the Lagrange Multiplier selection property of different background proportions of surveillance video, experiments for the statistics of different proportions of surveillance video coding results are conducted on LCU level.

2.1. Lagrange multiplier for surveillance and normal videos

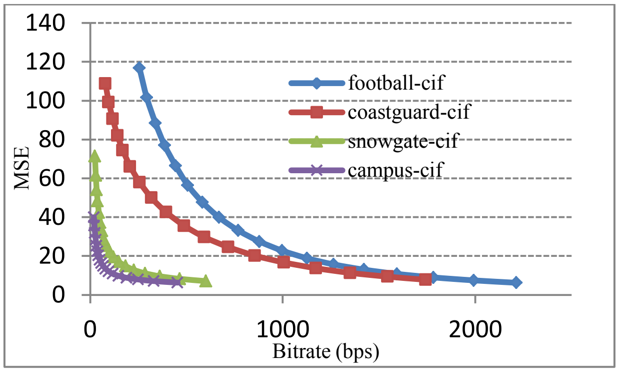
Surveillance videos of snowgate-cif and campus-cif and normal videos of football-cif and coastguard-cif are selected for the experiment. The snowgate-cif and campus-cif have a large region of background and the camera is static. On the contrary, the football-cif and coastguard-cif own a large region of fast moving objects and the camera also moves. The MSE-bitrate curves of these two sequences are shown in Fig.1 (a). From the RD-curves, we can see that the coding distortions of snowgate-cif and campus-cif at the same bitrate are better than those of football-cif and coastguard-cif. This is mainly because the snowgate-cif and campus-cif have larger static background regions and each prediction mode in these regions is more accurate to decrease the distortion for these regions. This will lead a frequent usage for large bit-saving modes.

As is described, $-\Delta D/\Delta R$ at each two QP points can approximate the tendency of Lagrange multiplier. Let any two adjacent points of (d_1, r_1, QP_1) and (d_2, r_2, QP_2) in the curve in Fig.1(a) denote the MSE and bitrate for QP_1 and QP_2, we can calculate the Lagrange multiplier value λ at QP_1 approximately in the following equation

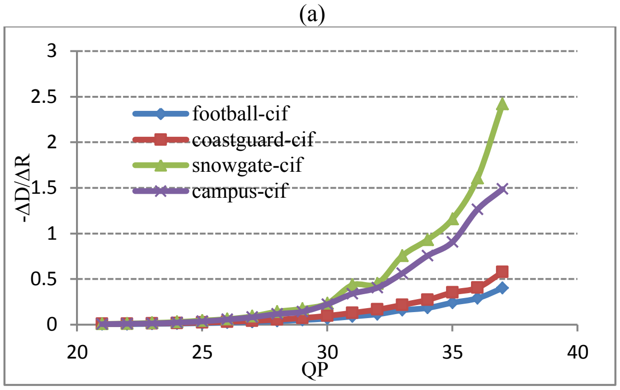
$$\lambda(QP_1) \approx -\Delta D/\Delta R = -(d_1 - d_2)/(r_1 - r_2) \qquad (4)$$

Following this for each QP, the Lagrange multiplier tendency curve is shown in Fig.1 (b). From the curve we can see, this exponent tendency between $-\Delta D/\Delta R$ and QP well copes with the relationship of the Lagrange multiplier and the QP value in HEVC reference software. Furthermore,

$-\Delta D/\Delta R$ of surveillance video at each QP point is larger than that of normal video. This reflects that a larger Lagrange multiplier should be selected for surveillance video than for normal video.



(a)



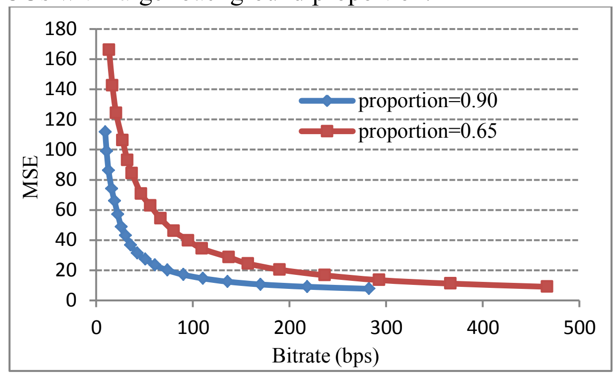
(b)

Fig. 1. Performance on frame level. (a) MSE-bitrate curve. (b) Lagrange multiplier tendency curve

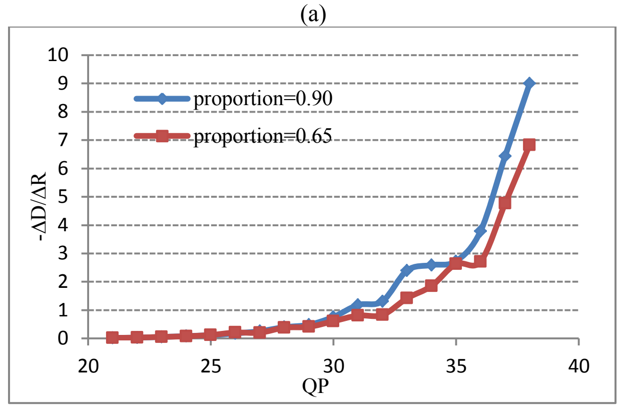
2.2. Lagrange multiplier for different background proportions

To analyze the relationship between Lagrange multiplier and background proportion, each LCU is firstly divided into a background proportion bin according to a background proportion calculation and classification method to be described in the Sec.4. When a LCU is classified into a background proportion bin, the MSE and coding bits of this LCU are also classified into that bin. In this way, we can get the MSE-bitrate curves of each background proportion bin of LCU. These MSE-bitrate curves represent the coding performance in the LCU level. The curves of the LCUs in the background proportion bin of 0.90 and 0.65 are respectively drawn in Fig.2 (a).

As stated in Sec.2.1, we can also see from the MSE-bitrate curves that the coding efficiency is better when the background proportion of LCU is larger. This is also because the prediction mode is more accurate in the background regions. The Lagrange multiplier tendency curves approximated by $-\Delta D/\Delta R$ at each two adjacent QP points are drawn in Fig.2 (b). From the curves we can see, $-\Delta D/\Delta R$ of the LCUs with larger background proportion is larger than that of the LCUs with smaller background proportion at each QP point. As a result, in surveillance video, a larger Lagrange multiplier should be adopted for LCUs with larger background proportion.



(a)



(b)

Fig. 2. Performance on LCU level. (a) MSE-bitrate curve. (b) Lagrange multiplier tendency curve.

From the analysis above, our conjecture is proved that the background region prefers low-bitrate modes to achieve large bit-saving and a modified Lagrange multiplier should be selected for surveillance video. In further, two conclusions can be drawn as follows: (i) a larger Lagrange multiplier should be selected for surveillance video; (ii) in surveillance video, a larger Lagrange multiplier should be selected for LCUs with larger background proportions.

3. THE MULTIPLIER SELECTION MODEL

In Sec.2, we have concluded that a LCU with a large background proportion has a large Lagrange multiplier at each QP point. To figure out the optimal Lagrange multiplier for each background proportion, we build up an optimized model of the computation of Lagrange multiplier for HEVC. As the Lagrange multiplier increases when the background proportion increases, a Lambda Factor (LF) dependent on the background proportion (p_b) is multiplied to the right side of Eq.3. Thereby the Lagrange multiplier is computed in the following Equation,

$$\lambda = f(p_b) \times \alpha \times W_k \times 2^{((QP-12)/3.0)} \qquad (5)$$

where p_b is the background proportion of the LCU, and $f(p_b)$ is the value of Lambda Factor. The Lambda Factor is a function of the background proportion of the LCU. In the following work, we train from different background's optimal Lagrange multipliers to get optimal Lambda Factor for LCUs with different background proportions

3.1. Experiments for training the model

As referred, background proportion of an LCU is a key factor to determine the value of Lagrange multiplier for this LCU in surveillance video. To find the optimal Lagrange multiplier for each background proportion, we have done a lot of experiments to build up the relationship between the Lambda Factor and the background proportion of LCU.

In our experiments, the optimal Lambda Factor for each background proportion of LCU is firstly found through the following testing procedure. 1) The Lambda Factor which ranges from 0.4 to 4.0 is firstly set as the input independent variable with an increasing step of 0.1; 2) for each Lambda Factor value, the BD-rate between its rate-distortion result and that of $LF = 0.4$ is set as the output; 3) The Lambda Factor with the smallest output value is the optimal Lambda Factor for the fixed background proportion.

For surveillance video snowgate-cif, the curves which show the relationship between BD-rate and the Lambda Factor at the background proportion of 0.80 and 0.65 are drawn in Fig.3. From this figure, we have following statements for the background proportion bin value of 0.65: when the Lambda Factor is smaller than 2.5, the curve is in the decreasing trend and when it is larger than 2.5, the curve is in the increasing trend. Thereby we get the valley value 2.5 as Lambda Factor value for the background proportion bin value of 0.65. Similarly for the background proportion bin value of 0.80, we get the optimal Lambda Factor of 3.5.

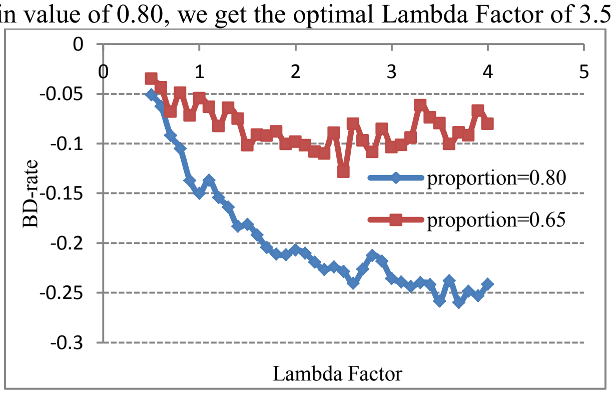


Fig. 3. Performance curves for different Lambda Factors.

Iteratively carrying out such experiment for each background proportion, the optimal Lambda Factor for each background proportion bin value of LCUs can be found. With such experiment methods, we can get the optimal Lambda Factor curve for different sequences. Fig.4 shows the relationship between the optimal Lambda Factor and the background proportion for sequences. Actually, when the background proportion is smaller than 0.6, the foreground is the dominant region in the LCU. The motion is large in this foreground dominant region and these regions do not obtain the background region's properties. Therefore, we set the Lambda Factor equal to 1.0 for these regions. For the background proportion ranging from 0.6 to 1.0, we utilize a step length of 0.05 for accurate modeling.

3.2. The built-up model

Through the experiment results as shown in Fig. 4, we find that the relationship between Lambda Factor and the background proportion is in the increasing tendency which can be described by a cubic function. So a function mapping from background proportion to Lambda Factor is built up by matching the optimal Lambda Factor curves. Through comparing different Pairs of (LF, p_b) in different training sequences, the function is built up as follows:

$$f(p_b) = \begin{cases} \alpha p_b{}^3 + \beta p_b{}^2 + \gamma p_b + \varepsilon, & p_b \geq 0.6 \\ \quad\quad 1.0 \quad\quad , & p_b < 0.6 \end{cases} \tag{6}$$

Further on, the Lagrange multiplier can be computed by substituting (6) into (5).

Different sequences have different parameters. Take sequence snowgate-cif for example, α equals to -84.85, β equals to 191.56, γ equals to 136.59 and ε equals to33.62. In practical, this is very reasonable because the coding relationship between different LCU categories affects the detailed value of the parameters. Nevertheless, the tendency is surely as Eq. 6 shows. To realize the best performance for coding each input sequence, we should employ a trainnig module in our encoder to obtain the applicable prameters.

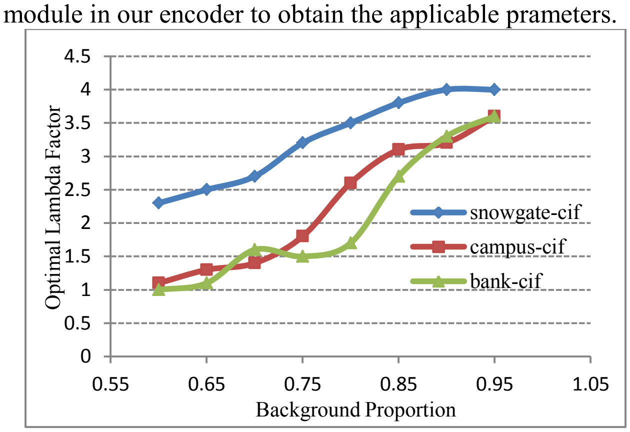


Fig. 4. The function mapping from background proportion to optimal Lambda Factor.

4. THE LAGRANGE MULTIPLIER OPTIMIZED ENCODER

4.1. Framework

Based on the Lagrange multiplier selection model derived in Sec.3, we introduce the LMO encoder in this section. The framework of the LMO encoder is shown in Fig. 5. The surveillance video coding method consists of the Lambda-Factor Parameter Training, the Background Proportion Classification, the Lagrange Multiplier Selection, the Multi-Modes Encoding and the Lagrange Multiplier Based Mode Decision modules.

The LMO encoder works as following steps.

1) For each sequence, the first *n* frames are utilized to train the Lambda Factor parameters of the referred α, β, γ and ε by the Lambda-Factor Parameter Training module.
2) For each LCU after the initial n training frames, the Background Proportion Classification module utilizes the existent four reference frames to calculate the background proportion and classifies the LCU into a background proportion bin.
3) With the LCU's background proportion bin value and the trained Lambda Factor parameters, the Lagrange Multiplier Selection module computes the value of optimal Lambda Factor for the LCU.
4) Meantime, the Multi-Modes Encoding module encodes the LCU with kinds of exiting intra-and-inter prediction modes in various CU partitions, when the corresponding rates and distortions are also obtained.
5) With the rates and distortions, the Lagrange Multiplier Based Mode Decision module finally selects the best prediction mode in the best CU partition.
6) Finally, the corresponding output bitstream of the best prediction mode is outputted.

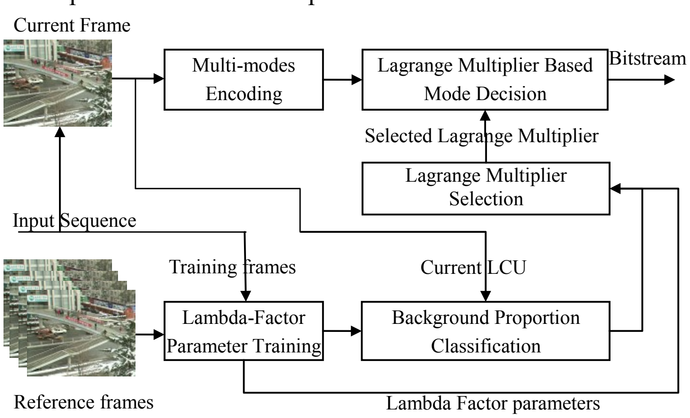


Fig. 5. The framework of the proposed method.

4.2 Lambda Factor Parameter Training

In the Lambda-Factor parameter training module, parameters of the referred α, β, γ and ε in Eq.6 are trained by encoding the first 10 frames repeatedly adopting different Lambda Factors in the range from 0.4 to 4.0. To save the training complexity, we practically utilize 0.3 as the Lambda Factor increasing step. Each LCU and its corresponding distortion and coding rate in the training frames is classified into a background proportion bin. Then the BD-rates of the background proportion bin encoded under the candidate Lambda Factors are calculated. By comparing the BD-rates, an optimal Lambda Factor for that background proportion bin can be figured out. With four pairs (LF, p_b), the four parameters in Eq.6 are solved. The training details are very similar to that referred in Sec.3.1. The difference is that all the comparing and calculating processes are programmed.

4.3. Background proportion classification module

In Background Proportion Classification module, each 4×4 sub-block's property $S(R)$ is firstly calculated to be a similar block *SB* or a different block *DB* for a reference frame *R* in the following criteria,

$$S(R) = \begin{cases} DB \quad if \quad \sum_{i=0}^{15}(L_i - LR_i) \geq Th \\ SB \quad if \quad \sum_{i=0}^{15}(L_i - LR_i) < Th \end{cases} \tag{7}$$

where L_i and LR_i respectively represents the i-th pixel of the sub-block of the current LCU L and the co-located LCU LR in the reference frame R, Th represents the classification threshold, which is practically set 160.

Secondly, denoting the i-th sub-block's property as $S_i(R)$, number of similar 4×4 sub-blocks for R is

$$bgNum(R) = \|\{i|S_i(R) = SB\}\| \tag{8}$$

and the background proportion $p(R)$ for the current reference frame is calculated by

$$p(R) = bgNum(R)/256 \tag{9}$$

Afterwards, the algorithm repeats the two steps above on four reference frames $R1$~$R4$ and finally chooses the minimum proportion p_{min} as the background proportion of this LCU by

$$p_{min} = \underbrace{p(R1), \ldots \ldots, p(R4)}_{min} \tag{10}$$

Finally, the LCU is classified into the background bin in the following criteria,

$$bin = \lfloor p_{min}/0.05 \rfloor \times 0.05 \tag{11}$$

4.4. Other modules

For each LCU, the Lagrange Selection module firstly computes the Lambda Factor by Eq.6 with the parameters and the background proportion of this LCU. And then, it computes the Lagrange multiplier by the algorithm in Eq. 5.

In the multi-modes encoding module, the current LCU is encoded with kinds of exiting intra-and-inter prediction modes in various CU partitions. Besides, the corresponding distortions, rates and bitstreams are also obtained. With the selected Lagrange multiplier, the best mode and the corresponding bitstream are selected.

In this section, an LMO encoder is proposed based on the Lagrange multiplier selection model and the algorithms for modules in this encoder are also described. The LMO encoder can be adopted to improve the coding efficiency for surveillance video.

5. EXPERIMENT RESULTS

5.1. Experiment setup

To verify our proposed method, the original Lagrange multiplier selection method as described in Eq.3 in HM8.0 is chosen as the anchor. The common testing parameters of our experimental platform HM8.0 are listed in Table 1. As usual, BD-rate and BD-PSNR of our proposed method compared with the anchor are chosen as the performance measure criteria. In our experiment, eight CIF and SD surveillance sequences from AVS workshop [8], which are shown in Fig.6, are adopted.

Table 1. The experimental configuration

Parameter	**Value**	**Parameter**	**Value**
Profile	Main	Framerate	25
Rate Control	Disable	Frame Structure	IBBB
Search Range	64	IntraPeriod	-1
RDOQ	0	SAO	0

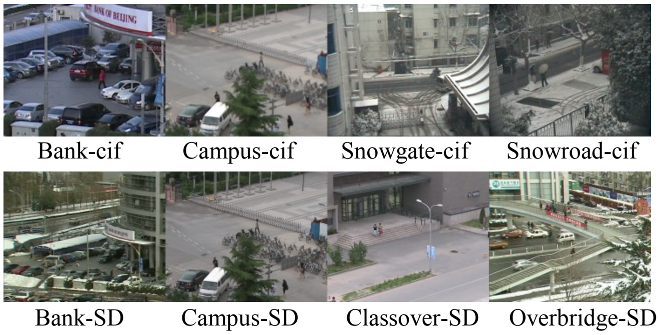


Fig. 6. Test sequences examples.

5.2. Experiment results

The experiment results are shown in Table.2 and the RD-curves of snowroad-CIF and classover-SD are shown in Fig.7. Compared with the original Lagrange multiplier selection method in HM8.0, our background proportion adaptive Lagrange multiplier method for surveillance video can get an average PSNR gain of 0.666 dB on CIF and 0.315 dB on SD, with an average bitrate saving of 18.07% on CIF and 11.88% on SD.

Table 2. Performance of the proposed method

Sequence	**BD-rate**	**BD-PSNR(ΔdB)**
CIF(352x288)		
Bank	-12.87%	0.106
Campus	-10.39%	0.290
Snowgate	-25.00%	0.853
Snowroad	-24.02%	0.895
Average	**-18.07%**	**0.666**
SD(720x576)		
Bank	-12.25%	0.347
Campus	-12.31%	0.308
Classover	-14.87%	0.333
Overbridge	-8.08%	0.271
Average	**-11.88%**	**0.315**

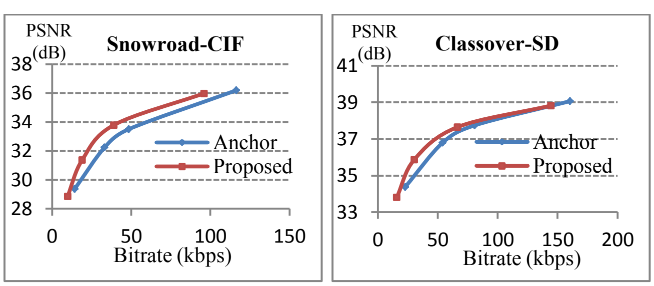


Fig. 7. The RD-curves of Snowroad-CIF and Classover-SD

6. CONCLUSION

In this paper, we propose a background proportion adaptive Lagrange multiplier selection method on HEVC. From an analysis on the relationship between the Lagrange multiplier and the background proportion, a Lagrange multiplier selection model for surveillance video is proposed. Based on this model, we propose an LMO encoder for surveillance video. In the proposed method, we get an average PSNR gain of 0.666dB on CIF and 0.315dB on SD, with an average bitrate saving of 18.07% on CIF and 11.88% on SD meanwhile. In the future, we will concentrate on the Lambda-Factor training algorithm to obtain better parameters for the Lagrange multiplier selection model in lower complexity.

7. ACKNOWLEDGEMENT

This work is partially supported by grants from National Basic Research Program of China under contract No. 2009CB320906, the Chinese National Natural Science Foundation under contract No. 61035001, 61121002 and 61176139.

8. REFERENCES

[1] Gary J. Sullivan and Thomas Wiegand, “Rate-Distortion Optimization for video compression,” Signal Processing Magazine, IEEE, vol. 15, issue 6, pp. 74-90, Nov. 1998.

[2] Thomas Wiegand and Bernd Girod, “Lagrange multiplier selection in hybrid video coder control,” Image Processing, vol. 3, pp. 542-545, 2001.

[3] Gary J. Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand, “Overview of the high efficiency video coding (HEVC) standard,” Circuits and Systems for Video Technology, issue 99, Dec. 2012.

[4] Il-Koo Kim, “HM7: High efficiency video coding (HEVC) test model 8 encoder description,” document JCTVC-J1002.doc, JCT-VC, 10th Meeting, Stockholm, SE, July 2012.

[5] Jun Zhang, Xiaoquan Yi, Nam Ling and Weijia Shang, “Context Adaptive Lagrange Multiplier (CALM) for rate-distortion optimal motion estimation in video coding,” IEEE Transactions on Circuits and Systems for Video Technology, 20(6), pp. 820-828, June. 2010.

[6] Xiang Li, Peter Amon, Andreas Hutter and Andre Kaup, “Lagrange multiplier selection for rate-distortion optimization in SVC,” Picture Coding Symposium, 2009 (PCS 2009), pp. 1-4, 2009.

[7] Pekka Sangi, Janne Heikkila and Olli Silven, “Selection of the Lagrange multiplier for block-based motion estimation criteria,” IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004 (ICASSP `04), vol. 3, pp. iii-325-8, May. 2004.

[8] ftp://124.207.250.92/public/seqs/video.