



HAL
open science

Analytical Distortion Aware Video Coding for Computer Based Video Analysis

Yuyang Liu, Ce Zhu, Min Mao, Fangliang Song, Frederic Dufaux, Xiang Zhang

► **To cite this version:**

Yuyang Liu, Ce Zhu, Min Mao, Fangliang Song, Frederic Dufaux, et al.. Analytical Distortion Aware Video Coding for Computer Based Video Analysis. 19th International Workshop on Multimedia Signal Processing (MMSP), Oct 2017, London-Luton, United Kingdom. 10.1109/mmsp.2017.8122253 . hal-01567381

HAL Id: hal-01567381

<https://hal.science/hal-01567381>

Submitted on 10 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analytical Distortion Aware Video Coding for Computer Based Video Analysis

Yuyang Liu¹, Ce Zhu¹, Min Mao¹, Fangliang Song¹, Frederic Dufaux², Xiang Zhang¹

¹University of Electronic Science and Technology of China, Chengdu, China

²Laboratoire des Signaux et Systèmes, Paris, France

{yuyangliu, minmao, songfangliang}@std.uestc.edu.cn, {eczhu, uestchero}@uestc.edu.cn, frederic.dufaux@12s.centralesupelec.fr

Abstract—With the development of artificial intelligence, more and more multimedia applications for various tasks have emerged in our daily life. Meanwhile, as one of the main information sources of the applications, a huge amount of video data has been being generated by portable or mounted cameras in daily basis for varying purposes including surveillance, in which case we may need computers to “watch” videos to save labor cost. However, most video coding standards are designed for the highest human perceptual quality given a bit rate by minimizing a fidelity cost function (e.g., mean squared error, MSE), assuming the content will be consumed by human beings. In view of the above considerations, this paper proposes a new rate-analytical-distortion optimization method (RADO) for video analysis. Specifically, we consider moving object detection as the analysis task. Accordingly, we develop a novel rate analytical distortion (RAD) model for video coding, where the analytical distortion is related to the object detection performance expressed in terms of F-measure. As shown in the experimental results, the performance of the video analysis task can be significantly improved (up to 40% reduction of analytical distortion) with a slight bit rate increase.

Keywords—Video coding, video analysis, rate-analytical-distortion optimization, HEVC.

I. INTRODUCTION

Recent advances in computer vision and the increasing availability of portable or installed cameras have enabled new emerging multimedia applications. In many application scenarios, especially for surveillance, computers have to deal with various analysis tasks, such as object detection, tracking and recognition. Consequently, computer is the first or maybe the only “audience” in these scenarios. At the same time, the considerable amount of generated videos also need to be compressed due to cost-effective storage and bandwidth limitation.

Video coding has been studied for decades, mainly concentrating on issues related to high coding efficiency. Most of the widely deployed video coding standards, such as H.264/AVC and HEVC (High Efficiency Video Coding) are designed for human visual properties by maximizing a

fidelity function (e.g. peak signal to noise ratio, PSNR) at a given bit rate, with human beings as the target audience. Besides, there are many developed rate-distortion optimization (RDO) schemes for HEVC standard [1]-[4]. However, applying the traditional RDO scheme during video compression may be suboptimal when the video is intended for machine analysis. The critical issue is that compression distortions may bring a negative impact on the video analysis performance.

In this context, recently research work focuses on feature-preserving coding, where only feature descriptors are transmitted to the server [5]. As only feature descriptors are transmitted, the video cannot be watched at the server side. Consequently, it is inapplicable in some scenarios (e.g. video surveillance) where it is necessary to visualize the video content. In contrast, some research work focuses on feature-preserving video coding. Baroffio *et al* [6] proposed a coding architecture designed for local features (e.g. SIFT, SURF) extracted from video sequences. The local feature descriptors were encoded by intra and inter-frame coding modes respectively. The final coding mode was determined by comparing the costs of intra-frame versus inter-frame coding. In their later work [7], the coding architecture was applied to code binary local features. In order to decrease the adverse effect of video compression on feature-matching performance, Chao and Steinbach [8] proposed a novel framework, in which keypoints extracted from a video were encoded and transmitted along with the compressed video to the server. However, this framework needs more bits since the keypoints are transmitted as side information.

Few studies have addressed video coding for video analysis tasks. Korshunov *et al.* [9] proposed a formal rate-accuracy optimization framework, where the encoding parameters in distributed video surveillance systems could be determined given a target bit rate or accuracy. Effrosyni and Pascal [10] proposed a supervised dimensionality reduction scheme which provides a tradeoff between compression and discriminant feature extraction. Liao *et al.* [11] proposed an analysis-oriented ROI based coding approach to reduce the effect of video compression on the

performance of video analysis. However, the approach in [11, 12] needs the prior knowledge of interest objects to detect ROI, which cannot always be obtained in practice.

In this paper, we develop a new rate-analytical-distortion optimization (RADO) method for video analysis, where the term “analytical distortion” represents the difference of the video analysis algorithms’ performance when the video quality degrades. Meanwhile, we introduce an analytical distortion prediction (ADP) model which uses the compression distortion to estimate the analytical distortion. Specifically, we choose to focus on one fundamental video analysis task, moving object detection and accordingly develop a novel rate analytical distortion (RAD) model. It is worthy to note that our proposed coding method is fully standard compatible and the encoded bit-stream can be decoded by any HEVC decoder.

The remainder of this paper is organized as follows. Section II presents our proposed RADO method and the corresponding coding framework. Section III presents the ADP model and parameter acquisition. Section IV experimentally evaluates the proposed method. Section V concludes the paper.

II. PROPOSED CODING METHOD FOR VIDEO ANALYSIS

In this section, we introduce our proposed coding method for video analysis. Specifically, we build up an analytical inconsistency model to represent the difference of video analysis performance when the video quality degrades, partly inspired by the work in [9]. We use the term “analytical distortion” hereafter to denote the difference of video analysis performance.

A. Analytical Inconsistency Model

Inspired by the fact that the video analysis algorithm’s performance may decline after a lossy video compression, we design an analytical inconsistency model to represent how the analysis algorithm behaves when the video quality degrades. Clearly, many evaluation criterions exist for different kinds of video analysis tasks. Even for a specific task, several evaluation metrics may be used to assess the performance. However, it is unmanageable to simply adopt all the evaluation metrics in the analytical inconsistency model, since the dimension of the model will be extremely high. In order to avoid this drawback, only one metric is chosen for each kind of video analysis task. Therefore, given a video analysis task T and the corresponding evaluation metric M , the analytical inconsistency model can be defined as $D_A(\Delta D|T, M)$, where ΔD can be obtained as

$$\Delta D = |D^o - D^c| \quad (1)$$

In (1), D^o and D^c denote the analysis results of the original video and the compressed video respectively.

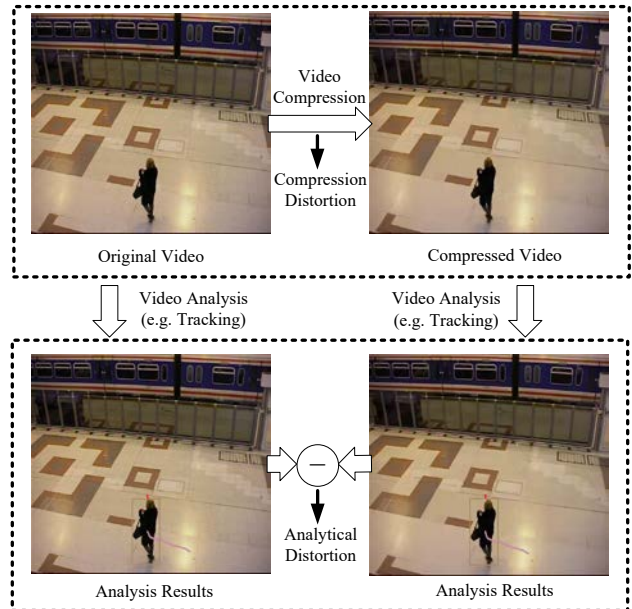


Fig. 1 Illustration for the analytical distortion acquisition procedure.

Fig. 1 illustrates how the analytical distortion can be obtained by (1). In this paper, we do not aim at studying the performance of a video analysis algorithm itself. Rather, we are more specifically concerned with how the analysis algorithm behaves when the video quality is degraded.

B. Rate Analytical Distortion Optimization

Video compression is a well-studied problem and is commonly formulated and optimized by the trade-off between the entropy of the discretized representation (rate) and the error arising from the quantization (distortion). In this context, many RDO techniques [13, 14] have been developed for modern video compression schemes. Typically, the objective of RDO is to minimize distortion D at a given rate R , under the available bit rate budget constraint R_T , which can be formulated as

$$\min \{D\} \quad s.t. \quad R \leq R_T. \quad (2)$$

To solve this constrained optimization problem, the Lagrangian multiplier method [15] is widely employed by converting it to an unconstrained form as

$$\min \{J\}, \quad \text{where } J = D + \lambda R, \quad (3)$$

where J is the Lagrangian cost function, and λ is the Lagrangian multiplier. According to the work in [16], λ in (3) can be calculated by

$$\lambda = -\frac{dD}{dR} = c \cdot q^2, \quad (4)$$

where c is a constant and q denotes the quantization step-size. Notably, distortion D in (2) is designed as a human perceptual related cost function D_p in pixel domain, such as MSE, PSNR or Sum of Absolute Difference (SAD).

Meanwhile, in some typical application scenarios such as face detection and recognition, the compressed video will not only be consumed by computers running video analysis algorithms, but also watched by human beings for some application-related purposes. Therefore, it is necessary to reduce the negative effect introduced by compression while maintaining video coding efficiency. For this purpose, we formulate the RADO problem by jointly minimizing the compression distortion D_p in pixel domain and the analytical distortion D_A at a given rate. It can be written as

$$\min D_p + \alpha D_A \quad s.t. \quad R \leq R_T, \quad (5)$$

where α is a weighting parameter. It is obvious that (5) is reduced to the traditional RDO formulation in (2) when $\alpha = 0$. To solve the constrained RADO problem in (5), the Lagrangian multiplier method is also employed by converting it to an unconstrained form, which can be written as

$$\min \{J_{new}\}, \text{ where } J_{new} = D_p + \alpha D_A + \lambda_{new} R. \quad (6)$$

It is worthy to note that the Lagrangian multiplier λ_{new} in (6) is quite different that the λ defined in (4). In order to distinguish the two different Lagrangian multipliers λ_{new} and λ , the Lagrangian λ in (4) will be denoted as λ_{HM} in the remaining of this paper. Besides, equation (4) is applied in the HEVC test model to calculate the Lagrangian multiplier λ_{HM} .

It is difficult to get the optimal solution in (6) without modeling the interaction between D_A and R . Therefore, we empirically design a derivable RAD model, which can be expressed as

$$R = C_1 e^{C_2 D_A}, \quad (7)$$

where C_1 and C_2 are constant parameters (the derivation of this model will be further discussed in section III). Besides, according to (4), R is also differentiable with respect to D_p .

Consequently, the minimal cost J_{new} is obtained when

$$\frac{\partial J_{new}}{\partial R} = \frac{\partial D_p}{\partial R} + \alpha \frac{\partial D_A}{\partial R} + \lambda_{new} = 0, \quad (8)$$

$$\lambda_{new} = -\frac{\partial D_p}{\partial R} - \alpha \frac{\partial D_A}{\partial R}. \quad (9)$$

Combining (4), equation (9) can be rewritten as

$$\lambda_{new} = \lambda_{HM} - \alpha \frac{\partial D_A}{\partial R}. \quad (10)$$

In conclusion, the new Lagrangian multiplier λ_{new} can be calculated by (10) for a given weighting parameter α .

C. Proposed Coding Framework

The analytical distortion D_A is needed during the coding process in our proposed RADO approach. In general, the accurate D_A can be obtained by two approaches: (i) integrating the analysis algorithm into video codec and (ii) running the analysis algorithm twice on the original video and the compressed version to compare the difference between the analysis results, as shown in Fig. 1. However, both approaches are very time consuming. To overcome this difficulty, we introduce an ADP model. Namely, the proposed coding framework consists of two steps: a training procedure and a coding procedure.

- **Training procedure:** The purpose of the training procedure is to obtain the parameters of the RAD model and the ADP model. During the training procedure, a set of reference video sequences are selected and compressed by HEVC codec with different quantization parameter (QP) values. Next, we run the analysis algorithm on the reference video and the compressed versions. In this way, both the compression distortion D_p (i.e. SAD) and the analytical distortion D_A can be calculated. Finally, from the statistics, the ADP model can be constructed and the parameters of the RAD model can be estimated.
- **Coding procedure:** In the coding procedure, each frame is encoded twice. Firstly, the frame is compressed using the traditional RDO method and then SAD can be calculated by comparison between the original and reconstructed frames. Secondly, given a video analysis task and the corresponding evaluation distortion D_A can be predicted from SAD using the designed ADP model. Furthermore, according to the RAD model in (7), we can express the derivation of D_A with respect to R . Then the new Lagrangian multiplier λ_{new} can be calculated by (10) for a given α . Finally, the same frame is compressed a second time using our proposed RADO method.

Note that, each frame needs to be compressed twice in our proposed framework, which will definitely increase the complexity of the encoder. However, if necessary, approached designed for estimating SAD before encoding [4] can be used to avoid the two pass encoding process.

III. VIDEO CODING FOR MOVING OBJECT DETECTION

To show the effectiveness of the proposed coding scheme, we consider a task of moving object detection. Indeed, moving object detection is a fundamental component in many application scenarios. For this purpose, we select the moving object detection algorithm in [17], and use F-measure [18] as the evaluation metric.

A. Analytical Distortion Prediction

In order to build the ADP model, we run some experiments on four reference video sequences (Video 1-4), including three indoor videos (Video 1-3) from PETS2006 [19] and one outdoor video (Video 4) from CAVIAR [20]. The four reference video sequences are compressed using the HEVC reference software (HM 16.7) [21] under Low-Delay P (LDP) configuration with QPs from 4 to 41. Thus, for each of the 4 original sequences, we get 38 compressed versions with different quality levels. The analytical distortion of one frame can be obtained by comparing the analysis algorithm performance between the reference video and its compressed version. Using F-measure, it can be expressed as

$$D_A^i = 1 - F_i, \quad (11)$$

where D_A^i represents the analytical distortion of the i -th frame in video sequence. Finally, the average analytical distortion D_A^S over the video sequence S can be denoted as

$$D_A^S = \frac{1}{N} \sum_{i=1}^N D_A^i, \quad (12)$$

where N is the total number of frames. In the ADP model and the RAD model, the analytical distortion is then represented as D_A^S .

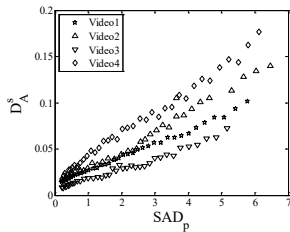


Fig. 2 Statistical results of D_A^S as a function of SAD_p

In order to reduce the influence of the spatial resolution and number of frames, we use SAD per pixel denoted as SAD_p to represent compression distortion, which can be written as

$$SAD_p = \frac{SAD^S}{N * W * H}, \quad (13)$$

where SAD^S denotes SAD of the entire compressed video and N , W and H represent the total number of frames, frame width and frame height respectively. The statistical results are shown in Fig. 2.

From Fig. 2, it can be seen that the average analytical distortion D_A^S increases linearly with the increase of SAD_p . Consequently, we design the ADP model as a linearly model

$$D_A^S = P_1 * SAD_p + P_2, \quad (14)$$

where P_1 and P_2 are constant parameters that are obtained by linear regression in the training procedure.

B. Rate Analytical Distortion Model

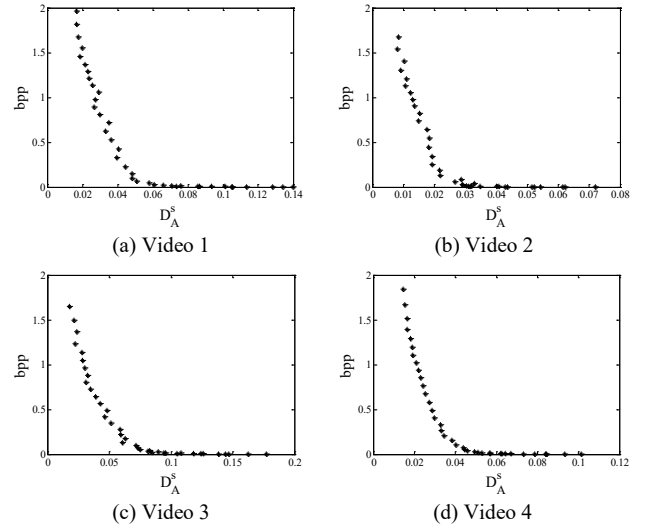


Fig. 3 Statistical results of the bit rate (bpp) as a function of D_A^S

Note that we choose bpp (bit per pixel) to denote the bit rate to reduce the influence of different frame rates and resolutions. From Fig. 3, it can be seen that the bit rate decreases exponentially with the increase of D_A^S . This observation gives a proof of the validity of RAD model expressed in (7). Additionally, the parameters of RAD model are obtained by a regression process according to the statistical results in the training procedure.

C. Weighting Parameter α Acquisition

In order to explore the impact of the weighting parameter α on the proposed framework, a set of α values ranging from 0.1 to 0.9 are tested. Fig. 4(a) shows the experimental results when QP = 27. It can be seen that the average analytical distortion D_A^S tends to increase smoothly along with α . Therefore we set α to be 0.1 as an initial value in this paper. Meanwhile, Fig. 4(b) implies that the object area of one frame may have a strong impact on the frame level analytical distortion denoted as D_A^i , since smaller object has less contribution to SAD of one frame.

Fig. 5 illustrates the detection results. From Fig. 5, it can be seen that the total detected object area in Video 3 is smaller than that in Video 1.

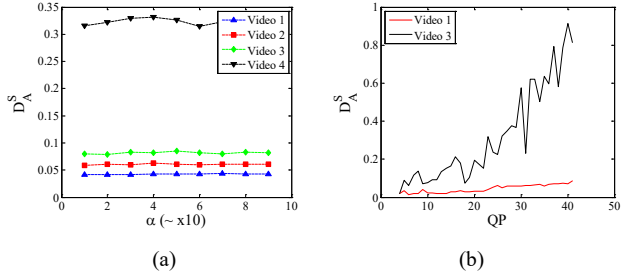


Fig. 4 Impact of α on the proposed framework. (a) In order to show the results, α is enlarged 10 times in the figure and 4 sequences are compressed using same QP (27). (b) Compressed with the same α (0.1), the analytical distortion of 599-th frame D_A^i in Video 1 and Video 3 are shown under different QPs.

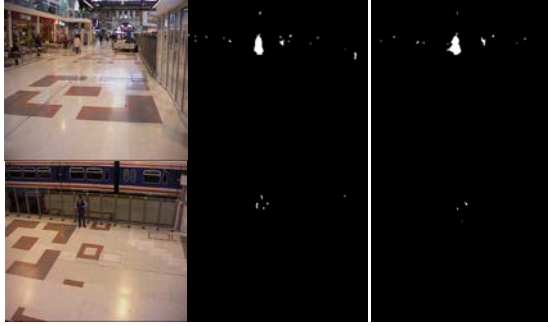


Fig. 5 The first column shows the 599-th original frames in Video 1 and Video 3 respectively. The second column represents the detection results of the original frames and the third column denotes the detection results of their compressed versions (QP = 37).

Inspired by the above observations, we calculate the variance of D_A^i under different QPs and collect the total detected object area of each original frame. The results are shown in Fig. 6. In Fig. 6, the analytical distortion variance is much bigger when the object area is smaller, which implies that smaller object area should be compressed with higher quality to avoid a high fluctuation of analytical distortion.

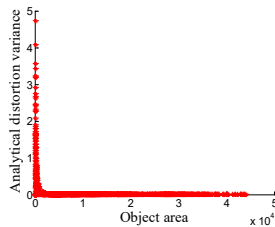


Fig. 6 Analytical distortion variance vs. object area

As a consequence, the weighting parameter α is updated according to the detected object area of each frame, which can be expressed as

$$\alpha = 0.1 * \exp\left(\frac{Area}{W * H}\right), \quad (15)$$

where W and H denote the frame width and frame height and $Area$ represents the total detected object area of each frame. In the coding procedure, the total object area can be derived by a frame subtraction method [22].

IV. EXPERIMENTAL RESULTS

In this section, simulations are conducted to evaluate the performance of the proposed method in the HEVC reference software HM 16.7. Low-Delay P (LDP) configuration is tested. Four QP values (22, 27, 32, 37) are used and BD-rate saving (BDBR) is employed to measure the RD performance of the proposed method. The RAD curves of our proposed approach and HEVC on each video sequence are shown in Fig. 7.

TABLE 1. PROPERTIES OF THE FOUR VIDEO SEQUENCES

Sequence	Resolution	Frame Number	Frame Rate
Video 5	720x576	600	25
Video 6	720x576	600	25
Video 7	720x576	600	25
Video 8	800x600	200	30

In order to show the performance of our proposed method, another 4 test video sequences Video 5-8 are chosen, including 3 indoor videos from PETS2006, and 1 outdoor video from CAVIAR. Detailed properties of the four sequences are given in Table 1.

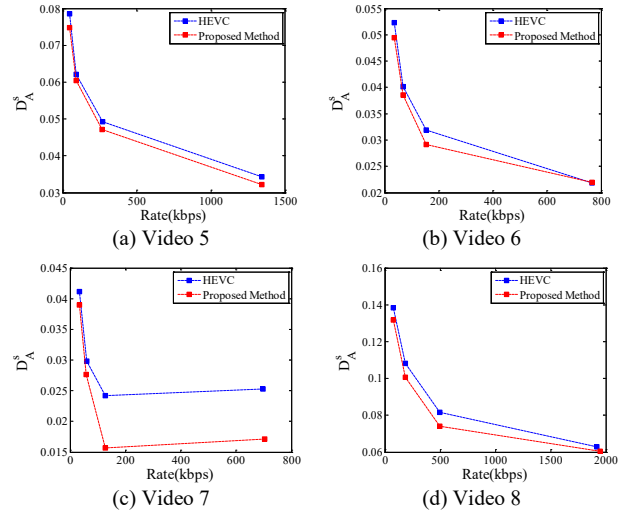


Fig. 7 Experimental results

TABLE 2. LUMA BDBR UNDER LDP CONFIGURATION

Video Sequence	BDBR
Video 5	0.5%
Video 6	0.6%
Video 7	1.1%
Video 8	2.4%
Average	0.9%

As a first observation in Fig. 7, it can be seen that our proposed approach can reduce the analytical distortion effectively under low bit rate (QP = 37) while the coding efficiency decreases slightly in terms of BD-rate (BDBR). In the best scenario, up to 40% reduction in terms of average

analytical distortion is achieved for QP 27 and 22 in Fig. 7(c), since the object area in Video 7 is smaller than that in other video sequences. In contrast, the analytical distortion increases at low bit rate (QP = 32, 37) in Fig. 7(c), since the background has a great impact on SAD which results in a lot of false positives. Table. 2 shows the performance in terms of BDBR between our proposed approach and HEVC. We can see that the coding performance of our approach increases about 0.9% in average. The reason is that our proposed RADO method may change the mode decision for coding units and the best chosen mode is no longer optimal in terms of RD. The complexity comparison between HM 16.7 and the proposed method are shown in Table 3 in terms of the encoding time. It can be seen that the encoding time of the proposed method is about 1.8 times compared with that of HEVC.

TABLE 3. ENCODING TIME RATIO

Sequence	Encoding Time (sec.)		Encoding Time Ratio
	HM	Proposed	
Video 5	7919.043	14286.955	180.413%
Video 6	7898.877	13849.762	175.338%
Video 7	8141.746	14894.478	182.940%
Video 8	1617.044	2865.203	177.188%

V. CONCLUSION

In this paper, we attempt to design a new video coding mechanism for video analysis and propose a new rate-distortion method. To illustrate the validation of our proposed method, we consider moving object detection as an example and propose a rate-analytical-distortion model for video coding. More specifically, the compression distortion is used to predict the analytical distortion. The experimental results show that our proposed model can effectively reduce the analytical distortion. However, our proposed method is a two-pass encoding method which increases the complexity. In our future work, we will aim at reducing the complexity by applying some SAD estimation approaches to avoid the two-pass encoding process. Besides, investigating a more adaption model to derive the weighting parameter α between the compression distortion and the analytical distortion is also an important aspect which will be addressed in the future.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grant 61571102 and in part by the Other Funds under Grant 2015AA015903 and Grant ZYGX2014Z003.

REFERENCES

[1] B. Li, J. Xu, D. Zhang, and H. Li, "QP Refinement according to Lagrange multiplier for high efficiency video coding," IEEE International Symposium on Circuits and Systems, May 2013, pp. 477-480.

[2] H. Zeng, K. N. Ngan, and M. Wang, "Perceptual adaptive Lagrangian multiplier for high efficiency video coding," IEEE Picture Coding Symposium, Dec. 2013, pp. 69-72.

[3] Y. Gao, C. Zhu, S. Li, T.W. Yang, "Temporally dependent rate-distortion optimization for low-delay hierarchical video coding," IEEE Transactions on Image Processing, vol. 26, no. 9, pp. 4457-4470, Sept. 2017. (DOI: 10.1109/TIP.2017.2713598)

[4] S. Li, C. Zhu, Y.B. Gao, Y.M. Zhou, F. Dufaux, M.T. Sun, "Lagrangian multiplier adaptation for rate-distortion optimization with inter-frame dependency," IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 1, pp. 117-129, Jan. 2016.

[5] A. Redondi, L. Baroffio, M. Cesana, and M. Tagliasacchi, "Compress-then-analysis vs. analysis-then-compress: Two paradigms for image analysis in visual sensor networks," IEEE International Workshop on Multimedia Signal Processing, Sep. 2013, pp. 278-282.

[6] L. Baroffio, M. Cesana, A. Redondi, M. Tagliasacchi, and S. Tubaro, "Coding visual features extracted from video sequences," IEEE Transactions on Image Processing, vol. 23, no. 5, pp. 2262-2276, May 2014.

[7] L. Baroffio, J. Ascenso, M. Cesana, A. Redondi, and M. Tagliasacchi, "Coding binary local features extracted from video sequences," in IEEE International Conference on Image Processing, Oct. 2014, pp. 2794-2798.

[8] J. Chao, E. Steinbach, "Keypoint encoding for improved feature extraction from compressed video at low bitrates," IEEE Transactions on Multimedia, vol. 18, no. 1, pp. 25-39, Mar. 2016.

[9] P. Korshunov and W. T. Ooi, "Critical video quality for distributed automated video surveillance," ACM Multimedia, Singapore, Nov. 2005, pp.151-160.

[10] E. Kokiopoulou and P. Frossard, "Semantic coding by supervised dimensionality reduction," IEEE Transaction on Multimedia, vol. 10, no. 5, pp. 806-818, Aug. 2008.

[11] L. Liao, R. Hu, J. Xiao, G. Zhan, Y. Chen, J. Xiao "An analysis-oriented ROI based coding approach on surveillance video data," Advances in Multimedia Information Processing - PCM 2016, Nov. 2016, pp. 428-438.

[12] B. Xiong, X. Fan, C. Zhu, X. Jing, Q. Peng, "Face region based conversational video coding," IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, no. 7, pp. 917-931, Jul. 2011.

[13] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression: An overview," IEEE Signal Processing Magazine, vol. 15, no. 11, pp. 23-50, Nov.1998.

[14] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," IEEE Signal Processing Magazine, vol. 15, no. 11, pp. 74-99, Nov.1998.

[15] H. Everett, "Generalized Lagrange multiplier method for solving problems of optimum allocation of resources," Operations Research, vol. 11, no. 3, pp. 399-417, Jun. 1963.

[16] H. Gish and J. Pierce, "Asymptotically efficient quantizing," IEEE Transactions on Information Theory, vol. 14, no. 5, pp. 676-683, Oct. 1968.

[17] B. Lei, A. Leonardis, and B. Schiele, "Robust object detection with interleaved categorization and segmentation," Journal of Computer Vision, vol. 77, no. 1, pp. 259-289, May 2008.

[18] Y. Dhome, N. Tronson, A. Vacavant, T. Chateau, C. Gabard, Y. Goyat, D. Gruyer, "A benchmark for background subtraction algorithms in monocular vision: a comparative study," International Conference on Image Processing, Theory, Tools and Applications, Jul. 2010, pp. 66-71.

[19] D. Thirde, L. Li, F. Ferryman, "Overview of the PETS2006 challenge," Proc. 9th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Jun. 2006, pp. 47-50.

[20] <http://homepages.inf.ed.ac.uk/homes/rbf/CAVIAR/>.

[21] https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/.

[22] N. Singla, "Motion detection based on frame difference method," International Journal of Information & Computation Technology, vol. 4, no. 15, pp. 1559-1565, Nov. 2014.