



HAL
open science

An adaptive Lagrange multiplier determination method for dynamic texture in HEVC

Chengyue Ma, Karam Naser, Vincent Ricordel, Patrick Le Callet, Chunmei
Qing

► **To cite this version:**

Chengyue Ma, Karam Naser, Vincent Ricordel, Patrick Le Callet, Chunmei Qing. An adaptive Lagrange multiplier determination method for dynamic texture in HEVC. IEEE International Conference on Consumer Electronics-China (ICCE-China), Dec 2016, Guangzhou, China. pp.1 - 4, 10.1109/ICCE-China.2016.7849749 . hal-01481686

HAL Id: hal-01481686

<https://hal.science/hal-01481686v1>

Submitted on 2 Mar 2017

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.

An Adaptive Lagrange Multiplier Determination Method for Dynamic Texture in HEVC

Chengyue Ma^{1,2}, Karam Naser², Vincent Ricordel², Patrick Le Callet² and Chunmei Qing¹

¹ South China University of Technology, Guangzhou, 510640, China

² LUNAM University, University of Nantes, IRCCyN UMR CNRS 6597
Polytech Nantes, Rue Christian Pauc BP 50609 44306 Nantes Cedex 3, France

Abstract—This research proposed a method for adaptive Lagrange multiplier determination for rate-distortion optimization with dynamic texture in High Efficiency Video Coding (HEVC). Inspired by the experimental results of the Lagrange multiplier selection test, the presented approach adaptively predicts the optimum Lagrange multiplier for different dynamic texture sequences, based on the features of the dynamic texture sequences such as normal flow and spatial-temporal information. The Lagrange multiplier among the given values will be chosen based on the Bjontegaard delta measurements. After that, the data of training dynamic texture will be used for Support Vector Machine (SVM) in machine learning for getting the predicting results. The proposed algorithm has been fully integrated into HEVC reference codec. The result shows that the proposed method can improve 0.5 in Structural Similarity Metric (SSIM) and 2 in Peak Signal-to-Noise Ratio (PSNR).

Keywords—Lagrange multiplier, rate-distortion optimization, Dynamic textures

I. INTRODUCTION

In recent year, since people does not satisfy with the normal fidelity in the traditional forms of videos, high definition (HD) and ultra HD (UHD) videos have been increasingly popular with the rapid development of cameras. However, it is a great challenge to transmit and store the huge amount of the videos data for the practical applications. The ITU-T Visual Coding Experts Group (VCEG) and the ISO-IEC Moving Picture Experts Group (MPEG) have developed a new video standard, namely, high efficiency video coding (HEVC). Compared with the previous video coding standards, HEVC achieves compression more efficiency by using variable size for coding units and prediction units, advanced motion vector prediction and adaptive loop filtering. With the improvement, HEVC is expected to be widely used in various applications, such as digital satellite broadcasting, entertainment, online education, and so on.

Following the classical coding framework, rate distortion optimization (RDO) techniques have been exploited in HEVC to identify the optimal encoding setting for achieving the optimal rate distortion (RD) performance. However, the Lagrange RDO technique can not accurately reflect the perceptual visual quality due to two reasons. Firstly, the Lagrange multiplier only relates to the quantization parameter but ignores the characteristic of the input video. Secondly, the

coding efficiency in RDO is measured by the objective criterion(i.e., Sum of square error, SSE), which can not accurately reflect the perceptual visual quality. In other words, HEVC does not fully consider the perceptual characteristics of the input video during the encoding process. Therefore, for improving the efficiency of video coding while keeping the high quality, it is required to use the characteristics of the Human Visual System (HVS).

For developing an adaptive Lagrange multiplier, some methods have already been proposed in the published literatures. [1] presented a Laplace-based Lagrange multiplier according to the Laplace distribution of transformed residuals. [2] proposed an adaptively adjusting Lagrange multiplier in the RDO process based on the perceptual sensitivity of the input CTU. [3] obtained an adaptive Lagrange multiplier based on the free-energy principle which represents the disorderly concealment effect in human eyes. [4] introduced an adjusting Lagrange multiplier based on the gradient magnitude of the input CTU.

Besides, another solution for the perceptual video coding is to incorporate the existing visual quality assessment (VQA) metrics which can more suitable describe the human perception into the video codec. For example, [5] proposed a normalization factor based on DCT domain-SSIM index to transform the DCT residual into the perceptually uniform space together with a new distortion model of improving the perceptual coding efficiency. [6] introduced the SSIM to replace the SSE as the quality metric and developed a SSIM-based RDO using the coding information of the key frame.

Although there are already some proposed methods for improving the Lagrange multiplier and the quality measurement in RDO, most of these methods only consider one perceptual characteristic while improving the rate distortion optimization techniques. It is not enough since human visual system would be affected by many factors. Therefore, it is necessary to incorporate many factors that would affect human visual and give each factor a suitable weight according to the experiment for improving the RDO more likely considering the characteristics of the HVS.

Moreover, dynamic texture is also a major research topic in the field of video coding. Dynamic textures are sequences of frames of moving scenes that exhibit temporal regularity, intended in a statistical sense, like sea-waves, smoke, foliage, whirlwind but also talking faces, traffic scenes etc. Since

dynamic texture usually cost much but not important in human visual system. Therefore, many algorithm for dynamic texture synthesis is presented as it is easy to implement and fast in computation. However, we can still improve these algorithm for making dynamic texture performed better in video coding.

The organization of the rest of the paper is given as follows: Section II briefly describes a brief background covering the encoding part in HEVC. Section III gives the proposed methods for adaptive Lagrange multiplier in HEVC. Section IV presents global view of my method for the predicting adaptive Lagrange multiplier factor and show the result of the experiment. Finally, Section V concludes the paper and give some thoughts for the future work.

II. LAGRANGIAN RATE DISTORTION OPTIMIZATION IN HEVC

In the HEVC, rate distortion optimization (RDO) plays an important role in the mode decision process, which is to find a good trade-off between the reconstructed video quality and the required bits. To adapt to various video content, there are various prediction modes in HEVC that can be roughly classified into Intra, SKIP, and Inter Modes. For example, the mode with larger size consumes less bits for head information and is efficient to code the picture block with homogeneous textures. In contrast, the mode with smaller size provides more accurate prediction and thus yields less residual, at the expense of higher head information. The mode decision process is to exhaustively compute the RD cost of all the prediction modes and find the one with the minimum RD cost as the optimal mode. In fact, this process is an optimization problem that minimizes the overall reconstructed video distortion D at a given rate R . The Lagrange RDO is defined as :

$$\min(J) \quad \text{where } J = D + \lambda_{HEVC} * R \quad (1)$$

where J is the Lagrange cost function, R denotes the total number bits for coding the headers, quantized coefficients, etc, measured in terms of bits per pixel and D means the distortion between the original block and its reconstructed block. The λ is the Lagrange multiplier in HEVC, which is defined as :

$$\lambda_{HEVC} = \alpha * 2^{\left(\frac{QP-12}{3}\right)} \quad (2)$$

where QP is the quantization parameter, and α is a constant that is empirically-determined from extensive experiments[2].

It is easily observed from equation (1) that HEVC optimizes the RD performance for a given CTU based on a empirically-determined Lagrange multiplier λ . Therefore, λ is very important, as it controls HEVC encoder to find the optimal trade-off between the distortion and the bit rate.

However, λ has one drawback that it is a only function of QP and does not consider the characteristic of the input video and the HVS perception. It means that λ cannot well adapt to various video contents and is not efficient for perceptual video coding. Therefore, it is very essential to develop a perceptual adaptive Lagrange multiplier for HEVC by incorporating HVS perception to improve the perceptual coding efficiency.

III. PROPOSED LAGRANGE MULTIPLIER DETERMINATION METHOD FOR HEVC

A. Lagrange multiplier factor

As it is well known, texture represented homogeneous area of video scenes with coherent statistics and appear in large areas of video streams while structures represented rather the semantics of the scene [7]. Therefore, human visual system usually focus on the structural area rather than textured area. However, textured area would be coded bit by bit as the structural area in the previous methods and the only difference of dealing with them is to give a different weight in Lagrange multiplier. But human visual system only focus on the semantic meaning of the textured area rather than the exact details of each pixel. Therefore, replacing them by an equivalent stochastic signal results in significant bitrate saving [8][9].

It is also noted that Lagrange multiplier in HEVC mainly depend on QP values as function (2), and are not directly related to the characteristics of the input video signal. Moreover, this model is based on the experiment in [10] that used an H.263 codec (without B frames) and four QCIF test sequences. With recent advances in video compression, especially with bi-directional prediction and intra coding in inter frames, the RD characteristic of codec with B frames is likely to be quite different.

In order to investigate the optimum Lagrange multiplier factor for encoding, an Lagrange multiplier selection experiment was applied to HEVC reference codec. The test values used are given in function (3).

$$\lambda_{test} = k_{factor} * \lambda_{orig} \quad (3)$$

Where the value of k_{factor} is chosen as, and this parameter was constant during encoding.

Primary coding settings used in this experiment include : HM16.2 for HEVC ; uniform QP used for all frames: QP 22, 27, 32, 37, 42 and 47 tested for HEVC ; Structural Similarity Index (SSIM) and Peak Signal-to-Noise Rate (PSNR) is used as distortion metric for RDO in the codec.

B. Determination of the best factor

For choosing the best value of Lagrange multiplier factor, we use the bjontegaard metric for deciding which value of Lagrange multiplier factor have the best performance in R-D curve. In figure 9, the figure has shown all of the different value of λ perform in rate distortion map with PSNR. The way to measure the compression quality with the bjontegaard metric, Bjontegaard delta bit rate(BDBR) and Bjontegaard delta peak signal-to-noise rate (BD-PSNR).

In BDBR, the basic idea is to measure the bit-rate saving with the same video quality. In our experiment as shown in Fig. 1. We take the λ_{best} and the λ_{orig} for all the Lagrange multiplier factor. It is easily to find out that the bit-rate is smallest while λ_{best} in these two figures. Therefore, we could prove that λ_{best} is the best Lagrange multiplier factor in this video.

In BD-PSNR or Bjontegaard delta structural similarity index (BD-SSIM), the basic idea is to measure the video

quality with the same bit-rate. In our experiment, we consider the video quality while bit-rate equal to 20. It can be seen that has the highest SSIM or PSNR value among them. Thus, we can make sure that it is the best Lagrange multiplier factor in this video once again.

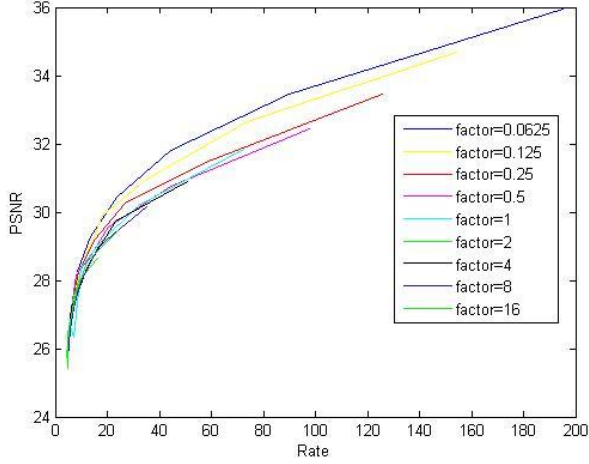


Fig. 1. Different k_{factor} performed in R-D curve with PSNR

C. Extracting the features

In this paper, we extract several features of the dynamic texture sequences: Normal flow, the standard Spatial Information (SI) and Temporal Information (TI), Gray Level Cooccurrence Matrix (GLCM).

The normal flow field is easy to compute and can be directly estimated without any iterative scheme used by regularization methods [11]. Moreover, it included both structural and temporal information on dynamic texture : spatial information is linked to edge gradient vectors, while temporal information is related to the moving edges. The drawback of its is the sensitivity to the noise, which can be reduced by smoothing or applying a threshold on the spatial gradients.

The Spatial perceptual Information, SI, is based on the Sobel filter. Each frame of the dynamic texture (luminance plane) at time $n(F_n)$ is firstly filtered with the Sobel filter [Sobel(F_n)]. The standard deviation over the pixels (std_{space}) in each Sobel-filtered frame is therefore computed. This operation is repeated for each frame in the dynamic texture sequence and results in a time series of spatial information of the scene. The maximum value in the time series (max_{time}) is chosen to represent the spatial information content of the scene.

The GLCM descriptor combines 4 features, that are contrast, correlation, energy and homogeneity. Similarly, the following descriptors are defined for normal flow vectors : Divergence, Curl, Peakness and Orientations.

Therefore, we have a list of the best Lagrange multiplier factor for each dynamic texture sequences and each corresponded features. Support Vector Machine (SVM) was used as a classification tool with the previous defined features as the attributes and the Lagrange multiplier factor as the result. To test the learning performance of SVM, we performed leave-

one-out, cross-validation and random sampling to test and measured the classification accuracy.

IV. EXPERIMENTAL RESULTS

The proposed predicting the best adaptive Lagrange multiplier factor method has been integrated into the HEVC reference models, and compared with the original codec. Identical test conditions are used here as in the experiment in Section 4.4, uniform test quantization parameters (QP) are used for all I, P and B frames from QP 22 to QP 47 with an interval of 5 for both codec.

The group of progressive format test sequences (YUV 4:2:0) are used here including 37 videos at in 30fps. These are all public video sequences from the standard test sequence group in the DynTex database.

Since we have 37 dynamic texture sequences in our dataset, we use the different uniform test quantization parameters (QP) which have been mentioned above for each dynamic texture sequences. Then we sum up all of the sequences in the same quantization parameter and compute the average SSIM and PSNR value. Finally, we also compute the total average SSIM and PSNR value among all the sequences with all test quantization parameters.

The compression performance of the proposed method for HEVC is compared with the corresponding anchor codec. The results are based on the Bjontegaard delta measurements with the two video quality metric SSIM and PSNR are shown in Table2. Besides, the proposed method compared with the original method with measurement of SSIM and PSNR over all frames are shown in Table I .

TABLE I. SUMMARY OF THE COMPRESSION RESULTS USING SSIM (A) AND PSNR (B)

(A)

QP	Proposed method	Anchor
22	0.8779	0.8123
27	0.8265	0.7565
32	0.7701	0.7045
37	0.7137	0.6542
42	0.6536	0.6007
47	0.585	0.5454
Avg	0.7378	0.6789

(B)

QP	Proposed method	Anchor
22	32.8893	30.0344
27	30.6499	28.3828
32	28.8581	27.0484
37	27.3798	25.9072
42	26.7491	24.813

47	24.8748	23.7938
Avg	28.5668	26.6633

It can be observed that the proposed method always performs better than the anchor codec in all test quantization parameters with average more than 0.1 better in SSIM. It should be noted that the proposed method is based on experimental results using uniform QP for all frames – one of the most commonly used scenarios. For cases with different GOP structure for different frame types, the optimum Lagrange multiplier factors may vary. In terms of computational complexity, even when including computation for scene cut detection, the increased complexity of our method is negligible in the overall context of HEVC compression.

Although the method presented here are inspired from the adaptive Lagrange multiplier selection method from [12], we want to dispel the misconception that our method is trying to predict the better Lagrange multiplier factor based on the features of dynamic texture sequences instead of selecting the best one among a number of values. In particular, we indeed choose the best Lagrange multiplier factor among some values in the training step. However, we use machine learning as solving the linear regression problem instead of the classification. Therefore, the techniques under consideration turn out to be simple predictor that show the predicted best Lagrange multiplier factor while we put in the features of the dynamic texture sequence.

V. CONCLUSION

In this paper, a method for predicting the adaptive perceptual Lagrange multiplier factor is proposed for HEVC to improve its perceptual coding performance. In our approach, it has trained by 37 homogeneous spatio-temporal patches, referred as dynamic texture. For a given test dynamic texture sequence, it was first extracted the features such as spatial information, temporal information and normal flow. Once we know the feature of the given dynamic texture, we can achieve predicting the best Lagrange multiplier factor in HEVC. Experimental results show that the proposed method is able to predict one Lagrange multiplier factor to obtain significant improvement on perceptual RD performance and visual quality of the reconstructed dynamic texture sequences, compared with the original rate distortion optimization in HEVC.

We derive a complete framework for predicting the best Lagrange multiplier factor in video compression standard (HEVC). We provide preliminary results indicating that the method can provide a predicted Lagrange multiplier factor that can improve the visual quality at equal bit-rate and should be further optimized.

The possible future outcome of this work is to improve the computation of the features of dynamic texture sequences. Although we extract a great number of features in this experiment for trying to using these features to describe the whole characteristic of the sequences better, we simply deal with these features with computing the average value of all the pixels. In this case, the average cannot well represent the characteristic of the dynamic texture sequences. Therefore, we

will try to compute the features of dynamic texture sequences in a more reasonable way for representing the sequences better.

Besides, we only have 37 dynamic texture sequences in our experiment. However, it is not enough for machine learning to make the precise prediction. Moreover, the 37 dynamic texture has the same format and the same size from the same database. In another word, the dataset we used is not variety enough. Therefore, the prediction cannot be very precise in predicting the best Lagrange multiplier factor. In the future work, we need to collect the dynamic texture sequences from many database as much as possible. Besides, using different length and GOP structure of the dynamic texture sequences may improve the prediction of the Lagrange multiplier factor.

ACKNOWLEDGMENT

This work was supported by the Marie Skłodowska-Curie under the PROVISION (Perceptually Optimised Video Compression) project bearing Grant Number 608231 and Call Identifier: FP7-PEOPLE-2013-ITN and the Fundamental Research Funds for the Central Universities in China under Grant 2015ZZ032.

REFERENCES

- [1] Li, Xiang, et al. "Laplace Distribution Based Lagrangian Rate Distortion Optimization for Hybrid Video Coding." *IEEE Transactions on Circuits & Systems for Video Technology* 19.2(2009):193-205.
- [2] Zeng, Huanqiang, K. N. Ngan, and M. Wang. "Perceptual adaptive Lagrangian multiplier for high efficiency video coding." *2013 Picture Coding Symposium (PCS) 2013*:69-72.
- [3] Jung, Cheolkon, and Y. Chen. "Perceptual rate distortion optimization for video coding using free-energy principle." *Electronics Letters* 51.21(2015):1656-1658.
- [4] Yang, Aisheng, et al. "Perceptual feature guided rate distortion optimization for high efficiency video coding." *Multidimensional Systems & Signal Processing* (2016):1-18.
- [5] Wang, S., et al. "Perceptual video coding based on SSIM-inspired divisive normalization." *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society* 22.4(2013):1418-29.
- [6] Huang, Yi Hsin, et al. "Perceptual Rate-Distortion Optimization Using Structural Similarity Index as Quality Metric." *IEEE Transactions on Circuits & Systems for Video Technology* 20.11(2010):1614-1624.
- [7] Naser, K., Ricordel, V., and Le Callet, P. "Experimenting texture similarity metric STSIM for intra prediction mode selection and block partitioning in HEVC." *International Conference on Digital Signal Processing 2014* : 882-887.
- [8] Balle, J., A. Stojanovic, and J. R. Ohm. "Models for Static and Dynamic Texture Synthesis in Image and Video Compression." *IEEE Journal of Selected Topics in Signal Processing* 5.7(2011):1353-1365.
- [9] Zhang, Fan, and D. R. Bull. "A Parametric Framework for Video Compression Using Region-Based Texture Models." *IEEE Journal of Selected Topics in Signal Processing* 5.7(2011):1378-1392.
- [10] Wiegand, Thomas, and B. Girod. "Lagrange multiplier selection in hybrid video coder control." *International Conference on Image Processing, 2001. Proceedings 2001*:542-545 vol.3.
- [11] Horn, Berthold K. P., and B. G. Schunck. *Determining optical flow.* International Society for Optics and Photonics, 1981.
- [12] Zhang, Fan, and D. R. Bull. "An adaptive Lagrange multiplier determination method for rate-distortion optimization in hybrid video codecs." *IEEE International Conference on Image Processing IEEE, 2015*:29138-29146.