J. Vis. Commun. Image R. 38 (2016) 824–837

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Machine learning based fast H.264/AVC to HEVC transcoding exploiting block partition similarity ☆

Linwei Zhu [a,b], Yun Zhang [a,*], Na Li [a], Gangyi Jiang [c], Sam Kwong [b,d]

[a] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[b] Department of Computer Science, City University of Hong Kong, Hong Kong, China
[c] Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China
[d] Shenzhen Research Institute, City University of Hong Kong, Shenzhen, China

ARTICLE INFO

Article history:
Received 1 September 2015
Revised 9 March 2016
Accepted 24 April 2016
Available online 25 April 2016

ABSTRACT

Video transcoding is to convert one compressed video stream to another. In this paper, a fast H.264/AVC to High Efficiency Video Coding (HEVC) transcoding method based on machine learning is proposed by considering the similarity between compressed streams, especially the block partition correlations, to reduce the computational complexity. This becomes possible by constructing three-level binary classifiers to predict quad-tree Coding Unit (CU) partition in HEVC. Then, we propose a feature selection algorithm to get representative features to improve predication accuracy of the classification. In addition, we propose an adaptive probability threshold determination scheme to achieve a good trade-off between low coding complexity and high compression efficiency during the CU depth prediction in HEVC. Extensive experimental results demonstrate the proposed transcoder achieves complexity reduction of 50.2% and 49.2% on average under lowdelay P main and random access configurations while the rate-distortion degradation is negligible. The proposed scheme is proved more effective as comparing with the state-of-the-art benchmarks.

1. Introduction

With the development of computing and multimedia technologies, various types of multimedia devices become available, including computers, laptop, smart phones, television, pad, set-top box, even the wearable devices, like glasses and watch. The diversities of capabilities of these multimedia devices, conditions of their accessed networks, multimedia data syntax/formats and user requirements unfortunately create a gap in end-to-end communication and sharing the multimedia contents among different terminals. Video transcoding is one of the proper solutions to bridge the gap for video communication among different applications and systems [1]. It is a process of converting one compressed stream to another required stream, in which properties of the bit stream may be changed, including coding syntax, bit rate, resolution, frame rate, coding structure and quality [2].

Coding technologies and standards are necessary to transcode one bit stream to another. With the advance of the video coding standards, the compression efficiency is significantly improved as adopting many novel and advanced coding tools and technologies. For example, the compression efficiency is doubled from MPEG-2 to H.264/Advanced Video Coding (AVC) [3] and H.264/AVC to High Efficiency Video Coding (HEVC) [4]. These video compression standards actually co-exist in a certain range of applications, which makes transcoding desirable. Meanwhile, the syntax and coding technologies vary significantly from standard to standard, which makes it challenging to transcode a bit stream of previous standard to that of the latest, *e.g.* HEVC.

Over the last few decades, many researches have been developing video transcoding algorithms for different system requirements and applications. Cock et al. [2] used motion refined rewriting of single-layer H.264/AVC streams to multiple quality layers for Scalable Video Coding (SVC) streams. Zhang et al. [5] proposed a novel multidimensional no-reference video quality metric for video transcoding, where frame rate and frame size are taken into account simultaneously. Liu et al. [6] proposed a Quality of Experience (QoE) oriented transcoding approach to enhance the quality of mobile 3D video service, where transcoding parameters are configured according to the feedbacks of both the network and

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author.
E-mail addresses: lwzhu2-c@my.cityu.edu.hk (L. Zhu), yun.zhang@siat.ac.cn (Y. Zhang), na.li1@siat.ac.cn (N. Li), jianggangyi@nbu.edu.cn (G. Jiang), cssamk@cityu.edu.hk (S. Kwong).

http://dx.doi.org/10.1016/j.jvcir.2016.04.020
1047-3203/

the user-end device information. A straightforward video transcoder is designed merely to cascade the decoder and encoder, noted as cascaded transcoder [1]. The incoming video stream is fully decoded, and then the target video bit-stream is generated with a new encoder by re-coding the former reconstructed video. In this transcoder, the bit rate, syntax and format can be fully changed. However, the information of the source compressed stream that has not been fully and effectively explored to facilitate thereafter encoding process, such as Motion Vector (MV) and block partition type. Though video coding technical details vary from standard to standard, there is still high correlation between the encoding outputs of two standards, *e.g.* block partition, reference frame indices, since they basically follow the same framework, block based hybrid structure.

To lower the coding complexity of the transcoder, Liu et al. [7] proposed a fast MPEG-2 to H.264/AVC transcoding method based on mode mapping and macroblock (MB) activity, in which skip and intra modes are directly mapped from MPEG-2 to H.264/AVC, and the MB mode of H.264/AVC is estimated according to the residual Discrete Cosine Transform (DCT) energy from MPEG-2 stream. Shu et al. [8] proposed a mode decision method by taking into account the error propagation to the following frame. It enhanced the overall robustness of the transcoded bit stream against the packet loss. Moreover, a fast Motion Estimation (ME) algorithm for MPEG-2 to H.264/AVC transcoding was proposed in [9], where MVs as well as MB mode information extracted from MPEG-2 bit-stream were utilized to speed up the ME of H.264/AVC encoder. An MV decomposition algorithm for H.263 to H.264/AVC transcoding was presented in [10]. Petljanski et al. [11] presented a MB mode estimation technique for MPEG-2 to H.264/AVC intra transcoding, in which the DCT coefficients of MPEG-2 stream were exploited to represent the video texture and predict intra MB mode of H.264/AVC. In summary, the above methods mainly focus on the transcoding among MPEG-2, H.263 and H.264/AVC.

In HEVC, quad-tree block partition is adopted and the Coding Unit (CU) size is from 8×8 to 64×64. In addition, more advanced and refined coding tools, such as flexible Prediction Unit (PU) and Transform Unit (TU) mode, are adopted to further reduce the prediction residue [4]. These techniques improve the coding efficiency, however, significantly increase the coding complexity. Meanwhile, the hike of the number of mode candidates in HEVC makes the optimization more challenging. Recent researches on video transcoding from previous hybrid compression standards to HEVC have also been reported. Peixoto et al. [12] proposed a fast H.264/AVC to HEVC transcoder, where the MVs of H.264/AVC stream were reused in HEVC coding process. In [13], power spectrum based Rate Distortion Optimization (RDO) model, input residue, modes and MVs were used to estimate the best CU, PU modes and their corresponding MVs. Shen et al. [14] proposed a parallelization optimization method for fast H.264/AVC to HEVC transcoder, which implemented fast mode decision with wavefront parallel processing and Single Instruction Multiple Data (SIMD) acceleration. Jiang and Chen [15] proposed a MV clustering based fast H.264/AVC to HEVC transcoding method. Chen et al. [16] developed a H.264/AVC to HEVC transcoder based on distributed multi-core processors, which claimed 720p@30fps real-time video transcoding. Generally, these above mentioned schemes use the information extracted from H.264/AVC bit-stream to form a mode mapping between H.264/AVC and HEVC with hard thresholds and categories, which will limit the transcoder's performance and adaptability. While, it jointly takes advantage of the current coding information in HEVC and video content to improve the transcoding performance.

To develop more advanced and reliable video coding or transcoding algorithms, learning algorithms have been introduced to solve the prediction or classification problems. Fernandez-Escribano et al. [17] presented an efficient MPEG-2 to H.264/AVC baseline profile transcoder, where machine learning tools were used to exploit the correlation between the MB mode in H.264/AVC and the distribution of motion compensation residue in MPEG-2. Furthermore, a dynamic motion estimation technique was proposed to further reduce the complexity of the decision process. Chiang et al. [18] proposed a statistical learning based fast coding algorithm for H.264/AVC, in which several representative features in H.264/AVC were analyzed and a statistical learning model was built. With the help of an off-line classification approach from statistical learning, the complexity of motion estimation and mode decision was significantly reduced. In [19], an on-line machine learning based solution was introduced for MPEG-2 to HEVC transcoding. After fully decoding and encoding a few frames, this transcoder analyzed the relationship between MPEG-2 stream information and CU depths in HEVC, so as to build the learning model. Then, with this learning model, the HEVC's CU depth of the rest frames can be predicted by classifiers. In [20], statistical thresholds are integrated in the machine learning based coding framework to help improve the accuracy. However, the features adopted are empirically determined. Thus, it is difficult to have a good trade-off between the complexity and RD performance for different sequences. Additionally, in [21], CU partition in HEVC was regarded as a binary classification problem solved by Support Vector Machine (SVM). In the SVM model training, the RD degradation caused by misclassification was introduced as weights, which reduced the number of misclassifications with higher negative impacts. Xiong et al. [22] proposed a fast CU decision scheme based on Markov Random Field (MRF) for fast HEVC inter coding, where the variance of the absolute difference is adopted as the key feature. In [23], Zhang et al. proposed an early termination structure and three-output joint classifier based on SVM, which can flexibly adjust the CU depth prediction accuracy and complexity reduction by changing the weighted factor in training. However, this model is trained off-line and the feature selection still could be further improved. Basically, machine learning based approaches are capable of integrating multiple features and learning from video content or previous coded information to improve the transcoding performance.

In this paper, a machine learning based fast H.264/AVC to HEVC transcoding method is proposed, where a fast CU decision algorithm is presented. Compared with our previous work in [23], this paper will address the problem in fast H.264/AVC to HEVC transcoding, and optimize the feature selection for the fast CU decision. Meanwhile, we propose an on-line probability based SVM method for CU depth prediction, in which an adaptive probability threshold decision algorithm is adopted to achieve a better trade-off between low complexity and high compression efficiency. The remainder of this paper is organized as follows: problems and motivations are presented in Section 2. Then, complexity redundancies are analyzed in Section 3. Section 4 describes the proposed machine learning based fast transcoding method. Experiments are implemented and the comparative results are analyzed in Section 5. Finally, conclusions are drawn in Section 6.

2. Problems and motivations

In conventional cascaded transcoder, the bit stream information is not effectively utilized to facilitate the encoding process. To tackle this problem and fully exploit the available information, the bit stream is not necessary fully decoded and some of the bit stream information can be shared to cascaded coding [17,19], such as MVs, reference frames and block partitions. However, there are