Signal Processing: *Image Communication* 38 (2015) 127–140

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Weight-based R-λ rate control for perceptual HEVC coding on conversational videos ☆

Shengxi Li [a], Mai Xu [a,b,*], Xin Deng [a], Zulin Wang [a]

[a] *School of Electronic and Information Engineering, Beihang University, Beijing 100191, China*
[b] *EDA Lab, Research Institute of Tsinghua University in Shenzhen, Shenzhen, China*

ARTICLE INFO

Available online 14 May 2015

Keywords:
HEVC
Perceptual video coding
Rate control

ABSTRACT

This paper proposes a novel weight-based R-λ scheme for rate control in HEVC, to improve the perceived visual quality of conversational videos. For rate control in HEVC, the conventional R-λ scheme is based on bit per pixel (bpp) to allocate bits. However, bpp does not reflect the visual importance variation of pixels. Therefore, we propose a novel weight-based R-λ scheme to consider this visual importance for rate control in HEVC. We first conducted an eye-tracking experiment on training videos to figure out different importance of background, face, and facial features, thus generating weight maps of encoding videos. Upon the weight maps, our scheme is capable of allocating more bits to the face (especially facial features), using a new term bit per weight. Consequently, the visual quality of face and facial features can be improved such that perceptual video coding is achieved for HEVC, as verified by our experimental results.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Supported by the recent advances in related techniques, the popularity of multimedia applications has been considerably increased. It has been pointed out in [1] that video applications with high resolutions, such as FaceTime and Skype, occupy a large proportion of data among the existing multimedia applications. The limited bandwidth issue thus becomes more and more serious, causing "spectrum crush". To better relieve the bandwidth-hungry issue, high efficiency video coding (HEVC) standard [1], also called H.265, has been formally established.

Rate control is a crucial module in HEVC, whose aim is to optimize visual quality via reasonably allocating bits to various frames and blocks, at a given bit-rate. An excellent rate control scheme is able to precisely allocate bits, and to output better visual quality of compressed videos. In other words, at the same visual quality, a better rate control scheme consumes less bit-rate and therefore achieves the goal of relieving the bandwidth bottleneck. There are many rate control schemes for different video coding standards (e.g. TM5 for MPEG-2 [2], VM8 for MPEG-4 [3] and JVT-N046 [4] for H.264). For HEVC, a pixel-wise unified rate quantization (URQ) scheme has been proposed in [5] to compute quantization parameter (QP) at a given target bit-rate. Since this scheme works at pixel level, it can be easily applied to blocks with various sizes. However, according to [6], Lagrange multiplier λ [7], which represents the bit cost of encoding a block, is more important than QP in allocating bits. Therefore, a new scheme, R-λ scheme, was proposed in [6] to better allocate the bits in HEVC.

Nevertheless, high resolution video delivery, especially at low bit-rate scenarios, still poses a great challenge to HEVC. In fact, according to the human visual system (HVS), there exists much perceptual redundancy that can be further exploited to greatly improve the coding efficiency of HEVC,

☆ This work was supported by NSFC under Grant no. 61202139 and 61471022 and China 973 program under Grant no. 2013CB29006.

* Corresponding author.
E-mail address: MaiXu@buaa.edu.cn (M. Xu).

http://dx.doi.org/10.1016/j.image.2015.04.011
0923-5965/© 2015 Elsevier B.V. All rights reserved.

thus relieving the bandwidth-hungry issue [8]. For instance, when a person looks at a video, a small region around a point of fixation, called region-of-interest (ROI), is concerned most [8] with high resolution, while the peripheral region is captured at a low resolution. Hence, in light of this phenomenon, a large amount of bits can be saved via reducing perceptual redundancy in the peripheral region, with little loss of perceived quality. Consequently, along with the development of the understanding of the HVS, perceptual video coding is able to more efficiently condense video data.

Rate control for perceptual video coding has received a great deal of research effort from 2000 onwards, due to its great potential in improving coding efficiency [9–12]. In H.263, a perceptual rate control (PRC) scheme [9] was proposed. In this scheme, a perceptual sensitive weight map of conversational scene (i.e., scene with frontal human faces) is obtained by combining stimulus-driven (i.e., luminance adaptation and texture masking) and cognition-driven (i.e., skin colors) factors together. According to such a map, more bits are allocated to ROIs by reducing QP values in these regions. Afterwards, for H.264/AVC, a novel resource allocation method [10] was proposed to optimize the subjective rate–distortion–complexity performance of conversational video coding, by improving the visual quality of face region extracted by the skin-tone algorithm. Moreover, Xu et al. [13] utilized a novel window model to characterize the relationship between the size of window and variations of picture quality and buffer occupancy, ensuring a better perceptual quality with less quality fluctuation. This model was advanced in [14] with an improved video quality metric for better correlation to the HVS. Most recently, in HEVC the perceptual model of structural similarity (SSIM) has been incorporated for perceptual video coding [15]. Instead of minimizing mean squared error (MSE) and sum of absolute difference (SAD), SSIM is minimized [15] to improve the subjective quality of perceptual video coding in HEVC. However, as pointed out by [16], assigning pixels with weights according to visual attention is much more accurate than SSIM for evaluating the subjective quality. To this end, a scheme [12] was proposed to improve the visual quality and meanwhile to reduce the encoding complexity, via considering the visual attention on ROIs (e.g., face and facial features). However, to our best knowledge, although larger weights are imposed on ROIs in the above approaches, their values are assigned in an arbitrary manner. Moreover, there is no perceptual approach for the latest R-λ rate control scheme [6] in HEVC.

Therefore, we propose a novel weight-based R-λ rate control scheme to improve the perceived visual quality of compressed conversational videos, based on the weights of face regions and facial features learned from eye-tracking data. To be more specific, similar to [12], we consider face regions as ROIs, and also consider facial features (e.g., mouth and eyes) as the most important ROIs. Different from [12], the weights allocated to background, face, and facial features are more precise and reasonable, as they are obtained upon the saliency distribution learnt from our eye-tracking data of several training videos. Based on these weights, the weight-based R-λ rate control scheme is proposed, using a new term, bit per weight (bpw), to enhance the quality of face regions, especially the facial features. Since the perceptual video coding is the main goal of our scheme, we review it in the following.

2. The related work on perceptual video coding

Generally speaking, main parts of perceptual video coding are perceptual models, perceptual model incorporation in video coding and performance evaluations, as illustrated in Fig. 1. Specifically, perceptual models, which imitate the output of the HVS to specify the ROIs and non-ROIs, need to be designed first for perceptual video coding. Secondly, on the basis of the perceptual models and existing video coding standards, perceptual model incorporation in video coding from perceptual aspects needs to be developed to encode/decode the videos, mainly by moving their perceptual redundancy. Rather than incorporating perceptual model in video coding, some machine learning based image/video

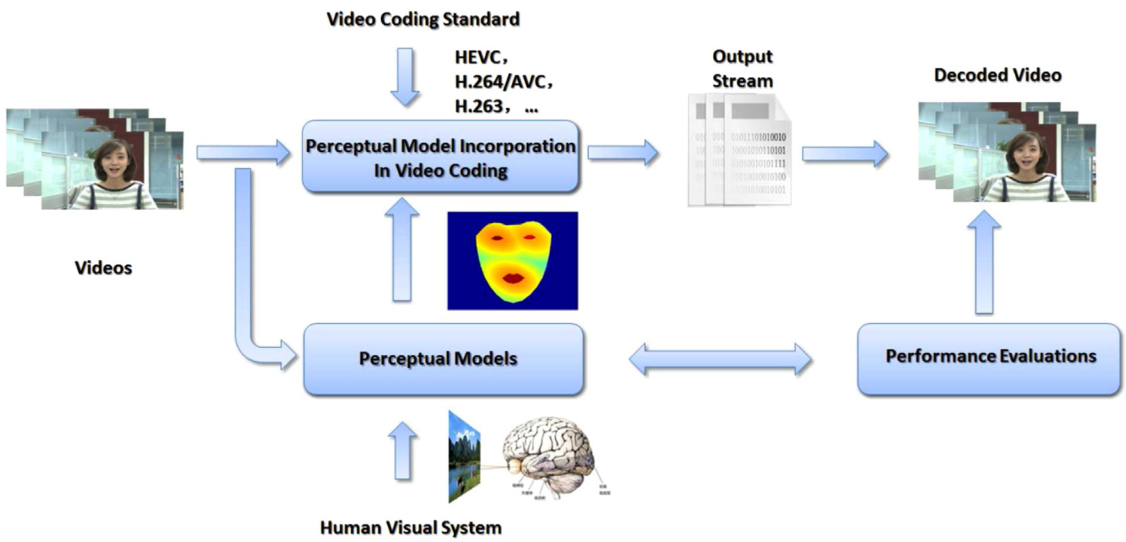

Fig. 1. The framework of perceptual video coding.

Download English Version:

https://daneshyari.com/en/article/538219

Download Persian Version:

https://daneshyari.com/article/538219

Daneshyari.com