J. Vis. Commun. Image R. 25 (2014) 1528–1542

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

An efficient algorithm to eliminate temporal pumping artifact in video coding with hierarchical prediction structure

Yanchao Gong [a,*], Shuai Wan [a], Kaifang Yang [a], Fuzheng Yang [b], Li Cui [a]

[a] School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China
[b] State Key Laboratory of ISN, Xidian University, Xi'an, China

ARTICLE INFO

Article history:
Received 7 January 2014
Accepted 22 June 2014
Available online 30 June 2014

Keywords:
Just noticeable temporal pumping artifact
Spatial–temporal masking effect
Hierarchical prediction structure
Hierarchical B-pictures structure
Quantization parameter cascading
Temporal noise
H.264/AVC
Video coding

ABSTRACT

At low bit-rates, hierarchical prediction structure using quantization parameter cascading strategy will introduce a new temporal distortion, i.e., the temporal pumping artifact (TPA) presenting itself as a stumbling effect which seriously affect the subjective video reconstructive quality. This paper analyzes one important key factor which greatly influence the critical perceptual of the TPA and proposes a metric for just noticeable temporal pumping artifact (JNTPA) based on spatial–temporal masking effects in human visual system (HVS). Finally, an efficient TPA eliminating algorithm in video coding based on the JNTPA has been proposed. The subsequent experiments demonstrate that the estimated JNTPA values using the proposed metric are in line with the HVS perception and comparable with JVT-P104, the proposed TPA eliminating algorithm achieves a more smooth subjective video quality and improves the observe comfort of the human eye. At the same bit-rate, better subjective video quality can be observed using the eliminating algorithm.

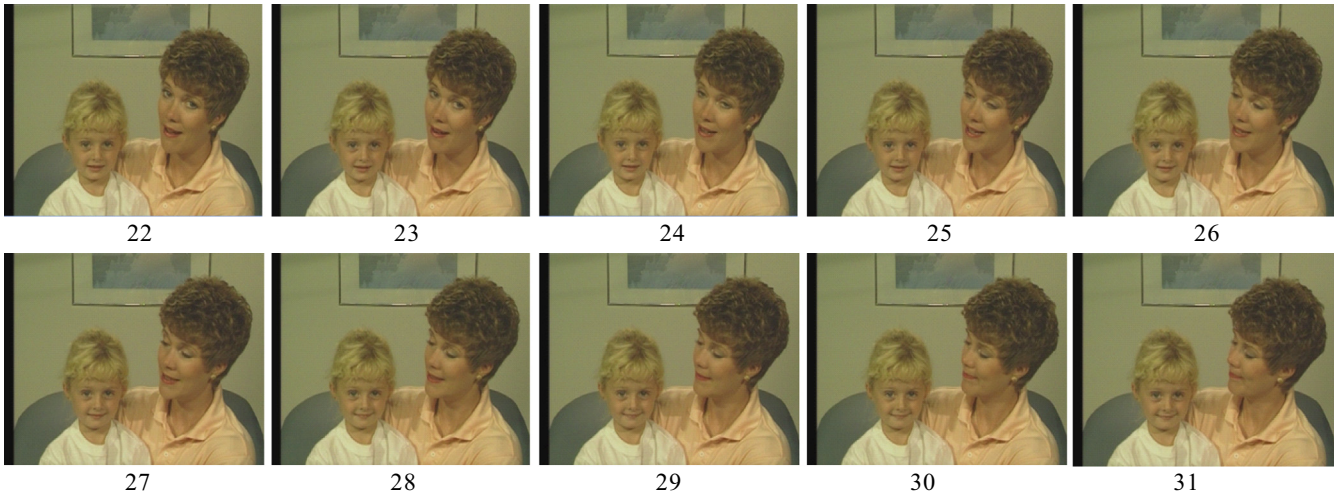
1. Introduction

Temporal noise or temporal distortion, which seriously affects the subjective quality of reconstructed videos, is generally defined as the temporal evolution or fluctuations of the spatial distortions in a video [1]. Commonly experienced temporal noises include flickering, jerkiness, and mosquito noise, as well as the temporal pumping artifact (TPA), which is a special type of temporal noise presenting itself as a stumbling effect [2].

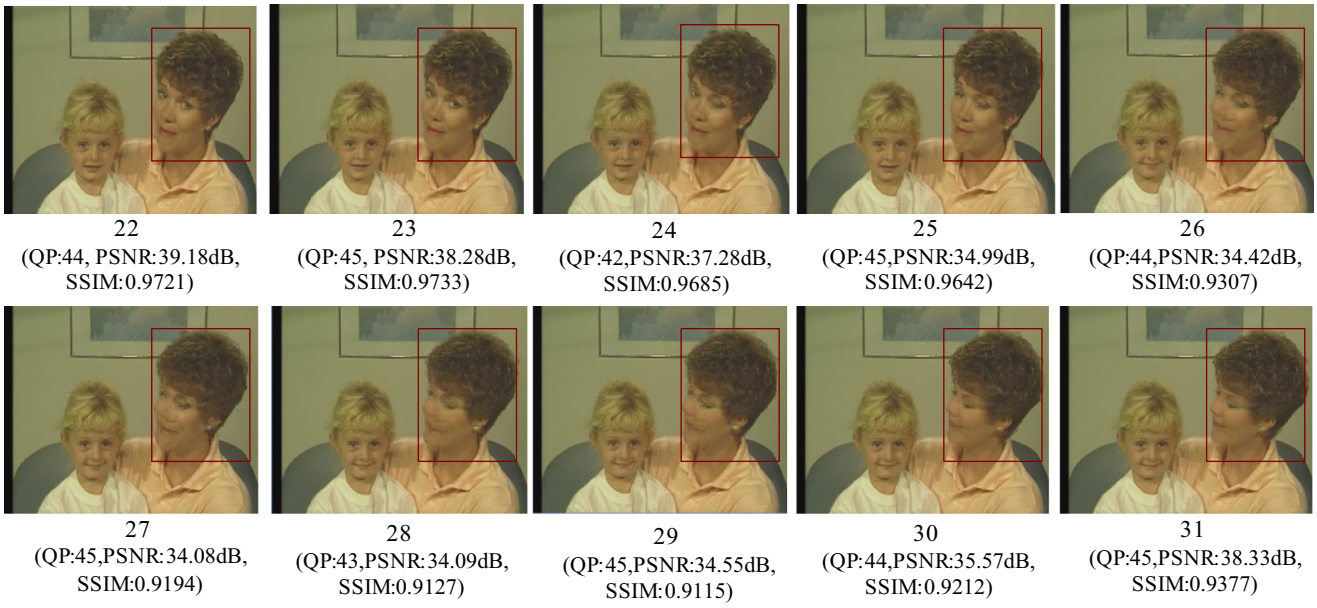
To understand how this special temporal artifact is formed, we need to deep into the hierarchical prediction structure (HPS). The HPS is recently employed in video coding to make inter-frame prediction more efficient since it can greatly improve the rate-distortion performance of video coding. The HPS is therefore included in H.264/AVC, and in the latest high efficiency video coding (HEVC) as the random-access configuration or low-delay configuration. When using the HPS, B frames or pictures are organized into temporal layers and pictures in different layers are with different importance in terms of prediction. The coding efficiency for the HPS is related to how the quantization parameter (QP) is chosen for pictures in different temporal layers [2]. To address this, it has been proposed by JVT-P104 [3] for H.264/AVC,/SVC and JCTVC-L1002 [4] for HEVC that the QP should be assigned with an increasing order according to the temporal layers, known as the QP cascading (QPC) technique. These techniques are efficient in coding and have been widely acknowledged. However, the QPC strategy may result in severe quality fluctuations among adjacent pictures inside a group of pictures (GOP) [2,3,5]. It has been noted that the related objective quality fluctuation leads to a perceptually stumbling effect, named as the TPA [2]. A general observation from subjective experiments shows that small quality fluctuations among frames may not be noticed sometimes, whereas the annoying pumping artifact is bound to be perceived when the quality fluctuation is large enough to reach a certain threshold. In Fig. 1, "Mother and Daughter" was used as a straightforward example to show what the TPA is. Fig. 1(a) presents the original frames and Fig. 1(b) shows the reconstructed frames. Here the objective quality matric peak signal-to-noise ratio (PSNR) and objective visual quality metric structural similarity (SSIM) [6] were used to indicate the quality of reconstructed frames. In Fig. 1(b), an obvious texture variation of "clear–blur–clear" can be observed in the region of the mother's head. When viewing the reconstructed videos, most subjects can perceive TPA in this region. As an identified temporal artifact, the TPA is caused by applying inappropriate quantization parameter selection strategies in the encoding process, which leads to severe quality fluctuations from frame to frame among adjacent pictures.

* Corresponding author.
E-mail address: ycgong@mail.nwpu.edu.cn (Y. Gong).

http://dx.doi.org/10.1016/j.jvcir.2014.06.015
1047-3203/



(a) The original frames



(b) The reconstructed frames which were coded using HPS and GOP=16

Fig. 1. An example of TPA.

Many researchers have analyzed temporal noises regarding to flickering [7–10], jerkiness [11–13], and mosquito noise [14–16]. However, the TPA has not gained enough attention by researchers, since coding efficiency is usually employed to evaluate the performance of a QP cascading strategy in HPS without considering its temporal visual effect [17,18]. In our previous paper [19], we have analyzed the fundamental reasons of TPA perception. Key factors influencing the TPA perception have been evaluated, including the quality of key frames, the GOP size and the content characteristics of different regions in the video.

The main work in [19] is a brief summary of the relevant factors influencing the perception of the TPA. In this paper, the TPA that just starts to be noticed is defined as just noticeable temporal pumping artifact (JNTPA). Based on the findings in [19], this paper firstly proposes a metric for the JNTPA, quantitatively determines the quantization step which causes the JNTPA when videos are encoded using the HPS. And based on the accurate prediction of the JNTPA, an efficient TPA eliminating algorithm is proposed which uses more appropriate quantization parameter selection strategies to achieve a more smooth subjective video quality in video encoding.

The rest of this paper is organized as follows. The fundamental reasons and key factors influencing the perception of the TPA are briefly summarized in Section 2. A metric for JNTPA is established in Section 3. An efficient TPA eliminating algorithm is proposed in Section 4. Performance evaluation is provided in Section 5. This paper closes with conclusions given in Section 6.

2. Fundamental reasons and key influence factors of perception of TPA

The dyadic hierarchical B-pictures structure (HBS) is a typical implementation of the HPS, which well applies in H.264/AVC as well as HEVC (i.e., the random access structure in HEVC). In this paper, the dyadic HBS as shown in Fig. 2 is addressed for evaluation. The other HPS of low delay structure in HEVC is currently under research. The HBS is usually implemented by organizing B