J. Vis. Commun. Image R. 38 (2016) 246–255

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Binocular perception based reduced-reference stereo video quality assessment method ☆

Mei Yu [a,b], Kaihui Zheng [a], Gangyi Jiang [a,b,*], Feng Shao [a], Zongju Peng [a]

[a] Faculty of Information Science and Engineering, Ningbo University, Ningbo, China
[b] National Key Lab of Software New Technology, Nanjing University, Nanjing, China

ARTICLE INFO

Article history:
Received 4 September 2015
Revised 3 January 2016
Accepted 7 March 2016
Available online 9 March 2016

Keywords:
Stereo video quality assessment
Binocular vision
Temporal characteristics
Reduced-reference frame
Motion intensity

ABSTRACT

A new reduced-reference (RR) stereo video quality assessment method is proposed in this paper by considering temporal characteristics of video and binocular perception in human visual system (HVS). Firstly, motion intensity is utilized to extract RR frames for the purpose of temporal characteristics in stereo video. Secondly, according to internal generative mechanism of HVS, fusion and rivalry in the process of binocular perception is modeled, and the RR frames are divided into binocular fusion portion and binocular rivalry portion. Then, RR frame quality indicators are computed for these two portions. Finally, the RR frame quality indicators of the original and distorted frames are compared. A temporal pooling strategy is utilized on these quality indicators to obtain final stereo video quality score, where the motion intensity is used for toning the pooling parameters. Experimental results show that the proposed method has better performances when compared to other state-of-the-art quality assessment methods.

1. Introduction

Three dimensional (3D) video becomes popular in digital multimedia in last decade [1,2]. These 3D contents can bring users stereo perception and immersive viewing experiences, and they have been displayed not only in theaters, but also on television, portable or other kinds of consumer electronic devices. Thus, for the purpose of higher quality of 3D video content generation and processing, stereo video quality assessment (SVQA) becomes more and more important [3–5].

In this new paradigm of image quality assessment, some works have been done in different approaches. For example, full-reference (FR) methods in classical manner with complete reference information from the original content were proposed [6,7]. However, the disadvantage of FR methods is that the reference images/videos are not usually accessible in most practical situations. As a solution, no-reference (NR) methods were proposed without the original reference [8]. But many existing NR methods may be only effective for some specific image/video contents or specific types of distortions. Reduced-reference (RR) method is an alternative of the above two kinds of methods, depending on partial information of the original contents [9]. Compared with the FR method, the RR method greatly reduces the reference data to be transmitted; and taking advantage of the reference data, the RR method ensures the evaluation effectively and universally in comparison with the NR method.

Different from viewing single viewpoint images/videos, human visual system (HVS) can perceive the difference between two retinal images to create a mental image with depth perception, which is the result of two binocular interactions, i.e., binocular fusion and suppression. Thus the perceptual quality of stereo image/video with two views is also different from the one obtained by using simple weighting of the two views' two dimensional (2D) quality in stereo image/video. The additional dimension, depth information which is generated from the left and right views in stereo image/video, needs to be carefully considered. Some SVQA methods were proposed through stereo image quality assessment (SIQA) models. Seo et al. proposed a FR objective SVQA metric by using blocking artifacts, blurring in edge regions and video quality difference between two views [10]. The first two terms describe the distortion in compressed video, while the third is used to represent 3D effect of stereoscopic video. Jin et al. proposed a FR SIQA model for mobile 3D video [11], which presents three quality components to measure the image quality, including the cyclopean view, binocular rivalry, and the scene geometry, respectively; the final 3D image quality is assessed through a machine learning

☆ This paper has been recommended for acceptance by Wiesi Lin.

* Corresponding author at: Faculty of Information Science and Engineering, Ningbo University, Ningbo, China.

E-mail address: jianggangyi@126.com (G. Jiang).

http://dx.doi.org/10.1016/j.jvcir.2016.03.010
1047-3203/

approach. Hewage et al. proposed a RR 3D image quality metric by extracting edges and contours of depth map [12]. Malekmohamadi et al. proposed a RR quality metric for 3D video by extracting side information from edge properties and gray level co-occurrence matrices from color and depth sections [13]. Taking into account the spatial information in video quality assessment (VQA), Ma et al. proposed a RR VQA model [14], which not only considers spatial statistical characteristics of video, but also temporal statistical characteristics. Soundararajan et al. proposed RR VQA models by utilizing spatial and temporal entropic differences [9], in which a Gaussian scale mixture model is used to measure the amount of spatial and temporal information differences. These RR models may improve the evaluation accuracy by utilizing temporal information.

In this paper, considering demands of practical applications and human visual perception on spatial and temporal characteristics of stereo video, a new RR-SVQA method is proposed, which consists of three parts. Firstly, according to the analysis of temporal characteristics of video, motion intensity is defined and used to extract the RR frame pairs from stereo video so as to reduce the amount of data to be processed in quality assessment while keep the temporal characteristic of video, and then these RR frames are decomposed into binocular fusion portion (BFP) and binocular rivalry portion (BRP). Secondly, the BFP is used to construct the cyclopean view, and then the generalized Gaussian distribution (GGD) features are extracted from the cyclopean view and the BRP. Finally, the quality indicators of BFP and BRP of stereo video are obtained. These quality indicators are pooled in spatial and temporal domains to get the final stereo video quality score. The proposed method is tested on the NAMA3DS1-COSPAD1 stereo video database, and the corresponding experimental results show that the proposed method has more excellent performance and better consistency with human visual perception compared with other start-of-art methods.

The remainder of this paper is organized as follows. In Section 2, the proposed RR SVQA method is described in detail. In Section 3, the experimental results are given and the performance indices are discussed. Finally, Section 4 concludes the paper.

2. The proposed RR-SVQA method

Considering human visual perception of temporal variability and binocular perception characteristics of stereo video, a new RR SVQA method is proposed, and its framework is shown as Fig. 1. The proposed method mainly consists of three parts, including (1) selection of RR frame for reflecting temporal variability of video, (2) extracting visual perception RR features in BFP and BRP of stereo video, and finally (3) quality indicator calculation and pooling for stereo video.

Since there are strong correlations among successive frames of stereo video, the RR frames are defined and selected from the original and distorted stereo videos to represent the videos, so as to decrease the amount of data to be processed in the following quality assessment while describe the temporal variability of the videos. In addition, according to binocular perception characteristics, the RR frames of stereo view are decomposed into two portions, that is, BFP and BRP, which resulted from binocular fusion and binocular rivalry of binocular vision. Then, multi-channel decomposition is applied on these portions and the GGD model [15] is used to normalize the coefficients to gain the RR features. Then, at the sender, by compression coding, some RR features in original video are sent to the receiver as the side information of H.264 standard compatible video stream. Finally, at the receiver, the quality of these two portions is computed and pooled as the final stereo video quality score.

2.1. Extraction of the RR frame

For a video, when a series of high correlation consecutive images are played continuously, if the correlation coefficient between the images is quite close or equal to 1, it can be understood as almost still image. But if the correlation between the frames before and after the current frame is weak, perceptual difference between the frames will be large, which implies that the played video is significantly varying. Thus, inter-frame difference can be used to represent the motion intensity of a video. The inter-frame motion is an important feature to describe temporal variability of video. In SVQA research, the existing methods pay little attention to temporal characteristics of video. However, it has been known that the temporal, spatial, and quantization variations will impact the perceived quality of video [16]. When the motion intensity in video is not strong, video's spatial features show higher significance. With the movement being stronger, the temporal features will play a more important role in the quality of video. Therefore, we put forward the following scheme for extracting the RR frame.

2.1.1. Motion intensity of video

As one of the most important temporal characteristics of video, motion intensity of video will be defined to describe temporal variability of video and considered in designing a new RR SVQA method. For videos with different motion intensity, the redundancy between the frames before and after the current frame is also different. If motion in video is slight, there are more temporal redundancies in the video; on the contrary, there are less temporal redundancies. Hence, the different RR frame selection strategy should be executed on video with different motion intensity. Considering HVS being sensitive to motion intensity, motion intensity

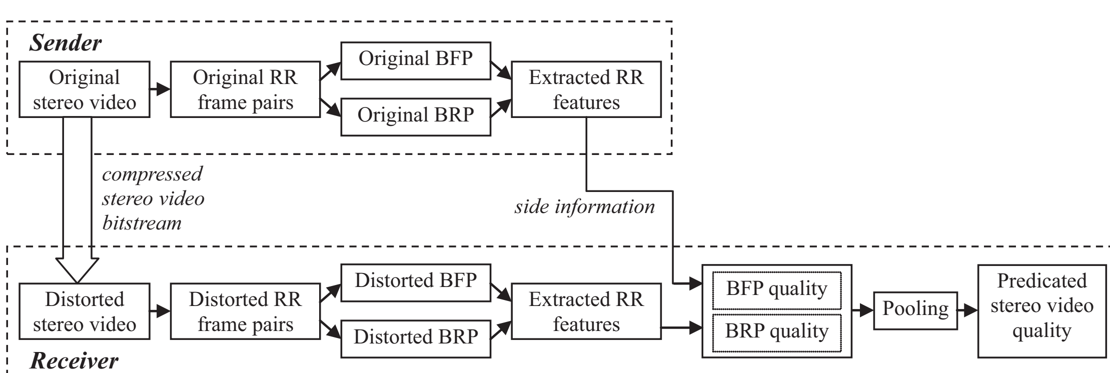

Fig. 1. Framework of the proposed RR SVQA method.