J. Vis. Commun. Image R. 31 (2015) 138–145

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

A perceptual stereoscopic image quality assessment model accounting for binocular combination behavior ☆

Jiachen Yang [a,1], Yun Liu [a,*,2], Zhiqun Gao [a], Rongrong Chu [a], Zhanjie Song [b]

[a] School of Electronic Information Engineering, Tianjin University, China
[b] School of Science, Tianjin University, China

ARTICLE INFO

Article history:
Received 29 December 2014
Accepted 2 June 2015
Available online 20 June 2015

Keywords:
Binocular vision
Image quality
Local amplitude
Gain control
Stereoscopic quality assessment
Difference channel
Summation channel
Symmetric distortion

ABSTRACT

Stereoscopic image quality assessment (SIQA) plays an important role in the development of 3D image processing. In this paper, a full-reference object SIQA model is built based on binocular summation channel and binocular difference channel. In our frame work, binocular combination behavior and how to experience the depth perception are thought to be the key factors to evaluate the quality of stereoscopic images. Differing from the current depth map methods, this method focuses on a new aspect, and an effective combination model is proposed based on the physiological findings in the Human Visual System (HVS). Experimental results demonstrate that the proposed quality assessment metric significantly outperforms the existing metrics and can achieve higher consistency with subject quality assessment when predicting the quality of stereoscopic images that have been symmetrically distorted.

1. Introduction

With the development of 3D technologies, more and more 3D contents for 3DTV and 3D cinema are produced. Being able to provide a high-quality 3D image and deliver to the consumer is both challenge and demanding. Differ from traditional 2D image quality assessment (2D-IQA) [1–3], many factors, such as depth perception, views' quality, and visual fatigue, could affect 3D perception. How to evaluate the quality of stereoscopic image quality draw much attention.

Many efforts have already been done to develop stereoscopic image quality assessment method (SIQA) over the last decade which can be categorized into two types: the subjective and objective models. Since subjective SIQA is time-consuming and impractical for online applications, the objective SIQA has been a fruitful area of work. In the past years, many researchers directly applied 2D-IQA metrics to measure 3D perception [4,5]. However, these methods based on 2D-IQA performed poorly in predicting the quality of stereoscopic images, since these metric were failed in considering strong correlation with standard disparity from two adjacent viewpoints. It shown that the perceptual quality of stereoscopic images cannot be predicted by simply averaging the quality of the left and right image. Hence, it is very necessary to develop an effective method to evaluate the objective quality of stereoscopic images.

Recently, a considerable amount of research have been conducted on SIQA [6,7]. Various factors that may affect the degree of visual comfort experience have been found and studied [8,9]. Considering that the depth information is the big difference between 2D image and stereoscopic images, many SIQA models have been proposed. For example, Benoit et al. [10] presented a linear combination of the disparity map distortion and the measure of the 2D image quality on both eyes. Lambooij et al. [11] proposed a 3D quality model by considering 2D image quality and perceived depth. Some other works have been conducted based on human visual perception. Kim et al. [12] proposed a visual fatigue prediction metric which can replace subjective evaluation for stereoscopic images. Chen et al. [13] explored a SQIA model by considering the image quality, depth quality and visual comfort. However, the above models needed to assess the depth quality using estimated disparity maps, and the ground truth disparity or depth is generally not available. The performance of the related SIQA model would be substantially affected by the accuracy of the used disparity estimation algorithm.

Later, many SIQA models based on HVS to stimulate binocular characteristics were proposed. Shao et al. [14] proposed a perceptual full-reference SIQA model by considering the binocular visual characteristics. Chen et al. [15] addressed binocular rivalry issues

☆ This paper has been recommended for acceptance by M.T. Sun.
* Corresponding author.
E-mail address: yunliu@tju.edu.cn (Y. Liu).
[1] In 2014, he is a visiting scholar in the Department of Computer Science, Loughborough University, United Kingdom.
[2] Now she is a visiting Ph.D. student in University of California at Berkeley.

http://dx.doi.org/10.1016/j.jvcir.2015.06.002
1047-3203/

by modeling the binocular suppression behaviors and developed a framework for assessing the quality of 3D images. Lin et al. [16] integrated the binocular integration behaviors, the binocular combination and the binocular frequency integration, into the existing 2D objective metrics for evaluating the quality of 3D images. These models were proposed by building a possible model of human binocular processing procedure when human viewing 3D content. The overall performance of these metrics indeed outperforms that of the former metrics in predicting the quality of stereoscopic images across different types of distortions. Remarkable progresses have been obtained in both theoretical and practical understanding about binocular perception of stereoscopic image [17,18]. What is more, with the development of human 3D visual system [19–21], more and more 3D visual quality assessment methods were proposed [22,23]. Tsai et al. [24] and Bosc et al. [25] proposed the quality assessment models of 3D synthesized views in the context of multi-view video, which enhances the correlation of the objective quality score to the 3D subjective scores. Later, Jiang et al. [26] proposed a 3D visual attention model for stereoscopic image quality assessment task, and the proposed 3D visual attention-based pooling scheme can achieve higher consistency with the subjective assessment of stereoscopic images. Battisi et al. [27] built an effective objective quality assessment model to evaluate the visual quality of DIBR-synthesized views. Park et al. [28] then developed an effective 3D visual discomfort predictor to predict the level of visual discomfort.

In this paper, a full reference quality assessment metric is proposed by applying properties of human binocular visual system. We take steps based on the biologically plausible visual processing which contains two channels: difference and summation channels. The principal advantage of this model is that we apply the absolute disparity to directly reflect human stereo sense instead of disparity map, which greatly reduce the algorithm complexity. Since the summation channel signal intrinsically reflect the image quality of 3D images, the proposed method of combing difference channel and summation channel (SDM) attains more accurate quality assessment.

The rest of this paper is organized as follows. Section 2 introduces some binocular combination models. The framework of the proposed metric is presented in Section 3. The experimental results are analyzed in Section 4, and finally conclusions are drawn in Section 5.

2. Binocular combination behavior

Human binocular vision is a complex visual process, and one of the most amazing properties of human binocular characteristics is the fusion of the left and right views of a scene into a 3D image. A large number of researches have been done on how two slightly different monocular images fuse to a "combined image" and generate depth perception.

2.1. Monocular channel

Traditionally, it is thought that the depth perception was achieved by simply combining the monocular information [29]. In 1968, Levelt [30] proposed a linear model to explain the perceived combination image when a stereoscopic image is present, as follows:

$$C = W_l \times E_l + W_r \times E_r \tag{1}$$

where C is the simulated cyclopean combined image, E_l and E_r are the signals from the left and right eyes, respectively. W_l and W_r are the weighting coefficients for the two eyes, respectively, where $W_l + W_r = 1$. Cogan et al. [31] then proposed a gain control model to stimulate human binocular visual experience, as shown in Fig. 1. They stated that each eye would exerts gain control on the other eye, not just simply added together, the model is given by:

$$C = \frac{1}{1+G_r}E_l + \frac{1}{1+G_l}E_r \tag{2}$$

where G_l and G_r are the visually weight for gain control of the left and right images, respectively.

With the further development of vision model, numerous biological models were proposed to describe binocular combination (e.g. vector summation model shown in Eq. (3), and neural network model shown in Eq. (4)). The detailed description and comparisons can be found in the study of [32].

$$C = N_l + N_r \tag{3}$$

where N_l and N_r are the responses of neural cells receiving strong excitation to one eye and weak inhibition from the other eye, respectively. As in work [32], we can apply E_l and E_r to calculate N_l and N_r, since E_l and E_r are mathematically equivalent ($i = l, r$).

$$C = \left(\frac{E_l}{E_l+E_r}\right)E_l + \left(\frac{E_r}{E_l+E_r}\right)E_r \tag{4}$$

2.2. Binocular summation and difference channels

An alternative view, however, seems more plausible. It proposed that there exist "Summation" and "Difference" channels used for stereopsis. Indeed, Silva et al. [33] demonstrated such mechanisms exist and accounted for binocular brightness perception. What is more, Li and Atick [34] presented the existence of a visual pathway of binocular visual system (shown in Fig. 2). They pointed that the signals from the left and right eyes will be firstly transformed into two uncorrelated signals: Summation (S) and Difference signals D, and then gains, G_+ and G_-, applied to the summation and difference channels dynamically optimized to the prevailing interocular correlation, respectively. Recently, Kingdom et al. [35], in detail, explained the theory procedure of summation and difference channel and how these two channels added and subtracted the neural signals from our two eyes.

Given the left image L and right image R, the binocular S and D signals can be calculated based on Eq. (5), shown as follows:

$$\begin{cases} D = |L - R| \\ S = L + R \end{cases} \tag{5}$$

Here, we present an example, in Fig. 3, to show the human binocular summation and difference signal experience. The first row shows the left images of the reference stereoscopic images and WN distorted stereoscopic images, respectively. The second row is

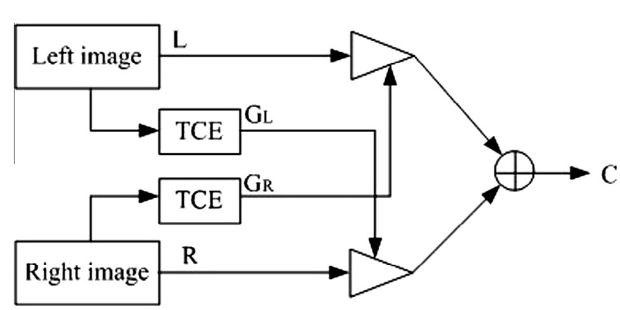

Fig. 1. The gain control model: Each eye exerts gain control on the other eye in proportion to its own total visually weighted contrast energy (TCE). Within a spatial-frequency-and-orientation channel, the input from each eye is divided by a gain-controlling signal from the other eye (1 + TCE) and the two dividends are summed linearly [31].