Signal Processing: Image Communication 58 (2017) 287–299

Contents lists available at ScienceDirect

Signal Processing: Image Communication

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

Binocular spatial activity and reverse saliency driven no-reference stereopair quality assessment

Lixiong Liu [a], Bao Liu [a], Che-Chun Su [b], Hua Huang [a,*], Alan Conrad Bovik [b]

[a] Beijing Laboratory of Intelligent Information Technology, School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China
[b] Laboratory for Image and Video Engineering, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712, USA

ARTICLE INFO

Keywords:
Stereopair quality assessment
No-reference
Binocular rivalry
Spatial activity
Reverse saliency

ABSTRACT

We develop a new model for no-reference 3D stereopair quality assessment that considers the impact of binocular fusion, rivalry, suppression, and a reverse saliency effect on the perception of distortion. The resulting framework, dubbed the S3D INtegrated Quality (SINQ) Predictor, first fuses the left and right views of a stereopair into a single synthesized cyclopean image using a novel modification of an existing binocular perceptual model. Specifically, the left and right views of a stereopair are fused using a measure of "cyclopean" spatial activity. A simple product estimate is also calculated as the correlation between left and right disparity-corrected corresponding binocular pixels. Univariate and bivariate statistical features are extracted from the four available image sources: the left view, the right view, the synthesized "cyclopean" spatial activity image, and the binocular product image. Based on recent evidence regarding the placement of 3D fixation by subjects viewing stereoscopic 3D (S3D) content, we also deploy a reverse saliency weighting on the normalized "cyclopean" spatial activity image. Both one- and two-stage frameworks are then used to map the feature vectors to predicted quality scores. SINQ is thoroughly evaluated on the LIVE 3D image quality database (Phase I and Phase II). The experimental results show that SINQ delivers better performance than state of the art 2D and 3D quality assessment methods on six public databases, especially on asymmetric distortions.

1. Introduction

In the past two decades there has been tremendous progress on both the theoretical and practical aspects of visual quality assessment (QA) [1]. The problem of 2D picture quality prediction has been intensively studied [2–7]; however, the field of 3D perceptual quality assessment is relatively nascent [8]. Although 3D capture and display technologies have greatly advanced and become widely affordable, progress still needs to be made on developing better tools that can effectively evaluate, maintain and improve the quality of 3D visual signals [8]. With the rapid growth of available 3D content, the need for accurate and efficient methods for assessing the quality of 3D visual signals is increasing.

Here we consider the problem of No-Reference Stereoscopic 3D (S3D) picture quality prediction, hence no reference signal is assumed present. Several studies have been presented to address the problem of NR 3D visual quality assessment. Akhter et al. proposed a NR 3D QA algorithm which extracts features from stereopairs and an estimated disparity map [9]. Ryu et al. discussed the relationship between stereo quality and distortion types [10] and Chen et al. proposed an NR method that extracts features from synthesized "cyclopean" images, computed disparity maps and uncertainty maps [11]. Su et al. modeled stereopairs with bivariate statistical models [12,13] and proposed a 3D blind image naturalness quality index based on both univariate and bivariate natural scene statistics (NSS) models [14]. Su et al. proposed an improved cyclopean image synthesis model, and extracted perceptually relevant NSS features using bivariate statistical models from spatially adjacent bandpass coefficients of the synthesized cyclopean images [14]. Natural scene statistics features measured on 3D visual coordinates propel the NR 3D IQA model in [15]. Lopez et al. [16] proposed an S3D video quality assessment method based on depth maps and video motion. While these early advances on the NR S3D IQA problem are promising, further improvements are possible by better accounting for human stereo perception mechanisms. Algorithms for conducting S3D IQA have been developed by modeling a variety of binocular vision characteristics, e.g., binocular energy responses [17],

* Corresponding author.
E-mail addresses: lxliu@bit.edu.cn (L. Liu), liubao@bit.edu.cn (B. Liu), ccsu@utexas.edu (C.-C. Su), huahuang@bit.edu.cn (H. Huang), bovik@ece.utexas.edu (A.C. Bovik).

http://dx.doi.org/10.1016/j.image.2017.08.011
Received 17 December 2016; Received in revised form 19 August 2017; Accepted 20 August 2017
Available online 30 August 2017
0923-5965/

gain-control mechanisms [18], models of visual information extraction [10] and of Gabor filter responses [19]. However, there is still room for improvement beyond available S3D IQA models with regards to correlation with human subjective judgments. Here we develop a way to account for both binocular perception and the differences in low-level content between the two views in a new NR S3D picture quality assessment model. The proposed model and resulting algorithm contains three main innovations. First, we use a simple measure of spatial activity to model the binocular perception process by calculating the differences between the two views arising from distortion, and use it to modify an existing method of synthesizing a cyclopean image. Second, based on recent evidence regarding the placement of 3D fixations by humans viewing S3D content, we deploy a model of visual reverse saliency to optimize the process of feature extraction. Lastly, we compute a novel disparity-corrected correlation image obtained by computing the products of corresponding pixels between the two stereopair images. The experimental results demonstrate the effectiveness of the proposed method in improving the prediction of the perceived quality of S3D picture content.

The rest of the paper is organized as follows. Section 2 reviews related S3D image quality assessment models. Section 3 describes the binocular perception model we use. We introduce the concept of the S3D product image or between-view empirical correlation in Section 4. The various features are integrated into a new NR stereopair quality assessment model, called SINQ. We evaluate the performance of SINQ against the state-of-the-art in Section 5 and conclude the paper in Section 6.

2. Related work

S3D image quality prediction models can be classified according to the types of 2D and 3D information that is computed from the stereoscopic pairs. The simplest 3D IQA algorithms do not incorporate depth/disparity information, instead using features drawn directly from the left and right view images. Yasakethu et al. [20] used various 2D IQA models to predict the quality of left and right view images independently, then synthesized both scores to obtain predictions of S3D quality. Most 2D-based 3D IQA algorithms belong to this category.

The deployment of models of stereoscopic perception and disparity computations can be used to improve on 2D-based models. You et al. applied a variety of 2D IQA algorithms on stereopairs as well as on disparity maps to predict perceptual quality [21]. Sazzad et al. [22] took disparity information into account to assess the quality of symmetrical and asymmetrical JPEG compression stereopairs. Zhao et al. proposed a binocular JND (BJND) model, which uses basic properties of binocular vision in response to asymmetric noise [23]. Lopez et al. utilized depth maps and video motion to measure the quality of a 3D video [16]. The authors of [24] proposed a theory that binocular perception is not uniformly distributed and can be classified into different types of perception patterns. Based on this theory, Shao et al. classified S3D images into non-corresponding, binocular fusion, and binocular suppression regions, and evaluated the quality of each region independently [24]. Lin et al. [25] assessed the quality of compression distortion considering binocular integration, and Silva et al. [19] proposed a novel video quality metric based on binocular suppression theory.

In addition to using depth/disparity information, accounting for the fused perception of a given S3D scene by computing an estimated cyclopean image is a useful way to improve S3D quality predictions. A growing number of algorithms have utilized a cyclopean image to construct better S3D IQA models. Maalouf et al. [26] utilized disparity information to compute a cyclopean image by averaging the left view and compensated right view. In [18], a new 3D image quality model specifically tailored for mobile 3D video was proposed, which considers a cyclopean view, binocular rivalry, and the scene geometry. More recently, Chen et al. developed a framework for assessing the quality of both symmetrically and asymmetrically distorted stereopairs [27]. In their framework, given the left and right view images, an estimated disparity map was computed using a stereo algorithm, while the perceptual responses are generated on the stereo images using a Gabor filter bank. Finally, a "cyclopean image" was synthesized from the stereo image pair, the estimated disparity map and the Gabor filter responses. Chen et al. further developed this framework by proposing an NR S3D IQA algorithm in [11]. Su et al. improved the cyclopean image synthesis model and proposed a 3D blind image naturalness quality index [14], which uses a parallel-viewing geometry to generate a convergent cyclopean image from the left and right views images, then extracts perceptually relevant NSS features from spatially adjacent bandpass coefficients of the cyclopean images. Yang et al. [28] proposed a saliency map to define an S3D model that processes a "cyclopean image". A "cyclopean image" synthesis framework is also adopted here in the proposed SINQ model, but with a different fusion strategy based on a measure of relative spatial activity.

3. Binocular spatial activity model

In order to analyze the mechanisms that drive binocular vision in an accessible way, we thoroughly consider the phenomena of binocular fusion, rivalry and suppression. The relationship between the amount of spatial activity, which is a measure of normalized image variance, contained in an S3D image pair [10] and perceived image quality is also studied. Levelt [29] demonstrated that low-level sensory factors can strongly influence binocular rivalry and suppression. He pointed out that strong stimuli, such as areas of high contrast or sharp edges tend to dominate rivalry between the two views. Following this idea, we construct a cyclopean fusion model based on the hypothesis that an image in a (grayscale) stereopair containing the greater amount of spatial activity will dominate perception.

3.1. *Binocular fusion, rivalry and suppression*

The human visual system (HVS) receives visual stimuli from the two eyes and combines them into a single combined percept [30]. When the two views share similar characteristics, binocular fusion can happen easily, driven by processes of vergence, accommodation, and other visual adaptations when viewing the real, natural world. However, binocular rivalry can occur when unusual dissimilar (dichoptic) monocular stimuli are presented to the corresponding retinal locations of the two eyes. Then, rather than perceiving a single, stable fused image, one either experiences alternations in perceptual awareness over time as the two stimuli compete for perceptual dominance, or diplopia, which is what usually occurs when viewing stereopairs containing asymmetric distortions. If one of the stimuli dominates the other, then binocular suppression will occur. A related challenging issue is how to model the process of binocular combination, when evaluating asymmetrically distorted stimuli [30]. Since the binocular perception of degrees or characteristics of distortion can greatly influence the perceived quality of S3D images [31], determining how to synthesize the two views to simulate the perception of a 3D scene is a key ingredient of a successful 3D images quality assessment model.

Normal viewing of stereo images in daily life usually results in binocular fusion rather than binocular rivalry, since the stereopairs share similar content. However, when viewing electronically displayed S3D content, the introduction of distortions may destroy this similarity, leading to binocular rivalry. This rivalry can modify the perception of the distortions [11,14,27,32], and can even lead to feelings of physiological discomfort [33]. Clearly, binocular rivalry is an important factor in S3D quality perception [18,19,24,25,27].