Int. J. Electron. Commun. (AEÜ) 94 (2018) 303–310

Contents lists available at ScienceDirect

Int. J. Electron. Commun. (AEÜ)

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

Regular paper

Toward an unsupervised blind stereoscopic 3D image quality assessment using joint spatial and frequency representations

Meixin Fang[a], Wujie Zhou[a,b,*]

[a] Zhejiang University, Hangzhou 310027, China

[b] School of Information and Electronic Engineering, Zhejiang University of Science & Technology, Hangzhou 310023, China

ARTICLE INFO

Keywords:
Stereoscopic 3D (S3D) image
Binocular visual mechanism
Unsupervised
Quality assessment

ABSTRACT

Existing blind stereoscopic 3D (S3D) image quality assessment (IQA) metrics usually require supervised learning methods to predict S3D image quality, which limits their applicability in practice. In this paper, we propose an unsupervised blind S3D IQA metric that utilizes the joint spatial and frequency representations of visual perception. The metric proposed in this work was inspired by the binocular visual mechanism; furthermore, it is unsupervised and does not require subject-rated samples for training. To be more specific, first, the various binocular quality-aware features in spatial and frequency domains are extracted from the monocular and cyclopean views of natural S3D image patches. Subsequently, these features are utilized to establish a pristine multivariate Gaussian (MVG) model to characterize natural S3D image regularities. Finally, with the learned MVG model, the final quality score for a distorted S3D image can be yielded using a Bhattacharyya-like distance. Our experimental results illustrate that, compared to related existing metrics, the devised metric achieves competitive prediction performance.

1. Introduction

Stereoscopic 3D (S3D) image processing techniques such as S3D scene capture, S3D compression, S3D transmission, S3D rendering, S3D vision enhancement, and S3D display are the focus of current research efforts [1–3]. Each processing stage is highly susceptible to being contaminated with various distortion artifacts. Therefore, approaches to precisely and faithfully predict the perceptual quality of distorted S3D content are in urgent demand and essential for S3D applications and services [4–7]. Human beings are the ultimate assessors of S3D image quality; however, their judgment is time-consuming, inconvenient, and cumbersome, and cannot be used in real-time online operations. Hence, objective assessment metrics offer many advantages such as low cost and easy operation, and can be embedded in S3D visual information processing systems. Therefore, substantial efforts have been made to develop and improve objective S3D IQA metrics. These efforts are a significant step towards mimicking the integral mechanisms of the monocular and binocular visual systems, and accurately and automatically predicting S3D image quality. Generally, existing objective S3D IQA metrics can fall into three distinct categories: full reference (FR), reduced reference (RR), and no reference (NR)/blind metrics, according to the availability of the original S3D images [8–13]. In this research, the discussion is focused on blind metrics.

Recently, numerous blind IQA metrics for conventional 2D images have been emphatically studied in detail, and can be categorized into two types [14,15]. The metrics in the first category extract effective quality-predictive representations from distorted images, and then learn a regression function using those representations. Therefore, different quality-predictive feature extraction schemes and machine learning methods lead to different blind metrics. The processes in the feature extraction stage are based on the following observations: (1) the image is properly normalized or transferred to some transform domain (e.g., DCT [14,15], wavelet [16,17], Gabor [18,19], shearlet [20,21], or spatial [22–24]), and (2) local descriptors can be modeled by various feature distributions. In the machine learning stage, different regression techniques such as support vector regression (SVR) [14,16,23], general regression neural networks (GRNN) [25], deep learning (DL) [26,27], k-nearest-neighbor (KNN) [28], and multiple kernel learning (MKL) [29] can be applied to learn the mapping function from feature distributions, in order to predict the visual's perceptual quality. The second category of blind metrics operates without requiring subject-rated images. For instance, in [30], Xue et al. presented a quality-aware clustering (QAC) metric that learns a set of quality-predictive centroids that are used as a dictionary to calculate the patches' quality in each image, and infer the overall quality score of the image. In [31], Li et al. presented a novel training-free blind IQA metric using several perceptually relevant and

* Corresponding author at: Zhejiang University, Hangzhou 310027, China.
E-mail addresses: wujiezhou@163.com, wujiezhou@zju.edu.cn (W. Zhou).

https://doi.org/10.1016/j.aeue.2018.07.010
Received 14 April 2018; Accepted 14 July 2018

complementary quality-predictive representations. In [32], Mittal et al. presented the natural image quality evaluator (NIQE), which does not need training with human-scored distorted images. Inspired by NIQE, in [33], Zhang et al. developed a blind IQA metric by integrating natural-image-feature statistics derived from multiple cues, without any transcendental knowledge of distortion types or image contents. In [34], Wu et al. developed the local pattern statistics index (LPSI), a highly efficient blind IQA metric that does not utilize a training process but exhibits surprisingly effective generalization capability. In [35], based on the basic concept of information maximization, Gu et al. proposed a training-free blind IQA metric for contrast distortion. However, the aforementioned 2D IQA metrics may not always be effective in measuring the quality of distorted S3D images (it is more challenging when an S3D image consists of two dichoptic views with varying quality levels), because the quality-aware representations in these metrics cannot sufficiently mimic the natural binocular visual mechanism.

Because blind S3D IQA is relatively less mature, only limited progress has been made in this research field. In [36], Chen et al. presented a cyclopean view-based blind S3D IQA metric that uses only blind 2D IQA metrics to assess disparity/depth information and cyclopean views. In [37,38], Gu et al. presented a blind S3D IQA metric that extracted 3D visual factors related to the nonlinear additive model, saliency-based parallax compensation, and ocular dominance model. In [39], Ryu and Sohn developed a blind S3D IQA method for distorted S3D images by employing the binocular visual perception model; however, the final quality score only included the perceptual quality scores of both views' blurriness and blockiness. In [40], Akhter et al. presented a blind S3D IQA method that independently extracts natural scene statistics (NSS) quality-aware features from depth/disparity information and both views; subsequently, a regression function is utilized to obtain the overall quality score via those NSS quality-aware features. In [41], Su et al. developed a blind S3D image naturalness quality method, which uses both correlated and bivariate NSS models to acquire the oriented structure information in distorted S3D images. In our previous work [42], we developed a blind quality metric that uses dictionary learning and machine learning methods to predict S3D images. Other relevant works can be found in [43–45]. However, the above-mentioned blind S3D IQA metrics are supervised methods. Specifically, they need a large number of subject-rated samples to learn the regression function; this diminishes their generalization capability, thereby limiting their practical applicability [46]. Therefore, there is an urgent need to develop unsupervised blind S3D IQA metrics. However, while the aim of unsupervised blind S3D IQA is quite intriguing, devising a practical method to achieve it is a significant challenge, owing to the fact that relevant information may be unavailable.

To overcome the shortcomings of previous methods, in this work we propose an unsupervised blind S3D IQA metric that exclusively utilizes computable deviations from statistics obtained in natural S3D images, without the need for training with human-rated distorted S3D images. To the best of our knowledge, this work is the first to introduce the joint spatial and frequency representations of binocular visual perception into the field of blind S3D IQA. The major contributions of this work include the following:

(1) We adopt joint spatial and frequency representations of binocular visual perception, which more comprehensively capture distortion artifacts.
(2) We exploit two different weighting schemes to simulate the binocular visual mechanism:
   (i) A generalized eye-weighting model is used to weight monocular features in order to obtain binocular features.
   (ii) A gain-contrast control theory model is used as a weighting scheme for the cyclopean view.
(3) A pristine multivariate Gaussian (MVG) model is established by using the joint spatial and frequency representations to obtain the final S3D perception, in place of machine learning methods.

Our experimental results demonstrate the superior performance of our metric.

The remainder of this work is organized as follows. Section 2 discusses theories related to the binocular visual mechanism. Section 3 illustrates the proposed S3D IQA metric. In Section 4, experimental results and comparisons are given and discussed. Finally, Section 5 concludes the work.

2. Binocular visual mechanism

The binocular visual mechanism is a complex visual process and plays an important role in depth perception. Owing to recent advances in visual cognition theories and neural science, numerous visual psychophysical and physiological discoveries have enabled us to more deeply understand the binocular visual mechanism; these findings are beneficial to the development of effective and efficacious S3D IQA metrics. Here we briefly describe recent findings on the binocular visual mechanism as they relate to this work.

(1) The generalized eye-weighting model: Unlike the conventional 2D visual mechanism, the binocular visual mechanism can perceive the discrepancy between two dichoptic views at the same visual retinal position, owing to two significant binocular interactions—binocular fusion and rivalry [47]. In binocular fusion interaction, each of two eyes provides its own nuanced visual content, and a single visual perception is obtained [48]. In particular, binocular rivalry interaction occurs when the two eyes perceive two mismatched views at the same visual retinal position in 3D space [49]. The main goal of S3D IQA is to characterize the binocular quality-aware features by considering binocular fusion and rivalry. In this work, the combining of binocular quality-aware features is based on Bayesian theory [50], in that the binocular quality-aware features can be modeled by a hybrid combination of monocular quality-aware feature distributions

$$P(\vartheta \mid I_l, \widehat{I_r}) = \frac{P(I_l, \widehat{I_r} \mid \vartheta)\cdot P(\vartheta)}{P(I_l, \widehat{I}\ )}$$
$$\approx \frac{P(I_l)}{P(I_l, \widehat{I_r})}\cdot P(\vartheta \mid I_l) + \frac{P(\widehat{I_r})}{P(I_l, \widehat{I_r})}P(\vartheta \mid \widehat{I_r}) \tag{1}$$

In Eq. (1), I_l and $\widehat{I_r}$ are the left view and the disparity-compensated right representation, respectively, and ϑ is the perceived quality. Clearly, it is essentially a generalized eye-weighting model for binocular features combination, and can be rewritten as

$$\mathbf{F}_{lr} = w_l \cdot f(I_l) + w_r \cdot f(\widehat{I_r}) \tag{2}$$

where $\mathbf{F}_{lr}$ denotes the binocular features, $f(\cdot)$ denotes the monocular features extraction function (this function will be discussed in Section 3, and w_l and w_r can be regarded as the visual-based weights that represent the binocular fusion and binocular rivalry processes. In this study, the weights are calculated as

$$w_\xi = e_\xi/(e_l + e_r)\ \xi \in \{l, r\} \tag{3}$$

where e_l and e_r respectively denote the local energy variance of the left and right views of S3D images to simulate binocular stimulus strength.

(2) The gain-contrast control theory model: In a binocular lightness combination, the images observed by the left and right retinas are combined to generate a single perceived "cyclopean" view (a binocularly fused percept) [51,52]. Recently, a number of studies have examined how two slightly different monocular views fuse to a combined view using disparity-based scene geometry and depth perception [53,54]. We use the gain-contrast control theory model to explain the perception of binocular lightness combinations and to