Signal Processing: Image Communication 52 (2017) 54–63

Contents lists available at ScienceDirect

Signal Processing: Image Communication

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

A stereoscopic image quality assessment model based on independent component analysis and binocular fusion property

Xianqiu Geng, Liquan Shen*, Kai Li, Ping An

School of Communication and Information Engineering, Shanghai University Shanghai 200072, China

ARTICLE INFO

Keywords:
Stereoscopic quality assessment
Independent component analysis
Image feature similarity
Local luminance consistency
Binocular combination

ABSTRACT

Quality assessment for stereoscopic image is an important research issue in three-dimensional researches. In this paper, an independent component analysis(ICA) and binocular combination-based full reference image quality assessment (FR-IQA) method is proposed for color stereoscopic image. Specifically, image features that reflect the responds of simple cells in the cortex are first extracted by an ICA-based feature detector, which models the functions of the receptive fields of simple cells in the primary visual cortex. Both image feature similarity (IFS) and local luminance consistency (LLC) of the right and left images are calculated to measure the structure and brightness distortions respectively. Image patches that the mean pixel values have been removed are used for IFS calculation, while the computation of LLC is based on the removed mean pixel value. To simulate the binocular fusion properties of the complex cells in the cortex, feature energy of the extracted image features is utilized to calculate the weighting factors of binocular combination, following which a single feature similarity index is obtained by fusing the right and left images IFS. Furthermore, global relative luminance information of the selected image patches is used to integrate the right and left images LLC into a single luminance consistency index. Finally, image quality is obtained by combining above two indexes. Experimental results demonstrate that the proposed algorithm achieves high consistency with subjective assessment on 3D image quality assessment databases.

1. Introduction

Recently, owing to the high-speed development of communication technologies and systems, image has become an efficient means for information transmission. The idiom, a picture is worth a thousand words, has perfectly demonstrated the important role of visual information for human perceptions. Meanwhile, as the final receiver of the visual information, human has put more attention to multimedia quality of experiences (QoE) and quality of services (QoS). Therefore, image quality assessment (IQA) is essential to various visual signal processing applications.

Researches on traditional 2D-IQA have been conducted for decades, and many metrics have been proposed [1,2]. Recently, the demand for stereoscopic image/video quality assessment (3D-IQA/VQA) has become urgent since the booming up of 3D technologies and the emerging of 3D image/video. However, 3D-IQA/VQA is still a complicated and challenging issue than it 2D counterparts with consideration of both 2D quality and 3D perceptual factors (i.e., disparity/depth perception [3–7], the properties of visual perception [8], and others [9–11]).

Since human eye is the final receiver of the visual information, the quality of stereoscopic image/video is ultimately determined by human eye. For stereoscopic image quality assessment, the most effective and accurate method is the subjective quality assessment [12,13], in which an image is firstly shown and then recorded a quality score by the observer. The public available 3D-IQA databases, such as LIVE 3D-IQA database phase I [14] and phase II [15] are established by subjective quality assessment. However, due to the applications and evaluation conditions, subjective quality assessment is not stable and generally time-consuming, which handicaps its value in real application.

In view of the drawbacks of subjective quality assessment, a large number of objective image quality assessment models, which automatically evaluate image quality with auxiliary of computer system have been proposed. These proposed objective methods including three categories: 1)no-reference stereoscopic image quality assessment (NR-SIQA) [16–19], in which no reference image information is available. Shao et al. [16] propose a no-reference image quality assessment for stereoscopic images by using phase-tuned quality lookup (PTQL) and phase-tuned visual codebook (PTVC) from the binocular energy responses. Ryu et al. [17] explore the relationship between the perceptual quality of stereoscopic images and visual information, based on which a no-reference quality metric for stereoscopic images is proposed; 2) reduced-reference stereoscopic image quality assessment (RR-SIQA)

* Corresponding author.

http://dx.doi.org/10.1016/j.image.2016.12.004
Received 11 May 2016; Received in revised form 19 December 2016; Accepted 19 December 2016
Available online 28 December 2016
0923-5965/

[20,21]. In these metrics, partial information of the reference images is available. Ma et al. [20] develop a novel reduced-reference stereoscopic image quality assessment by characterizing the statistical properties of the stereoscopic image in the reorganized discrete cosine transform (RDCT) domain. Qi et al. [21] propose a reduced-reference stereoscopic image quality assessment metric based on the distribution statistics of visual primitives in left and right images, which are extracted by sparse coding and representation; 3) full-reference stereoscopic image quality assessment (FR-SIQA) [22–29], where reference images with no distortion is available.

Our research aims emphasis on FR-SIQA metrics, and the existing objective FR-SIQA models can be classified into three types. The first type of metrics apply the existing 2D metrics on the left and right images, and combine two scores to yield a predicted 3D quality [24,25]. However, the quality of 3D viewed images is generally different from a simple combination of the qualities of the 2D viewed images.

The second type of FR-SIQA metrics are proposed with consideration of 3D perceptual properties, such as disparity/depth perception [3–7,26,29]. You et al. [4] apply various 2D quality evaluators on the stereopair and its disparity map, and find that optimal combination can yield the best performance. Benoit et al. [26] compute quality scores of both stereopair and the disparity map by 2D quality metrics and then combine them to produce a final score. Chen et al. [29] address binocular rivalry issues by modeling the binocular suppression [30] behaviors based on a cyclopean map, which formed by the left-right images and the estimated disparity map. However, it is difficult to evaluate the quality of the perceived disparity/depth information since ground truth disparity/depth is generally not available. Even though such models can assess the depth quality using estimated disparity maps based on a SSIM-based stereo match algorithm, the accuracy and high computational cost constrain its real-time application.

For the third type of metrics, the properties of binocular visual perception are simulated to obtain an ideal image quality model. In [31], Shao et al. classify the stereoscopic images into non corresponding, binocular fusion and binocular suppression regions, and then these regions are evaluated separately via the binocular just noticeable difference (BJND) to estimate the quality of stereoscopic image. Since the primary visual cortex (V1) is critical to the visual perception, both simple and complex receptive fields (RFs) should be appropriately characterized [32]. In [33], Bensalma et al. develop a Binocular Energy Quality Metric (BEQM), which first simulates the binocular visual signal by modeling simple and complex cells. The quality is then estimated based on the difference of the associated binocular energy. In [34], Shao et al. propose a K-SVD algorithm [35] based sparse coding model to learn binocular RFs properties to simulate simple and complex cells in the primary visual cortex. However, binocular rivalry [29] hasn't been effectively addressed in the metric, which resulting in a poor performance on the asymmetrically distorted image database.

Research has proved that the spatial receptive fields of simple cells in the cortex can be characterized as being localized, oriented, and band-pass [36]. Those three response properties of cortical simple cells can be sufficiently explained by a coding strategy that maximizes the sparseness [37]. Moreover, natural images contain sparse structures, and sparse coding can just extract these intrinsic structures in images [38–40]. When processed by sparse coding, a given image can be only represented by a set of basic vectors [37], and the difference between the basic vectors of the reference and distorted images can reflect the perceived difference in quality. Thus, measuring the similarity between the reference and a given image basic vectors is an effective way to qualify the given image quality. Due to sparse coding equivalent to independent component analysis (ICA) [41–43], a given image that represented by a set of basic vectors can be obtained when ICA is applied to the natural image. Therefore, the visual processing of cortical simple cells in the visual cortex can be simulated by ICA model.

In this paper, a model based on independent component analysis and binocular fusion for color stereoscopic image quality assessment is proposed. To simulate the responds of the simple cells in the cortex, both the right and left images features are extracted as the responds by a feature detector, which is obtained by an ICA-based training process. Image feature similarity and local luminance consistency are calculated to measure structure and luminance distortions between the reference and distorted images. As the latest biological model for binocular combination, Gain-Control Theory Model (GCTM) [44] is utilized to simulate the process of binocular combination. Moreover, binocular rivalry is simultaneously considered in binocular combination. Stimuli strength from the right and left images is used to determine the binocular rivalry for stereoscopic images with asymmetric distortions. When difference between two stimuli strength from the right and left images is large, human's eyes will focus on the image with higher stimuli strength. Therefore, stimuli strength, such as energy of image features and global relative luminance information of the distorted image are utilized to calculate the weighting factors of GCTM to integrate the right and left images feature similarity and local luminance consistency into a single IFS and LLC index. Finally, a final quality score will be obtained by integrating the single IFS and LLC index with a linear manner.

The rest of the paper is organized as follows. Section 2 describes the basic of independent component analysis for feature extraction. The proposed natural stereoscopic image quality assessment metric is detailed in Section 3. Experimental results are given and discussed in Section 4 and finally conclusions and the future work are drawn in Section 5.

2. Independent component analysis for feature extraction

In the primary visual cortex (V1), the lateral geniculate nucleus (LGN) can determine the amount of information to pass when visual signal transmit from the retina to cortex [45]. The important information is preserved through the process, while the redundant information is discarded. Mathematically, the functions of the retina and LGN can be explained by whitening and dimension reduction [46,47]. Whitening is a special processing, which is to attenuate the low frequencies and boost the high frequencies to yield a roughly flat power spectrum across all spatial frequencies. Since the input data can be whiten and data dimension will be reduced by ICA [41–43], it can be used to illustrate the perceptual process. Fig. 1 illustrates the basic model of ICA.

In the model of ICA, the observed data, x, is represented as a linear transformation of the latent independent components by

$$x = As \tag{1}$$

where $x = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$ is the vector of observed data and $s = [s_1, s_2, \ldots, s_n]^{\mathrm{T}}$ is the vector of the latent independent sources (i.e., independent components).

The goal of ICA is to merely use the observe data vector, x, to estimate the matrix A and the independent sources vector, s, which can be expressed as

$$y = Wx = WAs = \tilde{s} \tag{2}$$

$$W^{-1} = \widetilde{A} \tag{3}$$

where y is an estimate of s , W is a transformation matrix and the inverse matrix W^{-1} is seen as an estimate of A. When the transformation matrix, W, is estimated, the estimate of independent component, s, can be obtain by: $\tilde{s} = y = Wx$.

From a neurophysiological point of view, the function of each element of W models the simple cell spatial receptive field, and the independent component s can be interpreted as the responses of simple

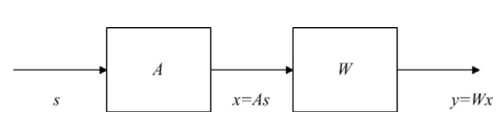

Fig. 1. The basic model of ICA.