Information Sciences 422 (2018) 110–121

Contents lists available at ScienceDirect

Information Sciences

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

Local and global sparse representation for no-reference quality assessment of stereoscopic images

Fucui Li, Feng Shao*, Qiuping Jiang, Randi Fu, Gangyi Jiang, Mei Yu

Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China

A R T I C L E I N F O

Article history:
Received 12 October 2016
Revised 16 July 2017
Accepted 3 September 2017
Available online 6 September 2017

Keywords:
Stereoscopic 3D image
No-reference quality assessment
Multi-modal sparse representation

A B S T R A C T

No-reference/blind quality assessment of stereoscopic 3D images is much more challenging than 2D images due to the poor understanding of binocular vision. In this paper, we propose a BLind Quality Evaluator for stereoscopic 3D images by learning Local and Global Sparse Representations (BLQELGSR). Specifically, at the training stage, we first construct a large-scale training set by simulating some common distortions that are likely encountered by stereoscopic images, and propose a multi-modal sparse representation framework to characterize the relationship between the feature and quality spaces for all sources of information from left, right and cyclopean views in local and global manners. At the testing stage, based on the derived 3D quality prediction framework, the local and global quality scores from different sources are predicted and combined to drive a final 3D quality score. Experimental results on three 3D image quality databases show that in comparison with the existing methods, the devised BLQELGSR can achieve better prediction performance to be in line with subjective assessment.

1. Introduction

With the great development in three-dimensional (3D) imaging and display technologies, various 3D applications and services, such as 3D movie, 3D mobile phone, and 3D television, etc., are emerging in recent years, to boost users quality of experience (QoE) with more vivid and realistic visual experiences. Among various factors affecting QoE, the quality of 3D content is particularly critical to guarantee the QoE. In practice, 3D content will undergo various quality degradations in a 3D signal processing chain, including capturing, compression, transmission, reconstruction, display, etc. Each stage will introduce different types of independent/hybrid distortions. Therefore, how to measure quality degradation of becomes a very challenging and urgent issue in the field of 3D visual signal processing.

Compared with traditional 2D image quality assessment (IQA), QoE in 3D involves not only evaluating contents image quality, but also evaluating the perceived depth perception and visual comfort. Visual comfort (also known as visual fatigue) assessment and depth perception (also known as depth sensation) assessment has been studied in [12,16,26,46]. Besides, different viewing environments may affect the QoE [15]. In this work, focusing on 3D content itself, we aim to design an objective 3D-IQA metric, but try to address the added information (e.g., depth sensation) during the evaluation.

Generally, based on the accessibility of reference image, current IQA methods can be divided into three categories: full-reference (FR), reduced-reference (RR) and no-reference (NR). Typical 2D FR methods include structural similarity index

* Corresponding author.
E-mail address: shaofeng@nbu.edu.cn (F. Shao).

http://dx.doi.org/10.1016/j.ins.2017.09.011

(SSIM) [41], visual information fidelity (VIF) [34], universal quality index (UQI) [40], feature similarity (FSIM) [48], etc. However, the perceptual issues involved in 3D-IQA are more complex than those in 2D-IQA, because some non-intuitive binocular visual properties, such as binocular fusion, binocular rivalry or binocular suppression, will also affect the 3D quality perception. Typical FR 3D-IQA methods include color information only based methods [4,17,18,30,39,45] and color plus disparity information based methods [3,6,31,50]. We have well reviewed these methods in our previous work, and readers can refer to [33]. Due to the difficulty in acquiring the original/reference images in practice, NR/blind IQA (BIQA) methods are particularly necessary but extremely challenging. State-of-the-art BIQA models may fall into two categories: distortion specific models and general purpose models. The first is devoted to specific distortion types, e.g., Blind Blur Metric (MBBM) [19], No-reference JPEG-quality Evaluator (WNJE) [42], Fast Image Sharpness (FISH) method [38]. For distortion specific 3D-IQA task, Ryu and Sohn [27] computed perceptual blurriness and blockiness scores of stereoscopic images. Gu et al. [7] evaluated JPEG-coded stereoscopic images based on the nonlinear additive, ocular dominance and parallax compensation models. Sazzad et al. [29] utilized disparity information to predict the quality of JPEG-coded stereoscopic images. However, the universality of these distortion specific models is comparatively limited due to the difficulty in identifying the distortion type and the limitation in extending one distortion specific model to other distortion types.

For general purpose BIQA metrics, the most straightforward way is based on the well-chosen handcrafted features, such as natural scene statistics (NSS), image entropies, and other statistics features. These features are generally represented by the parameters of the feature distributions, including generalized Gaussian distribution (GGD), Weibull distribution, etc. The most representative works in this category include Distortion Identification-based Image Verity and INtegrity Evaluation (DIIVINE) [28], Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [20], BLind Image Integrity Notator using DCT Statistics-II (BLIINDS-II) [23], gradient magnitude and Laplacian of Gaussian (GM-LOG) [43], FReeEnergy principle-based NR-IQA (FRENIQ) [8]. These methods rely on support vector machine (SVM) to train a regression function. Besides, Tang et al. [37] extracted Blind Image Quality Index (BIQI) features and used deep belief network (DBN) to train the model. Zhang et al. [47] derived NSS features from multiple cues and learned a multivariate Gaussian (MVG) model of image patches from a collection of pristine natural images. He et al. [9] extracted NSS features and represented the features via sparse representation, and weighted human option scores using the sparse coefficients to obtain the final quality value. Kang et al. [14] used a Convolutional Neural Network (CNN) to learn discriminant features for the NR-IQA task. Jiang et al. [11] proposed to apply the supervised dictionary learning framework to address the problem of BIQA using quality-constraint sparse coding.

Compared to FR 3D-IQA [3,4,6,17,18,30,31,39,45,50], researches on NR 3D-IQA are still in its initial development stage, and only limited researches have been concentrated on the advance of NR 3D-IQA models. For NR 3D-IQA task, information from cyclopean or disparity maps are needed to fully understand/characterize stereoscopic 3D vision (this make it differ from 2D vision). Chen et al. [5] proposed a NR-IQA metric for stereoscopic images based on the cyclopean model and NSS features. Su et al. [36] developed a Stereoscopic/3D BLind Image Naturalness Quality (S3D-BLINQ) index that extracts NSS features from the cyclopean view. Appina et al. [1] developed a Stereo QUality Evaluator (StereoQUE) based on NSS statistics. Zhang et al. [49] used multi-layer convolutional neural networks (CNNs) to learn the structures of stereoscopic image for no-reference quality assessment task. Mocanu et al. [22] developed a Quality of Experience for 3D Images through Factored Third Order Restricted Boltzmann Machine (Q3D-RBM) model, which is composed of three layers (a real valued visible layer by left eye, a real visible right layer by right eye and a binary hidden layer), and is connected using undirected weights. In our previous work [32], local receptive fields (LRFs) and global RFs (GRFs) were extracted from the pristine and distorted stereoscopic images, and the corresponding local quality lookups (LQLs) and global quality lookups (GQLs) were constructed. Followed by searching the optimal GRF and LRF indexes from the learned LQLs and GQLs, blind quality pooling can be easily achieved. In our another work [33], we devised a 3D deep blind quality evaluator (3D-DQE) for stereoscopic images by considering monocular and binocular interactions. In this work, two separate 2D DBNs from 2D monocular images and cyclopean images are trained to model the process of monocular and binocular quality predictions. A summary of BIQA methods cited in this paper can be found in Table 1.

In essence, previous 2D BIQA methods [14,20,23,28,37] and 3D BIQA methods [1,5,22,36,49] depend on machine learning tools to train the regression functions. However, when the handcrafted features cannot correctly predict the perceptual quality, the evaluation performance will drop significantly, e.g., one handcrafted feature may be only effective for one distortion type, but has a much worse performance on other distortion types. In addition, since these methods highly rely on human scored training examples to learn the regression/neural network models, different subjective evaluation methodologies and devices will acquire completely inconsistent human opinion scores, leading to poor generality capability of these methods in predicting the perceptual quality across different databases. To solve the aforementioned problems for both 2D and 3D BIQA methods, we aim to establish the intrinsic relationship between the feature and quality spaces via sparse representation. This work is a further extension of our previous work [32], but some important technical innovations related to the process of human vision perception are developed.

As experimental results of our previous work [32] show, sparse representation can be used effectively to learn useful visual perceptual information, e.g., defined as receptive fields (RFs) in our previous work [32]. However, there are certain limitations of this approach: 1) the semantic relationship between the learned RFs and the estimated quality lookups is ambiguous, because the candidate RFs are searched exhaustively on all training samples; 2) the relevance between the left and right views is in fact omitted in learning the RFs and quality lookups, leading to poor collaboration between the left and right views. In this paper, based on the 3D quality prediction model derived in our previous work [33], we propose a BLind Quality Evaluator using Local and Global Sparse Representation (BLQELGSR) for stereoscopic images. Our BLQELGSR model