Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Blind image quality assessment in multiple bandpass and redundancy domains

Yang Ma [a], Weixia Zhang [a], Jia Yan [a,*], Cien Fan [a], Wenxuan Shi [b]

[a] Electronic Information School, Wuhan University, Wuhan, 430072, China
[b] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, 430079, China

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Blind image quality assessment
Natural scene statistics
Image representation
Fisher vectors

ABSTRACT

Blind image quality assessment (BIQA) aims to automatically evaluate image quality without any prior knowledge of reference images and distortion types. Most of existing BIQA methods use certain probability distribution models to capture the natural scene statistics features of images in bandpass domains which can be viewed as a process of removing redundancy. There also exist methods which apply NSS features in redundancy domain. In this paper, we propose a novel method that employs both bandpass and redundancy domains to acquire the complementary features in multiple color spaces. Furthermore, hierarchical feature extraction strategy is adopted to make the image representation more powerful. Then we stack them as a multi-channel feature maps group, and use Gaussian mixture model to fit them. Finally, Fisher Vectors are used to encode them and a support vector regression model is trained as the quality predictor. Extensive experiments on four commonly evaluated image quality assessment benchmark databases show the proposed method is very competitive against other BIQA methods and has good generalization ability.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

With the rapid growth and near-ubiquitous presence of digital cameras, numerous digital images are produced everyday all over the world. Images are often distorted in the procedure of acquisition, transmission and compression, etc., which may hinder people understanding the content of images and getting information. Thus, researches on accurately evaluating the quality of images without human assistance become increasingly important yet very challenging in computer vision and image processing. While the ultimate criterion of these researches is subjective assessment by humans, it is time-consuming, troublesome, expensive and hence not scalable to large-scale tasks. Therefore, it is highly desired to develop objective assessment systems that can automatically evaluate image quality.

Objective image quality assessment (IQA) methods can be classified into three categories: full-reference (FR), reduced-reference (RR) and no-reference (NR). While FR and RR IQA methods usually show superior performance, e.g., Liu et al. proposed a FR-IQA method based on the parallel boosting (ParaBoost) idea, which had made great achievements [1], full or part information of the reference image is needed to evaluate the quality of an image in these methods. However, in most scenarios, neither full nor partial prior knowledge of reference images is accessible. Hence, NR-IQA attracts much more research interests. Early NR-IQA approaches aim to measure the quality of images with specific distortion types, which limit their applications in real world. Recent researches on NR-IQA have focused on distortion-generic methods, which are also known as blind image quality assessment (BIQA).

A majority of existing BIQA approaches consist of two primary components: feature extraction and regression module. In feature extraction stage, the quality-aware features are generated to measure the distortion degree of images. These methods are usually based on natural scene statistics (NSS), which depend on an assumption that pristine natural images hold certain stable statistical regularities that will be disturbed when introducing distortions. These NSS features are usually extracted from different domains including both the spatial domain and the transform domain such as discrete cosine transform (DCT) domain and discrete wavelet transform (DWT) domain. BLIINDS [2] and BLIINDS-II [3] employ the generalized Gaussian distribution (GGD) parameters of the DCT coefficients of images. BIQI [4] and DIIVINE [5] focus on the DWT domain. SRNSS is based on the sparse representation of NSS features in the wavelet domain [6]. BRISQUE computes the locally normalized luminance coefficients of images as the quality aware

* Corresponding author.
E-mail address: jiayan0080@outlook.com (J. Yan).

https://doi.org/10.1016/j.dsp.2018.05.010
1051-2004/© 2018 Elsevier Inc. All rights reserved.

features [7]. The M3 model extracts features from the gradient magnitude (GM) map and the Laplacian of Gaussian (LoG) map and uses the joint statistics of them to measure images quality [8]. FRIQUEE utilizes a set of feature maps in several bandpass domains and color spaces to assess image quality [9]. Li et al. proposed a BIQA method named NRSL, which employs the probability distribution of local binary pattern (LBP) on the normalized luminance coefficients of images to capture structural information [10]. The aforementioned BIQA approaches focus on the bandpass transform domains which can be viewed as a process of removing the redundancies. In contrast to these methods, Yan et al. proposed the so-called natural redundancy statistics evaluator (NRSE) method, which captures the statistical naturalness of images directly on redundancy images [11]. The TCLT metric develops a multichannel fusion scheme by combining the NSS features in multiple-domain such as the DWT, DCT and spatial domain from all YCbCr color channels to simulate the hierarchical structure and the trichromatic property of human visual system (HVS) [12]. Yang et al. proposed a method for authentically distorted images with perceptual features (ADIPF, for short) including some widely used NSS features as well as some other quality-aware features such as blurriness and dynamic range of image [13]. CORNIA is a learning-based method that uses features learned from raw normalized local image patches by unsupervised clustering [14]. Zhang et al. introduced the semantic obviousness metric (SOM) into perceptual quality assessment [15].

Recently, several BIQA models using convolutional neural network (CNN) emerge. Kang et al. proposed the Le-CNN model by extending CORNIA into a shallow CNN [16]. BIECON first computes local quality scores as proxy patch labels using a full-reference algorithm to remedy the lack of adequate local ground truth scores, and then train a deep CNN model using these labeled image patches to measure image quality [17]. Bosse et al. proposed a deeper CNN model-based algorithm called deepIQA which uses raw local image patches as input and the average of patchwise scores as the overall quality of image and introduces another strategy that the image score is calculated by weighted aggregation of patchwise scores [18].

Other than the above-mentioned methods which need regression models to map the extracted features to human subjective perceptual quality scores, another kind of BIQA methods do not need any regression model, which are also known as opinion-unaware methods. NIQE [19] and IL-NIQE [20] extract a set of features from pristine images and fit the feature vectors to a multivariate Gaussian (MVG) model. The quality score of a test image is then calculated by a Bhattacharyya-like distance of the test image's MVG and the acquired clean model.

Natural images are high dimensional signals which comprise lots of redundancies. Barlow proposed that sensory systems can extract signals of high relative entropy from the highly redundant sensory input [21]. Each stage of processing in sensory system attempts to eliminate as much redundancy as possible, which can be regarded as bandpass filters [22]. From the perspective of information theory, the extraction of image statistical features can decompose the image information $H(\text{Image})$ into two parts: $H(\text{Image}) = H(\text{Bandpass}) + H(\text{Redundancy})$ [23], where $H(\text{Bandpass})$ denotes the remaining useful part, $H(\text{Redundancy})$ is the redundant information that can be suppressed by a bandpass filtering system. However, flawless bandpass filter that can keep all useful information and remove all useless information is non-existent, which means that $H(\text{Redundancy})$ may always contain valuable components that shouldn't have been discarded. To address the aforementioned problem, we propose a novel BIQA model from three aspects: 1) we extract quality-aware features by different bandpass filters to obtain more sufficient and complementary information, so as to handle complex distortions and contents of images. It is verified that the neurons of the primary visual cortex are sensitive to local structure features [24]. Here, we employ several Gaussian derivative features including mean subtracted contrast normalized (MSCN) coefficients, LoG and GM to capture the local image structure., which can be viewed as the classical 'edge' and 'bar' types of features [25]. 2) We also apply redundancy features on IQA to complement bandpass features. Inspired by the NRSE model which computes statistics on image redundancy acquired by using the singular value decomposition (SVD) and reconstruction [11], we introduce another approach, i.e., DCT and its inverse transform to obtain image redundancy. 3) We proposed a hierarchical feature extraction strategy based on multiple color spaces to make the image representation more powerful. In a nutshell, the first contribution of this paper is to extract quality aware features in different bandpass and redundancy domains from multiple levels on various color spaces to make use of as much useful information as possible.

The NSS-based metrics commonly utilize one or more univariate probability distribution models such as GGD, asymmetric GGD (AGGD) and Weibull distribution models to describe the structure information of natural images, and estimates different parameters including mean, variance, skewness, kurtosis and goodness to predict the quality of images. However, univariate probability distribution models are not sufficient to represent the complex properties of natural image quality and model parameters as perceptual quality features may be inaccurate to represent the image quality because of the fitting errors. To solve this problem, we stack the extracted features as feature maps, and employ multivariate Gaussian mixture model (GMM) to describe the image feature maps, then use Fisher Vectors [27] to represent image quality. Furthermore, to ensure the fairness and efficiency of experiments, we build an extra database, contents of which are not overlapped with all images in the evaluated IQA databases, for off-line GMM clustering. Hence, the second contribution of this paper lies in using GMM to describe image features so as to sufficiently represent the complex properties of image quality. The flowchart of the proposed method is illustrated in Fig. 1. Extensive experiments show the proposed method is very effective and has good generalization ability.

The remainder of the paper is organized as follows. Section 2 describes the details of the proposed method. Experimental results on public IQA benchmark databases and the corresponding analysis are presented in Section 3. Section 4 concludes the paper.

2. The proposed method

2.1. Feature maps

2.1.1. Gaussian derivative feature maps in bandpass domain

It is verified that the neurons of the primary visual cortex are sensitive to local structure features [24], e.g., edges, bars, corners and blobs, etc. The principal components of natural images resemble derivatives of Gaussian operators, which is similar to those findings in visual cortex and inferred from psychophysics [28]. The researches on Independent Components Analysis (ICA) also show that Gaussian derivative functions can be used to represent perceptual features of images since the retino-cortical receptive fields respond selectively to stimulus from different orientations and scales [26].

For natural images, luminance changes comprise most of the structural information. Bandpass image responses, especially Gaussian derivative responses, can well describe semantic structures of images [8]. Here, we employ several Gaussian derivative features to sufficiently capture the local image structure. Center-surrounded mean subtracted contrast normalized (MSCN) coefficients measure the interactions between neighboring pixels with contrast masking

Download English Version:

https://daneshyari.com/en/article/6951656

Download Persian Version:

https://daneshyari.com/article/6951656

Daneshyari.com