J. Vis. Commun. Image R. 33 (2015) 123–133

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Supervised dictionary learning for blind image quality assessment using quality-constraint sparse coding ☆

Qiuping Jiang, Feng Shao *, Gangyi Jiang, Mei Yu, Zongju Peng

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

ARTICLE INFO

Article history:
Received 17 June 2015
Accepted 15 September 2015
Available online 25 September 2015

Keywords:
Blind image quality assessment (BIQA)
Supervised dictionary learning
Sparse coding
Sparse representation
Quality-constraint
Opinion free
Gabor
Histogram of Oriented Gradient (HOG)

ABSTRACT

Blind image quality assessment (BIQA) involves predicting the perceptual quality of distorted images without using their corresponding reference images as benchmark. Especially, it is desirable and meaningful to design effective opinion-free BIQA (OF-BIQA) model to predict image quality without depending on human subjective score. Toward this end, we propose a supervised dictionary learning framework for OF-BIQA using quality-constraint sparse coding. The prominent advantage of the proposed model is that "ground truth" quality scores derived from existing full-reference IQA (FR-IQA) metrics are incorporated into the traditional dictionary learning framework so that a quality-aware sparse model can be learnt. Since the goal of BIQA is to predict the quality score, the introduction of quality information into dictionary learning can be regard as a supervised dictionary learning framework. In the detailed implementation, a quality-aware regularization term is added to the traditional dictionary learning formulation, such that a feature-aware dictionary and a quality-aware dictionary can be learned jointly. Especially, these two dictionaries share the same sparse coefficients, so that the reconstruction errors from the image feature vectors and quality score vectors are both minimized. Once the feature-aware and quality-aware dictionaries are jointly learned, given a testing sample, we first abstract its feature vector and then compute the corresponding sparse coefficients w.r.t. the learnt feature-aware dictionary, finally its quality score can be directly reconstructed based on the learnt quality-aware dictionary and the estimated sparse coefficients w.r.t. the learnt feature-aware dictionary. The reconstructed quality score is expected to well approximate to the "ground truth" quality score. Thorough validation experiments on three publicly available IQA benchmark databases demonstrate the promising performance of the proposed OF-BIQA model both on the prediction accuracy and generalization capability.

1. Introduction

Digital imaging or digital image acquisition is regularly used in our daily lives, which has been becoming an important carrier to deliver and share the impressive visual information. Unfortunately, digital images may suffer from various types of distortions during the processes of acquisition, compression, transmission, storage, display, etc. Hence, it has become especially urgent to design an effective image quality assessment (IQA) algorithm to faithfully evaluate the perceived quality in line with human vision system (HVS). Early studies on IQA are based on full-reference models, in which the quality score of a distorted image is measured as the fidelity or similarity with its reference one (without any distortion) [1–6]. However, since the reference image is usually unavailable in many practical situations, blind image quality assessment (BIQA), which targets to evaluate the perceived quality without reference image as benchmark, is particularly important.

There have been many efforts on designing BIQA algorithms over the past several years. Based on the prior knowledge of distortion types, state-of-the-art BIQA models fall into two categories: distortion specific models [7–11] and distortion independent models [12–19]. Models belonging to the former category typically extract distortion-specific features related to the particular artifacts, and are capable of performing BIQA task only when the distortion type is known beforehand. Models belonging to the latter category are much more challenging than the former one since they are lack of the prior knowledge of distortion types. Recent studies on BIQA mainly focus on the latter category. A common architecture for the latter general purpose BIQA involves to learn a mapping function from the quality-relevant features to human subjective scores which are available in the existing IQA

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author at: Faculty of Information Science and Engineering, Ningbo University, 818 Fenghua Road, Ningbo 315211, China. Fax: +86 574 87600582.

E-mail addresses: shaofeng@nbu.edu.cn, shaofeng@126.com (F. Shao).

http://dx.doi.org/10.1016/j.jvcir.2015.09.009
1047-3203/

benchmark databases, such as LIVE [20], CSIQ [21], TID2008 [22], and TID2013 [23]. In this regard, we term these algorithms as opinion-aware BIQA (OA-BIQA) models. In general, they share a similar two-stage framework: feature representation and opinion score based regression. In the feature representation stage, handcrafted quality-aware features are extracted to capture the factors on which the distortions may have impact. Then, a regression model is learned to establish the latent relationship between quality-aware features and the associated human subjective scores (e.g. mean opinion score or differential mean opinion score (MOS/DMOS)). As such, arbitrary regression algorithm (e.g., support vector machine, random forest, Naive Bayes and so on) can be used here. In the testing stage, a feature vector is similarly extracted from the testing image and then fed into the learnt regression model to predict its quality score. The representative two-stage OA-BIQA models include: Blind Image Quality Indices (BIQI) [12], Distortion Identification-based Image Verity and INtegrity Evaluation (DIIVINE) [13], Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [14], BLind Image Integrity Notator using DCT Statistics-II (BLIINDS-II) [15], and other related works [16–19].

Although powerful, the limitations of these OA-BIQA models are apparent since they always require a large number of distorted images associated with their human opinion scores to train a reliable quality evaluator, which makes their generalization capability limited. The reasons are twofold: (1) the quality scales of different databases are inconsistent so that a quality model trained on one specific database may not take effect on other databases and (2) according to state-of-the-art subjective quality assessment standards, generating human opinion scores (MOS/DMOS) is generally cumbersome and time-consuming. Gradually, there has raised an increasing interest in deriving effective opinion-free BIQA (OF-BIQA) models to predict image quality without using human opinion scores. Towards this end, Mittal et al. [24] conducted probabilistic Latent Semantic Analysis (pLSA) on the quality-aware visual words extracted from a large set of pristine and distorted image patches. Then, the uncovered latent quality factors are applied to infer the quality score for the testing image. The Natural Image Quality Evaluator (NIQE) proposed by Mittal et al. [25] extracts a set of local features from an image, and fits the feature vectors to form a multivariate Gaussian (MVG) model. The distance between the distorted MVG model and its corresponding original version is computed for quality estimation. Xue et al. [26] proposed a Quality-Aware Clustering (QAC) method to learn a set of quality-aware centroids and use them as the codebook to infer the quality score of an image patch. The final quality score for the testing image is the weighted average of the patch-level quality scores. Ye et al. [27] proposed a simple yet effective OF-BIQA model named BLISS (Blind Learning of Image Quality using Synthetic Scores) by training on the synthetic scores derived from full-reference IQA (FR-IQA) measures. Specifically, unsupervised rank aggregation is applied to combine different FR-IQA metrics to generate a "golden standard" synthetic score. Most recently, inspired by the learning to rank algorithm, Gao et al. [28] designed a novel BIQA framework by training on a collection of preference labels instead of DMOSs. The underlying principle is that preference labels, representing the relative quality of two images, are generally more reliable and easy to obtain. One distinct property of these OF-BIQA models is that they typically have higher generalization capability than the opinion-aware counterparts because they do not require training on the distorted samples associated with their human opinion scores. However, the performance of state-of-the-art OF-BIQA models is usually inferior to the currently available OA-BIQA models. Hence, there is still a tremendous potential for further investigation in this field.

From the perspective of biological physiology, the goal of IQA is to exploit the sophisticated visual processing mechanism of HVS [29]. As an important part of HVS, simple cells in primary visual cortex (also known as striate cortex and V1) are responsible for most of our cognitive understanding of visual scenes. Therefore, an effective model for IQA is expected to be highly correlated with the neural response properties of simple cells. It has been widely evidenced that the receptive fields (RFs) of simple cells can be characterized as being spatially localized, oriented and band-pass [30]. Theoretically, a powerful mathematical tool to characterize this inherent property of RFs is known as the sparse coding principle, which attempts to represent a high-dimensional signal as a linear combination of only a small portion of basis functions from a dictionary [31]. In general, this dictionary can be either predefined or learned from a corpus of training samples. Different from the predefined dictionaries (e.g., basis functions of Discrete Cosine Transform (DCT), Contourlet transform, Wavelet transform, etc.), each basis function in the learnt dictionary is tailored to one specific type of structure or a particular feature presented in the given training samples. The linear decomposition of a signal using a few basis functions of a learnt dictionary has recently led to state-of-the-art performance for numerous low-level image processing tasks [32]. We term the output of the sparse coding model with a learnt dictionary as sparse feature which is generally considered to be highly correlated with visual perception. Based on this fact, some sparse coding-based FR-IQA models have been developed to evaluate the perceived quality by measuring the fidelity relative to the reference image (distortion free) in terms of the sparse features [33,34]. However, although the sparse feature fidelity (SFF) metric can well reflect the perceived quality, how to utilize these sparse features for OF-BIQA still remains a quite challenging problem thus far.

In this paper, we attempt to solve the above problem by supervised dictionary learning. As stated above, in order to exploit the relationship between the sparse features and quality scores, two types of methods are commonly used. The first method follows a general SFF computation framework in which the reference images are needed. The second method follows a two-stage training and testing framework in which the human subjective scores are required to learn a quality regression model so as to establish the latent relationship between the sparse feature and quality scores. One common property of these methods is that the used dictionary for sparse coding are learned in an unsupervised manner and acted as an unsupervised cortex-like visual feature detector, hereby making the relationship between the sparse features and quality scores implicit. To address this problem, we propose a supervised dictionary learning framework for OF-BIQA by using quality-constraint sparse coding. In the framework, a feature-aware dictionary and a quality-aware dictionary are first jointly learned under the quality constraint. Then, the predicted quality score of a testing image patch is reconstructed based on the quality-aware dictionary and its corresponding sparse coefficients w.r.t. the learnt feature-aware dictionary. In summary, we make the following technical contributions:

(1) We propose a new general purpose OF-BIQA model without training on the human subjective scores. Thorough experimental results on three large-scale publicly available benchmark databases demonstrate the promising performance of the proposed model both on the prediction accuracy and generalization capability.
(2) To fill the semantic gap between the learnt dictionary and image quality, we propose a supervised dictionary learning framework by using quality-constraint sparse coding. In the framework, a feature-aware dictionary and a quality-aware dictionary are jointly learned such that the final quality score of a testing image patch can be directly