Displays 44 (2016) 21–26

Contents lists available at ScienceDirect

Displays

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

Pairwise comparison and rank learning for image quality assessment

Long Xu [a,*], Jia Li [b], Weisi Lin [c], Yun Zhang [d], Yongbing Zhang [e], Yihua Yan [a]

[a] Key Laboratory of Solar Activity, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China
[b] State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing 100191, China
[c] Nanyang Technological University, Singapore 639798, Singapore
[d] Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[e] Graduate School at Shenzhen, Tsinghua University, Shenzhen 518005, China

ARTICLE INFO

Article history:
Received 19 January 2016
Accepted 20 June 2016
Available online 21 June 2016

Keywords:
Image quality assessment
Machine learning
Rank learning
Pairwise comparison

ABSTRACT

To know what kinds of image features are crucial for image quality assessment (IQA) and how these features affect the human visual system (HVS) is still largely beyond human knowledge. Hence, machine learning (ML) is employed to build IQA by simulating the HVS behavior in IQA processes. Support vector machine/regression (SVM/SVR) is a major member of ML. It has been successfully applied to IQA recently. As to image quality rating, the human's opinion about it is not always reliable. In fact, the subjects cannot precisely rate the small difference of image quality in subjective testing, resulting in unreliable Mean Opinion Scores (MOSs). However, they can easily identify the better/worse one from two given images, even their qualities do not differ much. In this sense, the human's opinion on pairwise comparison (PC) of image quality is more reliable than image quality rating. Thus, PC has been exploited in developing IQA metrics. In this paper, a rank learning optimization framework is firstly developed to model IQA. Particularly, the PCs of image quality instead of numerical ratings are incorporated into the optimization framework. Then, a novel no-reference (NR)-IQA is proposed to infer image quality in terms of image quality ranks. By importing rank learning theory and PC into IQA, a fundamental and meaningful departure from the existing framework of IQA could be expected. The experimental results confirm that the proposed Pairwise Rank Learning based Image Quality Metric (PRLIQM) can achieve comparable performance over the state-of-the-art NR-IQA approaches.

1. Introduction

There have been dramatically increased interest in image quality assessment (IQA) recently. The most accurate and reliable way of IQA is to ask the subjects who are shown a group of images for their opinions about the quality of these images. This way called subjective IQA, however, is highly time-consuming, human labor consuming, and impractical in real-time application. Thus, a plenty of objective IQA approaches have been developed during last decade. Based on the availability of references, IQA approaches can be classified into full-reference (FR), no-reference (NR) and reduced-reference (RR) approaches. In FR category, structural similarity (SSIM) [1] has been investigated extensively by the researchers due to its simple philosophy and mathematical form, as well as good performance.

Concerning real-world application, NR approaches are more general and applicable than FR approaches. We categorize the NR approaches of the literatures into three categories. The ***first*** one analyzes the behavior of specific distortion for IQA. In [2], Sheikh et al. employed wavelet statistical model to capture JPEG compression distortion. Liang et al. [3] combined the sharpness, blurring, and ringing measurements together to evaluate images distorted by JPEG 2000. In [4], Ferzli et al. introduced just noticeable blur into probability summation model to measure sharpness/blurriness. In [5], Brandao et al. exploited the DCT statistics of JPEG compression to establish a NR-IQA approach for assessing quality of images coded by JPEG. The ***second*** one uses quality aware clustering which arranges image patches of training set into several clusters according to certain local image features, such as histogram of oriented gradients (HoG), difference of Gaussian (DoG) and Gabor filter. Each cluster centroid is assigned quality by averaging the qualities of image patches in this cluster. By associating cluster centroid with its quality, codebook can be established. It performs like a dictionary. Each time, given a new image patch, we look up codebook to find the mostly matched codeword, and then retrieve

* Corresponding author.
E-mail addresses: lxu@nao.cas.cn (L. Xu), jiali@buaa.edu.cn (J. Li), WSLin@ntu.edu.sg (W. Lin), yun.zhang@siat.ac.cn (Y. Zhang), zhang.yongbing@sz.tsinghua.edu.cn (Y. Zhang), yyh@nao.cas.cn (Y. Yan).

http://dx.doi.org/10.1016/j.displa.2016.06.002
0141-9382/

its corresponding quality. In [6], visual codebook is formed on Gabor filter based local appearance descriptors. In [7], Wu et al. used FSIM [8] to compute quality of image patch instead of MOS to establish codebook. The ***third*** one utilizes machine learning (ML) tools, such as support vector machine/regression (SVM/SVR), Adaboost and Clustering [9–12], to map image features onto image quality ratings. In [13], Moorthy et al. employed SVM and SVR to learn a classifier and an ensemble of regressors for distortion classification and quality rating respectively. In [14], Tang et al. proposed an approach similar to [13] but with more elaborate features, including distortion texture statistics, blur/noise statistics and histogram of each subbands of image decomposition.

In [13,14], the distance between MOS and predicted image quality was optimized. Such an optimization objective cannot address IQA very well for the reasons: (1) *the numerical image quality, e.g., with rate of 1–5, is not exactly with a strong confidence for measuring real image quality. The small difference of image quality ratings may not truly reflect the real difference of image qualities*; (2) *to assess image quality, pairwise competition is more reliable/reasonable than numerical quality rating. The subjects are only requested to indicate the binary opinion (better or worse) to two compared images. This kind of comparison is less taxing and confusing than numerical rating system*; (3) *the diversity of image content and distortion types also make it difficult to rate image quality numerically under complex scenarios, but pairwise comparison (PC) is not that difficult*. To address these issues, PC of image quality has been introduced into IQA for assisting image quality rating. Since PC concerns $n \times (n-1)/2$ times of comparisons given n images, it is very labor consuming for acquiring MOSs in subjective experiment.

Two related works have been reported in [19,20]. In [19], image quality preference in pairs were exploited to lead to a rank learning optimization problem, and SVM with multiple kernel were adopted to solve this optimization problem. In [20], an approach was developed for ranking image enhanced algorithms, where image quality ranking rather than giving physical quantity of image quality was investigated. Both [19,20] were associated with a pairwise rank learning (PRL) [15,16] framework. Since PC of image quality only concerns binary option of image quality competition, PRL optimizations were realized by a binary classifier in both [19,20], and SVM/SVR was employed to do classification.

In this work, PC of image quality is formulated into a new PRL framework [21] which was originally used for saliency model. This framework forms PRL task as a general optimization problem instead of a binary classifier as mentioned above. In addition, it uses steepest descent method to solve this optimization problem, which would be faster than SVM, so it would be suitable for large-scale database. Moreover, we additionally take the quality difference intensity into consideration besides binary competition (better or worse) of image quality by introducing scaling factors which account for image quality difference of each pair of images.

The rest of this paper is organized as follows. Section 2 describes the proposed PRLIQM in detail. Section 3 presents the experimental results. And, the final section concludes this paper.

2. Proposed pairwise rank learning based image quality metric

Recently, there is a new trend to establish NR-IQA models by using ML [6,13,14,22–25]. Inspired by the development of rank learning in information retrieval (IR) [15–18], we make a fundamental departure from the family of existing ML based approaches. And, a new PRL framework is proposed with two distinct characteristics from previous ones: (1) *it is established on a rank learning framework*; (2) *only logical comparison instead of numerical rating of image quality is concerned.* The proposed PRL only requires the variable of MOS to be ordinal, while the conventional ML based approaches need an assumption of interval variable for MOS since the numerical computing and statistics are used there (please refer to [26] for the definitions of “ordinal” and “interval”). Therefore, it is more applicable in real-world applications.

Regarding rank learning, the deduced computer model targets at ranking objects instead of assigning a physical quantity of image quality (like PSNR) to each object. Usually, in IR, it ranks the retrieved items by their relevance with the query. To our concerned IQA, we measure image qualities by their ranks instead of physical quantities. Thus, the computer model derived from rank learning rank images firstly. Then, a relation between MOSs and ranks can be established by using polynomial curve fitting. In addition, the pairwise approach as stated in [26] is employed to establish optimization objective function, where the binary comparisons of MOSs are to be ground-truth for training computer model, and the risk function [26] is based on indicator (0–1) loss function which has the binary outputs of 0 and 1, representing inconsistence and consistence between predicted rank of image quality and ground-truth respectively. The related issues of rank learning based optimization are to be detailed in this section.

2.1. Training data for rank learning

We carry out our work on image quality rating databases, such as LIVE image database [27], with numerical ratings of image qualities, i.e., MOSs given by subjects. For conventional ML based training task, we assume the feature vectors $\{x_i\}(i = 1, 2, \ldots, n)$, and labels $\{y_l\}(l = 1, 2, \ldots, k)$ given by MOS. Generally, feature vector concerns high level information of a visual scene, which is extracted from image by using some local/global image descriptors, such as difference of Gaussian (DoG), Gabor filter, wavelet coefficients, Fourier coefficients and newly developed deep learning techniques [28].

To establish the pairwise rank learning task on image quality rating system for IQA, the ground-truth is given by comparing images in pair with respect to their MOSs. Given MOSs $\{y_i\},\ i = 1 \ldots n$, a binary label $\{+1, -1\}$ is assigned to $y_i \geqslant y_j$ and $y_i < y_j$ respectively.

2.2. Pairwise rank learning model

SVM is a supervised learning tool of ML category. The objective of SVM is a little sophisticated relative to general optimization objective of least square error. It optimizes the maximum margin between two classes of samples. There are some variants of SVM, such as L1-SVM, L2-SVM and least squares (LS) SVM. We explore the intrinsic principle of ML for IQA, by optimizing the numerical distance between predicted image quality ($\varphi_\omega(x_i)$) and MOS (y_i) as

$$\begin{aligned} \omega^* &= \arg\min_{\omega} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i + \hat{\xi}_i), \\ & s.t.\quad y_i - \varphi_\omega(x_i) \leqslant \varepsilon + \xi_i, \forall i \\ & \quad \varphi_\omega(x_i) - y_i \leqslant \varepsilon + \hat{\xi}_i, \forall i \\ & \quad \xi_i \geqslant 0, \hat{\xi}_i \geqslant 0 \end{aligned} \tag{1}$$

where φ_ω is a model parameter learned by resolving (1), and used to compute image quality for unknown input image; x_i represents image features of the i-th image, y_i is the label of x_i given by MOS, and $\|\cdot\|_p$ represents p-norm operation. The linear form of $\varphi_\omega:\ \varphi_\omega = \omega^T x$, is widely used in the literature. For fitting MOS curves more generally, nonlinear functions are employed, which explore the nonlinear relationship between image features and MOS. By using kernel functions, nonlinear problems can be converted into linear problems. Observing the optimization objective of (1), the p-norm is optimized, while a new optimization objective