J. Vis. Commun. Image R. 32 (2015) 205–216

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

No reference image quality assessment metric via multi-domain structural information and piecewise regression ☆

Qingbo Wu [a,*], Hongliang Li [a], Fanman Meng [a], King Ngi Ngan [a,b], Shuyuan Zhu [a]

[a] School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu, China
[b] Department of Electronic Engineering, The Chinese University of Hong Kong, ShaTin, Hong Kong

ARTICLE INFO

Article history:
Received 17 October 2014
Accepted 15 August 2015
Available online 28 August 2015

Keywords:
No reference image quality assessment
Human vision system
Image representation
Quality aware feature
Multi-domain structural information
Gradient of wavelet domain
Piecewise regression
HEVC

ABSTRACT

The general purpose no reference image quality assessment (NR-IQA) is a challenging task, which faces two hurdles: (1) it is difficult to develop one quality aware feature which works well across different types of distortion and (2) it is hard to learn a regression model to approximate a complex distribution for all training samples in the feature space. In this paper, we propose a novel NR-IQA method that addresses these problems by introducing the multi-domain structural information and piecewise regression. The main motivation of our method is based on two points. Firstly, we develop a new local image representation which extracts the structural image information from both the spatial-frequency and spatial domains. This multi-domain description could better capture human vision property. By combining our local features with a complementary global feature, the discriminative power of each single feature could be further improved. Secondly, we develop an efficient piecewise regression method to capture the local distribution of the feature space. Instead of minimizing the fitting error for all training samples, we train the specific prediction model for each query image by adaptive online learning, which focuses on approximating the distribution of the current test image's k-nearest neighbor (KNN). Experimental results on three benchmark IQA databases (i.e., LIVE II, TID2008 and CSIQ) show that the proposed method outperforms many representative NR-IQA algorithms.

1. Introduction

No reference image quality assessment (NR-IQA) aims at designing a computational model to automatically predict the perceptual quality of a test image without its undistorted reference version [1–3]. In many practical applications, the reference images are unaccessible, which makes the NR-IQA algorithm more desirable in comparison with the full-reference (FR)-IQA [4–7] and the reduced-reference (RR)-IQA [8–12] metrics. For example, many denoising algorithms require the manual parameters to obtain a good result. Since there is no reference image, the NR-IQA is very desirable for parameter tuning, such as [13]. Similarly, the NR-IQA based image enhancement applications have also been discussed in recent works [14,15]. To date, many NR-IQA methods have been developed, which are usually composed of the quality aware feature extraction and perceptual quality regression modules.

The quality aware feature corresponds to an efficient image representation, which could capture the variation of the perceptual image quality caused by the distortion. Many efficient features have been proposed in existing algorithms. Seo et al. introduce the visual saliency into the perceptual quality metric in [16]. Ghanem et al. [17] utilize the inter/intra-segment interactions to measure the image quality variation. In [18,19], Moorthy et al. employ the parametric statistical model to describe the natural scene statistics (NSS) of an image in the wavelet domain. Meanwhile, some nonparametric wavelet statistics are also introduced in [20]. In [21,22], the NSS information from the multi-scale DCT domain is utilized. In [23], Tang et al. propose the α-stable model in the wavelet domain to describe the image, and Sang et al. [24] introduce the singular value curve to measure the image degradation caused by blur. After extracting the image features, the regression module is used to map these quality aware features to a subjective quality score. Many different regression schemes have been discussed in existing methods, such as, the general regression neural network (GRNN) [25], the support vector regression (SVR) [18,19] and the multiple kernel learning (MKL) [20].

Although the aforementioned methods have achieved good result in capturing the human perception, some important local

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

* Corresponding author. Fax: +86 028 61830064.
E-mail addresses: wqb.uestc@gmail.com (Q. Wu), hlli@uestc.edu.cn (H. Li), fmmeng@uestc.edu.cn (F. Meng), knngan@ee.cuhk.edu.hk (K.N. Ngan), eezsy@uestc.edu.cn (S. Zhu).

http://dx.doi.org/10.1016/j.jvcir.2015.08.009
1047-3203/

properties in both the image representation and perceptual quality regression are still underutilized. Firstly, in the feature extraction module, the NSS is usually extracted from some specific domains [18–22]. When the transform coefficients are quantized into different bins, only the global frequency distribution is reserved and the local spatial neighboring relationship is lost. In [19,20], the neighboring coefficients' joint distribution were introduced to address this issue. However, the similar problem still exists in the coefficients' quantization process. Secondly, in the regression module, the existing methods usually learn a map function by off-line training, which focuses on minimizing the average fitting error over all training samples. For some local training data, its performance may not be very well [26].

Based on these analysis, we develop our method from two aspects: (1) For the image representation, we introduce a new local image representation, which is then combined with a widely-used global image feature. As discussed in [27,28], the visual cortex integrates both the spatial-frequency and spatial information. Most existing methods don't consider this characteristic and only extract the image features from single domain, such as, wavelet domain [18,19], DCT domain [21,22] or spatial domain [29]. In contrast, we describe the local image structures from both the spatial-frequency and spatial domains. Particularly, the spatial-frequency information is derived from a novel orientation statistics of the gradients in the local patch of each wavelet subband. The spatial information is captured with the classic local binary pattern (LBP) [30]. Since the human vision captures both the local and global information from the natural image [31], we further introduce the global distribution of the wavelet coefficients to compensate our local features. (2) For the perceptual quality regression, we design an efficient local regression method to further improve the prediction accuracy. Inspired by our previous work [32], the piecewise regression criterion [33] is employed in our method. Unlike previous single-phase regression [33] which uses all training data to learn a regressor, we try to build the appropriate segmented data or training sample subset for each test image. Then, the specific regression parameters can be learned from these training sample subset by online training. Experimental results on three IQA databases show that the proposed method is highly efficient and robust.

The rest of this paper is organized as follows. Sections 2 describes the quality-aware features in our method and Section 3 presents the proposed piecewise quality regression model. The experimental results are shown in Section 4. Finally, we draw the conclusion in Section 5.

2. Quality aware features

For the image representation, the spatial domain, transform domain [34,35], saliency [36,37] and segmentation [38–42] information have been widely used. Among these analysis tools, the wavelet transform is particularly popular due to its abilities in multiresolution analysis and spatial-frequency information representation [35]. For clarity, the spatial-frequency structure of four scales wavelet transform is illustrated in Fig. 1. The label in the top-right corner of each sub-image denotes its scale and direction information. The HLx, LHx and HHx represent the horizontal, vertical and diagonal details for the scaled image under the xth level. The block label in the top-left corner denotes the subbands with the same frequency from the fine scale to the coarse one. In this section, we describe the quality aware local and global features in details.

2.1. Local multi-domain structural information

In order to represent the local spatial-frequency structure in the wavelet domain, we partition each subband into non-overlapped cell units and count the distribution of the gradient in some quantized orientations in the form of histogram of oriented gradient (HOG) [43]. Here, the wavelet coefficient's gradient is first calculated by convolving the 1-D derivatives ($[-1;0;1]$) in the horizontal and vertical directions, where g_x and g_y denote the horizontal and vertical gradient in each location of a cell. Then, the gradient orientation G_o and the gradient magnitude G_m can be calculated as,

$$
\begin{aligned}
G_o &= \arctan(g_y/g_x) \\
G_m &= \sqrt{g_y^2 + g_x^2}
\end{aligned} \tag{1}
$$

As shown in Fig. 2, the gradient orientation histogram statistic is applied in each cell, where the histogram bins indicate the quantized gradient orientation ranging from -180° to 180°. The gradient magnitude voting scheme is employed in counting the histogram. In our method, we count the wavelet domain gradient

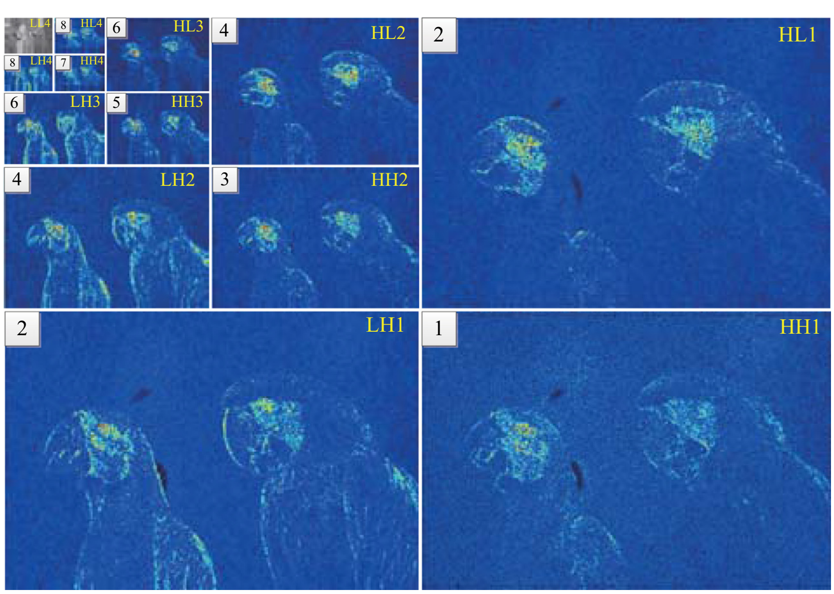

Fig. 1. The wavelet subband and frequency band locations.