Signal Processing 128 (2016) 186–197

Contents lists available at ScienceDirect

Signal Processing

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

Full-reference image quality assessment by combining global and local distortion measures

Ashirbani Saha *, Q.M. Jonathan Wu

Department of Electrical and Computer Engineering, University of Windsor, Canada N9B 3P4

ARTICLE INFO

Article history:
Received 23 March 2016
Accepted 30 March 2016
Available online 31 March 2016

Keywords:
Image quality assessment
Saliency
Image gradient
Local standard deviation

ABSTRACT

Full-reference image quality assessment (FR-IQA) techniques compare a reference and a distorted/test image and predict the perceptual quality of the test image in terms of an objective score. Evaluation of FR-IQA techniques is carried out by comparing the objective score with the image's subjective score obtained through human observer ratings. The goal of an observer is to rate the distortion present in the test images. The goal oriented tasks are processed by the human visual system (HVS) through top-down processing which actively searches for local distortions driven by the goal. Therefore local distortion measurement is important for the top-down processing. Simultaneous bottom-up processing also takes place signifying spontaneous visual functions in the HVS. To account for this, global perceptual features can be used. Therefore, we hypothesize that the objective score for an image can be derived from the combination of local and global distortion measures calculated from the reference and test images. We calculate the local distortion by measuring the local correlation differences from the gradient and contrast information. For global distortion, dissimilarity of the saliency maps computed from a bottom-up model of saliency is used. Experimental analysis conducted in six benchmark databases suggest the effectiveness of the proposed approach.

1. Introduction

Automatic evaluation of the perceptual quality of a distorted image with respect to the original high quality image is technically called full-reference image quality assessment or FR-IQA. Though methods for evaluating subjective image quality by directly employing human observers are available, they are time consuming and less economic compared to an automatic evaluation using an FR-IQA technique. Given the applicability of FR-IQA in image acquisition, watermarking [1], restoration and denoising [2], fusion [3] etc., it has gained immense popularity and attention since the last decade. Image quality and video quality evaluations are important elements of multimedia processing. They can be used together with text processing to serve for contextual quality evaluation and searching in large databases [4–8]. There are FR-IQA techniques that employ the properties of Human Visual System (HVS) directly, and there are others, which have hypotheses supporting some properties of HVS. Some methods of FR-IQA are combined approaches involving both of the aforesaid techniques [9]. Though peak-signal-to-noise ratio (PSNR) follows none of the approaches, it is one of the best perceptual evaluators for mainly content-independent noise [10,11]. Still, the general applicability of PSNR is very low when considered with other types of distortions or across several types of distortions. This is where the development and enhancement of FR-IQA techniques became necessary [9]. Therefore, from the last decade, a number of different techniques has been developed for FR-IQA. We have the popular SSIM [12], which is based on the property of HVS to extract structural similarity. Several SSIM based or inspired approaches and analyses [13,14] have been developed since then. MSSSIM [15] incorporates some flexibility to SSIM by embedding multiscale information in it. Also, information content weighted SSIM, called IWSSIM [10] uses pooling based on information content calculated using mutual information from three pairs of images selected from reference, distorted, perceived reference and perceived distorted images. However, the first methods to use mutual information in FR-IQA are Information Fidelity Criterion (IFC) [16] and Visual Information Fidelity (VIF) [17]. Hence, they are called information theoretic approaches. Among other approaches are phase congruency based technique feature similarity (FSIM/FSIM$_c$ for color images) [11] and its modifications to form phase deviation sensitive energy features based similarity technique (PDSESIM) for FR-IQA [18] to improve the general applicability. In a similar framework, a faster technique called spectral residual based similarity (SR-SIM) [19] has been developed and shown to perform well in three databases. SR-SIM uses saliency maps derived from spectral

* Corresponding author.
E-mail address: sahai@uwindsor.ca (A. Saha).

http://dx.doi.org/10.1016/j.sigpro.2016.03.026
0165-1684/

residual technique to find perceptual similarity between pixels as well as to pool the quality map obtained from the reference and distorted images. The work presented in [20] points out the parameter dependence of the said framework by experimental analysis. A number of techniques show interesting steps towards separating distortions in an image according to some pre-defined criteria. Most Apparent Distortion (MAD) [21] segregates the distortions present in an image as near threshold or supra-threshold and uses dual strategies for evaluating those distortions. In the same context, another approach is based on defining the distortions as additive or detail loss based and combining their measurements (ADM) [22]. A recent approach [23] uses the principal of internal generative mechanism (IGM) of human brain and dissolves the given images into predicted and disorderly parts. Then the distortions on these two parts are separately evaluated by different techniques and combined to form the objective score. MAD, ADM and Visual-Signal-to-Noise ratio (VSNR) [24] employ the direct applications of the HVS properties like contrast sensitivity and visual masking. Most of the methods mentioned in the previous discussions form the state-of-the-art techniques of FR-IQA. Some recent approaches are developed that compute full-reference image quality based on gradient similarity [25], visual saliency based modeling [26] and sparse representation [27]. Adaptively combining the inter-patch and intra-patch similarity using gradient magnitude and isophote curvature comparison [28], resulted in an improved FR-IQA technique. A FastICA based technique designed to work on color distortion of images is proposed in [29]. The existence of so many techniques validates the general observation that each of them has some shortcomings for certain distortions and databases on which they are evaluated. An FR-IQA technique that performs well for all degradations (provided in the benchmark databases) does not exist. Thus, development of an FR-IQA technique that performs well with all possible types of distortions is still an open problem.

As already discussed, MAD, ADM and IGM apply different strategies to separately evaluate different aspects of distortion. Inspired by all of these, the aspect we focus on is the automatic evaluation of image quality based on global and local perceptual visual cues and their combination. A recent possible way to explore global and local structures in images is graphlets and they have been used in different applications such as photo cropping [30], photo aesthetics evaluation [31,32] and vehicular communication [33]. In photo aesthetics evaluation [32], the methodology used combines local and global structural cues for computing the score of the photo. Graphlets [34] (used for capturing the interaction of spatially neighboring atomic regions) are used as local aesthetics descriptor and a manifold embedding technique is used to encode global spatial layout information of the photo into the graphlets. However, the goal of our work is to quantify the perceptual quality rather than photo aesthetics and hence we have used global and local distortion features being motivated by the following. As explained in [35], visual processing is a simultaneous combination of bottom-up and top-down processes. Bottom-up processing focuses on highlighting the relatively important regions in an image [36]. On the other hand, top-down processing caters to goals and it searches actively for local features [37] based on contextual information. In HVS, these processes occur simultaneously and quickly. Whenever, a human subject is given the task of evaluating perceptual quality, he has the final goal to rate the distortions present. Based on this, we hypothesize that three main events taking place during this evaluation: 1. Initially, the global image content determines the varying attention (region-wise) of the evaluator; 2. existing regional distortions are used to assign local quality evaluations; 3. finally, based on the attention, a refinement on the previous evaluation is carried out and final rating is provided. To serve the first event, we resort to the global perceptual features of the image. For the second step, local attributes between the reference and test/distorted images are compared. Finally, third step involves a pooling strategy taking place based on the global perceptual features. In the present work, the perceptual quality of an image is expressed in terms of image saliency maps (as global perceptual features) obtained from bottom-up saliency models, gradient information and local standard deviation (local features) also termed as RMS (root-mean-squared) contrast. The saliency map of an image represents global information about how often a particular pixel is gazed at. On the other hand, local standard deviation and gradient information signify the local and contextual information of any pixel. The local correlation between the global information obtained from the reference and distorted images is computed. The local correlation between the gradient information obtained from the reference and distorted images is also calculated. These local correlations are combined with the local RMS contrast between the images. From the experimental results, we find that the integration of simple visual details of global perceptual difference information and local information may result in an effective FR-IQA technique. It differs from its predecessors in terms of treatment of local and global features by using the regional correlation. It has been shown in [38], gradient is structure-variant as well as contrast-variant. Thus similar variations in gradient magnitude values and standard deviation are expected for a pixel. However, change in standard deviation may not be caused by change in gradient magnitude only. Gradient orientation is also affected by presence of distortion. Thus, the proposed approach applies all of these visual details to arrive at the quality score. The performance analysis of the technique in six benchmark databases shows the promise of the proposed method as a competitive technique in FR-IQA. Also, we carry out analysis on distortion dependent performance of the FR-IQA techniques using color based representation. This representation of the results clearly shows that with certain distortions, the FR-IQA techniques fail to perform well.

The remaining parts of the paper are arranged in the following manner. The motivating factors for the proposed approach are presented in Section 2. The details of the proposed method are presented in Section 3. The performance of the proposed method is analyzed in Section 4. The concluding remarks are presented in Section 5.

2. Motivating factors

The motivation behind the proposed approach is to combine simple visual cues, representing the global and the local information present in an image, to formulate a competitive FR-IQA technique. We have selected saliency maps as global perceptual features for an image. For local features, block based gradients and standard deviations are used. The proposed technique obtains global distortion information by comparing global features obtained from the reference and distorted images using local correlation. The local distortion information is obtained by comparing local features using local correlation and local difference. In the following sections, we describe the importance of each of the global features and the roles they are expected to play as parts of an FR-IQA technique. Based on their properties, we hypothesize that the effective combination of these simple features can result in a competitive FR-IQA technique.

2.1. Global features

The saliency map of an image is chosen as a representation of the global perceptual features of an image. Since saliency maps have perceptual information contained in them, they have been