Signal Processing: *Image Communication* 29 (2014) 1138–1148

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Sparse representation-based image quality assessment

Tanaya Guha*, Ehsan Nezhadarya, Rabab K. Ward

Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada

A R T I C L E I N F O

Article history:
Received 12 November 2013
Received in revised form
30 September 2014
Accepted 30 September 2014
Available online 13 October 2014

Keywords:
Dictionary learning
Image quality assessment
Sparse representation
Structural similarity

A B S T R A C T

A highly promising approach to assess the quality of an image involves comparing the perceptually important structural information in this image with that in its reference image. The extraction of the perceptually important structural information is however a challenging task. This paper employs a sparse representation-based approach to extract such structural information. It proposes a new metric called the *sparse representation-based quality* (SPARQ) *index* that measures the visual quality of an image. The proposed approach learns the inherent structures of the reference image as a set of basis vectors. These vectors are obtained such that any structure in the image can be efficiently represented by a linear combination of only a few of these basis vectors. Such a sparse strategy is known to generate basis vectors that are qualitatively similar to the receptive field of the simple cells present in the mammalian primary visual cortex. To estimate the visual quality of the distorted image, structures in the visually important areas in this image are compared with those in the reference image, in terms of the learnt basis vectors. Our approach is evaluated on six publicly available subject-rated image quality assessment datasets. The proposed SPARQ index consistently exhibits high correlation with the subjective ratings of all datasets and overall, performs better than a number of popular image quality metrics.

1. Introduction

Digital images incur a variety of distortions during their acquisition, compression, transmission, storage or reconstruction. Such processes often degrade the visual quality of images. In order to monitor, control and improve the quality of images produced at the various stages, it is important to *automatically* quantify the image quality. As the end-users of the majority of image-based applications are humans, the automatic quantification of image quality requires the understanding of human perception of image quality, so as to mimic it as closely as possible.

The *mean squared error* (MSE) and the *peak signal-to-noise ratio* (PSNR) have been traditionally used to measure the image quality degradations. These metrics are mathematically convenient to use but they do not correlate well with human perception of image quality [1]. A considerable amount of research effort has been put towards quantifying the quality of images as *perceived* by humans, and a number of *objective* image quality assessment algorithms that agree with the subjective judgment of human beings have been developed. The objective quality assessment methods, depending on how much information about the original undistorted image they use, are broadly classified into three categories: no-reference, reduced-reference and full-reference. This paper concentrates on the *full-reference* quality estimation approach.

Earlier focus of the full-reference image quality assessment research has been on building a comprehensive and accurate model of the *human visual system* (HVS) and its

* Corresponding author.
E-mail addresses: tanaya@ece.ubc.ca (T. Guha), ehsann@ece.ubc.ca (E. Nezhadarya), rababw@ece.ubc.ca (R.K. Ward).

http://dx.doi.org/10.1016/j.image.2014.09.010
0923-5965/

psychophysical properties, such as the contrast sensitivity function. In this approach, errors between the distorted and the reference images are quantized and pooled according to the HVS properties [2]. These methods require precise knowledge of the viewing conditions and are computationally demanding.

The interest in modern image quality estimation research has since shifted to modeling the visual content of images based on certain significant properties of the HVS. This *visual fidelity-based* approach is more attractive because of its practicality and mathematical foundation [3,4]. The majority of these fidelity-based methods attempt to quantify the perceptual quality either in terms of the *statistical information* or in terms of the *structural information* of images. The statistical approaches [5,6] hypothesize that the HVS has over the years evolved to extract information from natural scenes and therefore, use the natural scene statistics to estimate the perceptual quality of images. The structural approaches [7–13] on the other hand operate on the basis of a rather important aspect of the HVS – its sensitivity towards the image structures for developing cognitive understanding. In this approach, image quality is estimated in terms of the *fidelity of the structures* in the distorted image relative to that in the reference image.

An image quality metric that is representative of the class of structural information-based metrics is the *structural similarity index* (SSIM) [7]. SSIM treats the structural distortions separately from the non-structural distortions (such as luminance and contrast change). The visual quality of a patch in the distorted image is measured by comparing it with the corresponding patch in the reference image in terms of three components: luminance, contrast and structure. A global quality score is computed by combining the effects of the three components over all image patches. SSIM achieved much success because of its simplicity, and its ability to tackle a wide variety of distortions. Due to its pixel-domain implementation, SSIM is highly sensitive to geometric distortions like scaling, translation, rotation and other misalignments [2]. To improve the performance of SSIM, multiscale extension [9], wavelet transform-based modification [12], gradient-domain implementation [10] and various pooling strategies [11,14] have been proposed.

The underlying assumption behind utilizing the structural information is that the HVS uses the structures it extracts from the viewing field for its cognitive understanding. Therefore, for an image to be considered of high-quality, all the structural information present in its reference image should be well preserved. From this viewpoint, the efficient capture of the structural information of images is the key to developing a successful image quality assessment algorithm. But extracting or analyzing the structural information in a perceptually meaningful way is a non-trivial task. Although the use of predefined, data-independent basis functions is prevalent in analyzing image structures, their success is often limited by the degree as to how suitable the basis functions are in capturing the structural information of the signals under consideration.

We propose the use of a more generalized approach to analyzing image structures in the context of image quality assessment. This involves *learning* from the training data a set of basis elements that could be adapted to represent the inherent structures of the signal in question. These learnt basis elements are collectively known as a *dictionary*. As each basis vector could be tailored to represent a significant part of the structures present in the given data, a learnt dictionary is more efficient in capturing the structural information compared to a predefined sets of bases. In the last few years, several practical dictionary learning algorithms have been developed [15,16]. It has been shown that the data-dependent, learnt dictionaries, due to their superior ability to efficiently model the inherent structures in the data, can outperform predefined dictionaries like wavelets in several image processing tasks [15,17,18]. Dictionary learning ideas have also been explored in the context of measuring generic image similarity, applicable to problems like clustering or retrieval [19,20]. The problem of perceptual image quality assessment, although related to image similarity measures, is more specific and needs to be addressed as a different problem.

This dictionary learning approach is connected to an important result obtained by Olshausen and Field in 1996 [21]. They proposed to learn a set of basis elements from the input itself instead of designing a new basis function that could sparsely represent the input signal. They enforced (i) a *sparsity* prior – an assumption that it is possible to describe the input using a small number of basis elements, (ii) *overcompleteness* – the number of basis elements is greater than the vector space spanned by the input. They showed that this strategy results into a set of basis elements that are localized, oriented, and bandpass in nature, which resemble the properties of the receptive field of simple cells in the primary visual cortex [21].

In this paper, we develop a full-reference image quality assessment metric which we call the *sparse representation-based quality* (SPARQ) *index*. This metric relies on capturing the inherent structures in the reference image as a set of basis vectors which collectively form an overcomplete dictionary. These vectors are obtained such that any structure (patch) in the image can have a sparse representation w.r.t. the dictionary so as to resemble the visual cortex. To estimate the visual quality of the distorted image the visually important regions in this image are compared with those in the reference image. These regions are detected by using a saliency detection algorithm. Each patch in the visually important region in the distorted image and its corresponding patch in the reference image are decomposed in terms of the learnt dictionary. Their sparse coefficients are then compared to yield a measure of quality of the distorted image. Since our method analyzes image structures by building a cortex-like model of the stimuli, we expect the extracted structural information to be important to the HVS, and perceptually more meaningful compared to the structural information used in existing methods.

To evaluate the efficacy of the proposed metric, we perform various experiments on six publicly available, subject-rated image quality assessment datasets. The proposed SPARQ index shows great promise as it consistently exhibits high correlation with the subjective scores and often outperforms its competitors.

The rest of the paper is organized as follows. Section 2 describes the proposed quality assessment approach in detail; this is followed by experimental results and discussion in Section 3. Section 4 concludes the paper and suggests possible directions to future work.