Signal Processing: Image Communication 58 (2017) 212–227

Contents lists available at ScienceDirect

Signal Processing: Image Communication

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

Towards the design of a consistent image contrast enhancement evaluation measure

Muhammad Ali Qureshi [a,b,*], Azeddine Beghdadi [c], Mohamed Deriche [a]

[a] King Fahd University of Petroleum and Minerals (KFUPM), Dhahran 31261, Saudi Arabia
[b] The Islamia University of Bahawalpur, 63100, Pakistan
[c] L2TI, Institut Galilée, Université Paris 13, Sorbonne Paris Cité, France

A R T I C L E I N F O

Keywords:
Image enhancement
Image quality assessment
Subjective experiments
Image contrast
Contrast enhancement evaluation measures
Contrast changed database

A B S T R A C T

Contrast Enhancement Evaluation (CEE) is a very challenging problem. In this work, we provide a detailed performance analysis of CEE measures. The study was conducted on our newly developed database dedicated to psychophysical Contrast Enhancement (CE) quality evaluation. The database contains 30 original images and 180 enhanced images obtained using six different CE methods as a representative set of the most common approaches used in the literature. The CE methods were subjectively evaluated and ranked by 23 observers using a Pairwise Comparison (PC) protocol. The correlation analysis between the subjective preferences and objective evaluations of the enhanced images show that most of the existing CEE metrics are not well consistent with the human judgment of quality. We also present in this paper a thorough discussion on the available CEE metrics, their strengths, their weaknesses, and their inter-correlation. In addition to the individual metrics, we show that by fusing different metrics together, a significant increase in correlation performance can be achieved. This study reveals that there is a clear need to develop more robust CEE measures which are perceptually motivated and correlated well with the quality of enhancement a given image is subjected to. The new database introduced in this paper is expected to contribute substantially promoting such research area, which is of primary importance to diverse multimedia applications.

1. Introduction

Image Quality Assessment (IQA) has attracted a lot of interest during the last three decades, and a plethora of efficient and advanced IQA measures have been proposed [1–3]. However, some simple and practical image quality measures, such as Mean Square Error (MSE) based measures, are still in use in some multimedia applications such as bit-rate optimization for video coding [4]. This is mainly due to their mathematical tractability and the absence of a well-established standard for measuring image quality. The notion of visual information fidelity or image quality is highly related to the way humans perceive distortions that may affect the quality of the observed image [5,6]. Therefore, the IQA dilemma, in its traditional sense, has been long considered as a distortion estimation problem [7]. This, of course, is an important problem as it is desirable to have ready to use techniques to evaluate the quality of images subject to distortions or artifacts that may result from processing, lossy compression, or transmission. On the other hand, very few studies have been done on the performance evaluation of image enhancement methods (better quality images rather than distorted images). Indeed, performing a quantitative evaluation of image quality enhancement methods is a very challenging task. This is due to the absence of any objective measures able to account for some high-level vision tasks and their interaction with low-level image analysis when assessing the perceptual quality of image enhancement [2]. This is also due to the difficulty in determining the most appropriate visual features to be used in the design of an overall image enhancement quality measure. Therefore, subjective evaluation is still the most reliable approach to assess the quality of enhanced images.

Enhancing image contrast is of major interest in many applications ranging from medical imaging [8], remote sensing [9], underwater imaging [10], image forensics [11], defogging [12], etc. A plethora of Contrast Enhancement (CE) methods has been proposed in the literature, and it becomes rather difficult to provide a comprehensive and complete survey of published work in this area. Moreover, there is no study to test the reliability of these measures themselves. Given the importance of CE in different applications, there is a need to

* Correspondence to: Department of Telecommunication Engineering, The Islamia University of Bahawalpur, 63100, Pakistan.
E-mail addresses: ali.qureshi@iub.edu.pk (M.A. Qureshi), azeddine.beghdadi@univ-paris13.fr (A. Beghdadi), mderiche@kfupm.edu.sa (M. Deriche).

http://dx.doi.org/10.1016/j.image.2017.08.004
Received 16 August 2016; Received in revised form 3 August 2017; Accepted 4 August 2017
Available online 14 August 2017
0923-5965/

investigate the performance of these measures in terms of robustness and consistency with human judgment.

One of the first studies on Contrast Enhancement Evaluation (CEE) has been proposed in [13]. However, it was only restricted to images containing two classes of pixels (i.e., one object on a uniform background or many similar objects on a uniform background). The CE evaluation was based on the bimodality analysis of the gray-level distribution. Thereafter, some simple and interesting CEE measures have been proposed in [14–17]. These measures are not inspired by the classical approaches of IQA. The proposed measures are based on the computation of a global index derived from some local measures related to contrast. These are inspired originally by Michelson and Weber–Fechner contrast measures. These measures are based on min–max operations that make them more noise sensitive. The authors proposed some improvements to overcome these limitations by using entropy of local contrast, or by introducing logarithmic arithmetic operations inspired by the non-linear Human Visual System (HVS) response. We should note, that in the study conducted in [14–17], no complete subjective experiments were performed, and the performance analysis was only based on the perceptual judgment of output images. Moreover, the tests were conducted on a limited set of images (very often grayscale images), and the measures were not evaluated on any dedicated database but only on few images from the TID2013 database that has been built for traditional IQA purpose [18]. Furthermore, the statistical analysis of these measures and comparison with some representative CE methods were also missing.

In contrast, Vu et al. [19], proposed another study based on a database containing processed images obtained by changing color, saturation, brightness, sharpness, and their combinations. The subjective evaluation was performed to assess the quality of processed images. The use of classical IQA approaches in a reverse order was proposed, i.e., the given image (enhanced image) is considered as the reference and the original image as the distorted one. It has also been reported that the Visual Information Fidelity (VIF) [20] measure offers better performance as compared to many of the classical IQA measures. The authors in [19] improved the results by proposing a more efficient measure combining contrast, sharpness, and color in an empirical manner.

Following the approach of Vu et al. [19], another study of contrast change evaluation was discussed in [21] using a database consisting of 15 original and 633 enhanced images. The global contrast of images is modified using non-linear mapping functions. The conventional IQA measures designed for degradations assessment were then used to assess the quality of the processed images from the database. For this purpose, a reduced reference metric was derived combining the entropy of phase congruency image and other higher-order statistics of local features computed from the histogram of the observed image.

Another recent study, by Fang et al. [22] on quality assessment of contrast distorted images was carried using the natural scene statistics model. The contrast problem is considered only in terms of distortion. However, in our case, we enhance the contrast and, for CEE, we try to account for the undesirable side effects that may result from CE process.

Besides these works, predicting visual quality of enhanced/modified images for different applications has also been investigated in some interesting studies [23–29]. Ledda et al. [23] proposed a database for only subjective evaluation of six tone mapping methods. The Pairwise Comparison (PC) was performed in a subjective experiment to rank these methods in accordance with the perceived quality. But the authors did not perform CEE performance analysis. Virtanen et al. [24] provided another database related to tone-mapping applications. It contains images degraded with different types of distortions and images with variation of contrast due to gamut mapping. The main objective of the database was to validate the performance of existing IQA metrics designed mainly for degraded images. Another similar database was also proposed in [28] to evaluate gamut mapping, blurring, and other distortions. Similarly, sharpening algorithms are also used to enhance the perceived quality of a given image. In [29], quality evaluation of sharpened images is investigated. A new subjective experiment framework, specifically adopted for the quality assessment of sharpened images was introduced.

In addition to the above, Chen et al. [27] developed a database for CEE of images in bad visibility (i.e., haze, underwater, and low light environment). The images were enhanced through different dehazing methods and the performance of various enhancement algorithms was discussed. In this work, the original and pair of enhanced images were shown on the same screen to allow the observer to compare the enhanced images with respect to the original image.

Another less studied application, namely image retargeting quality assessment, has been addressed in [25,26]. Here, subjective and objective quality evaluation of retargeted images was performed using dedicated databases. In [25], the authors provided a database containing images by different retargeting methods. The subjective quality of the retargeted images was measured in terms of rank in a pairwise subjective experiment, and the performance of different retargeting evaluation measures were assessed in terms of correlation analysis. Similarly, Ma et al. [26] also carried out the same study except, instead of ranking, they provided the rating scores on a different proposed database. Although image retargeting, Gammut Mapping, Dehazing, and Tone Mapping have no direct relation with CE, the main purpose of discussing these works in this paper is to provide information about different subjective experiments performed with the same goal, i.e., performance evaluation of objective quality measures designed for diverse applications. To summarize the related works carried to date, we provide, in Table 1, our own perspective on the main contributions made in this field of research.

It is worth noting, that our methodology differs from previous works in many aspects; (1) The objectives are not the same. We aim here to analyze the performance of CEE measures in contrast to the work in [19,21,24] in which the performance analysis of classical IQA measures (i.e., IQA for distortions) was discussed, (2), The database is not the same compared with classical IQA databases like TID2013 [18], CSIQ [30] containing contrast images and few existing contrast databases [19,21]. These databases contain simulated changes of global contrast using a simple pixel value mapping function so as to produce a decrease in contrast. The authors consider these transformations as contrast distortion. Whereas, in our framework, we deal with the artifacts and distortion that may happen when applying CE operations. The distortion that might appear in the enhanced images after processing with CE methods are, for example, color saturation, color loss, blocking and ringing amplification in the case of compressed images, noise amplification, and halo effects and some others. The common databases did not contain any of these after effects due to CE. Moreover, in our case, we use different representative CE methods. (3), In contrast of all the databases, we do not want to estimate distortion in terms of decrease in quality like in classical IQA, rather our goal is to assess and quantify, subjectively and objectively, the increase in quality. (4), The application is entirely different compared to the CEE of tone mapped and retargeted images [23,25,26].

To the best of our knowledge, there are only two dedicated databases related to contrast manipulation [19,21], where the processed images are obtained using simple artificial pixel-based transformations. Whereas, in our proposed database, some realistic CE artifacts are considered and provided with subjective ranking of different CE methods, which can be used to validate the performance of new CEE measures. The proposed database will help in preliminary validation of new image CEE measures without performing dedicated subjective experiments. The main contributions of this work are:

- To provide a comprehensive performance analysis of the state-of-the-art CEE measures in terms of correlation with the subjective evaluation provided in the developed database as well as on other existing contrast manipulated databases.