Computers in Biology and Medicine 90 (2017) 68–75

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

No-reference quality index for color retinal images

Lamiaa Abdel-Hamid [a,*], Ahmed El-Rafei [b], Georg Michelson [c,d]

[a] Misr International University, Faculty of Engineering, Dept. of Electronics and Communication, Cairo, Egypt
[b] Ain Shams University, Faculty of Engineering, Dept. of Engineering Physics and Mathematics, Cairo, Egypt
[c] Friedrich-Alexander University of Erlangen-Nuremberg, Dept. of Ophthalmology, Erlangen, Germany
[d] Talkingeyes & More GmbH, Erlangen, Germany

ARTICLE INFO

Keywords:
Retinal image quality assessment
Quality index
Image sharpness
Image homogeneity
Wavelet transform

ABSTRACT

Retinal image quality assessment (RIQA) is essential to assure that the images investigated by ophthalmologists or automatic systems are suitable for reliable medical diagnosis. Measure-based RIQA techniques have several advantages over the more commonly used binary classification-based RIQA methods. Numeric quality measures can aid ophthalmologists in associating a degree of confidence to the diagnosis performed through the investigation of a certain retinal image. Moreover, a numeric quality index can provide a mean for identifying the degree of enhancement required as well as to evaluate and compare the improvement achieved by enhancement techniques. In this work, a no-reference retinal image sharpness numeric quality index is introduced that is computed from the wavelet decomposition of the images. In order to account for the obscured retinal structures in unevenly illuminated image regions, the quality index is modified by a homogeneity parameter calculated from the previously introduced retinal image saturation channel. The proposed quality index was validated and tested on two datasets having different resolutions and quality grades. A strong (Spearman's coefficient > 0.8) and statistically highly significant (p-value < 0.001) correlation was found between the introduced quality index and the subjective human scores for the two different datasets. Moreover, multiclass classification using solely the devised retinal image quality index as a feature resulted in a micro average F-measure of 0.84 and 0.95 using the high and low resolution datasets, respectively. Several comparisons with other retinal image quality measures demonstrated superiority of the proposed quality index in both performance and speed.

1. Introduction

Retinal images are being increasingly used by ophthalmologists as well as in automatic systems for medical diagnosis and follow-up of retinal diseases such as diabetic retinopathy, glaucoma, and age-related macular degeneration. However, some retinal images can be unsuitable for reliable medical analysis and diagnosis due to their insufficient quality, most commonly due to reduced sharpness and/or uneven illumination. Poor quality retinal images can occur due to several factors including inadequate imaging conditions (e.g., insufficient illumination, poor focus) or patient related issues (e.g., pupil dilation, patient fixation, media opacity) [1,2]. Several studies have shown that practical datasets can include a large number of poor quality retinal images that can be as high as 60% of the total images [3,4].

Various shortcomings can result from employing bad quality retinal images in medical investigations. If a poor quality retinal image is not immediately identified, a recapture would be requested by the ophthalmologist costing himself and the patient both time and money, especially in cases when the imaging procedure and medical facility are in distant locations. Nevertheless, a worse scenario can occur if automatic screening systems are used for the medical diagnosis. Automatic retinal screening systems capture and process retinal images without any human intervention. Based on this analysis, the patient is advised to or against the need for further physical investigation depending on whether or not early disease symptoms were detected in the processed images. If poor quality retinal images are used within the automatic systems, misdiagnosis could occur leading to delayed treatment. As a result, further disease progressions can occur causing irreversible visual impairments that could lead to blindness.

No-reference retinal image quality assessment (RIQA) algorithms are being increasingly integrated as a preliminary preprocessing step in medical retinal image analysis in order to assure reliability of the performed diagnosis. RIQA algorithms automatically determine whether the captured images are suitable for reliable medical analysis in the absence

* Corresponding author.
E-mail address: lamiaa.a.hamid@miuegypt.edu.eg (L. Abdel-Hamid).

https://doi.org/10.1016/j.compbiomed.2017.09.012
Received 20 July 2017; Received in revised form 9 September 2017; Accepted 18 September 2017

of a gold standard reference image. Good quality retinal images are then stored and passed on for further processing, whereas poor quality images are either enhanced, if possible, to improve their quality or discarded altogether and an image recapture is performed. Generally, RIQA algorithms can be categorized into classification- and measure-based methods [5].

Classification-based RIQA algorithms rely on supervised learning methods to classify images into a specific quality class. Binary classification-based RIQA methods are very widely implemented in literature [3,4,6–15] where retinal images are considered to be either of excellent quality making them readily suitable for direct medical investigation or of severely poor quality rendering them unsuitable for medical diagnosis. However, captured retinal images may be of adequate quality which would require the application of proper image enhancement techniques (e.g., sharpening, luminosity improvement) before they are suitable for medical analysis. In this study, a practical dataset manually graded by human experts into good, adequate, and poor quality was found to have ∼45% of its images of adequate quality. In a previous work, Katuwal et al. [16] presented a multi-classification RIQA algorithm that categorized images into one of five different quality groups using several features related to the symmetry of the wavelet segmented blood vessels. The micro-averaged F-measure was found to be 0.6. The relatively low F-measure was attributed to the close similarity between the neighboring quality classes leading to increased misclassifications.

Measure-based RIQA algorithms compute a numeric quality index that is related to the quality of the retinal image. Numeric measures can help ophthalmologists associate a certainty degree to the diagnosis performed through inspection of a certain retinal image based on the value of the image's quality index [17]. Furthermore, numeric quality measures can facilitate evaluating the necessity and effectiveness of image enhancement methods performed to improve an image's quality making it more suitable for reliable medical diagnosis [17]. As a result, more adequate quality retinal images can be efficiently used, after proper enhancement, in medical analysis and diagnosis thus reducing the need for increased image recaptures. Numeric-based quality metrics have been commonly implemented in several fields including quality assessment of satellite images [18], stereoscopic images [19–21], underwater images [22,23], generic images [24,25], as well as for the evaluation of retinal image registration algorithms [26]. However, only few measure-based RIQA approaches exist in literature.

In the early work of Lee and Wang [17], a quality index was assigned to a test retinal image based on the similarity between its histogram and a template histogram created from a group of excellent quality retinal images. Although it was mentioned that their quality measure agreed with human perception, no correlation measure was given to quantify the degree of this agreement. In a more recent work, Bartling et al. [27] introduced a retinal image quality measure computed as the product of wavelet-based sharpness and illumination features. Retinal images were then classified into one of three quality categories based on intervals of the presented quality measure. The kappa value between the automatic evaluation and six human graders was found to be in the range of [0.52, 0.68]. Köhler et al. [5] presented a numeric RIQA index based on an adaptation of the general approach in which the retinal vessel tree was taken as guidance to determine a global quality score from local estimates in anisotropic patches [28]. Spearman's rank correlation coefficient between their proposed measure and each of the peak-signal-to-noise ratio (PSNR) and the structural similarity (SSIM) [29] full reference metrics was found to be 0.89 and 0.91, respectively. However, the PSNR has been reported to be inefficient in matching human judgement of image quality [30,31] whereas the SSIM index was shown to be less competitive in assessing image quality related to its sharpness [32]. Measure-based retinal image quality techniques are thus still in their early stages which is indicated by the limited research and the relatively low statistical correlation, if any was given, between the existing measures and human experts.

In this work, a no-reference wavelet-based quality index is introduced to assess the quality of color retinal images based on their overall sharpness. Wavelet transform separates the sharpness and background information of an image in its detail and approximation subbands, respectively. Wavelet transform thus has the advantage of being consistent with the theory of human visual processing indicating that the eye has different optical paths for high and low frequencies [33]. Furthermore, wavelet multiresolution tends to bring out finer image details in the subsequent wavelet levels which were shown in previous work by Nirmala et al. [34] to be related to the different retinal structures. Recently, wavelet-based RIQA algorithms have been introduced to overcome the limitations in the more commonly utilized generic and segmentation based methods features by considering information related to the retinal structures and being computationally inexpensive while giving superior results [35]. Moreover, it was shown that transform-based retinal image quality features can be adapted to maintain reliable performance in practical scenarios in which the train and test datasets had significantly different image resolutions [36].

The introduced wavelet-based quality index is modified using a homogeneity parameter to account for reduced retinal image quality due to hidden structures in unevenly illuminated regions, thus increasing the reliability of the introduced measure. The homogeneity parameter was computed from the retinal saturation channel (S_{retina}) which was previously introduced by the authors to assess retinal image homogeneity [37].

In order to validate and test the proposed retinal image quality index, several analyses were performed. Two manually graded retinal image quality datasets having different resolutions, number of quality groups, and degree blurring of the bad quality images were included in these analyses. Initially, Spearman's rank correlation coefficient between the introduced quality index and the human graders was computed for the different datasets. Next, the retinal image quality index was input to a classifier to categorize the images into good, adequate, and bad quality retinal images. Finally, the presented quality index was employed to compare the performance of contrast limited adaptive histogram equalization (CLAHE), which is commonly implemented for retinal image contrast enhancement [38,39], when it is applied to different color models. Several comparisons with other retinal image quality measures from literature are also presented showing superiority of the proposed index in reliability, performance, and computational time.

The rest of the paper is organized as follows: Section 2 summarizes the datasets used for devising and validating the proposed retinal image quality index. Section 3 presents the details of the retinal image quality measure and the theory behind its formation. Section 4 includes the results and discussion of the statistical tests and classification experiments performed to validate the introduced quality index. Moreover, several comparisons with other quality measures from literature are presented. Finally, Section 5 summarizes the conclusions of the presented work.

2. Materials

Several quality graded datasets were included in this study. The details of these datasets are as follows:

Dataset-1 (DS1) [40]: consists of 301 optic disc centered retinal images acquired with a Kowa nonmyd fundus camera and having a resolution of 1600 × 1212 pixels. Three human graders evaluated the quality of the images, then a majority vote determined their final quality class resulting in 236 good and 65 bad quality retinal images.

Dataset-2 (DS2): includes 190 optic disc centered retinal images taken with a Kowa nonmyd alpha fundus camera and having a resolution of 960 × 845 pixels. Two human graders classified the images into good, adequate, and bad quality based on their overall sharpness such that it finally consisted of 80 good, 85 adequate, and 25 bad quality retinal images.