J. Vis. Commun. Image R. 40 (2016) 76–84

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Improved visual information fidelity based on sensitivity characteristics of digital images ☆

Tien-Ying Kuo [a], Po-Chyi Su [b,*], Cheng-Mou Tsai [a]

[a] Department of Electrical Engineering, National Taipei University of Technology, Taiwan
[b] Department of Computer Science and Information Engineering, National Central University, Taiwan

ARTICLE INFO

Article history:
Received 12 November 2015
Revised 18 March 2016
Accepted 16 June 2016
Available online 16 June 2016

Keywords:
Image quality assessment
Log-Gabor filter
Visual information fidelity
Visual sensitivity

ABSTRACT

Digital images may lose certain information during transmission or transcoding processes. Since the lost information can influence the visual quality perceived by the human eyes, several quality assessment metrics have been proposed. The structural similarity index (SSIM) and visual information fidelity (VIF) are two of the most common methods that take characteristics of the human perceptual system into account. Although many improved metrics based on SSIM have been developed, the methods related to VIF, which outperforms SSIM-based approaches in certain image databases, have rarely been discussed. This research aims at improving VIF to increase the effectiveness and reduce its computational complexity. The enhanced VIF employs the Haar wavelet transform, log-Gabor filter, and spectral residual approach to emphasize the visual sensitivity in image quality assessment. The experimental results demonstrate the superior performance of the proposed method, when compared to various popular or latest assessment indices.

1. Introduction

A significant amount of imagery information is digitized nowadays to facilitate storage and transmission. Digital images are easily influenced by various types of processing and noises during acquisition and conversion processes. Whether a distorted image demonstrates reduced visual quality or not is always a focus of study. Image quality assessment plays a crucial role in many applications, such as image enhancement [1,2], acquisition [3], watermarking [33], compression [34] and transmission [35]. The assessment methods are divided into subjective and objective quality measures. More accurate and reliable ways of assessing the image quality should be through subjective evaluation by the human visual system (HVS). However, such methods are restricted by the environment and specific conditions. In addition, subjective measures are not only time consuming but also costly, rendering them improper in many situations. Objective measures involve a set of assessment algorithms that can automatically evaluate the image quality so the feasibility is significantly increased. Generally, when the results of objective image assessment are more consistent with the subjective scores, they are considered closer to the quality perceived by the human eyes. Therefore, the methods of objective quality measurement are usually designed to seek consistency with the subjective scores in experiments. Various algorithms have been developed in this manner to pursue more accurate objective quality evaluation [1–4,29].

Regarding the use of reference images or undistorted original images, objective image quality measures can be classified as full-reference, reduced-reference, and no-reference methods. When the complete original image is available, the full-reference method yields the most accurate results. The reduced-reference method employs partial information in the original image to assess the quality such that the necessary data for quality evaluation can be reduced. Reduced-reference methods are thus more suitable to video quality evaluation. The no-reference method, although preferred, is relatively difficult to conduct since the distorted areas in images cannot be located easily. This research focuses on improving quality assessment for digital images to produce the results better corresponding with the visual quality perceived by the human eyes, so the full-reference methodology is adopted.

The traditional full-reference image quality measures are mean squared error (MSE) and peak signal-to-noise ratio (PSNR), which use the amount of pixel errors in the reference image and the distorted one to determine degree of distortion. The major advantage of MSE/PSNR is a low computational complexity but the errors without considering the HVS characteristics do not yield a good

☆ This paper has been recommended for acceptance by Zicheng Liu.
* Corresponding author.
E-mail address: pochyisu@csie.ncu.edu.tw (P.-C. Su).

http://dx.doi.org/10.1016/j.jvcir.2016.06.010
1047-3203/

correspondence between the results and subjective scores. Therefore, current image assessment methods usually incorporate the models of HVS to enhance the performance. Such methods can be roughly divided into two stages. Local information of the two images is first gathered and compared to obtain local assessments, which are combined in the second stage into an overall quality assessment. The structural similarity index (SSIM) [5] is the most commonly employed image assessment method. SSIM separately assesses the local brightness, contrast, and structure of both images, and then averages all local assessments to acquire the overall assessment. Unlike traditional image assessment methods, which tend to compare differences pixel by pixel, SSIM adopts a patch-based approach because the human eyes can easily perceive local information differences in an area, instead of individual pixel differences.

Several SSIM-based methods are proposed to improve the effectiveness. The information content weighted SSIM (IW-SSIM) [7] and feature similarity index (FSIM) measures [8] are considered very effective image assessment methods in the recent literature. The IW-SSIM is an improved image assessment method based on SSIM. FSIM simultaneously assesses the phase congruency [9] and gradient magnitude of images. Compared to other SSIM-based methods, both IW-SSIM and FSIM perform well as the reported results correspond to the subjective scores more closely, with the cost of additional complexity.

Visual information fidelity (VIF) [6] is another image assessment method based on the characteristics of HVS. VIF includes the concept of information theory and simulates image signals to extract cognitive information through the HVS channel. Fidelity refers to the similarity between signals from the reference and distorted images. VIF applies the wavelet decomposition to calculate the mutual information between the two images. The compiled mutual information will determine the ratio for generating the overall assessment results. Compared with SSIM-based methods, the improvement of VIF is rarely discussed probably because SSIM is more intuitional and possesses less computational complexity. However, by considering the three most frequently used image databases, TID2008 [10], LIVE [11], and CSIQ [12], SSIM-based methods only perform better in TID2008 and is outperformed in LIVE and CSIQ by VIF. Therefore, this research aims at enhancing the effectiveness of VIF in TID2008 to achieve overall superior performance and reducing its computational complexity to increase the scope of applications.

The proposed method is basically composed of three parts, including the Haar wavelet transform [13], log-Gabor filter [14], and spectral residual approach [15]. The Haar wavelet transform is first applied to both the reference and distorted images to filter out high-frequency components, which are less visually sensitive, leaving low-pass images that contain important contents. The log-Gabor filter is used to decompose the spectra of two low-pass images into multiple-scale horizontal and vertical subbands. According to the spatial domain responses of each band, the features of image distortion that the human eyes are sensitive to will be captured, along with the corresponding reference image features. The VIF method then calculates the local mutual information between both images and estimate the amount of local image information via the HVS channel. Finally, the spectral residual approach is adopted to detect the object regions of the reference image that the human eyes care, which will act as the weighting basis of the integrated local information. The local information of both images is compiled to determine the ratio, which is used to compute the overall image assessment score.

The organization of the rest of the paper is as follows: Section 2 describes the VIF image assessment principles. Section 3 explains the theory, structure, and implementation of the proposed image assessment method, S-VIF. Section 4 presents the experimental results to demonstrate the feasibility of S-VIF. The conclusion is presented in Section 5.

2. VIF

The main structure of VIF is shown in Fig. 1. First, the Gaussian scale mixture (GSM) model [16] is used to transform natural images into the Gaussian vector random field. Image signals without distortion can pass directly through the HVS channel and enter the brain, where perceptual information is extracted. However, if images are distorted, it can be assumed that the reference image signals have passed through other distortion channel before entering the HVS channel. In this structure, VIF separately calculates the following two types of mutual information: the mutual image information before and after transmission through the HVS channel, which is called reference image information; and the mutual image information before and after transmission through the distortion and HVS channels, which is called distorted image information. Each piece of information of the reference and distorted images will be used in the quality assessment process.

2.1. Source model

A wavelet transform is first applied to decompose the reference image into multiple-scale horizontal and vertical subbands, and then the GSM model is used to transform each band of the reference image into the Gaussian vector random field. The reference image bands are divided into 3×3 non-overlapped blocks. The nine coefficients in each block form vector $\boldsymbol{c}$. The overall band can thus be regarded as the vector random field $\boldsymbol{C}$. Using the GSM model, we can further represent $\boldsymbol{C}$ as the product of two independent random fields, specifically, the positive scalars random field S and the zero-mean Gaussian vector random field $\boldsymbol{U}$ with a covariance matrix C_U, as shown in (1). The original vector random field $\boldsymbol{C}$ transformed through the GSM model possesses Gaussian distribution characteristics with a zero mean and the covariance matrix $S^2 C_U$ [16–18]. I denotes the set of spatial indices for the random field.

$$\boldsymbol{C} = S \cdot \boldsymbol{U} = \{s_i \cdot \boldsymbol{u}_i : i \in I\} \tag{1}$$

2.2. Distortion channel

The wavelet transform is also applied to decompose distorted images into multiple-scale horizontal and vertical subbands. The distortion channel through the reference image band and corresponding distorted image band is calculated. The distortion channel mainly uses the signal attenuation and noises in the wavelet domain to simulate image distortion, as shown in (2), where $\boldsymbol{C}$ is the reference image band random field, $\boldsymbol{D}$ is the corresponding distorted image band random field, g is the scalar gain field for simulating signal attenuation, and $\boldsymbol{V}$ is the zero-mean additive Gaussian noise with a covariance matrix $C_V = \sigma_v^2 \mathrm{I}$, which can be used to simulate image signal noises. The distorted image random field $\boldsymbol{D}$ can be represented as the reference image random field $\boldsymbol{C}$

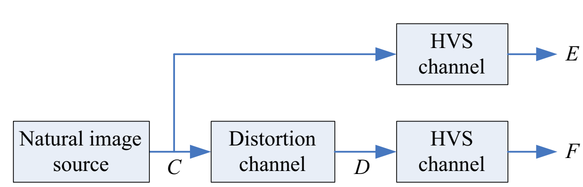

Fig. 1. The VIF schema.