Signal Processing 145 (2018) 99–105

Contents lists available at ScienceDirect

Signal Processing

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

Full-reference image quality assessment based on image segmentation with edge feature

Zaifeng Shi[a,b,*], Jiaping Zhang[a], Qingjie Cao[a], Ke Pang[c], Tao Luo[c]

[a] School of Microelectronics, Tianjin University, Tianjin 300072, PR China
[b] Tianjin Key Laboratory of Imaging and Sensing Microelectronic Technology, Tianjin, PR China
[c] School of Computer Science and Technology, Tianjin University, Tianjin, PR China

ARTICLE INFO

Article history:
Received 9 June 2017
Revised 23 October 2017
Accepted 20 November 2017
Available online 22 November 2017

Keywords:
Full-reference
Perceptual image quality assessment
Visual masking effect
Image segmentation

ABSTRACT

Full-reference image quality assessment is widely used in many applications, such as image compression, image transmission and image mosaic. The visual masking effect has a significant impact on the perception of the human visual system, which is ignored in previous image quality assessments. Combined with the visual masking effect, a full-reference image quality assessment method by edge-feature-based image segmentation (EFS) was proposed. First, the image is segmented into three parts: contour regions, edge-extension regions and slowly-varying regions. The pixels in different regions are then described by different low-level features in the light of the visual masking effect. Finally, the low-level features in each part are pooled by two complementary aspects: visual saliency and visual masking effect. Experimental results on four large-scale benchmark databases show that the proposed method has a better prediction accuracy in all distortion types than other state-of-the-art image quality assessment indices.

1. Introduction

Image quality assessment (IQA) plays an important role in image and video processing applications. Replacing subjective IQA methods with machine assessment methods is a basic and challenging technology in vision research. According to the availability of the reference image, the objective IQA methods can be divided into three categories: full-reference (FR), reduced-reference (RR) and no-reference (NR). FR metric needs distorted signal and the complete reference signal, RR metric only needs distorted signal and part of the reference signal and NR metric is a single-ended metric that uses only the distorted signal [1]. In this paper, the discussion is focused on FR IQA metric. The conventional metrics of FR IQA are mean squared error (MSE) and peak signal-to-noise ratio (PSNR) that associate to simple models which are computational efficient. However, mathematical statistics fails to correlate with the human visual system (HVS), and hence, causing MSE and PSNR to be not reliable [2]. Recently, researches on IQA start to take visual perception characteristics into consideration [3]. In an IQA flow, the first step is to extract independent low-level features of the image which are sensitive to the HVS, followed by the calculation of corresponding similarity matrixes. Combined with visual physiology and psychology, the final IQA result is obtained through pooling the similarity matrixes of different low-level features. It is important to select the appropriate low-level visual features for assessment. Wang et al. proposed a structural similarity method (SSIM) [4] which assume that HVS is sensitive to the structure of the image. Because of the success of SSIM, several SSIM based or inspired approaches have been proposed such as multi-scale SSIM (MS-SSIM) [5] and information content weighted SSIM (IW-SSIM) [6]. In [7], Yang et al. proposed a space similarity decomposition model based on the Weber-Fechner law and SSIM's structure parameter. Previous works show that the structural feature is one of the most important visual features in IQA. Most of the existing structure based IQA metrics adopt different methods to extract structure feature, such as image gradient magnitude (GM), phase congruency and difference of Gaussian (DoG). Inspired by the success of DoG used in the scale-invariant feature transform (SIFT) algorithm [8,9] extracted the local structure changes from several DoG bands. Zhang et al. [10] proposed a FSIM model that applies phase congruency and GM in color images quality assessment. Observing that the phase congruency algorithm was less sensitive to noise, in [11] a perceptual image quality assessment using phase deviation sensitive energy features was proposed. In [12], Xue et al. proposed simple gradient magnitude similarity deviation (GMSD), where the GM similarity was used to capture the image distor-

* Corresponding author at: School of Microelectronics, Tianjin University, Tianjin 300072, PR China.
E-mail address: shizaifeng@tju.edu.cn (Z. Shi).

https://doi.org/10.1016/j.sigpro.2017.11.015
0165-1684/

tion. Since GM can effectively reflect the structural loss of images, [13–15] also used GM to measure the structural differences between the reference image (I_R) and distorted image (I_D). In addition to structure similarity based IQA methods, some other works have been proposed. Sheikh et al. proposed the information fidelity criteria (IFC) [16] and its extension visual information fidelity (VIF) [17], which introduced the information theory into image fidelity measurement and predicted the IQA score by computing the information shared within I_R and I_D. Larson et al. proposed the most apparent distortion (MAD) that used different strategies on high-quality image and low-quality image [18]. For the color images, only spatial structures cannot correctly reflect the HVS of the color perception. The visual degradation caused by the reduction of saturation in the color image may cause an over optimistic result [19]. Therefore, it is necessary to introduce a metric for color distortion in the IQA. Image can be decomposed into different color spaces such as RGB, CIE [20], YCbCr [21], YIQ [15] and HIS [22]. Different color spaces have different color characteristic variables. The metrics mentioned above use the same low-level similarity matrices for each pixel, which is inconsistent with the HVS observation. According to the psychological and physiological characteristics of the HVS, different regions have different visual masking effect, which means that different regions should use different low-level features to describe.

Since the HVS has different degrees of attention to each pixel, it is necessary to apply a weighting function to indicate the importance of a local image region for quality score pooling. We called this weighting function as a pooling strategy. The pooling strategy is closely related to the HVS. An image quality assessment based on sparse learning way was proposed in [23], where the final global quality score is obtained from a regression based machine learning method which needs large training samples. According to the visual saliency (VS), weighted full-reference IQA metrics were proposed in [13,14,24]. These methods ignore the absolute impact of the surrounding pixels on the central pixel.

According to the analysis above, in this paper, the image is separated into three parts: contour regions, edge-extension regions and slowly-varying regions by edge feature. Considering the texture masking effect, the edge-extension regions only extract color as the low-level feature. Similarly, considering the color masking effect, the contour regions only extract structural characteristics. The remaining regions consider both color and structure distortion. Moreover, the pooling strategy combines two complementary parts, the visual masking effect and visual saliency. Experimental results on four benchmarks prove that our method can achieve better performance than other state-of-the-art IQA methods.

The rest of the paper is organized as follows. Section 2 introduces related works. Section 3 presents the details of the EFS algorithm. Section 4 reports the experimental results and associated discussions. Finally, Section 5 concludes the paper.

2. Related works

Since the HVS is the ultimate receiver of visual information, an increasing number of researchers try to incorporate HVS models to IQA metrics to improve the performance. The information in an image is often redundant and the HVS understand an image based on several low-level features [10]. In other words, the low-level features convey crucial information of the image to the HVS to interpret the scene. What's more, the pooling strategy which can reflect how perceptually important a local region is to the HVS is also an important part of the IQA. Visual masking effect and visual saliency are the important psychological and physiological characteristics of the HVS, which play important roles in low-level features extraction and pooling strategy.

2.1. Visual masking

Through the long-term observation of human visual phenomena, people found the visual masking effects. Visual masking effect means that one stimulus may make other stimulus invisible [25]. In the step of low-level features extraction, image distortion cannot be well characterized by a single feature. As mentioned in the Section 1, we assume that the HVS's perception of the image is based on two low- level features, structure and color. Unlike previous papers, according to the visual masking features, we hold the point that the different regions segmented by edge feature can be described by different low-level features. Texture masking and color masking are considered as two factors that influence the selection of low-level features in different regions.

Texture masking represents the effect that a strong variation signal can mask other image details around [26]. Texture masking is a local effect which means that people are insensitive to structural features of the pixels around the image edge. In other words, even if the intensity of pixels in the surrounding region of the edge changes greatly, it does not affect the visual effect of the image, therefore the structure distortion is not considered in these regions. The color masking effect is the resolution of HVS on color changes. Where the background luminance changes dramatically, the sensitivity to the color changes decreases significantly, therefore the color difference between I_R and I_D is not considered on the edges of the images.

The visual masking model also plays an important role in the pooling strategy. The visual masking model can reflect how "no-salient" a local region is to HVS. It is only based on the changes of the intensity of the surrounding pixels, and has no bearing on the center pixel. The visual masking model stress the absolute impact of the surrounding pixels on the central pixel. Entropy masking (EM) which refers to the decrease of visibility to a visual signal imposed on a mask signal that is unfamiliar or uncertain to human eyes [27]. EM is used as a visual masking model in the step of the pooling strategy. According to the concept of entropy masking, the HVS is not sensitive to visual distortions in unstructured visual signals.

2.2. Visual saliency

For the pooling strategy, visual saliency (VS) model can reflect how "salient" a local region is to the HVS based on the relative characteristic of the center pixel and its surrounding pixels. VS models and human-model agreements are summarized in [28] and [29]. In this paper, SDSP [30] is considered as the VS model. SDSP combines three preconditions. First of all, HVS detects salient objects in a visual scene can be well modeled by band-pass filtering. Secondly, the HVS is sensitive to the changes in the center of an image. Thirdly, HVS pays more attention on the warm colors than cool colors. The VS value is defined as:

$$SDSP(x) = SDSP_F(x) \cdot SDSP_D(x) \cdot SDSP_C(x) \tag{1}$$

where $SDSP_F(x)$, $SDSP_D(x)$ and $SDSP_C(x)$ are three priors, respectively.

3. EFS IQA index

According to the analysis above, combined with the visual characteristics of the HVS, we propose a new FR IQA method by edge-feature-based image segmentation. The procedures of our method are illustrated by an example in Fig. 1. The details of the algorithm are as follows.