Lu et al. Applied Informatics (2015) 2:9
DOI 10.1186/s40535-015-0010-x

Applied Informatics
a SpringerOpen Journal

RESEARCH Open Access

Blind image quality assessment via semi-supervised learning and fuzzy inference

Wen Lu, Ning Mei, Fei Gao, Lihuo He and Xinbo Gao*

* Correspondence:
xbgao@mail.xidian.edu.cn
School of Electronic Engineering, Xidian University, Xi'an 710071, China

Abstract

Blind image quality assessment (BIQA) is a challenging task due to the difficulties in extracting quality-aware features and modeling the relationship between the image features and the visual quality. Until now, most BIQA metrics available try to extract statistical features based on the *natural scene statistics* (NSS) and build mapping from the features to the quality score using the supervised machine learning technique based on a large amount of labeled images. Although several promising metrics have been proposed based on the above methodology, there are two drawbacks of these algorithms. First, only the labeled images are adopted for machine learning. However, it has been proved that using unlabeled data in the training stage can improve the learning performance. In addition, these metrics try to learn a direct mapping from the features to the quality score. However, subjective quality evaluation would be rather a fuzzy process than a distinctive one. Equally, human beings tend to evaluate the quality of a given image by first judging the extents it belongs to "excellent," "good," "fair," "bad," and "poor," and estimating the quality score subsequently, rather than directly giving an exact subjective quality score. To overcome the aforementioned problems, we propose a semi-supervised and fuzzy framework for blind image quality assessment, S^2F^2, in this paper. In the proposed framework, (1) we formulate the fuzzy process of subjective quality assessment by using fuzzy inference. Specially, we model the membership relation between the subjective quality score and the truth values it belongs to "excellent," "good," "fair," "bad," and "poor" using a Gaussian function, respectively; and (2) we introduce the *semi-supervised local linear embedding* (SS-LLE) to learn the mapping function from the image features to the truth values using both the labeled and unlabeled images. In addition, we extract image features based on NSS since it has led to promising performances for image quality assessment. Experimental results on two benchmarking databases, i.e., the LIVE database II and the TID2008 database, demonstrate the effectiveness and promising performance of the proposed S^2F^2 algorithm for BIQA.

Keywords: Blind image quality assessment; Fuzzy logic; Semi-supervised LLE; Natural scene statistics

Background

Image quality assessment (IQA) is a practical research project and has been attracting increasing attentions during the past decades due to the dramatic development of visual equipment, such as TVs, digital cameras, and mobile phones. The quality of these equipments and the images we obtained by using these equipment affects the information perception of human beings. However, we cannot obtain the undistorted version of these images in most cases. Therefore, it is necessary to develop blind IQA (BIQA) algorithms to estimate the visual quality of these images to help us choose a better equipment or image.

Due to the limited exploration of human visual system (HVS) and the mechanism of subjective quality assessment, it is a challenging task either to extract quality-aware features or build the relationship between the image features and the visual quality. It is therefore of great difficulty to develop effective BIQA metrics, especially universal BIQA (UBIQA) metrics which can work for various types of distortions.

Until now, most BIQA metrics available try to extract statistical features based on the natural scene statistics (NSS) (Brandao and Queluz 2008) and learn the mapping function from the features to the quality score using the supervised learning technique based on a large amount of labeled images (Moorthy and Bovik 2010; Jung et al. 2002; Charrier et al. 2006). Although several promising BIQA metrics have been proposed based on this methodology, there are two drawbacks of these algorithms. First, only the labeled images are adopted for machine learning. However, it has been proved that using unlabeled data in the training stage can improve the learning performance (Yang et al. 2006). In addition, these metrics try to learn the direct mapping function from the image features to the quality score. However, subjective quality evaluation would rather be a fuzzy process than a distinctive one. Equally, human beings tend to evaluate the quality of a given image by first judging the extents it belongs to "excellent," "good," "fair," "poor," and "bad," and estimating the quality score subsequently, rather than directly giving an exact subjective quality score. This is consistent with the subjective experiments conducted for constructing the IQA databases (Sheikh et al. 2003).

To overcome the aforementioned problems, we propose a semi-supervised and fuzzy framework for blind image quality assessment, S^2F^2, in this paper. In the proposed framework, we formulate the fuzzy process of subjective quality assessment by using fuzzy inference. Specially, we model the membership relation between the subjective quality score and the truth values it belongs to "excellent," "good," "fair," "poor," and "bad" using a Gaussian function, respectively. Secondly, we introduce the semi-supervised local linear embedding (SS-LLE) (Yang et al. 2006) to learn the mapping function from the image features to the truth values using both the labeled and unlabeled images. In addition, we extract image features based on NSS since it has led to promising performances for IQA. Experimental results on two benchmarking databases, i.e., the LIVE database II (Sheikh et al. 2003) and the TID2008 database (Ponomarenko et al. 2009), demonstrate the effectiveness and promising performance of S^2F^2.

The rest of this paper is organized as follows. In the Related works section, we introduce some related works, including blind image quality assessment, semi-supervised learning, and fuzzy logic inference. The proposed S^2F^2 metric is detailed in the Semi-supervised and fuzzy framework for blind image quality assessment section. The

Experiments and analysis section presents the experiments conducted on the LIVE database II and the TID2008 database. Finally, the Conclusions section concludes this paper.

Related works

Blind image quality assessment

The past five years have witnessed the emergence of various new BIQA algorithms (Saad et al. 2012; Moorthy and Bovik 2011; Mittal et al. 2012a; Mittal et al. 2012b; Mittal et al. 2012c; Mittal et al. 2013; He et al. 2012; Ye and Doermann 2012; Moorthy and Bovik 2010; Sheikh et al. 2005; Wang et al. 2002; Ciancio and da Costa 2011; Gao et al. 2009; Brandao and Queluz 2008; Jung et al. 2002; Charrier et al. 2006). These BIQA metrics can be broadly divided into two categories: the distortion-specific metrics and the universal metrics. The former means the BIQA metrics can only work on a specific type of distortion or the type of distortion should be given before using the metrics (Sheikh et al. 2005; Wang et al. 2002; Ciancio and da Costa 2011). In contrast, the universal BIQA metrics can work on various types of distortions without given the information of distortions (Saad et al. 2012; Moorthy and Bovik 2011; Mittal et al. 2012a; Mittal et al. 2012b; Mittal et al. 2012c; Mittal et al. 2013; He et al. 2012; Ye and Doermann 2012; Moorthy and Bovik 2010). Since we aim to construct a universal BIQA metric in this paper, a compact introduction of the state-of-the-art universal BIQA metrics is given below.

Saad et al. (2012) introduced a NSS-based model, blind image integrity notator using DCT statistics (BLIINDS-II). It uses the univariate generalized Gaussian density (GGD) model (Varanasi and Aazhang 1989) to formulate the distribution of discrete cosine transform (DCT) coefficients and uses the parameters of univariate GGD as features to predict the quality score. Specially, BLIINDS-II relies on a simple Bayesian probabilistic inference model to predict image quality scores.

The distortion identification-based image verity and integrity evaluation (DIIVINE) proposed by Moorthy et al. (Moorthy and Bovik 2011) is based on a two-stage framework of a NSS-based model. In DIIVINE, a classifier is first utilized to estimate the distortion type contained in the given image, and then, a distortion-specific BIQA regression metric learned for each possible type of distortion is adopted to estimate the image quality. In addition, DIIVINE extracts image features based on NSS by integrating the wavelet transform and Gabor transform.

Mittal et al. (2012a) introduced a new perspective of image features using the empirical distribution of locally normalized luminance which is under a spatial NSS model to quantify possible losses of "naturalness" in the image and proposed the blind/ referenceless image spatial quality evaluator (BRISQUE) metric. The quality score is estimated by using a support vector regression (SVR) module in BRISQUE. An improved version of BRISQUE (Mittal et al. 2012b) is proposed afterward, which utilizes a robust statistics approach based on the L-moments (Hosking 1990). The L-moments method makes BRISQUE less sensitive to empirical variations in NSS statistics.

In (He et al. 2012), He et al. proposed a sparse representation of natural scene statistics (SRNSS) model based on the hypothesis that the feature space and subjective quality space share an almost same intrinsic manifold. In SRNSS, a dictionary which

includes the image features and subjective quality scores is first constructed. For a given test image, its NSS features are then extracted and encoded in the dictionary via sparse representation, and the coding coefficients are adopted to weight the corresponding subjective quality for estimating its quality. Similarly, Ye and Doermann (2012) used Gabor-filter-based features extracted from local image patches to construct visual codebook and then learned the mapping function from the quantized feature space to the image quality score using both sample-based and learning-based methods.

All these aforementioned metrics need a large amount of human-rated images for training or constructing the dictionary/codebook. However, the subjective quality evaluation is time consuming and very expensive. To overcome this problem, Mittal et al. has developed two metrics, topic model of image quality assessment (TMIQA) (Mittal et al. 2012c) and natural image quality evaluator (NIQE) (Mittal et al. 2013). In TMIQ, patches extracted from natural and distorted images are adopted to construct a set of visual words. For a test image, probabilistic latent semantic model (pLSA) (Hofmann 2001) is first introduced to estimate the associated visual word distribution. Afterward, the quality of the test image is estimated by comparing the estimated distribution and the average distribution of the natural images. NIQE is based on the distribution of the statistical features proposed in BRISQUE. And the distribution is modeled by the multivariate Gaussian (MVG) (Eaton 1983). In NIQE, the average distribution of the features computed from patches of a set of natural images is utilized as the benchmark. For a test image, its feature distribution is first estimated, and then, its quality is predicted as the distance between the MVG fit to its feature distribution and the MVG fit to the benchmark.

Yet, the performances of these metrics are still not adequate enough. Most of the BIQA metrics above construct a black-box mapping from the features to the quality score which fail to take not only the manifold ways of human visual perception (Seung and Lee 2000) but also the unlabeled images into account. Meanwhile, the intuitive knowledge of human perception would rather be a fuzzy process than a discriminative one, which further limits the performance of these metrics.

Semi-supervised learning

The interest in semi-supervised learning has increased in recent years, particularly because of application domains where unlabeled data are plentiful in large volumes of high-dimensional data, such as images, text, and bioinformatics. Semi-supervised learning addresses this problem by using the unlabeled data, together with the labeled data, to build better classifier (Zhu et al. 2009). Because semi-supervised learning requires less human effort and gives higher accuracy, it is of great interest both in theory and in practice.

Many machine learning researchers have found that unlabeled data, when used in conjunction with a small amount of labeled data, can produce considerable improvement in learning accuracy (Cozman et al. 2003). Semi-supervised learning also shows potential as a quantitative tool to understand human category learning, where most of the input is self-evidently unlabeled (Zhu et al. 2009). In recent years, many researchers have considered that real word data may live on or close to a lower dimensional space (Seung and Lee 2000). If a pristine image has a continuous family of distorted images, the visual memory of them is hypothesized to be stored as a manifold of stable states,

or represented by a continuous attractor (Seung and Lee 2000). However, more natural learning problems may also be viewed as instances of semi-supervised learning. The learning process of image in human mind involves a small amount of direct instruction combined with a large amount of unlabeled experience.

Until now, there have been various popular semi-supervised learning models, e.g., semi-supervised support vector machines (SVM) (Zhu 2006), transductive SVMs (Vapnik 1998), semi-supervised locally linear embedding (SS-LLE) (Yang et al. 2006), and so on. Zhu et al. conducted an experiment that demonstrates semi-supervised learning behavior in humans (Zhu et al. 2007). Yang et al. (2006) extend the traditional locally linear embedding (LLE) (Roweis and Saul 2000) to SS-LLE which is of great practical value to reduce human annotators and improve accuracy. The essence of SS-LLE is to use both the labeled and unlabeled data points to approximate manifold structure as smoothly as possible, learn compact representation of images for data visualization, and recover global nonlinear structure from locally linear fits. This kind of "Think globally, fit locally" manner (Saul and Roweis 2003) makes the SS-LLE better in formulating the process of human perception.

It has been proved in many applications that SS-LLE reduces the computational complexity and outperforms other data mining or machine learning methods, such as dimension reduction (Yang et al. 2006) and image processing (Xiao et al. 2011; Zhang et al. 2009). SS-LLE has the advantage of yielding global low-dimensional coordinates that bear the same physical meaning. SS-LLE shows potential as a quantitative tool to understand human category learning, where most of the input is self-evidently unlabeled. In BIQA problem, we cannot grade an image directly but can easily sequence it in a large amount of images. The clue to this phenomenon is that perception is manifold, deriving from the studies of neurons (Seung 1998). Therefore, we inherit the advantages of SS-LLE to BIQA which is capable and competent in learning the mapping function from the features to the truth values.

Fuzzy logic

The term "fuzzy logic" was introduced in fuzzy set theory proposed by Zadeh (1965). In a nutshell, the basic principle of fuzzy logic is a matter of degree and it deals with reasoning that the real world is approximate rather than fixed and exact (Zadeh 1988). Fuzzy logic handles the concept of partial truth, where the truth value may range between completely true and completely false. This leads to a system for computing with linguistic variable (Zadeh 1996).

Fuzzy logic has been applied to many fields, e.g., control theory (Sugeno and Yasukawa 1993), artificial intelligence (Sathacopoulou et al. 1999), and image processing (Choi and Krishnapuram 1997; Ahmed et al. 2002), and has led to promising performances. It is becoming abundantly clear that the role model for fuzzy logic is the human mind and there is much to be gained by exploiting the tolerance for imprecision in dealing with real word problems, such as IQA.

Subjective quality assessment should be a kind of practical research topic computing with words rather than an exact score, because human employ linguistic variables or words, such as "bad" and "good," to describe the quality of an image in general. The intuitive knowledge of our mind about image quality is about classification. In human

visual system, instead of exact score, the subconscious feeling of an image the moment we see it is about good feasibility. Therefore, it is practical and more convenient to make a fuzzy description or classification of image in IQA.

More recent work has shown that infants and children take into account not only the unlabeled examples available but also the sampling process from which labeled examples arise (Gweon et al. 2010). This reminds us of the criterion for image quality assessment. Since human is the termination of all multimedia and images are analyzed and comprehended in visual system, subjective quality assessment is the most reasonable criterion for IQA.

In the recommendation of subjective quality assessment, (ITU-R BT.500-11, 2002) is the Methodology for the subjective assessment of the quality for television pictures, the categorical judgment method is adopted. The grading defined in ITU-R BT.500-11 is a five-grade impairment scale: 5 imperceptible, 4 perceptible but not annoying, 3 slightly annoying, 2 annoying, and 1 very annoying, corresponding to "excellent," "good," "fair," "poor," and "bad," shown in Table 1. A word is viewed as a label of a granule, that is, a fuzzy set of points drawn together by similarity. For example, "good image" can be the combination of large amounts of images which has the character of "good" but expressed in other form. However, the quality is defined as to consist of specific perceptual attributes and expressed in numbers or symbols. Therefore, the fuzzy modeling of image quality is somewhat intermediate, which can reduce the granularity of the image's characterization and suit with linguistic approximation. This tolerance for imprecision in BIQA can be exploited to achieve tractability, robustness and better rapport with reality (Zadeh 1996).

Methods

Semi-supervised and fuzzy framework for blind image quality assessment

The semi-supervised and fuzzy framework proceeds as follows. Firstly, we extract image features based on NSS because NSS is a comprehensive description of natural images and has led to promising performances for image quality assessment. Secondly, we formulate the fuzzy process of subjective quality assessment by using fuzzy logic. Specially, we model the membership relation between the subjective quality score and the truth values it belongs to "bad," "poor," "fair," "good," and "excellent" using a Gaussian-based membership function, respectively. Thirdly, we introduce SS-LLE to learn the mapping from the image features to the truth values using both the labeled and unlabeled images. The quality score is finally estimated based on the truth values. The framework of the S^2F^2 for BIQA is illustrated in Fig. 1.

Table 1 ITU-R quality and impairment scales in BT.500-11

Level	Impairment	Quality
5	Imperceptible	Excellent
4	Perceptible, but not annoying	Good
3	Slightly annoying	Fair
2	Annoying	Poor
1	Very annoying	Bad

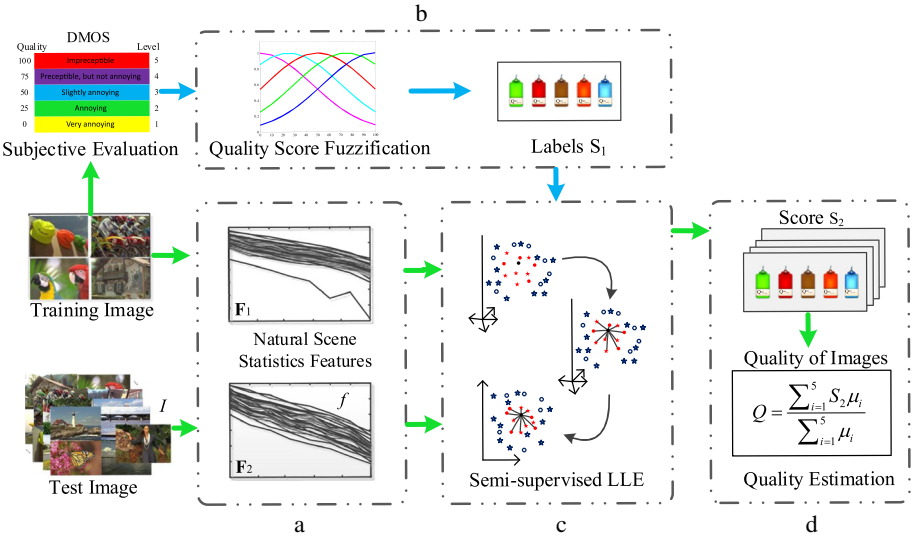

Fig. 1 The proposed semi-supervised and fuzzy logic framework for BIQA. **a** is natural scene statistics features. **b** is quality score fuzzification. **c** is semi-surpervised local linear embedding. **d** is qualtiy estimation

A. Natural scene statistics features

Learning the compact representation of the images is helpful for data visualization and quality assessment. Images are naturally multiscale, and therefore, there exists a decomposition module in the early visual system. The wavelet transform performs mirror models of spatial decomposition that occurs in area V1 of the primary visual cortex (Buccigrossi and Simoncelli 1999) and has been previously used for many reference IQA methods with success (Moorthy and Bovik 2011; He et al. 2012).

In order to capture the statistical properties of natural scenes that hold across different contents, we utilize the mature wavelet transform approach which is scale, space, and orientation selectivity. An input image I is decomposed into wavelet coefficients. After taking some attempts in the selection of scales, we found that three-scale decomposition provides compact and adequate information, which would be 10 wavelet subbands in total. However, the vertical and horizontal subbands in the same scale are approximately the same. Thus, we combine them through averaging which makes the number of subbands into 6. For each subband, we extract the magnitude feature m_k which signifies the size information of the coefficients to encode the generalized spectral behavior, and the entropy feature e_k which shows the distribution and relation of the coefficients to represent the generalized information, as follows.

$$m_k = \frac{1}{N_k \times M_k} \sum_{j=1}^{N_k} \sum_{i=1}^{M_k} \log_2 |C_k(i,j)| \tag{1}$$

$$e_k = \sum_{j=1}^{N_k} \sum_{i=1}^{M_k} p[C_k(i,j)] \ln p[C_k(i,j)] \tag{2}$$

where M_k and N_k $(k = 1, 2, \ldots, 6)$ are the length and width of the kth subband, respectively, and $C_k(i,j)$stands for the(i,j) coefficient of the kth subband, and $p[\cdot]$is the

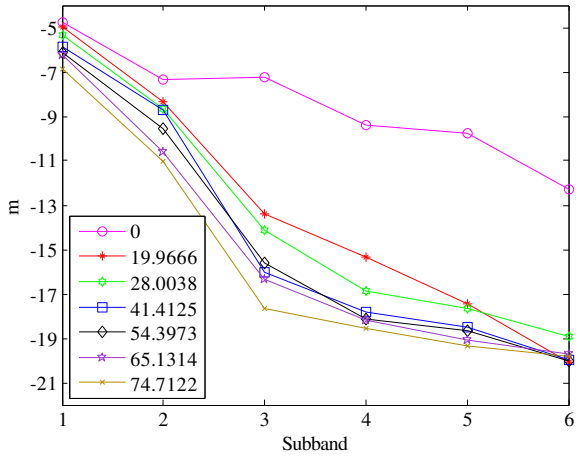

Fig. 2 Magnitudes of the wavelet coefficients for different DMOS decay exponentially across scales

probability density function. As features go, stack these 12 statistics to form a single vector which is called NSS.

$$f = [m_1, m_2, ..., m_6, e_1, e_2, ..., e_6]^{\mathrm{T}} \tag{3}$$

The relationship between the features and the quality can be visualized by the feature we present above. The magnitude spectra of the reference images have the similar exponential decay characteristics across scales while the distorted ones have different downtrend which is certainly less specific in their representation of a particular image, so they can be suitable for representing the generalized behaviors of natural scenes. The distribution of magnitudes m_k follows the law that the increase of distortion degree brings out the sharper decrease and stronger persistence at fine scales, shown in Fig. 2 (He et al. 2012). Therefore, the features extracted by NSS are quality aware. In this paper, we apply this speciality to conduct semi-supervised manifold which can help learn the mapping from the features to the quality.

B. Quality score fuzzification

In ITU-R BT.500-11, observers are required to assign an image to one of five categories which are defined in semantic terms, which reminds us of using linguistic variables in place of or in addition to numerical variables in practical application (Zadeh 1988). Meanwhile, Gaussian function is commonly used to model the fuzzy membership. As a result, we design the Gaussian-based fuzzy membership functions which use five primary terms "excellent," "good," "fair," "poor," and "bad" as primary terms and treat other words as the modification of primary terms. The form of functions is shown in Fig. 3.

In our Gaussian-based fuzzy membership functions, the meanings of the words are represented by functions mapping to five quality scales. The mean of each Gaussian distribution represents the primary term and the variance details the possible distribution. As a result, each difference mean opinion scores (DMOS) corresponds to five truth values. It is easy to get that relative large variance gets closer to human

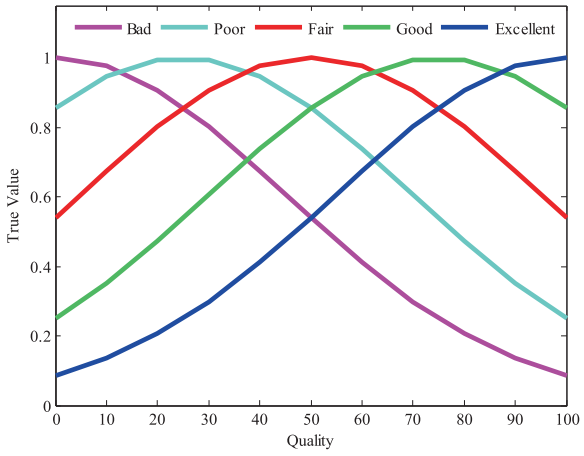

Fig. 3 The proposed fuzzy membership functions for five categories of image qualities

perception, because many "good" images can be kind of "poor." These five truth values are expressed in terms of fuzzy membership function, shown as

$$s_l = f(\text{DMOS}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\text{DMOS}-\mu_l)^2}{2\sigma^2}} \tag{4}$$

$$l = 1, 2, \ldots, 5 \text{ and}$$

$$s = [s_1, s_2, \ldots, s_5]^{\mathrm{T}} \tag{5}$$

The fuzzy expression models the process of human perception and represents the quality of an image in five truth values instead of an absolute score. In addition, the fuzzification of the quality score increases the volumes of information. And the experimental results demonstrate the effectiveness of the fuzzification approach. The characterization of the image is better in five truth values than in one quality score which makes the granularity of image smaller.

C. Semi-supervised local linear embedding

We combine features and truth values of both labeled and unlabeled images. f_i is the feature vector of the ith image. Stack all the features to form the feature space $\mathbb{R}^D$.

$$\begin{aligned} &\mathbf{F} = [f_1, f_2, \ldots, f_N], \text{ and} \\ &f_i \in \mathbb{R}^D (i = 1, 2, 3, \ldots, N) \end{aligned} \tag{6}$$

Divide $\mathbf{F}$ as $[\mathbf{F}_1, \mathbf{F}_2]$, where $\mathbf{F}_1$ represents the features of labeled images and $\mathbf{F}_2$ represents the unlabeled ones for test.

$$\begin{aligned} &\mathbf{F}_1 = (f_1, f_2, \ldots, f_c), \text{ and} \\ &\mathbf{F}_2 = (f_c, f_{c+1}, \ldots, f_N) \end{aligned} \tag{7}$$

Let s_{il} be the lth truth value for the ith image. All the truth values form the label space $\mathbb{R}^d$.

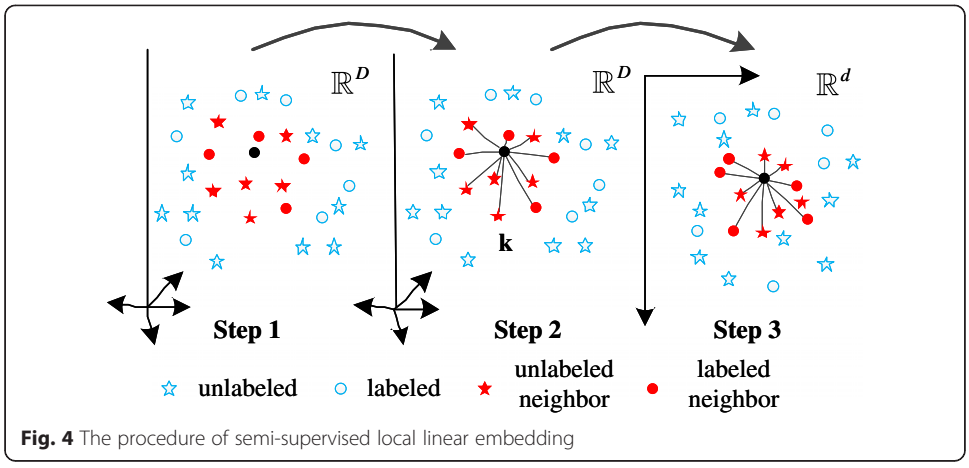

Fig. 4 The procedure of semi-supervised local linear embedding

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{15} \\ s_{21} & s_{22} & \cdots & s_{25} \\ \vdots & \vdots & \cdots & \vdots \\ s_{N1} & s_{N2} & \cdots & s_{N5} \end{bmatrix}, \text{ and} \quad (8)$$

$$s_{il} \in \mathbb{R}^d (i = 1, 2, 3, ..., N;\ l = 1, 2, ..., 5)$$

Partition $\mathbf{S}$ as $[\mathbf{S}_1, \mathbf{S}_2]$, where $\mathbf{S}_1$ represents the truth values for the labeled images and $\mathbf{S}_2$ corresponds to the other ones. N is the total number of images in our experiment, and c is the number of labeled images.

$$\mathbf{S}_1 = (s_1, s_2, ..., s_c), \text{ and} \quad (9)$$

$$\mathbf{S}_2 = (s_c, s_{c+1}, ..., s_N) \quad (10)$$

Given the input feature points and the output labeled points, the SS-LLE consists of three steps (Yang et al. 2006) shown in Fig. 4:

Step 1: Find the k nearest neighbors for each feature point f_i based on the Euclidean distance.

Step 2: Compute the reconstruction coefficient by minimizing the reconstruction error which is measured as

$$\varepsilon(W) = \sum_{i=1}^{N} \left\| f_i - \sum_{j=1}^{k} W_{ij} f_j \right\|^2 \quad (11)$$

Subject to constraint: $\sum_{j=1}^{N} W_{ij} = 1\ (i = 1, 2, \cdots N)$.

Step 3: Compute the low-dimensional embedding.

The low-dimensional embedding is found through the following minimization.

$$\phi(\mathbf{S}) = \sum_{i=1}^{N} \left\| \gamma_i - \sum_{j=1}^{k} W_{ij} \gamma_j \right\|_2^2 = \mathbf{SMS}^{\mathrm{T}} \quad (12)$$

Subject to two constraints: $\sum_{i=1}^{N}\gamma_i = 0$ and $\frac{1}{N}\sum_{i=1}^{N}\gamma_i^{\mathrm{T}}\gamma_i = \mathbf{I}$, matrix **M**is constructed based on the matrix **W**: $\mathbf{M} = (1 - \mathbf{W})^{\mathrm{T}}(1 - \mathbf{W})$. The resulting problem is equivalent to finding the smallest $d + 1$ eigenvectors of matrix **M. M** is partitioned into four parts:

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{12}{}^{\mathrm{T}} & \mathbf{M}_{22} \end{bmatrix} \tag{13}$$

$\mathbf{M}_{11}$ is a matrix of size $c \times c$, referred to labeled images. The minimization problem can be written as

$$\min_{\mathbf{S}_2}[\mathbf{S}_1, \mathbf{S}_2]\begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{12}{}^{T} & \mathbf{M}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{S}_1^{\mathrm{T}} \\ \mathbf{S}_2^{\mathrm{T}} \end{bmatrix} \tag{14}$$

Equivalently,

$$\min_{\mathbf{S}_2} \mathbf{S}_2\mathbf{M}_{22}\mathbf{S}_2^{\mathrm{T}} + 2\mathbf{S}_1\mathbf{M}_{12}\mathbf{S}_2^{\mathrm{T}} \tag{15}$$

Set the gradient of the above objective function to 0

$$\mathbf{M}_{22}\mathbf{S}_2{}^{\mathrm{T}} = \mathbf{M}_{12}\mathbf{S}_1^{\mathrm{T}} \tag{16}$$

Therefore, the $\mathbf{S_2} \in R^d$ of the unlabeled images is

$$\mathbf{S}_2 = \left(\frac{\mathbf{M}_{12}\mathbf{S}_1{}^{\mathrm{T}}}{\mathbf{M}_{22}}\right)^{\mathrm{T}} \tag{17}$$

$\mathbf{S}_2$ includes the five truth values associated with the unlabeled images. As a part of unlabeled images, the test image participates in SS-LLE. Its features f are contained in $\mathbf{F_2}$, and truth values s_t are contained in $\mathbf{S}_2$.

D. Quality estimation

In this section, we introduce two different approaches for quality estimation. The first one concludes a defuzzification module to keep in step with traditional IQA metrics. However, practical applications call for non-numeric descriptors rather than one quality score. We consider expressing image quality in words or linguistic variables. By comparison, the second approach takes the five truth values obtained from SS-LLE for the final result of image quality. For clarity, the first approach is referred as S^2F^2-I and the second is referred as S^2F^2-II in the following part of this paper.

S^2F^2-I: To obtain the traditional "quality" of images, we should defuzzify the five truth values obtained in the learning module. Defuzzification is the process of producing a quantifiable result in fuzzy logic, given the fuzzy sets and corresponding membership degrees. Defuzzification is interpreting the membership degrees of the fuzzy sets into a specific decision or real value. For example, five truth values deciding how good are the test images might result in "excellent (0.35), good (0.56), fair (0.64), bad (0.92), poor (0.63)," but finally expressed in one score "28." By calculating the area under the scaled membership functions and then within the range of the output variable, we adopt the center of area (CoA) defuzzification method. The formula of CoA is

$$Q = \text{defuzz}(s_t) = \frac{\sum_{i=1}^{5} s_t \mu_i}{\sum_{i=1}^{5} \mu_i} \tag{18}$$

As stated earlier, $\mathbf{S}_2$ are the output labels of SS-LLE for the test image. The outcome$Q$, ranging from 0 to 100, is the final quality score of the test image. The CoA defuzzification method effectively calculates the best compromise between multiple output truth values.

S^2F^2-II: The S^2F^2-I method defuzzify the five truth values obtained from SS-LLE to get one quality score which is the tradition of BIQA. However, the subconscious feeling of an image the moment we see it is about good feasibility which is expressed in words rather than exact score. In view of the practical application, we directly adopt the estimated five truth values as the presentation of the image quality.

$$Q = s_t \tag{19}$$

The five truth values for each of the five functions represent the degrees of truth they belong to "excellent," "good," "fair," "poor," and "bad." This estimation method makes the assessment much closer to human behavior.

Results and discussion

Experiments and analysis

To verify the effectiveness of the proposed S^2F^2-I and S^2F^2-II metrics, we test them on two benchmarking databases: the LIVE database II (Sheikh et al. 2003) and the TID2008 database (Ponomarenko et al. 2009). The LIVE database II consists of 29 reference images and 779 distorted images that span various distortion types—JPEG2000 compression (JP2K), JPEG compression (JPEG), additive white Gaussian noise (WN), Gaussian blurring (Gblur), and fast fading (FF), along with the associated subjective human DMOS, which are representative of the perceived quality of the image. The TID2008 database contains 1700 test images and 25 reference images over 17 distortion categories, 4 different levels. We test our algorithm only on the four distortions JPEG, JP2K, WN, and Gblur of TID2008 as done in Saad et al. (2012), Moorthy and Bovik (2011), Mittal et al. (2012a) and Mittal et al. (2012b).

The indices considered in the experiment are the Spearman's rank ordered correlation coefficient (SROCC) and the linear (Pearson's) correlation coefficient (LCC). A value close to 1 for SROCC and LCC indicates superior correlation with human perception. In original LLC calculation for S^2F^2-I, each point in the group of quality represents an image for test. It is meaningful to declare that the LCC for S^2F^2-II is formulated as:

$$\text{LCC} = \frac{\sum_{i=1}^{N-c} (s_{ti} - \bar{s}_t)(s_i - \bar{s})}{\sqrt{\sum_{i=1}^{N-c} (s_{ti} - \bar{s}_t)^2 (s_i - \bar{s})^2}} \tag{20}$$

s_t is the predict quality in form of five truth, values and $s$is the original fuzzy subjective DMOS which is also in five words. In the LCC for S^2F^2-II, a point is no longer the representation of an image, but five points cooperate in presenting an image. This makes the LLC change from a 1-to-1 calculation to 5-to-5 calculation.

We conduct several experiments to verify the consistency between the proposed BIQA metrics and subjective IQA, and their robustness to the training set, the selection of the parameters, and the databases. Details are introduced in the following subsections.

Consistency experiments

S^2F^2-I and S^2F^2-II approaches require a training stage in order to learn the mapping function from the features to the subjective image quality score. We randomly select part of the LIVE database II for labeled images set and the rest for unlabeled set. The manifold module of SS-LLE learns the mapping from the image features to the truth values using both the labeled images set and unlabeled images set. In order to ensure that the proposed approaches S^2F^2-I and S^2F^2-II are robust across content and are not governed by the specific train-test split utilized, we repeat this random selected train 1000 times on the LIVE database II and evaluate the average performance.

We compare the proposed metrics with the state-of-the-art BIQA metrics, i.e., the natural scene statistics (NSS) (Ciancio and da Costa 2011), the distortion identification-based image verity and integrity evaluation (DIIVINE) (Moorthy and Bovik 2011), the blind image quality index (BIQI) (Moorthy and Bovik 2010), the blind image integrity notator using DCT statistics (BLIINDS-II) index (Saad et al. 2012), the blind/ reference-less image spatial quality evaluator (BRISQUE) (Mittal et al. 2012a), the sparse representation of natural scene statistics (SRNSS) (He et al. 2012), the natural image quality evaluator (NIQE) (He et al. 2012), and the codebooks image quality (CBIQ) (He et al. 2012). In addition, we adopt several classic full-reference (FR) IQA metrics as the benchmarks, the peak signal-to-noise ratio (PSNR), the structural similarity (SSIM) (Wang et al. 2004), and the visual information fidelity (VIF) (Sheikh and Bovik 2006).

Table 2 Median LCC across 1000 train-test on the LIVE database II

Metric	Type	JP2K	JPEG	WN	Gblur	FF	Entire database
PSNR	FR	0.8962	0.8596	0.9858	0.7834	0.8895	0.8240
SSIM	FR	0.9367	0.9283	0.9695	0.8740	0.9428	0.8634
IFC	FR	0.9027	0.9047	0.9581	0.9608	0.9614	0.9106
VIF	FR	0.9615	0.9430	0.9839	0.9744	0.9618	0.9501
NSS	Blind	0.9210	0.3661	0.8217	0.7007	0.7224	0.4946
BIQI	Blind	0.8086	0.9011	0.9538	0.8293	0.7328	0.8205
DIIVINE	Blind	0.9220	0.9210	0.9880	0.9230	0.8880	0.9170
BLIINDS-II	Blind	0.9630	0.9793	0.9854	0.9481	0.9436	0.9232
SRNSS	Blind	0.9359	0.9391	0.9404	0.9356	0.9473	0.9318
BRISQUE	Blind	0.9229	0.9734	0.9851	0.9506	0.9093	0.9424
NIQE	Blind	0.9370	0.9564	0.9773	0.9525	0.9128	0.9147
CBIQ	Blind	0.912	0.963	0.959	0.918	0.885	0.896
S^2F^2-I	Blind	*0.9578*	*0.9489*	*0.9668*	*0.9556*	*0.9370*	*0.9464*

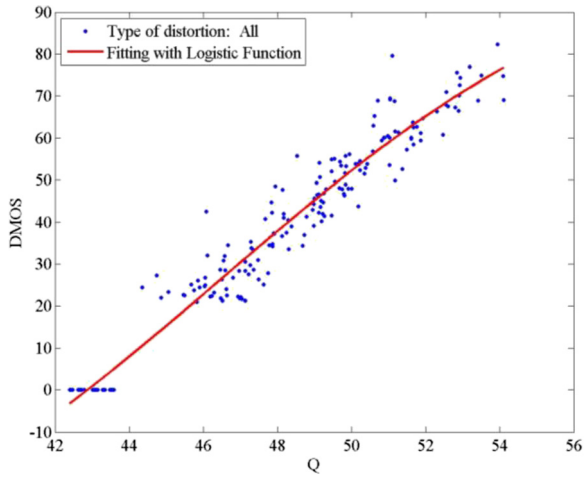

Fig. 5 Predicted Q vs. subjective DMOS on the entire LIVE database

The realizations of DIIVINE, BIQI, BLIINDS-II, and BRISQUE are available online, and the results of SRNSS, NIQE, and CBIQ are obtained in Mittal et al. (2013), He et al. (2012), and Ye and Doermann (2012). The index comparison results are shown in Table 2.

The number of the nearest neighbors k selected in the first step of SS-LLE is 70. The scale of fuzzy membership function l is determined to be 5 in view of the ITU-R BT.500-11. The variance of Gaussian distribution σ is 90. The size of the training set is selected to be 23 groups of the LIVE database II.

In this subsection, we present the performances of S^2F^2-I and S^2F^2-II. As an improved version of S^2F^2-I, S^2F^2-II makes the index calculation to be a 5-to-5 comparison. It is unreasonable to compare S^2F^2-II with other available BIQA methods, but feasible with S^2F^2. Therefore, we present them separately.

The performance of S^2F^2-I

Figure 5 shows the scatterplots and the nonlinear curve fittings between the estimated quality score by S^2F^2-I and DMOS across the entire test set in one trial. Table 2 shows

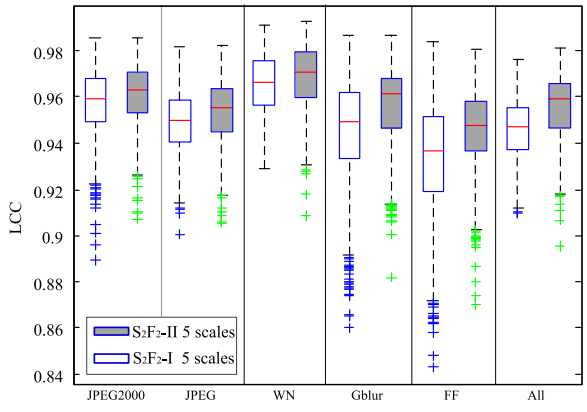

Fig. 6 Boxplot of LCC distributions of S^2F^2-I and S^2F^2-II in five scales on LIVE database

that S^2F^2-I achieves the highest accuracy of 0.9464, outperforming the other methods. More crucially, S^2F^2-I makes about 2 % improvement over BLIINDS-II and is superior to other holistic BIQA approaches. S^2F^2-I approach is also competitive with PSNR, the most popular FR-IQA metric. This demonstrates that features we extracted are quality aware, that SS-LLE formulates the manifold perception well, and that the concept of fuzzy logic reduces the granularity of the image's characterization.

In general, the S^2F^2-I proposed in this paper has a good consistence with human subjective perception. In comparison with other metrics, this proposed S^2F^2-I BIQA algorithm obtains better performances.

The performance of S^2F^2-II

In this part, we present the comparative performance of S^2F^2-I and S^2F^2-II to demonstrate the advancement of S^2F^2-II. To visualize the statistical significance of the comparison, we compare the boxplots of the LCC values during the 1000 experimental trials, as shown in Fig. 6. It is notable that the performance of S^2F^2-II is much better than S^2F^2-I for every distortion type. The reason behind this improvement is that linguistic variables reduce the granularity of an image in calculation and better imitate the human visual perception. Five truth values represent five degrees of memberships which is not a hard division but a cooperation of fuzzy language.

Another attempt at fuzzy logic is to change the fuzzy scales from five to three, with the final result in the form of fuzzy words. The three scales stands for three primary terms "good," "fair," "bad," which is clearer than the five scales introduced in ITU-R 500-11. Because of the difference in index calculation, this 3-to-3 S^2F^2-II metric cannot be compared to state-of-the-art general purpose BIQA, but comparable to three-scale S^2F^2-I, as shown in Fig. 7. As shown in the Table 3, 3-to-3 S^2F^2-II outperforms three-scale S^2F^2-I with an accuracy of over 0.9602 on the LIVE database II, broke though the record in BIQA. This reminds us of changing the criterion for quality.

Robustness to the size of the training set

In this section, we verify the performance of the proposed BIQA metrics when only a small amount of data is available. To conduct this experiments, we choose K group(s)

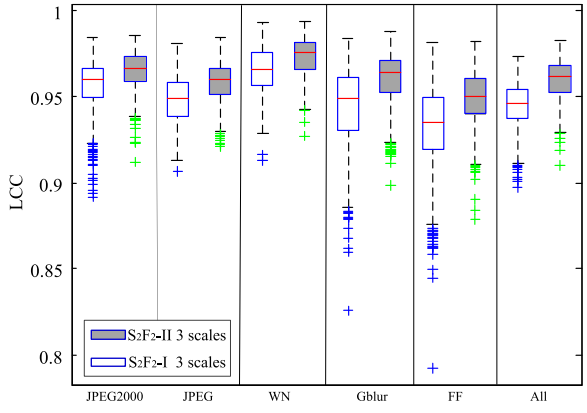

Fig. 7 Boxplot of LCC distributions of S^2F^2-I and S^2F^2-II in three scales on LIVE database

Table 3 Median linear correlation across 1000 train-test on the LIVE database II

Metric	LCC
S^2F^2 5 scale	0.9464
S^2F^2-II 5-5	0.9560
S^2F^2 3 scale	0.9453
S^2F^2-II 3-3	*0.9602*

of images for training and the rest for test, $K = 1, 2, \ldots, 28$, and run the training-test procedure 1000 times. One group means one reference image and the distorted images derived from it. Figure 8 shows the median LCC of S^2F^2-I and S^2F^2-II during the 1000 trials when we choose different sizes of training sets. It is obvious that LCC increases with the size of the training set. And even when only three groups of images are chosen for training, the LCC of S^2F^2-I is as large as 0.81, and the LCC of S^2F^2-II is 0.86. In addition, the LCC of both S^2F^2-I and S^2F^2-II become larger than 0.90 when only 10 groups of images are for training, which only takes a part of 1/3 of the entire data set. Clearly, the proposed metrics are robust to the size of the training set.

Robustness to parameters

To verify how the parameters affect the performance of the proposed method, we conduct experimental studies on influential factors of S^2F^2-I and S^2F^2-II in this section. We repeat the training-test procedure 1000 times on the LIVE database II and calculate the median performance. In our experiments, four parameters utilized in the proposed framework are considered: the form of membership function, the scale of fuzzification l, the variance of the Gaussian membership function σ, and the neighborhood selection k. The experiments are divided into four parts.

We only change one of the parameter to testify the effectiveness in each part, with other parameters unchanged: $k = 70$, $l = 5$, and $\sigma = 90$. The size of the training set is 23. Both S^2F^2-I and S^2F^2-II are conducted in five scales and three scales, respectively, if there is no special to declare.

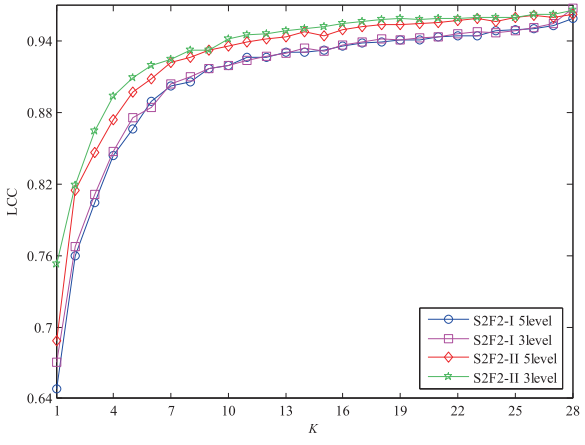

Fig. 8 The median LCC of S^2F^2-I and S^2F^2-II when choosing different sizes of training sets

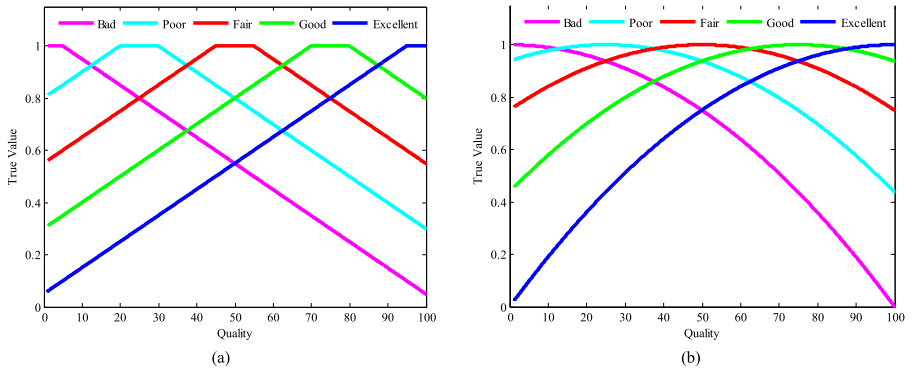

Fig. 9 Membership function. **a** Linear function. **b** Quadratic function

Membership function

Apart from the Gaussian distribution, we also try two other functions in the design of fuzzy logic: piecewise linear function and quadratic function, shown in Fig. 9.

$$s' = g(\mathrm{DMOS}) = \begin{cases} \frac{1}{\sigma}(\mathrm{DMOS}-\mu+5)+1, & \mu-\mathrm{DMOS} > 5 \\ 1, & |\mathrm{DMOS}-\mu| \le 5 \\ -\frac{1}{\sigma}(\mathrm{DMOS}-\mu-5)+1, & \mathrm{DMOS}-\mu > 5 \end{cases}$$

$$s'' = h(\mathrm{DMOS}) = -\frac{1}{\sigma^2}(\mathrm{DMOS}-\mu)^2 + 1 \tag{21}$$

With other parameters remain unchanged, Fig. 9a plots the comparison experimental results. It can be seen that results of the three functions are almost the same. In

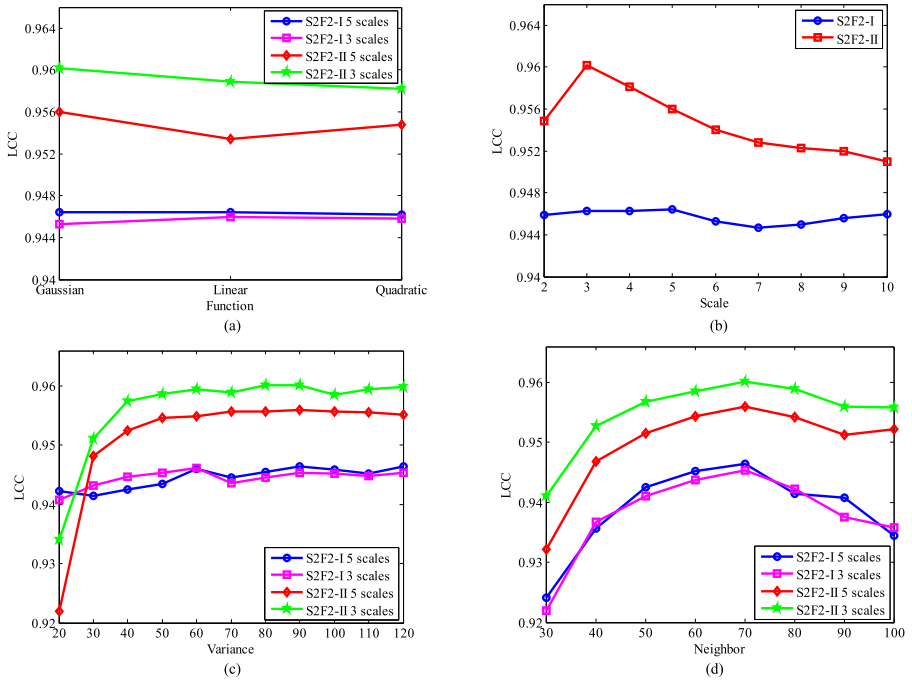

Fig. 10 Parameter comparison. **a** Function. **b** Scale. **c** Variance. **d** Neighbor

Table 4 Median SROCC across 1000 of different metrics trained on the LIVE database II and tested on TID2008 database

	JP2K	JPEG	WN	Gblur	All
PSNR	0.8250	0.8760	0.9230	0.9342	0.8700
SSIM	0.9603	0.9354	0.8168	0.9544	0.9016
BLIINDS-II	0.9157	0.901	0.6600	0.8500	0.8442
DIIVINE	0.924	0.966	0.851	0.862	0.889
BRISQUE	0.832	0.924	0.82	0.881	0.896
S^2F^2-I	0.9115	0.9143	0.7503	0.8422	0.8761

aggregate, Gaussian distribution is a little bit better than other functions in our fuzzy logic framework. That is to say, our fuzzy framework is not sensitive to the form of membership function.

Fuzzy scales

To illustrate our designation, we conduct totally nine choices for the fuzzy scales. As can be seen from Fig. 10b, five-scale S^2F^2-I wins just a minor victory over three-scale S^2F^2-I, mostly due to the five-grade subjective assessment taken in the LIVE database II, and three-scale S^2F^2-II is better than five-scale S^2F^2-II. However, the distinction between them is not always readily apparent. Therefore, the proposed frameworks keep performing well for different selections of the number of fuzzy scales.

Variance

"Excellent" image can be "poor" image to some extent in human mind. However, we do not know to what extent the distribution varies. On the premise of Gaussian distribution, a test for the influence of the variance on quality assessment is shown in Fig. 10c. The curves of five-level and three-level S^2F^2-I are almost the same, but S^2F^2-II drops with the decreasing of variance. We can draw the conclusion that the proposed BIQA framework is robust to the variance of the Gaussian function.

Neighborhood selection

The neighbor selection step in SS-LLE is simple to implement, but it can be time consuming for large data sets if performed without any optimizations. Meanwhile, SS-LLE is somewhat sensitive to the selection of the number of the nearest neighbors. Too much neighbors will cause elimination of small-scale structures in the manifold. In contrast, too small neighbors may falsely divide the continuous manifold into disjointed sub-manifolds. Figure 10d demonstrates the effect of neighborhood variety. It is obvious to choose 70 neighbors for our semi-supervised framework. In addition, although the performances decrease when the number of the neighborhood is very small or large, they are still comparative to state-of-the-art BIQA metrics.

In general, the parameters in our framework are insensitive. And S^2F^2-II is better than S^2F^2-I under multiple circumstances of parameters. Therefore, we can decide the parameters in accordance with actual condition of applications.

Robustness to databases

In order to demonstrate the algorithm is database independent, we train S^2F^2-I on the LIVE database II and test on TID2008. It needs emphasizing that we change the details of Gaussian distribution in fuzzy membership function to fit new marking standard of TID2008. And although there exists 17 distortion categories, we tested S^2F^2-I only on these distortions that it is trained for: JPEG, JP2K, WN, and Gblur. FF distortion does not exist in the TID database. Parameter remains the same: $k = 70$, $l = 5$, and $\sigma = 90$. The size of the training set is 23.

We also list the performance of PSNR, SSIM, BLIINDS-II, and BRISQUE for comparison purposes. The SROCC of S^2F^2-I metric drops because of the differences in simulated distortions present in databases and objective evaluation which make the fuzzy module not precise. However, the correlations are still consistently high. The SROCC results are shown in Table 4.

Note that it is difficult to perform directly a comparison with all the previously reported work on available database due to the different experimental settings. Nevertheless, the performance of S^2F^2-I on TID2008 database is still very encouraging compared with the available methods. Therefore, the proposed S^2F^2-I is robust against the test data and can be applied to other different databases.

Conclusions

In this paper, we propose a new semi-supervised and fuzzy framework for blind image quality assessment, called S^2F^2. Experimental results on the two benchmarking databases demonstrate that S^2F^2 not only makes an obvious improvement over state-of-the-art BIQA metrics but also is robust against the selection of parameters contained in the proposed model. Nevertheless, the proposed framework is still limited while compared with the best FR-IQA metrics. Improvement and development of S^2F^2 will be our future work.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

XG participated in the design of the proposed method and helped to revise the manuscript. WL carried out the methods and drafted the manuscript. NM participated in the design of the experiment and performed the statistical analysis. FG and LH participated in its design and helped to revise the manuscript. All authors read and approved the final manuscript.

Received: 17 April 2015 Accepted: 7 June 2015
Published online: 25 July 2015

References

Ahmed MN, Yamany SM, Mohamed N, Farag AA, Moriarty T (2002) A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data. IEEE Trans Med Imag 21(3):193–199

Brandao T, Queluz MP (2008) No-reference image quality assessment based on DCT-domain statistics. Signal Process 88(4):822–833

Buccigrossi RW, Simoncelli EP (1999) Image compression via joint statistical characterization in the wavelet domain. IEEE Trans Image Process 8(12):1688–1701

Charrier C, Lebrun G, Lezoray O (2006) A machine learning-based color image quality metric. In: Proc 3rd Euro Conf Color Graphics Imag Vis., pp 251–256

Choi YS, Krishnapuram R (1997) A robust approach to image enhancement based on fuzzy logic. IEEE Trans Image Process 6(6):808–825

Ciancio A, da Costa A (2011) No-reference blur assessment of digital pictures based on multifeature classifiers. IEEE Trans Image Process 20(1):64–75

Cozman FG, Cohen I, Cirelo MC (2003) Semi-supervised learning of mixture models. In: Proc Int Conf Mach Learn., pp 99–106

Eaton ML (1983) Multivariate statistics: a vector space approach. Wiley, New York, ch. 10

Gao X, Lu W, Tao D, Li X (2009) Image quality assessment based on multiscale geometric analysis. IEEE Trans Image Process 18(7):1409–1423

Gweon H, Tenenbaum JB, Schulz LE (2010) Infants consider both the sample and the sampling process in inductive generalization. Proc Natl Acad Sci USA 107(20):9066–9071

He L, Tao D, Li X, Gao X (2012) Sparse representation for blind image quality assessment. In: Proc IEEE Int Conf Comput Vis Pattern Recog., pp 1146–1153
Hofmann T (2001) Unsupervised learning by probabilistic latent semantic analysis. Mach Learn 42(1):177–196
Hosking JRM (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. Jour Royal Stat Soc 52(1):105–124
ITU-R BT. 500–11 (2002) Methodology for the subjective assessment of the quality for television pictures.
Jung M, Léger D, Gazalet M (2002) Univariant assessment of the quality of images. J Electron Imag 11(3):354–364
Mittal A, Moorthy AK, Bovik A (2012a) No-reference image quality assessment in the spatial domain. IEEE Trans Image Process 21(12):4695–4708
Mittal A, Moorthy AK, Bovik AC (2012b) Making image quality assessment robust. In: Proc Int Conf Signals, Systems and Computers., pp 1718–1722
Mittal A, Muralidhar GS, Ghosh J, Bovik AC (2012c) Blind image quality assessment without human training using latent quality factors. IEEE Signal Process Lett 19(2):75–78
Mittal A, Soundararajan R, Bovik AC (2013) Making a completely blind image quality analyzer. IEEE Signal Process Lett 22(3):209–212
Moorthy AK, Bovik AC (2010) A two-step framework for constructing blind image quality indices. IEEE Signal Process Lett 17(5):513–516
Moorthy AK, Bovik AC (2011) Blind image quality assessment: from natural scene statistics to perceptual quality. IEEE Trans Image Process 20(12):3350–3364
Ponomarenko N, Lukin V, Zelensky A, Egiazarian K, Carli M, Battisti F (2009) Tid 2008—a database for evaluation of full-reference visual quality assessment metrics. Advances of Modern Radioelectronics 10:30–45
Roweis S, Saul L (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500):2323–2326
Saad MA, Bovik AC, Charrier C (2012) Blind image quality assessment: a natural scene statistics approach in the DCT domain. IEEE Trans Image Process 21(8):3339–3352
Sathacopoulou R, Magoulas GD, Grigoriadou M (1999) Neural network-based fuzzy modelling of the student in intelligent tutoring systems. Int Joint Conf Neural Netw 5:3517–3521
Saul L, Roweis S (2003) Think globally, fit locally: unsupervised learning of low dimensional manifolds. J Mach Learn Res vol. 4(no.xx):119–155
Seung HS (1998) Learning continuous attractors in recurrent networks. Adv Neural Info Proc Syst 10:654–660
Seung HS, Lee DD (2000) The manifold ways of perception. Science 290(5500):2268–2269
Sheikh HR, Bovik AC (2006) Image information and visual quality. IEEE Trans Image Process 15(2):430–444
Sheikh HR, Wang Z, Cormack L, and Bovik AC (2003) Live image quality assessment database release II. Available online: http://live.ece.utexas.edu/research/quality.
Sheikh HR, Bovik AC, Cormack LK (2005) No-reference quality assessment using natural scene statistics: JPEG2000. IEEE Trans Image Process 14(11):1918–1927
Sugeno M, Yasukawa T (1993) A fuzzy-logic-based approach to qualitative modeling. IEEE Trans Fuzzy Syst 1(1):7–31
Vapnik V (1998) Statistical learning theory. Wiley, New York
Varanasi MK, Aazhang B (1989) Parametric generalized Gaussian density estimation. J Acoust Soc Amer 86(4):1404–1415
Wang Z, Sheikh HR, Bovik AC (2002) No-reference perceptual quality assessment of JPEG compressed images. In: Proc IEEE Int Conf Image Process., pp 477–480
Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
Xiao R, Zhao Q, Zhang D, Shi PF (2011) Facial expression recognition on multiple manifolds. Pattern Recognition 44(1):107–116
Yang X, Fu H, Zha H, Barlow JL (2006) Semi-supervised nonlinear dimensionality reduction. In: Proc Int Conf Mach Learn., pp 1065–1072
Ye P, Doermann D (2012) No-reference image quality assessment using visual codebooks. IEEE Trans Image Process 21(7):3129–3138
Zadeh LA (1965) Fuzzy sets. Information and control 8(3):338–353
Zadeh LA (1988) Fuzzy logic. IEEE Comput Mag 21(4):83–93
Zadeh LA (1996) Fuzzy logic = computing with words. IEEE Trans Fuzzy Syst vol. 4(no. 2):103–111
Zhang S and Chau KW (2009) Dimension reduction using semi-supervised locally linear embedding for plant leaf classification. In:Proc. Int. Conf. Intell. Comput., pp.948-955.
Zhu X (2006) Semi-supervised learning literature survey. Technical Report 1530, Univ. of Wisconsin-Madison
Zhu X, Rogers T, Qian R, Kalish C (2007) Humans perform semi-supervised classification too. In:Proc. 22nd AAAI Conf Artif Intell., pp 864 – 869
Zhu X, Goldberg A, Brachman R, Dietterich T (2009) Introduction to semi-supervised learning. Morgan and Claypool, San Rafael, CA