www.elsevier.com/locate/sigpro

Biologically Inspired Image Quality Assessment

Fei Gao, Jun Yu

PII: S0165-1684(15)00285-6
DOI: http://dx.doi.org/10.1016/j.sigpro.2015.08.012
Reference: SIGPRO5891

To appear in: *Signal Processing*

Received date: 21 May 2015
Revised date: 14 August 2015
Accepted date: 23 August 2015

Cite this article as: Fei Gao and Jun Yu, Biologically Inspired Image Quality Assessment, *Signal Processing,* http://dx.doi.org/10.1016/j.sigpro.2015.08.012

Biologically Inspired Image Quality Assessment

Fei Gao[1] and Jun Yu[1]

[1] School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, P.R. China

***Abstract*—Image quality assessment (IQA) aims at developing computational models that can precisely and automatically estimate human perceived image quality. To date, various IQA methods have been proposed to mimic the processing of the human visual system, with limited success. Here, we present a novel IQA approach named biologically inspired feature similarity (BIFS), which is demonstrated to be highly consistent with the human perception. In the proposed approach, biologically inspired features (BIFs) of the test image and the relevant reference image are first extracted. Afterwards, local similarities between the reference BIFs and the distorted ones are calculated and then combined to obtain a final quality index. Thorough experiments on a number of IQA databases demonstrate that the proposed method is highly effective and robust, and outperform state-of-the-art FR-IQA methods across various datasets.**

***Index Terms*—Biologically inspired feature, full-reference, image quality assessment, percentile pooling, structural similarity.**

I. INTRODUCTION

The past decades have witnessed a dramatic increase in the number of images with the tremendous development of social networking websites, smartphones, and cameras. And various systems have been developed to deal with such a large scale of images. In these systems, image quality usually plays a significant role. For example, images of poor quality may lead to obstacles in learning or applying such systems for practical applications, e.g. scene recognition [1], image retrieval [2], and so on. In addition, image quality can be adopted as a criterion for evaluating the performance of image processing systems [3]-[5], optimizing image processing algorithms, and monitoring the working condition of devices [6]. Thus it is meaningful to develop image quality assessment (IQA) methods that can precisely and automatically estimate human perceived image quality.

In recent years, many IQA methods have been developed and we can classify them into three classes

[6]: full-reference (FR) IQA [7], reduced-reference (RR) IQA [8][9], and no-reference (NR) or blind IQA [10][11]. FR-IQA methods need all the information of the reference image, i.e. the undistorted version of the test image, is needed. In contrast, RR-IQA and NR-IQA methods only need part of or none of the information about the reference image. Consequently, the quality prediction accuracies of FR-IQA methods are usually better than present RR-IQA and NR-IQA methods.

Generally speaking, the intrinsic idea of FR-IQA is to estimate the quality of a test image by measuring the similarity or difference between the test image and the corresponding reference image. For example, in peak signal-to-noise ratio (PSNR) and root mean squared error (RMSE), the most widely used two IQA methods, the differences between the reference image and the test image are calculated pixel by pixel, and then combined into a single value. Because PSNR and MSE do not always consistent well with human perception [12], great efforts have been paid to develop progressive methods for quality assessment in the past decades [13]-[28]. And many of them have shown impressive and inspiring consistency with human perception over a large range of datasets [26][28].

Since the goal of IQA is to approximate human beings' judgments of image quality, it is meaningful to develop IQA methods that mimic the perception mechanism of the human visual system (HVS). Although many attempts have been proposed, most of them only consider some particular properties of HVS, e.g. contrast sensitivity function (CSF) [18][29], just noticeable difference (JND) [30], and visual attention (VA) [31], etc. Usually they do not perform as well as state-of-the-art FR-IQA methods. To date, only limited IQA methods have been proposed to formula the processing in the visual cortex [27]. And the properties of primary visual cortex, V1, have not been well explored for IQA, although neuroscientists have demonstrated that V1 plays a much significant in visual processing [32].

In this paper, we utilize biologically inspired feature (BIF) models [33] to mimic the properties of (S1) and complex (C1) cells in V1, and construct a novel IQA index by measuring the similarity between the BIFs of the reference image and those of the test image. Although BIFs have been introduced to FR-IQA before, it was adopted for estimating visual attention [34]. In contrast, in the proposed method, BIFs are deplored for representing the input image in the primary visual cortex and directly utilized for quality prediction. Thorough experiments conducted on various IQA databases demonstrate that the proposed method is in highly consistency with human perception and outperform state-of-the-art FR-IQA methods across a number of datasets. The highlights of the proposed method are summarized below:

a) We explore BIF for FR-IQA by employing it to mimic the processing in the primary visual cortex;

b) We construct a novel FR-IQA framework by measuring the similarity between the BIFs of the test image and the BIFs of the corresponding reference image; and

c) Thorough experiments on existing databases demonstrate that the proposed method is highly comparable with state-of-the-art FR-IQA methods.

The rest of the paper is organized as follows. Section II introduces the calculation of BIFs. In Section III, we present the framework of the proposed quality evaluation method. Extensive experiments conducted on standard IQA datasets are presented and analyzed in Section IV. Section V concludes the paper.

II. Biologically Inspired Features

Biologically inspired feature models mimic the tuning properties of the simple and complex cells in V1 and have been demonstrated to be effective and efficient for solving various image processing problems, e.g. scene classification [33], object recognition [35][36], visual attention detection [32][37], and so on. We thus choose to use BIF for representing an image in the proposed research. Specially, we follow the work presented in [33], and adopt the C1 units, intensity units, and color units as the unified BIFs for quality prediction. For more details please refer to [33].

A. C1 Units

The C1 units correspond to the complex cells in V1. Because these complex cells show tolerance to shift and size, the C1 responses are set as the max response of a local area of the S1 responses from the same orientation. And the S1 responses are obtained by applying Gabor functions to the input image. The Gabor mother function is given by:

$$G(x,y)=\exp(-(x_0+\gamma^2 y_0^2)/(2\delta^2))\times\cos(2\pi x_0/\lambda), \quad (1)$$

where $x_0=x\cos\theta+y\sin\theta$ and $y_0=-x\sin\theta+y\cos\theta$ are the range of x and y decides the scales of Gabor filters, and θ decides the orientations; δ is effective width, and λ is wavelength.

Following the settings in [33], we arrange the Gabor filters to form a pyramid of scales, spanning a range of sizes from 7×7 to 21×21 pixels in steps of two pixels. Besides, we consider four orientations: $0°$, $45°$, $90°$, and $135°$. Consequently, we have 32 Gabor filters, and 32 feature maps in S1 units. Afterwards, we pool over S1 units by using a maximum operation to obtain 16 C1 feature maps. In particular, a box of size $k\times k$ slides on two adjacent scales (with an identical orientation) of S1 units, and the maximum intensity value is adopted as the

corresponding pixel in C1 feature map. In the experiments, the box sizes range from 8×8 to 14×14 with a step of two pixels.

B. Intensity Units

There are a large number of cells which are sensitive to bright centers on dark surrounds or dark centers on bright surrounds [37]. In spired by this conclusion, Song *et al*. compute the intensity units by using a center-surround operation [33]. The procedure of computing the intensity units is shown in Fig. 1. First, dyadic Gaussian pyramids are used to convolve on the intensity channel of an input image, resulting in nine spatial scales with a ratio from 1:1 (level 0) to 1:256 (level 8). Afterwards, we choose some center and surround levels and perform the center-surround operation by:

$$I(c,s) = \left|I(c) - Interp_{s-c}(I(s))\right|. \tag{2}$$

where $c = 2,3,4$ indicates the center levels; $s = c + d$, with $d = 3,4$, indicates the surround levels; and $Interp_{s-c}$ indicates the operation of interpolating the surround feature map to the same size of the relevant center feature map. Finally, we have 6 intensity feature maps.

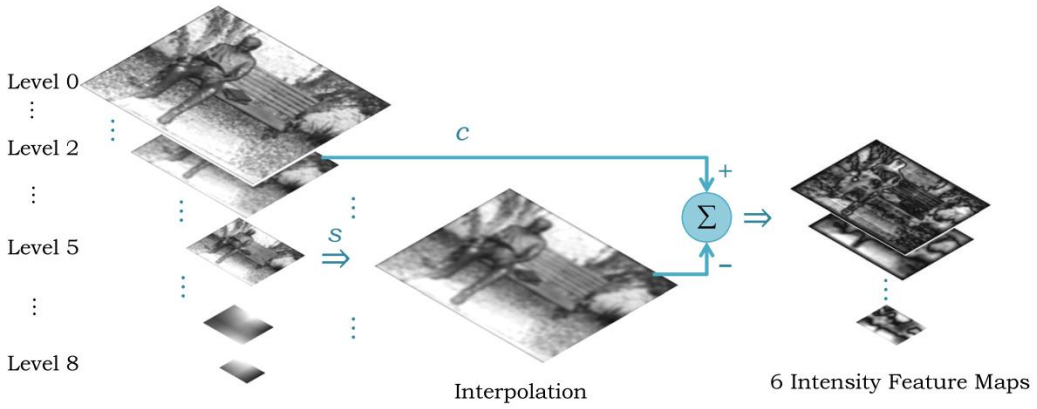

Fig. 1 Flowchart of computing the intensity units.

C. Color Units

Various neurophysiological results have demonstrated that double-opponent cells, which are excited by one color (e.g. red) and inhibited by another color (e.g. green) in the center of receptive field, so are cells in the surround [37]. Inspired by this discovery, Song *et al.* proposed to calculate the color units as follows.

Given a color image, four color channels are adopted:

$$R = (g+b)/2, G = g - (r+b)/2, B = b - (r+g)/2, Y = r + g - 2(|r-g| + b) \tag{3}$$

For each color channel, dyadic Gaussian pyramids are used to generate nine spatial scales with a ratio from 1:1 (level 0) to 1:256 (level 8). According to the property of double-opponent cells,

two color pairs are adopted: red-green (R-G) and blue-yellow (B-Y). Afterwards, the feature maps are calculated by:

$$\begin{aligned} RG(c,s) &= \left|\left(R(c)-G(c)\right)-Interp_{s-c}\left(R(s)-G(s)\right)\right|, \\ BY(c,s) &= \left|\left(B(c)-Y(c)\right)-Interp_{s-c}\left(B(s)-Y(s)\right)\right| \end{aligned} \tag{4}$$

where c, s, and $Interp_{s-c}$ are the same as previously introduced. Finally we obtain 12 feature maps (6 *RG* feature maps + 6 *BY* feature maps) by subtracting the interpolated surround map from the relevant center map. And such color units have been demonstrated to be valuable for color image representation [32].

D. Image Representation

Based on the previous introduction, a color image is presented by 34 feature maps (16 C1 units + 12 Color units + 6 intensity units), and a grayscale image is presented by 22 feature maps (16 C1 units + 6 intensity units). In the proposed IQA framework, these feature maps are adopted as the image representation in the primary visual cortex and utilized for quality prediction.

III. Quality Assessment Framework

The proposed IQA approach simulates the process in the visual cortex and can be divided into three main components: the biologically inspired feature maps, the similarity maps between the distorted feature maps and the relevant reference feature maps, and the percentile pooling based quality prediction. And we term the proposed IQA method as biologically inspired feature similarity (BIFS). The flowchart of the BIFS is shown in Fig. 2. Details will be discussed below.

A. Biologically Inspired Features Extraction

In this framework, both the test image and the corresponding reference image are presented by the biological inspired features introduced in Section II. For clarity, we denote the feature maps of the reference image and the test image by $\{\mathbf{F}_1^{ref},\ldots,\mathbf{F}_N^{ref}\}$ and $\{\mathbf{F}_1^{test},\ldots,\mathbf{F}_N^{test}\}$, respectively. N is the number of feature maps. As introduced in Section II, $N=34$ if the relevant image is colorful, $N=22$ if the relevant image is gray. It is notable that the test image and reference image should be uniformly colorful or gray, thus they will correspond to the same number of feature maps.

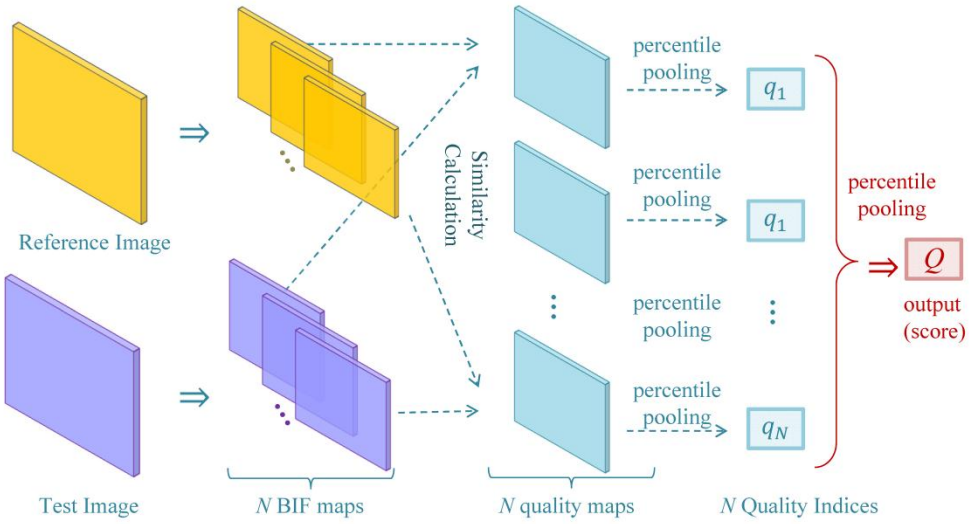

Fig. 2 Framework of the biologically inspired feature similarity (BIFS) approach.

B. Similarity Maps Estimation

Because BIFs are precise and robust representations of images, it is reasonable for us to estimate the quality of the test image by computing the similarities or differences between its BIFs and the BIFs of the relevant reference image. In the proposed framework, we first measure the local similarity between one reference feature map $\mathbf{F}_i^{ref}$ and the relevant test feature map $\mathbf{F}_i^{test}$, $i=1,\ldots,N$. The local similarity can be viewed as the local quality, thus we can obtain N quality maps: $\{\mathbf{Q}_1,\ldots,\mathbf{Q}_N\}$.

Let $\mathbf{f}_i^{ref}(x,y)$ be the area of size $w\times w$ and centered at (x,y) in $\mathbf{F}_i^{ref}$, and $\mathbf{f}_i^{test}(x,y)$ be the relevant area in $\mathbf{F}_i^{test}$. The magnitude, contrast, and structure similarities between $\mathbf{f}_i^{ref}(x,y)$ and $\mathbf{f}_i^{test}(x,y)$ are computed by

$$\begin{aligned}
\mathbf{m}_i(x,y) &= \frac{2\mu_i^{ref}(x,y)\mu_i^{test}(x,y)+C_1}{(\mu_i^{ref}(x,y))^2+(\mu_i^{test}(x,y))^2+C_1},\\
\mathbf{c}_i(x,y) &= \frac{2\sigma_i^{ref}(x,y)\sigma_i^{test}(x,y)+C_2}{(\sigma_i^{ref}(x,y))^2+(\sigma_i^{test}(x,y))^2+C_2},\\
\mathbf{s}_i(x,y) &= \frac{2\sigma_i^{ref,test}(x,y)+C_3}{\sigma_i^{ref}(x,y)+\sigma_i^{test}(x,y)+C_3},
\end{aligned} \tag{5}$$

where $\mu_i^{ref}(x,y)$ and $\mu_i^{test}(x,y)$ are the mean value of $\mathbf{f}_i^{ref}(x,y)$ and $\mathbf{f}_i^{test}(x,y)$, respectively; $\sigma_i^{ref}(x,y)$ and $\sigma_i^{test}(x,y)$ are the standard variance of $\mathbf{f}_i^{ref}(x,y)$ and $\mathbf{f}_i^{test}(x,y)$, respectively; $\sigma_i^{ref,test}(x,y)$ is the covariance between $\mathbf{f}_i^{ref}(x,y)$ and $\mathbf{f}_i^{test}(x,y)$; C_1, C_2, and C_3 are positive constants and adopted to

prevent the denominators from being zero.

Finally, the similarity index associated with location (x, y) is estimated by:

$$\mathbf{Q}_i(x,y) = \mathbf{m}_i(x,y) \cdot \mathbf{c}_i(x,y) \cdot \mathbf{s}_i(x,y). \quad (6)$$

In our experiments, we chose $C_1 = C_2 = C_3 = 0.001$ and $w = 11$ for simplicity. Preliminary experimental results show that the performance of the proposed IQA method is not sensitive to the values of them.

C. Quality Score Prediction

Given a test image, the areas of poorer quality usually have greater impacts on subjects' judgments of perceived image quality than those of better quality. We therefore utilize a percentile-pooling method to pool over all the relevant quality maps to obtain the final quality index. Specially, we first compute the average of the lowest $p\%(0 < p \le 100)$ quality values for each quality map, and denote them by $\{q_1, \ldots, q_N\}$. Afterwards, the average of the lowest $K(1 \le K \le N)$ quality values in $\{q_1, \ldots, q_N\}$ is computed and adopted as the final quality score of the test image.

IV. Experimental Results

To evaluate the performance of the proposed method, we test it on several existing IQA databases and compare it with a number of state-of-the-art FR-IQA methods.

A. Databases

In the experiments, we utilized eight IQA databases, i.e. the A57 database[1] [18], Categorical Subjective Image Quality (CSIQ) database [29], IRCCyN/IVC database[2], Laboratory of Image and Video Evaluation (LIVE) database[3] [7], and LIVE multiple distorted (LIVEMD) database [39], Media Information and Communication Technology Laboratory (MICT) database[4], Tampere Image Database 2013 (TID2013) [40], and Wireless Imaging Quality (WIQ) database[5].

Each dataset contains a number of distorted images generated from various original images by

[1] [Online]. Available: http://foulard.ece.cornell.edu/dmc27/vsnr/a57_db.zip/

[2] IRCCyN stands for Institut de Recherche en Communications et Cybernétique de Nantes, and IVC stands for Image and Video Communications. [Online]. Available: http://www.irccyn.ec-nantes.fr/ivcdb/

[3] [Online]. Available: http://live.ece.utexas. edu/research/quality

[4] [Online]. Available: http://mict.eng.u-toyama.ac.jp/mictdb.html

[5] [Online]. Available: http://www.bth.se/tek/rcg.nsf/pages/wiq-db

corrupting them with different types of distortions. And the mean opinion score (MOS) or difference mean opinion score (DMOS) of each image is available [7]. The MOS of an image is the mean of the relevant quality scores reported by different subjects. DMOS is the difference between the perfect quality score and the MOS. Thus a higher MOS value and a lower DMOS value indicate better quality. The compositions of these datasets are summarized in Table I and briefly introduced below.

Table I The compositions of present IQA databases.

	A57	CSIQ	IVC	LIVE	LIVEMD	MICT	TID2013	WIQ
# of distortion categories	6	6	5	5	2	2	24	2
# of reference images	3	30	10	29	12	14	25	7
# of images	54	866	185	982	450	196	3000	80
color images ?	×	√	√	√	√	√	√	×
hybrid distortion ?	×	×	×	×	√	×	×	×

The A57 database includes 54 gray images which are generated from three original gray images by corrupting them with six types of distortions. The distortions includes: quantization of the LH subbands of a 5-level DWT of the image using the 9/7 filters (QLH), additive Gaussian white noise (WN), baseline JPEG compression (JPEG), JPEG2000 compression (JP2k), JPEG2000 compression with the dynamic contrast-based quantization algorithm (DCQ), and Gaussian Blurring (Gblur). And the subjective rating of each image is available.

The CSIQ database consists of 866 color images which are derived from 30 original images. Six types of distortion are considered in CSIQ: WN, JPEG, JP2k, additive Gaussian pink noise (PN), Gblur, and global contrast decrements (GCD). The DMOS of each image is released.

The IRCCyN/IVC database includes 10 original images and 235 distorted images derived from them. This database includes four types of distortions: JPEG, JP2k, locally adaptive resolution (LAR) coding, and Blurrring. And the JPEG compressed images are further divided into two subsets: JPEG compression on the luminance channel and the monochrome channels (JPEG_lumi+chr) and JPEG compression on the luminance channel (JPEG_lumi).The DMOS of each image is available.

The LIVE database includes 982 color images, which are generated from 29 original images. And it contains five types of distortion, i.e. JP2k, JPEG, WN, Gblur, and fast fading Reilly Channel distortion (FF). Besides, DMOS of each image is available.

The TID2013 includes 3,000 color images in sum. These images are generated from 25 original images

by corrupting them with 24 types of distortion at 5 different levels. The distortion types include: WN (#1), additive white Gaussian noise which is more intensive in color components than in the luminance component (#2), additive Gaussian spatially correlated noise (#3), masked noise (#4), high frequency noise (#5), impulse noise (#6), quantization noise (#7), Gblur (#8), image denoising (residual noise, #9), JPEG (#10), JP2k (#11), JPEG transmission errors (#12), JPEG2000 transmission errors (#13), non-eccentricity pattern noise (#14), local block-wise distortion of different intensity (#15), mean shift (#16), contrast change (#17), change of color saturation (#18), multiplicative Gaussian noise (#19), comfort noise (#20), lossy compression of noisy images (#21), image color quantization with dither (#22), chromatic aberrations (#23), and sparse sampling and reconstruction (#24). The MOS of each image is available.

The images in LIVEMD database are distorted by multiple types of distortion. Specially, there are two subsets, one of which conlude images corrupted by Gblur followed by JPEG (GblurJPEG), and the other conclude images corrupted by Gblur followed by WN (GblurWN). And each subset includes 225 color images. The DMOS of each image is available.

The MICT database (i.e. the Toyama database) includes 196 images which are generated from 14 original images by corrupting them with JPEG or JP2k. The MOS of each image is available. And the WIQ database includes 7 reference images and 80 distorted images derived from them. Similarly, the subjective scores of the distorted images are released.

We adopt two widely used indices as the criteria for evaluating the performance of IQA metrics, i.e. the Pearson's linear correlation coefficient (PLCC) and Spearman's rank-order correlation coefficient (SRCC) between the predicted quality scores using a FR-IQA method and the subjective scores. Greater PLCC and SRCC values indicate better performance.

Let $\{x_1,...,x_n\}$ denote the predicted quality scores of a given set of n images, and $\{y_1,...,y_n\}$ the corresponding subjective quality scores. The formula for PLCC is:

$$\mathrm{PLCC}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}, \tag{7}$$

where, x_i and y_i are related to the i-th image; $\bar{x}$ and $\bar{y}$ are the mean values of $\{x_1,...,x_n\}$ and $\{y_1,...,y_n\}$, respectively. If we convert the scores to ranks and denote them as $\{X_1,...,X_n\}$ and $\{Y_1,...,Y_n\}$

accordingly, SRCC is computed from:

$$\mathrm{SRCC} = 1 - \frac{6\sum_{i=1}^{n}\left(X_i - Y_i\right)^2}{n\left(n^2-1\right)}, \tag{8}$$

where, X_i and Y_i are the ranking orders of the i -th image in $\{x_1,...,x_n\}$ and $\{y_1,...,y_n\}$, respectively.

We perform a nonlinear mapping using a logistic function as described in [41] before computing these indices. The logistic function is shown below:

$$Q = \beta_1\left(\frac{1}{2} - \frac{1}{1+e^{\beta_2(q-\beta_3)}} + \beta_4 x + \beta_5\right), \tag{9}$$

where, q is the raw quality score estimated by using an IQA method, Q is the final quality score adopted in calculating PLCC and SRCC, and $\{\beta_1,...,\beta_5\}$ are the parameters of the logistic function. And we estimate $\{\beta_1,...,\beta_5\}$ by using the MATLAB function *nlinfit* .

B. Variation with Algorithm Parameters

In this subsection, we focus on the impact of choosing different algorithm parameters: 1) the percents of quality values utilized to estimate the quality index for each quality map, $p\%(0 < p \le 100)$; and 2) the number of feature maps utilized for estimating the final quality score, $K(1 \le K \le N)$. N is the number of feature maps associated with a given image. We conduct experiments on the LIVE database. As previously discussed, we can obtain $N = 34$ feature maps for each color image in the LIVE database. Thus we considered the following parameters: $p \in \{10, 20, ..., 100\}$ and $K = 1, 2, ..., 34$.

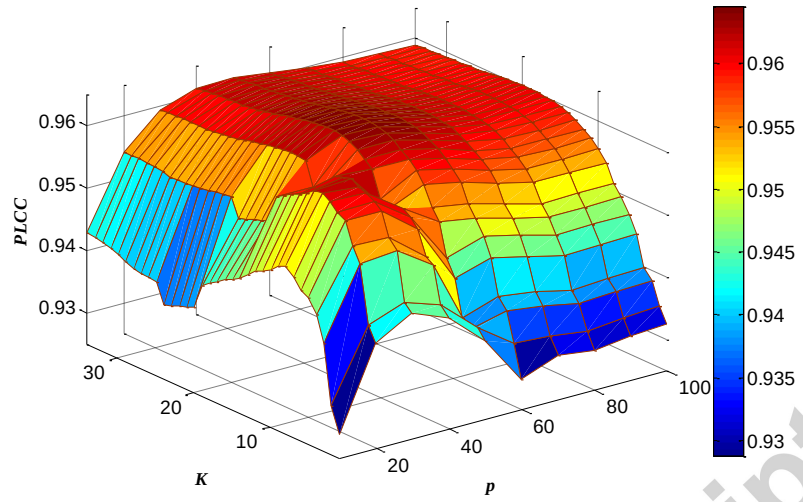

Fig. 3 The variation of PLCC with algorithm parameters: p and K .

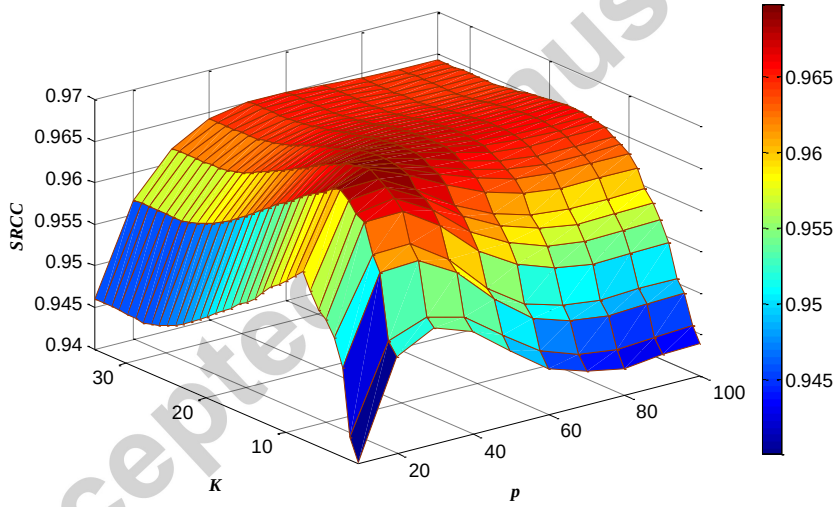

Fig. 4 The variation of SRCC with algorithm parameters p and K .

The PLCC and SRCC values on the entire LIVE database while adopting different values of p and K are shown in Figs. 3 and 4, respectively. Clearly, the performance of BIFS first increases and then slightly decreases when the value of p or K becomes larger. Recall that the areas with poorer quality have a greater impact on subjects' judgments of the image perceived quality. Thus when p and K are too large, the areas with better qualities may cause biases in the estimated quality score. In contrast, when p and K are small, insufficient information about the human perceived quality is considered, resulting in impressive estimations of quality scores. And BIFS obtains the best performance while $p = 40$ and $K = 12$.

C. Consistency with Human Perception

In this section, we compare the performances of BIFS with a number of state-of-the-art FR-IQA metrics, i.e., visual information fidelity measure (VIF) [17], feature-similarity (FSIMc) [26], gradient magnitude similarity deviation (GMSD) [28], and the quality metric with internal generative mechanism (IGM) [27], as well as two widely used FR-IQA methods, i.e. RMSE and structural similarity (SSIM) [15]. For all of these methods but RMSE, we used the implementations released by the original authors, which are online available. For BIFS, we set $p = 40$ and $K = 12$..

For each database, the PLCC and SRCC values associated with each distortion subset as well as the whole database (denoted by "All") are calculated and shown in Tables II and III, respectively. Besides, the weighted averages of the PLCC and SRCC values over all the databases for each method are calculated and listed in the last columns of Tables II and III, respectively. The average SRCC/PLCC is calculated by weighting the SRCC/PLCC values according to the number of images contained in each database. The best performance for each dataset is highlighted in boldface.

It is rather inspiring that our approach statistically yields the best result across all databases according to SRCC. And the proposed method performs either better than or highly comparable to state-of-the-arts across all the datasets. It is impressed that BIFS obtain a remarkable higher SRCC on the largest database, TID2013, than all the existing FR-IQA methods. In addition, BIFS works well for distortion #18 (change of color saturation) in the TID2013 database, while all the other FR-IQA methods fail to predict human perceived quality scores correctly. This demonstrates the strong power of BIFs for representing color images.

Based on the previously presented experimental results, we summarize the advantages of the proposed IQA framework as follows:

- Statistically, the proposed method is highly consistent with human perception and outperforms existing FR-IQA methods across a large number of distortion categories and databases;
- The high consistency between the predicted quality scores and subjective quality judgments corroborates the expectation that BIFs is a versatile representation for image quality and that the proposed framework is insensitive to different databases; and
- It is easy and efficient to extend the proposed framework to other types of image features, e.g. wavelet

features. And we can obtain good performance after some possible slight tuning of the parameters.

Table II PLCC values associated with BIFS and present FR-IQA methods on existing IQA databases.

Database	Distortion	RMSE	SSIM	VIF	FSIMc	GMSD	IGM	BIFS
A57	Gblur	0.692	0.914	0.950	**0.966**	0.963	0.923	0.931
	DCQ	0.659	0.873	0.967	**0.997**	0.995	0.957	0.854
	QLH	0.880	0.913	0.847	**0.995**	0.980	0.994	0.857
	JP2k	0.795	0.952	0.860	0.910	**0.970**	0.911	0.793
	JPEG	0.746	**0.991**	0.953	0.951	0.972	0.968	0.860
	WN	0.880	0.863	0.899	0.885	**0.982**	0.959	0.931
	All	0.657	0.415	0.614	**0.939**	0.907	0.923	0.839
CSIQ	WN	0.941	0.927	0.962	0.930	**0.968**	0.967	0.932
	JPEG	0.866	0.936	**0.964**	0.960	**0.964**	**0.964**	0.951
	JP2k	0.933	0.929	0.963	0.974	0.974	**0.976**	0.952
	PN	0.916	0.896	**0.959**	0.931	0.958	0.955	0.918
	Gblur	0.919	0.913	0.955	0.957	0.961	**0.962**	0.958
	GCD	0.899	0.759	**0.949**	0.935	0.928	**0.949**	0.931
	All	0.658	0.819	0.915	0.903	**0.947**	0.919	0.930
IRCCyN/IVC	Blur	0.831	0.918	0.989	0.988	0.972	0.978	**0.992**
	JP2k	0.760	0.863	0.936	0.942	0.942	0.937	**0.944**
	JPEG_lumi+chr	0.520	0.800	0.919	0.947	0.924	0.914	**0.950**
	JPEG_lumi	0.500	0.837	0.940	**0.962**	0.955	0.944	0.934
	LAR	0.602	0.764	0.906	0.938	**0.948**	0.921	0.901
	All	0.613	0.792	0.903	**0.938**	0.923	0.913	0.915
LIVE	JP2k	0.944	0.967	0.986	**0.987**	0.986	0.985	0.981
	JPEG	0.934	0.970	**0.992**	0.988	0.990	0.991	0.984
	WN	**0.993**	0.970	**0.993**	0.982	0.983	**0.993**	0.987
	Gblur	0.900	0.932	0.985	0.977	0.976	0.979	0.948
	FF	0.926	0.962	**0.979**	0.961	0.959	0.945	0.916
	All	0.926	0.942	**0.976**	0.973	0.975	0.974	0.960
LIVEMD	GblurJPEG	0.746	0.761	**0.905**	0.819	0.875	0.893	0.887
	GblurWN	0.783	0.747	0.849	0.818	0.852	0.883	**0.887**
	All	0.742	0.668	0.872	0.818	0.863	**0.886**	0.884
MICT	JPEG	0.575	0.757	0.924	0.918	0.868	0.871	**0.940**
	JP2k	0.887	0.936	**0.968**	**0.968**	0.954	0.946	0.958
	All	0.740	0.847	0.932	0.927	0.890	0.899	**0.936**
TID2013	#1	**0.954**	0.859	0.908	0.904	0.949	0.949	0.879
	#2	**0.923**	0.808	0.875	0.859	0.910	0.898	0.805
	#3	**0.951**	0.871	0.889	0.886	0.939	0.927	0.863
	#4	**0.870**	0.843	0.880	0.833	0.755	0.854	0.739
	#5	**0.973**	0.883	0.950	0.935	0.955	0.969	0.936
	#6	**0.872**	0.796	0.852	0.798	0.755	0.853	0.735
	#7	0.887	0.813	0.823	0.874	**0.911**	0.895	0.869
	#8	0.917	0.960	0.953	0.955	0.909	**0.977**	0.970
	#9	0.964	0.939	0.930	0.963	**0.976**	0.973	0.952
	#10	0.917	0.933	0.955	0.971	**0.984**	0.977	0.960
	#11	0.919	0.916	0.966	0.973	0.980	**0.982**	0.972
	#12	0.788	0.871	0.908	0.915	0.892	0.903	**0.925**
	#13	0.890	0.866	0.869	0.892	**0.908**	**0.908**	0.877
	#14	0.673	0.753	0.806	0.804	0.812	0.792	**0.820**
	#15	0.141	0.621	0.569	0.546	**0.637**	0.572	0.487
	#16	0.796	**0.802**	0.706	0.788	0.771	0.687	0.761
	#17	0.553	0.542	**0.899**	0.749	0.754	0.747	0.729
	#18	0.238	0.435	0.344	0.172	0.363	0.399	**0.827**
	#19	**0.920**	0.783	0.862	0.851	0.891	0.897	0.821
	#20	0.889	0.856	0.933	0.945	**0.956**	0.951	0.934
	#21	0.943	0.913	0.932	0.954	**0.970**	0.958	0.937
	#22	**0.931**	0.791	0.868	0.888	0.920	0.918	0.897
	#23	0.949	0.966	0.973	0.977	0.973	**0.982**	0.970
	#24	0.931	0.906	0.959	0.976	**0.985**	0.981	0.965
	All	0.682	0.686	0.772	**0.859**	**0.859**	0.856	0.851
WIQ	All	0.766	0.674	0.761	0.855	**0.874**	0.826	0.811
Weighted Average		0.725	0.754	0.843	0.887	**0.896**	0.891	0.888

Table III SRCC values associated with BIFS and present FR-IQA methods on existing IQA databases.

Database	Distortion	RMSE	SSIM	VIF	FSIMc	GMSD	IGM	BIFS
A57	Gblur	0.467	0.683	0.817	**0.883**	0.850	0.667	0.700
	DCQ	0.500	0.767	0.900	**0.967**	**0.967**	0.850	0.883
	QLH	0.900	0.767	0.817	**1.000**	0.983	0.983	0.800
	JP2k	0.800	0.883	0.900	0.817	**0.967**	0.833	0.683
	JPEG	0.633	**0.950**	0.900	0.933	**0.950**	**0.950**	0.833
	WN	0.950	0.817	0.950	0.900	**0.967**	0.950	0.917
	All	0.619	0.407	0.622	**0.918**	0.912	0.898	0.819
CSIQ	WN	0.936	0.926	0.958	0.927	**0.968**	0.964	0.930
	JPEG	0.888	0.920	0.955	0.951	0.953	**0.954**	0.943
	JP2k	0.936	0.925	0.959	0.966	0.971	**0.973**	0.940
	PN	0.933	0.894	**0.952**	0.925	0.951	0.943	0.910
	Gblur	0.925	0.931	0.959	**0.966**	0.963	0.963	0.961
	GCD	0.874	0.754	**0.940**	0.935	0.909	0.950	0.938
	All	0.805	0.839	0.911	0.917	**0.950**	0.932	0.935
IRCCyN/IVC	Blur	0.805	0.869	0.973	0.967	0.950	0.955	**0.974**
	JP2k	0.850	0.849	0.936	0.940	0.937	0.934	**0.945**
	JPEG_lumi+chr	0.562	0.746	0.878	**0.937**	0.908	0.873	0.933
	JPEG_lumi	0.660	0.827	0.926	**0.958**	0.948	0.930	0.926
	LAR	0.699	0.712	0.888	0.884	**0.897**	0.875	0.860
	All	0.688	0.779	0.896	**0.926**	0.914	0.903	0.906
LIVE	JP2k	0.956	0.973	0.979	**0.988**	**0.988**	0.986	0.984
	JPEG	0.942	0.970	0.980	**0.987**	0.985	0.986	0.981
	WN	0.992	0.978	0.989	0.980	0.985	**0.993**	0.991
	Gblur	0.873	0.939	**0.982**	0.983	0.975	0.973	0.955
	FF	0.936	0.966	**0.978**	0.971	0.966	0.953	0.932
	All	0.936	0.954	0.977	**0.981**	0.979	0.978	0.968
LIVEMD	GblurJPEG	0.662	0.744	**0.879**	0.854	0.847	0.856	0.836
	GblurWN	0.709	0.702	**0.881**	0.864	0.837	0.855	0.868
	All	0.677	0.646	**0.882**	0.864	0.845	0.856	0.855
MICT	JPEG	0.501	0.739	0.917	0.917	0.865	0.867	**0.928**
	JP2k	0.884	0.922	**0.947**	0.949	0.936	0.930	0.941
	All	0.722	0.840	0.917	0.919	0.883	0.891	**0.922**
TID2013	#1	0.929	0.853	0.900	0.898	0.946	**0.937**	0.875
	#2	**0.898**	0.774	0.843	0.821	0.868	0.879	0.774
	#3	0.920	0.862	0.889	0.875	**0.935**	0.924	0.859
	#4	0.832	0.810	**0.845**	0.794	0.717	0.805	0.747
	#5	0.918	0.848	0.897	0.900	0.916	**0.926**	0.909
	#6	**0.897**	0.799	0.854	0.807	0.764	0.859	0.742
	#7	0.883	0.806	0.816	0.872	**0.905**	0.891	0.862
	#8	0.916	0.963	0.965	0.955	0.911	**0.977**	0.974
	#9	0.948	0.910	0.906	0.930	**0.952**	0.949	0.926
	#10	0.919	0.910	0.919	0.938	**0.951**	0.949	0.923
	#11	0.890	0.905	0.952	0.958	0.966	**0.968**	0.959
	#12	0.777	0.818	0.844	0.846	0.849	0.845	**0.901**
	#13	0.902	0.877	0.876	0.891	**0.914**	0.918	0.876
	#14	0.686	0.759	0.772	0.792	**0.814**	0.803	0.810
	#15	0.155	**0.617**	0.531	0.552	0.663	0.527	0.384
	#16	0.767	**0.777**	0.628	0.753	0.735	0.609	0.741
	#17	0.440	0.364	**0.852**	0.486	0.621	0.460	0.512
	#18	0.099	0.406	0.310	0.275	0.302	0.322	**0.825**
	#19	**0.891**	0.775	0.847	0.847	0.889	0.883	0.814
	#20	0.858	0.819	0.895	0.912	**0.930**	0.920	0.897
	#21	0.914	0.911	0.923	0.947	**0.963**	0.949	0.931
	#22	**0.927**	0.789	0.847	0.876	0.910	0.907	0.881
	#23	0.887	0.889	0.885	0.871	0.853	**0.914**	0.899
	#24	0.907	0.903	0.937	0.956	**0.968**	0.967	0.955
	All	0.674	0.627	0.668	0.802	0.804	0.810	**0.832**
WIQ	All	0.626	0.524	0.692	**0.801**	0.799	0.771	0.726
Weighted Average		0.739	0.724	0.788	0.863	0.866	0.866	**0.876**

V. CONCLUSIONS

In this paper, biologically inspired feature is introduced to mimic the processing in the primary visual cortex. Afterwards, the similarity between the BIFs of the reference image and the BIFS of the distorted image is calculated for quality prediction. The comparison of both proposed algorithm with state-of-the-art FR-IQA metrics on a number of databases shows it has an impressive consistency with human perception and overwhelming superiority over the state-of-the-art FR-IQA metrics for various types of distortions. Nevertheless, they are still inferior to the best performances for some distortions. In addition, both metric learning [42]-[44] and multi-view learning [45]-[47] techniques have shown great advantages in solving computer vision problems. It is valuable to integrate BIFs with them to develop effective IQA algorithms. Besides, it is meaningful to explore BIFs in IQA for other types of images, e.g. HDR images [48] and sketches [49][50]. How to improve the efficiency of computing BIFs will be another future work.

REFERENCES

[1] J. Yu, C. Hong, D. Tao, and M. Wang, "Semantic embedding for indoor scene recognition by weighted hypergraph learning," *Signal Processing*, 112: 129-136, 2015.

[2] J. Yu, D. Tao, M. Wang, Y. Rui, "Learning to rank using user clicks and visual features for image retrieval," *IEEE Transactions on Cybernetics*, 45(4): 767-779, 2015.

[3] Y. Wang, D. Tao, X. Li, M. Song, J. Bu, and P. Tan, "Video Tonal Stabilization Color States Smoothing," IEEE Transactions on Image Processing, vol. 23, no. 11, pp. 4838-4849, 2014.

[4] M. Song, D. Tao, J. Bu, C. Chen, and Y. Yang, "Color-to-Gray based on Chance of Happening Preservation," Neurocomputing, vol. 119, pp. 222-231, 2013.

[5] M. Song, D. Tao, C. Chen, X. Li, and C.W. Chen. "Color to Gray: Visual Cue Preservation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1537-1552, 2010.

[6] Z. Wang and A. C. Bovik, Modern Image Quality Assessment. New York: Morgan and Claypool Publishing Company, 2006.

[7] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, 15(11): 3440-3451, 2006.

[8] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E. Yang, and A. C. Bovik, "Quality-aware images." *IEEE Transactions on Image Processing*, 15(6): 1680-1689, 2006.

[9] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment based on multiscale geometric analysis", *IEEE Transactions on Image Processing*, 18(7): 1409-1423, 2009.

[10] L. He, D. Tao, X. Li, and X. Gao, "Sparse representation for blind image quality assessment", in *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (CVPR), 1146-1153, 2012.

[11] X. Gao, F. Gao, D. Tao, and X. Li, "Universal blind image quality assessment metrics via natural scene statistics and multiple kernel learning," IEEE Transactions on Neural Networks and Learning Systems, 24(12): 2013-2026, 2013.

[12] Z. Wang and A. C. Bovik, "Mean squared error: love it or leave it? - A new look at signal fidelity measures," *IEEE Signal Processing Magazine*, 26(1): 98-117, 2009.

[13] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Transactions on Image Processing*, 6(8): 1164-1175, 1997.

[14] A. P. Bradley, "A wavelet difference predictor," *IEEE Transactions on Image Processing*, 8(5): 717-730, 1999.

[15] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, 13(4): 600-612, 2004.

[16] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing*, 14(12): 2117-2128, 2005.

[17] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, 15(2): 430-444, 2006.

[18] M. D. Chandler and S. S. Hemami, "VSNR: A Wavelet-based visual signal-to-noise ratio for natural images," *IEEE Transactions on Image Processing*, 16(9): 2284-2298, 2007.

[19] X. B. Gao, T. Wang, and J. Li, "A content-based image quality metric," D. Slezak et al. (Eds.): RSFDGrC2005, *Lecture Notes in Artificial Intelligence*, LNAI 3642, 231-240, 2005.

[20] Z. Wang, and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, 20(5): 1185-1198, 2011.

[21] L. He, W. Lu, X. Gao, D. Tao, and X. Li, "A novel metric based on MCA for image quality," *International Journal of Wavelets, Multiresolution and Information Processing*, 9(5): 743-757, 2011.

[22] L. He, X. Gao, W. Lu, X. Li, and D. Tao, "Image quality assessment based on S-CIELAB model," *Signal, Image and Video Processing*, 5(3): 283-290, 2011.

[23] L. He, D. Wang, X. Li, D. Tao, X. Gao, and F. Gao, "Color fractal structure model for reduced-reference colorful image quality assessment," *ICONIP*, Part II, LNCS 7664: 401-408, 2012.

[24] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Transactions on Image Processing*, 21(4): 1500-1512, 2012.

[25] T.J. Liu, W. Lin, and C.C.J. Kuo, "Image quality assessment using multi-method fusion," *IEEE Transactions on Image Processing*, 2013, 22(5): 1793-807.

[26] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, 2011, 20(8): 2378-2386.

[27] J. Wu, W. Lin, G. Shi, and A. Liu, "Perceptual quality metric with internal generative mechanism," IEEE Transactions on Image Processing, 22(1): 43-54, Jan. 2013.

[28] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Transactions on Image Processing*, 2014, 23(2): 684-695.

[29] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, 19(1): 011006-1-011006-21, 2010.

[30] W. Lin, L. Dong, and P. Xue, "Visual distortion gauge based on discrimination of noticeable contrast changes," *IEEE Transactions on Circuits and Systems for Video Technology*, 15(7): 900-909, 2005.

[31] U. Engelke, H. Kaprykowsky, H.-J. Zepernick, and P. Ndjiki-Nya, "Visual attention in quality assessment," *IEEE Signal Processing Magazine*, 28(6): 50-59, 2011.

[32] S. Christian and L. Itti, "Rapid biologically-inspired scene classification using features shared with visual attention," *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, 29(2): 300-312, 2007.

[33] D. Song and D. Tao, "Biologically inspired feature manifold for scene classification," *IEEE Transactions on Image Processing*, 19(1): 174-184, Jan. 2010.

[34] C. Deng, J. Li, Y. Zhang, D. Huang, and L. An, "An image quality metric based on biologically inspired feature model," *International Journal of Image and Graphics*, 11(2): 265-279, April 2011.

[35] Y. Huang, K. Huang, D. Tao, T. Tan, and X. Li, "Enhanced biologically inspired model for object recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 41(6): 1668–1680, Dec. 2011.

[36] S. Thomas, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(3 (2007): 411-426.

[37] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11): 1254–1259, Nov. 1998.

[38] A. Ninassi, P. Le Callet, F. Autrusseau, "Pseudo No Reference image quality metric using perceptual data hiding", in *SPIE Human Vision and Electronic Imaging*, 6057-08, San Jose, CA, USA, January 2006.

[39] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik, "Objective quality assessment of multiply distorted images," *IEEE Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, California, USA: IEEE, 2012: 1693-1697.

[40] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. J. Ku, "Color image database TID2013: Peculiarities and preliminary results," *4th Europian Workshop on Visual Information Processing (EUVIP2013)*, Paris, France: IEEE, 2013, 106-111.

[41] Final report from the Video Quality Experts Group (VQEG) on the Validation of Objective Models of Video Quality Assessment, Phase II VQEG, 2003. [Online]. Available: http://www.vqeg.org/.

[42] D. Tao, L. Jin, Y. Wang, Y. Yuan, and X. Li. "Person re-identification by regularized smoothing kiss metric learning," IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 10, pp. 1675-1685, 2013.

[43] D. Tao, L. Jin, Y. Wang, and X. Li. "Person reidentification by minimum classification error-based KISS metric learning," IEEE Transactions on Cybernetics, vol. 45, no. 2, pp. 242-252, 2015.

[44] D Tao, X. Lin, L. Jin, and X. Li. "Principal Component 2-D Long Short-Term Memory for Font Recognition on Single Chinese Characters" IEEE Transactions on Cybernetics, accepted, 2015.

[45] C. Xu, D. Tao, and C. Xu. "Large-margin multi-view information bottleneck," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 8, pp. 1559-1572, 2014.

[46] C. Xu, D. Tao, and C. Xu. "A survey on multi-view learning," arXiv preprint arXiv:1304.5634, 2013.

[47] C. Xu, D. Tao, C. Xu, "Multi-view Intact Space Learning," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.PP, no.99, pp.1,1.

[48] M. Song, D. Tao, C. Chen, J. Bu, J. Luo, and C. Zhang. "Probabilistic Exposure Fusion," IEEE Transactions on Image Processing, vol. 21, no. 1, pp. 341-357, 2012.

[49] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "A comprehensive survey to face hallucination," International Journal of Computer Vision, vol. 106, no. 1, pp. 9-30, 2014.

[50] N. Wang, D. Tao, X. Gao, X. Li, and J. Li. "Transductive face sketch-photo synthesis," IEEE Transactions on Neural Networks and Learning Systems, 24, 2013.

Highlights

We propose a novel IQA approach named biologically inspired feature similarity (BIFS), which is demonstrated to be highly consistent with the human perception.

In the proposed approach, biologically inspired features (BIFs) of the test image and the relevant reference image are first extracted.

Afterwards, local similarities between the reference BIFs and the distorted ones are calculated and then combined to obtain a final quality index.

Thorough experiments on a number of IQA databases demonstrate that the proposed method is highly effective and robust, and outperform state-of-the-art FR-IQA methods across various datasets.