LEARNING A BLIND IMAGE QUALITY INDEX BASED ON VISUAL SALIENCY GUIDED SAMPLING AND GABOR FILTERING

Zhongyi Gu, *Lin Zhang*[*1], *and Hongyu Li*
School of Software Engineering, Tongji University, Shanghai, China

ABSTRACT

The goal of no-reference image quality assessment (NR-IQA) is to estimate the quality of an image consistent with the human perception of the image automatically without any prior of the reference image. In this paper, we present a simple yet efficient and effective approach to learn a blind *I*mage *Q*uality index based on *V*isual saliency guided sampling and *G*abor filtering, namely *IQVG*. Given an image, we at first randomly sample a sufficient number of image patches guided by the image's visual saliency map and convolve each patch with Gabor filters to get a bag of features. Then, the image is represented by using a histogram to encode the bag of features. Support vector regression (SVR) is used to learn the mapping from feature space to image quality. Extensive experiments conducted on the LIVE IQA database demonstrate the overall superiority of our IQVG over the other state-of-the-art NR-IQA algorithms evaluated. The Matlab source code of IQVG and the evaluation results are available online at http://sse.tongji.edu.cn/linzhang/IQA/IQVG/IQVG.htm

***Index Terms*—**No-reference image quality assessment, visual saliency, Gabor filter

1. INTRODUCTION

Rapid proliferation of digital imaging and communication technologies has rendered the image quality assessment (IQA) an important issue in numerous applications. Recent years have seen growing interests in developing no-reference image quality assessment (NR-IQA) approaches. Most NR-IQA approaches follow one of the three trends: distortion-specific (DS) approaches, natural scene statistics (NSS) based approaches and training-based approaches. DS approaches assume that the distortion type is known as a prior knowledge. These methods mainly measure the impact of one distortion type on the image quality such as blur [1-2] or ringing [3]. One important feature of DS approaches, which limits their application domain, is that they are distortion-specific. NSS-based approaches assume that natural scene images occupy a small subspace of all possible images and the presence of distortion will affect the distance between the distorted images and the subspace of the natural images [4-5]. For example, in [4], Moorthy *et al.* proposed an NSS-based approach in the wavelet domain. In [5], Saad *et al.* proposed another NR-IQA approach using a NSS model of discrete cosine transform coefficients. Training-based approaches [6-7] at first design a large number of features to describe the factors that may affect the image quality. Then, these features can be used to train a model using some regression techniques such as support vector regression (SVR) or general regression neural network (GRNN). For example, in [7], Li *et al.* use the mean value of phase congruency image, the entropy of phase congruency image, the entropy of the distorted image and the gradient of the distorted image as inputs to GRNN to train a learning model. In training-based approaches, one important issue is how to design proper features, while one feature may be only sensitive to one kind of distortion. In [8], Moorthy *et al.* proposed a two-stage framework, which is a combination of NSS-based approaches and training-based approaches, using support vector machine (SVM) to classify an image into a distortion class and then using a distortion specific quality metric to predict its quality. In [9, 10], Peng *et al.* proposed a training-based approach, namely CBIQ, using Gabor filter to extract local features and visual codebook to quantize the feature space.

As an alternative, in this paper, we propose a novel training-based NR-IQA approach using visual saliency and Gabor filtering (IQVG). Our idea is inspired by the success of CBIQ. CBIQ has following advantages. First, CBIQ uses Gabor filters to extract patch level features to get bags of features instead of designing a large number of different features, which is often used in other training-based approaches. Second, CBIQ uses visual codebook to quantize the feature space instead of building a statistical model for image patches in high-dimensional feature space. CBIQ relies on the hypothesis that patches with similar quality should share similar Gabor-filter-based local features, and it uses Euclidean distance to measure the similarity. But, we observe that though some patches with similar quality share similar local features, the Euclidean distance between them may be large. Hence, using the Euclidean distance to construct the codebook or quantize the feature space is not so convincing. Our approach, named IQVG, has following advantages over CBIQ. First, IQVG

[1] Corresponding author: cslinzhang@tongji.edu.cn

uses the image's visual saliency map to guide sampling instead of sampling randomly. Second, IQVG uses the distribution of local features directly to quantize the feature space instead of using visual codebooks, and hence to reduce the computational cost. Third, IQVG uses the distribution of local features which is more accurate, instead of Euclidean distance to measure the similarity between local features. Our algorithm is tested on the LIVE IQA database and is compared with two full-reference IQA and four state-of-the-art NR-IQA algorithms. Spearman rank-order correlation coefficient (SROCC) and Pearson linear correlation coefficient (PLCC) are used to evaluate the performance of various IQA metrics. Experimental results demonstrate that IQVG has advantages both in accuracy and efficiency compared with the other IQA indices evaluated.

The rest of this paper is organized as follows. Section 2 discusses four steps in IQVG. Experiment results and discussions are shown in Section 3 and Section 4 concludes the paper.

2. IMAGE QUALITY BASED ON VISUAL SALIENCY GUIDED SAMPLING AND GABOR FILTERING

The proposed NR-IQA approach IQVG relies on the hypothesis that the image quality can be measured through sufficient number of local features extracted from image patches. In the following, we present our IQVG method mainly from four aspects, the image patch sampling strategy, the local feature extraction strategy, the image representation strategy and the regression strategy. Specifically, for each image, we at first sample a sufficient number of patches of which the mean visual saliency is greater than a threshold. Then, we convolve each patch with Gabor filters to get a bag of patch level features. Histograms are employed to encode the bag of features for image representation. Finally, some regression techniques are used to train a model. The quality of an unseen test image can be automatically predicted by using the trained model. Fig. 1 shows the overall working pipeline of IQVG.

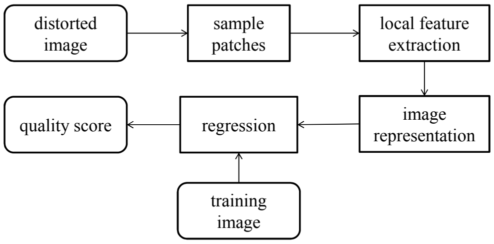

Fig. 1: Overall working pipeline of IQVG.

2.1. Sampling strategy

The main problem of sampling strategy lies in how to sample patches. There are mainly two sampling strategies: sampling randomly and sampling based on interest points. Here we combine these two strategies.

Given an image, human have the ability to perceive the content of the image quickly and change the points of visual fixation rapidly. Estimating the distribution of such points, or computing a visual saliency map [11, 12] has been proven to be of great help in improving the performance of FR-IQA algorithms. As Zhang *et al*. has proposed in [13], an image's visual saliency map is closely related to its perceived quality and it can be used as a weighting function indicating the importance of a local region to the human visual system. There are already many different kinds of algorithms proposed for visual saliency computation in the literature. For a recent comprehensive survey of this field, readers can refer to [11, 12]. In this paper, with respect to the visual saliency computation, we adopt the model proposed by Hou and Zhang [14], which has been proved to be simple, effective, and can generate satisfactory results in most cases. An image from LIVE IQA database and its corresponding visual saliency map are shown in Fig. 2.

Here we use the image's visual saliency map to guide sampling. When sampling patches, we choose the patches of which the mean visual saliency is not so small because the patches of which the visual saliency is small play little role in human perception of the image quality. Specifically, given an image, we randomly sample *M* patches of which the mean visual saliency is bigger than a threshold.

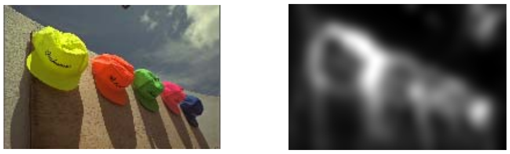

Fig. 2: An image from LIVE IQA database and its corresponding visual saliency map computed by using the saliency model proposed in [14].

2.2. Local feature extraction

Local features extracted from image patches should have such characteristics. First, they should be highly correlated with distortion. Second, they should be adaptable to different distortion types. As Peng *et al*. proposed in CBIQ, Gabor-filter-based local feature is a good choice. Here we follow the same setting of CBIQ in extracting local features. In the following part we briefly introduce how to use Gabor filter for feature extraction.

A 2-D Gabor filter has a real component and an imaginary component. They are defined as

$$g_e(x,y) = \exp\left(-\frac{1}{2}\left(\frac{x'^2}{\sigma^2}+\frac{y'^2}{(\gamma\sigma)^2}\right)\right)\cos\left(\frac{2\pi}{\lambda}x'\right) \tag{1}$$

$$g_o(x,y) = \exp\left(-\frac{1}{2}\left(\frac{x'^2}{\sigma^2}+\frac{y'^2}{(\gamma\sigma)^2}\right)\right)\sin\left(\frac{2\pi}{\lambda}x'\right) \tag{2}$$

where $x' = x\cos\theta + y\sin\theta$, $y' = -x\sin\theta + y\cos\theta$. λ is the frequency of the sinusoid factor, θ is the orientation of the normal to the parallel stripes of the Gabor function, σ and γ are the sigma of the Gaussian envelope and the spatial aspect ratio respctively.

Given an image patch $\Omega(x, y)$, the response of the convolution between $\Omega(x, y)$ and Gabor filter $g(x, y, \lambda, \theta)$ can be written as

$$G(x, y; \lambda, \theta) = \Omega(x, y) * g(x, y, \lambda, \theta) \qquad (3)$$

With the changing of λ and θ, for each point in $\Omega(x, y)$, we can get a pixel-level matrix. Specifically, given a point (x_0, y_0), we can get such a matrix:

$$G(x_0, y_0) = \begin{pmatrix} G(x_0, y_0; \lambda_0, \theta_0) & \dots & G(x_0, y_0; \lambda_0, \theta_{n-1}) \\ \vdots & \ddots & \vdots \\ G(x_0, y_0; \lambda_{m-1}, \theta_0) & \cdots & G(x_0, y_0; \lambda_{m-1}, \theta_{n-1}) \end{pmatrix} \qquad (4)$$

where m is the number of frequencies and n is the number of orientations that we use.

Instead of using pixel level feature, patch level feature is obtained through calculating the mean and variance of all the features extracted from each pixel of the patch. Specifically, for each patch we extracted, we can obtain a $2mn$-by-1 feature vector to describe it.

2.3. Image representation

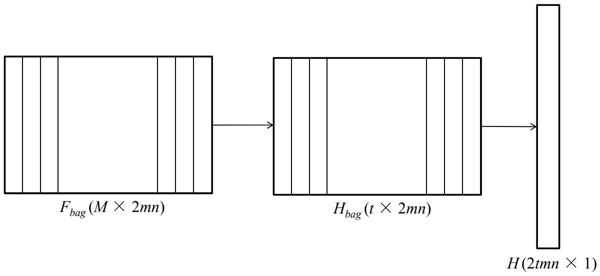

Fig. 3: Illustration of using histograms to encode a bag of features.

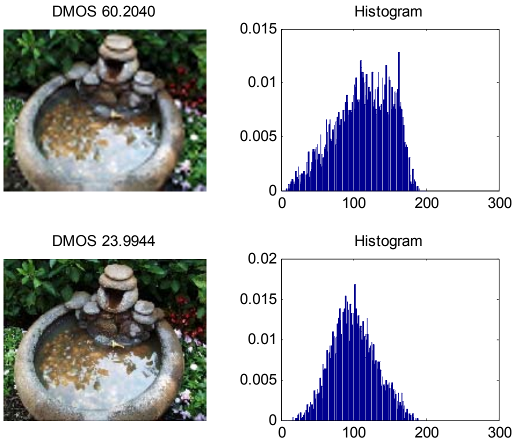

Fig. 4: Two images from LIVE IQA database and their corresponding histograms which encode part of the bags of words.

CBIQ [10] uses a codebook, consisting of local features, to quantize the feature space and encode each image by counting appearance of patch level features. In IQVG, each image is represented directly using histogram to encode the bag of patch level features. Given an image, we sample M different patches and convolve each patch with Gabor filters to get a bag of features. The bag of features can be given by

$$F_{bag} = \begin{pmatrix} f_{0,0} & \cdots & f_{0,2mn} \\ \vdots & \ddots & \vdots \\ f_{M,0} & \cdots & f_{M,2mn} \end{pmatrix} \qquad (5)$$

Each row of F_{bag} represents a patch level feature. We use a histogram to encode each column of F_{bag} and use a combination of these histograms as the image representation. Fig. 3 illustrates the process of using histograms to encode a bag of features, where t represents the number of bins.

Fig. 4 presents an example of how the distribution of the first column of F_{bag} varies with the level of distortions. The left column of Fig. 4 shows two images from LIVE IQA database with DMOS 60.2040 and 23.9944. The right column of Fig. 4 shows their corresponding histograms which encodes the first column of each image's bags of features. Here the number of bins is set to 200.

2.4. Regression

Since each image is represented by a single feature vector, the NR-IQA problem can be solved as a regression problem. A lot of regression techniques such as SVR and random forest can be used to learn the mapping. Here we use SVR with RBF kernel to do the regression.

3. EXPERIMENTS AND DISCUSSIONS

3.1. Determination of parameters

Parameters involved in our approach are empirically determined. In the sampling stage, we randomly sample 5000 patches of which the mean saliency should be greater than 0.1 from each image. Theoretically, the more patches that we sample from the image, the more accurate our approach will be, but we find 5000 patches are enough to catch sufficient information. When extracting Gabor-filter-based local features, the size of the patch was 11-by-11, five frequencies (1, $1/\sqrt{2}$, 1/2, $1/2\sqrt{2}$ and 1/4) and four orientations (0^o, 45^o, 90^o and 135^o) were used. The number of bins when constructing image representation was set to 200.

3.2. Database and methods for comparison

We evaluated our approach on the LIVE IQA [15] database. LIVE IQA database contains 29 reference images associated with their distorted images with five distortion types: JPEG2000, JPEG, Gaussian blur, white noise and fast fading. Each image is given an associated DMOS ranging from 0 to 100, where 0 indicates that the image has no distortion, namely reference image and 100 indicates that the image quality is the most possible poor.

We compared the results of our approach with two FR-IQA algorithms, PSNR [16] and SSIM [17] and four NR-IQA algorithms, CBIQ [10], LBIQ [6], BLIINDS-II [5], and DIIVINE [4]. PSNR and SSIM are not the FR-IQA algorithms that can achieve best results, but they are most commonly used and can serve as baselines. The four NR-IQA algorithms selected are state-of-the-art ones in this field nowadays.

3.3. Experiment strategy and evaluation metrics

In order to compare with other methods properly, we followed the experiment strategy in [10]. We randomly selected 23 reference images associated with their distorted copies for training, and 6 reference images with their distorted copies for testing to ensure the content of the training images and the testing images do not overlap. The reported result is the median of 1000 train-test runs.

Two metrics, Spearman rank-order correlation coefficients (SROCC) and Pearson linear correlation coefficients (PLCC) were used to evaluate the accuracy of all the algorithms. A value close to 1 for SROCC and PLCC indicates a good performance for quality estimation.

3.4. Experimental results

We first carried out our approach on each type of distortion. Then, we used the whole database to test our approach to verify that our approach was non-distortion-specific. Table 1 and Table 2 show the median SROCC and PLCC results across 1000 train-test runs on the LIVE IQA database respectively.

Based on Table 1 and Table 2, we can have the following findings. First of all, IQVG can achieve higher SROCC and PLCC than all the other methods evaluated. Secondly, the comparison between the results of CBIQ and IQVG proves that using histograms to encode local features instead of using visual codebooks to quantize the feature space can greatly improve the estimation accuracy.

3.5. Computational complexity

The speed of NR-IQA algorithms should be as fast as possible, and hence can be used in real-time applications. Our approach consumes about 60s to extract local features from 5000 patches of which the size is 11-by-11 on an Intel Pentium 2.13-GHz machine. The time consumption for computing the visual saliency map and making use of histograms to encode local features is relatively negligible. Compared with other NR-IQA methods which need a domain transform or use a codebook, our approach has relatively a low computational complexity.

Table 1: Median SROCC coefficients

	JP2K	JPEG	WN	BLUR	FF	All
PSNR	0.8646	0.8831	0.9410	0.7515	0.8736	0.8636
SSIM	0.9389	0.9466	0.9635	0.9046	0.9393	0.9126
CBIQ	0.8935	0.9418	0.9582	0.9324	0.8727	0.8945
LBIQ	0.9040	0.9291	0.9702	0.8983	0.8222	0.9063
BLIINDS-II	0.9323	0.9331	0.9463	0.8912	0.8519	0.9124
DIIVINE	0.9123	0.9208	0.9818	0.9373	0.8694	0.9250
IQVG	0.9190	0.9003	0.9622	0.9430	0.9377	0.9420

Table 2: Median PLCC coefficients

	JP2K	JPEG	WN	BLUR	FF	All
PSNR	0.8762	0.9029	0.9173	0.7801	0.8795	0.8592
SSIM	0.9405	0.9462	0.9824	0.9004	0.9514	0.9006
CBIQ	0.8898	0.9454	0.9533	0.9338	0.8951	0.8955
LBIQ	0.9103	0.9345	0.9761	0.9104	0.8382	0.9087
BLIINDS-II	0.9386	0.9426	0.9635	0.8994	0.8790	0.9164
DIIVINE	0.9233	0.9347	0.9867	0.9370	0.8916	0.9270
IQVG	0.9266	0.9203	0.9785	0.9527	0.9404	0.9424

4. CONCLUSION

In this paper, we presented a novel training-based NR-IQA algorithm, IQVG, without any prior of the reference image and the distortion type. In our method, we at first sample patches from the given image following the guidance of its visual saliency map. Then, Gabor filters were used to extract local features from each image patch. Histograms were used to quantize the feature space. IQVG can be adaptable to different distortion type. Compared with the other state-of-the-art NR-IQA algorithms, in addition to the higher quality prediction accuracy, IQVG also has advantages of simplicity and low computational complexity.

ACKNOWLEDGEMENT

This work is supported by the Fundamental Research Funds for the Central Universities under grant no. 2100219033, the Natural Science Foundation of China under grant no. 61201394, and the Innovation Program of Shanghai Municipal Education Commission under grant no. 12ZZ029.

5. REFERENCE

[1] Z.M.P. Sazzad, Y. Kawayoke, and Y. Horita, “No-reference image quality assessment for jpeg2000 based on spatial features,” *IEEE Signal Process. Image Commun.*, vol. 23, pp. 257-268, 2008.

[2] X. Zhu and P. Milanfar, "A no-reference sharpness metric sensitive to blur and noise," *IEEE Quality Multim. Exp. Int. Workshop*, pp. 64-69, 2009.

[3] X. Feng and J.P. Allebach, "Measurement of ringing artifacts in JPEG images," *Process. SPIE*, vol. 6076, pp. 74-83, Jan. 2006.

[4] A.K. Moorthy and A.C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. IP*, vol. 20, pp. 3350-3364, 2011.

[5] M. Saad, A.C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. IP*, vol. 21, no. 8, pp. 3339-3352, Aug. 2012.

[6] H. Tang, N. Joshi, and A. Kapoor, "Learning a blind measure of perceptual image quality," *CVPR'11*, pp. 305-312, 2011.

[7] C. Li, A.C. Bovik, and X. Wu, "Blind image quality assessment using a general regression neural network," *IEEE Trans. Neural Netw.*, vol. 22, no. 5, pp. 793-799, May 2011.

[8] A.K. Moorthy and A.C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 513-516, May 2010.

[9] P. Ye and D. Doermann, "No-reference image quality assessment based on visual codebook," *ICIP'11*, pp. 3129-3138, Jul. 2011.

[10] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks," *IEEE Trans. IP*, vol. 21, no. 7, pp. 3129-3138, Jul. 2012.

[11] A. Toet, "Computational versus psychophysical bottom-up image saliency: a comparative evaluation study," *IEEE Trans. PAMI*, vol. 33, pp. 2131-2146, 2011.

[12] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Trans. PAMI*, vol. 35, pp. 185-207, 2013.

[13] L. Zhang and H.Y. Li, "SR-SIM: A fast and high performance IQA index based on spectral residual," *ICIP'12*, pp. 1473-1476, 2012.

[14] X. Hou and L. Zhang, "Saliency detection: a spectral residual approach," *CVPR'07*, pp. 1-8, 2007.

[15] H.R. Sheikh, M.F. Sabir, and A.C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. IP*, vol. 15, no. 11, pp. 3440-3451, Nov. 2006.

[16] Z. Wang and A.C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98-117, Jan. 2009.

[17] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. IP*, vol. 13, no. 4, pp. 600-612, Apr. 2004.