NO-REFERENCE IMAGE QUALITY ASSESSMENT BASED ON VISUAL CODEBOOK

Peng Ye and David Doermann

University of Maryland, College Park
Language and Media Processing Laboratory
College Park, Maryland, 20742

ABSTRACT

In this paper, we propose a new learning based No-Reference Image Quality Assessment (NR-IQA) algorithm, which uses a visual codebook consisting of robust appearance descriptors extracted from local image patches to capture complex statistics of natural image for quality estimation. We use Gabor filter based local features as appearance descriptors and the codebook method encodes the statistics of natural image classes by vector quantizing the feature space and accumulating histograms of patch appearances based on this coding. This method does not assume any specific types of distortion and experimental results on the LIVE image quality assessment database show that this method provides consistent and reliable performance in quality estimation that exceeds other state-of-the-art NR-IQA approaches and is competitive with the full reference measure PSNR.

Index Terms— no-reference image quality assessment, visual codebook, texture analysis, Gabor filter

1. INTRODUCTION

The goal of no-reference image quality assessment (NR-IQA) is to develop some computational model that can predict the quality of distorted images accurately and automatically without prior knowledge of reference images. The NR-IQA problem is widely considered as an extremely difficult problem [1, 2, 3, 4, 5]. NR-IQA algorithms can further be classified into distortion-specific (DS) and non-distortion-specific (NDS) based on the prior knowledge on the distortion type used in the assessment. Most of the existing NR-IQA algorithms are DS and assume the type of distortion is known. They are limited typically to one or two specific types of distortion. For example, [1] develops blur and ringing metrics for JPEG2000 compressed images, and [2] introduces blockiness measure for JPEG compressed images. This underlying assumption limits the application domain of these approaches. NDS algorithms do not consider prior knowledge of distortion type, but instead they assume access to samples with the same types of distortion as those in the target population. The quality estimation procedure typically involves machine learning techniques. In these approaches, the IQA problem is usually transformed to a regression or classification problem, where a learning algorithm is trained using features related to image quality. Examples of such NR-IQA algorithms are [3], which estimates image quality score by training a two class classifier using Ada-boost for estimating the probability of an image patch being "good" or "bad"; [4], which predicts image quality using a probabilistic prediction model which is trained on DCT statistics-based feature and [5], which proposes a two-step frame work for constructing blind image quality indices.

In this paper, we present a learning-based approach for NR-IQA. The proposed algorithm differs from all the previous learning based methods in that it approaches the problem from a texture analysis [6] prospective and uses the visual codebook method and Gabor filter based features to effectively capture image statistics. Unlike [5], which first estimates the probability of the occurrence of each distortion in the image, then computes the quality metric using an algorithm specific to each possible distortion, our algorithm integrates the distortion classification into a general codebook framework. The blind IQA is achieved by constructing a codebook using training images with all possible types of distortion.

This method is motivated by the observation that images with the same type of distortion and with similar quality share similar "texture". People recognize texture when they see it, but no precise, general definition of texture exists [6]. A definition in [7] summarizes the general idea about texture: "we may regard texture as what constitutes a macroscopic region. Its structure is simply attributed to the repetitive patterns in which elements or primitives are ranged according to a placement rule." Visual codebook constructed from invariant descriptors extracted from local image patches has been widely used in texture analysis and visual recognition [8]. They can effectively capture the complex statistics of real images in a convenient local form. There are various methods for modeling textures and extracting texture features, and we choose to use Gabor filter based feature which has good invariance properties (with respect to illumination, rotation, scale and translation) and efficient implementation.

Our algorithm is tested on the LIVE image quality assessment dataset [9] and is compared to one state-of-the-art NR-IQA algorithm described in [5] and the *full reference* (FR) measure peak signal-to-noise ratio (PSNR). The pre-

978-1-4577-1303-3/11/$26.00 ©2011 IEEE

dicted quality scores given by our method is evaluated by the Spearman rank-order correlation, the Pearson linear correlation coefficient and the root mean squared error with respect to subjective *differential mean opinion score* (DMOS), as well as the variance of the three metrics. Experimental results show the predicted DMOS is highly correlated with true DMOS from subjective survey and this method provides consistently reliable performance on quality estimation.

The rest of the paper is organized as follows. Section 2 describes feature extraction using Gabor filters. In Section 3, we introduces visual codebook based method for IQA. Experimental results and a thorough analysis of our results are presented in Section 4. Finally, Section 5 concludes with a summary of the proposed method.

2. FEATURE EXTRACTION USING GABOR FILTERS

Good features for quality estimation should be able to capture structural information of image and be invariant to manipulations of images which do not affect image quality. In this paper, we use simple Gabor features [10] for quality estimation. The use of Gabor filters is motivated by the fact that 1) they are optimal in time and frequency or space and spatial-frequency in two dimension [11], i.e., they achieve the theoretical lower limit of joint uncertainty in space and spatial-frequency; 2) the frequency and orientation representations of Gabor filters are similar to those of human visual system; 3) simple operations on Gabor filters can be established to achieve illumination, rotation, scale and translation invariance [10]. With these properties, Gabor filters are particularly appropriate for texture representation and discrimination, and in our case, quality estimation.

A normalized 2-D Gabor filter function in the continuous spatial domain is defined as follows:

$$\begin{aligned} &\psi(x,y,f,\theta) = \tfrac{f^2}{\pi\gamma\eta} exp[-(\tfrac{f^2}{\gamma^2}x'^2 + \tfrac{f^2}{\eta^2}y'^2) + j2\pi f x'] \\ &x' = xcos\theta + ysin\theta \\ &y' = -xsin\theta + ycos\theta \end{aligned} \tag{1}$$

where f is the frequency of sinusoidal plane wave, θ is the rotation of the Gaussian envelop and the sinusoidal, γ and η are the spatial widths of the filter along the major and the minor axis respectively.

Suppose the image function is $\xi(x, y)$, the response of Gabor filter to ξ is given by the convolution between ξ and ψ:

$$r(x, y; f, \theta) = \psi(x, y, f, \theta) * \xi(x, y) \tag{2}$$

The feature matrix for point (x_0, y_0) in ξ is given by:

$$G = \begin{bmatrix} r(x_0, y_0; f_0, \theta_0) & \cdots & r(x_0, y_0; f_0, \theta_{n-1}) \\ r(x_0, y_0; f_1, \theta_0) & \cdots & r(x_0, y_0; f_1, \theta_{n-1}) \\ \vdots & \ddots & \vdots \\ r(x_0, y_0; f_{m-1}, \theta_0) & \cdots & r(x_0, y_0; f_{m-1}, \theta_{n-1}) \end{bmatrix} \tag{3}$$

where a typical choice of the group of frequencies is $f_m = f_0/(\kappa^m)$ ($\kappa > 1$ is a constant) and f_0 is the highest frequency. Illumination invariance is achieved by using a normalized feature matrix G'.

$$G' = \frac{G}{\sqrt{\sum_{i,j} |g_{i,j}^2|}} \tag{4}$$

Rotation invariance is achieved by column-wise circular shift in Eq.(3). By choosing different f_0, scale invariance can be achieved. G' is concatenated into an mn-by-1 vector as the feature vector for point (x_0, y_0) in ξ. Given an image patch, simple Gabor features are extracted from each point in the patch, then the mean and variance of elements in the feature vector over all points in the patch are computed to form a $2mn$-by-1 vector, which is referred to as *Gabor feature vector*.

3. ASSESSMENT USING A VISUAL CODEBOOK

Our approach for learning to estimate image quality consists broadly of the following stages

(1) Codebook Construction The first stage consists of building a "codebook" of image patches which can be used to represent "quality information". Given one training image, we divide the image into $B \times B$ patches. Constant patches are removed since they do not contain any information for quality estimation. *Gabor feature vectors* are computed for the rest of image patches and they are labeled by the true DMOS of the training image. By repeating this operation on all the available training images, we obtain a large set of *Gabor feature vectors* Ω. A codebook C are then created from this set using a clustering algorithm such as k-means or by uniformly and randomly downsampling Ω. A clustering algorithm is preferred if downsampling factor $|\Omega|/|C| >> 1$. However, in order to capture the complex distortion patterns, we choose to use a large redundant codebook, so the downsampling factor is usually less than 10 and the latter approach is used in our experiment.

Codewords in C are denoted as $C(i), i = 1, \ldots, N$, where N is the size of codebook. The corresponding DMOS for codeword $C(i)$ is denoted as $DMOS(C(i))$.

(2) Image Representation Input images are represented by the distribution of codewords from the codebook obtained in the first stage. This requires determining how many times a codeword from the codebook has appeared. Specifically, given an input image I, we divide it into $B \times B$ patches and extract a *Gabor feature vector* for each (removing all constant patches). For each extracted feature vector, the count of its nearest neighbor in the codebook is increased by one. If several, say s, codewords have the same smallest distance to the feature vector, each of their counts is increased by $1/s$. If the distance between the feature vector to its nearest neighbor is larger than some predetermined threshold, it is considered an outlier. If a large percentage of patches in the input image are outliers, this input image may contain some type of distortion that was not seen in training images. In this paper, we assume a closed test set. Finally, the feature vector for I is the normalized (scaled to total sum 1) histogram H_I of occurrence counts for the different codewords. The probability of the occurrence of a codeword $C(i)$ can be approximated by $H_I(i)$. And the quality measure of I is given by:

$$Qm(I) = \sum_{i=1}^{N} H_I(i) \times DMOS(C(i)) \tag{5}$$

(3) Nonlinear Mapping In the Video Quality Experts Group (VQEG) Phase I FR-TV test [12], a nonlinear mapping between the objective model outputs and the subjective quality ratings was used. In the final stage, we approximate DMOS using Qm by nonlinear regression using the following function

$$DMOS_p(I) = \alpha + \beta \times Qm(I)^{\gamma} \tag{6}$$

where α, β, γ are optimized by a nonlinear regression routine. In [12], logistic function is used for nonlinear mapping, we use the function in Eq.(6) because we find it works more stable than logistic function in our case.

4. EXPERIMENTAL RESULTS

We tested the proposed algorithm on the LIVE image quality assessment database [9] and compared it to result in [5] and a FR measure PSNR. The LIVE IQA database consists of 29 reference images and their degraded version with five different types of distortion, i.e, JPEG2k, JPEG, white noise (WN), Gaussian blur and fast fading (FF). A total of 808 distorted images together with their reference image and associated DMOS values are available. DMOS values are in the range [0,100], where 0 is quality score for non-distorted reference image and 100 is the worst possible quality score. We randomly select 10 reference images and their associated distorted images (*group I*) for codebook construction. The remaining 19 reference images and their associated distorted images (*group II*) are used for estimating the parameter α, β, γ in Eq.(6) and testing.

As mentioned previously, for Gabor feature extraction, rotation and scale invariance can be realized by a columnwise circular shift in Eq.(3) and by using different highest frequencies. For the LIVE image quality database, where resolutions of images with different levels of degradation are the same and these images are not rotated, we do not need to perform these matrix manipulations for invariance properties. Five filter frequencies (1, $1/\sqrt{2}$, $1/2$, $1/(2\sqrt{2})$, $1/4$) are used for extracting multi-scale features and the filters are in four orientations (0°,45°,90°,135°). The block size B used in our experiment is 11. We construct five different codebooks of size 60000-70000 for different types of distortion. They are denoted as C_{jp2k}, C_{jpeg}, C_{wn}, C_{blur}, C_{ff} and the combination of these codebook is $C_{total} = \{C_{jp2k}, Cjpeg, C_{wn}, C_{blur}, C_{ff}\}$. Given the distortion type, its associated codebook is used for quality estimation, if we don't know which one of the five different types of distortion exists in the testing image, the codebook C_{total} is used for quality estimation.

Cross-validation is done by randomly selecting 5 reference images (from *group II*) along with their distorted version for parameter estimation and the rest 14 (different) reference images along with their distorted version for testing and repeating this procedure for 500 times.

Three metrics [12] are used in our experiment to evaluate the performance of the objective quality assessment model. The first metric is Spearman rank-order correlation coefficient between $DMOS_p$/$DMOS$. It is related to prediction monotonicity of a model. The second metric is Pearson linear correlation coefficient between $DMOS_p$/$DMOS$ after non-linear mapping. It is considered as a measure of prediction accuracy of a model. The third metric is root mean squared error (rmse) between $DMOS_p$/$DMOS$ after non-linear mapping. The non-linear mapping function used for obtaining PSNR result is logistic function of the form $\alpha/(1 + exp(-\beta(PSNR(I) - \gamma)))$. The evaluation results of our algorithm (CB), a recent NDS-NR-IQA algorithm BIQI [5] and the FR measure PSNR are given in Table 1-3. A scatter plot of the predicted DMOS versus the subjective DMOS from on experiment is shown in Fig.1. The result of PSNR is obtained by using 15 reference images and their distorted version for estimating parameters in logistic regression and the rest 14 reference images and their distorted version for testing and repeat the experiments for 500 times with different choices of training set and testing set. The result of BIQI is directly obtained from [5], where the experimental setting is the same as ours. The variance of the three metrics is shown in Table 4 and it can be considered as a measure of the consistency of model performance.

From Table 1-3, we can see that according to the three metrics our NDS-NR-IQA algorithm performs better than BIQI and is competitive with PSNR which is one of the most widely used FR quality measures. For blur and WN distortion, our method has better performance than PSNR. As is shown in Table 4, variances of three metrics on all data of our method is smaller than PSNR's result, which demonstrates the consistency of our model performance.

5. DISCUSSION AND LEAVE CONCLUSION

There are a number of issues that are worth noting. First, we use approximate nearest neighbor (ANN) search [13] for finding codebook distribution. Using ANN, the query time of one patch over the entire codebook is about $O(dN^{1+\varepsilon})$, where d is the dimension of feature vector, N is the size of codebook and ε is a parameter in ANN search. It implies that the computational cost increases only sublinearly with the increase of codebook size. Furthermore, the proposed method has the potential to be implemented efficiently using parallel computing techniques since most computations (Gabor feature extraction and nearest neighbor search) are performed on different image patches independently. Second, the codebook based method is modular in that it can be extended to any number of distortions. To deal with a new type of distortion, we only need to add codewords extracted from examples of this distortion to the codebook. Third, when the number of examples is much larger than the size of codebook, clus-

Table 1. MEDIAN SPEARMAN CORRELATIONS (DMOS VERSUS PREDICTED DMOS)

	JP2K	JPEG	WN	BLUR	FF	ALL
CB	0.9137	0.9348	0.9833	0.9387	0.8729	0.8954
BIQI	0.7995	0.8914	0.9510	0.8463	0.7067	0.8195
PSNR	0.9471	0.9422	0.9660	0.8937	0.9123	0.9022

Table 2. MEDIAN LINEAR CORRELATIONS (DMOS VERSUS PREDICTED DMOS)

	JP2K	JPEG	WN	BLUR	FF	ALL
CB	0.9167	0.9332	0.9591	0.9415	0.8824	0.8904
BIQI	0.8086	0.9011	0.9538	0.8293	0.7328	0.8205
PSNR	0.9559	0.9511	0.9583	0.9232	0.9263	0.9221

tering algorithms may be applied for selecting representative or information-rich feature vectors as codeword to form the codebook.

We have presented a novel NDS-NR-IQA algorithm which estimates image quality without knowing a reference image and without any assumption on the types of distortion in testing image. The proposed method is based on visual codebook, which is a collection of Gabor feature extracted from local image patches. Our experimental results show that this method provides consistent performance in quality prediction on LIVE image quality assessment database and is competitive to the most widely used FR measure PSNR.

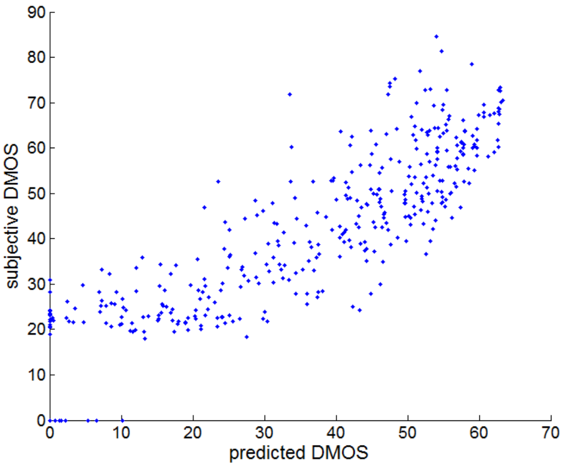

Fig. 1. Scatter plot of subjective DMOS versus predicted DMOS (Entire LIVE database)

Acknowledgment

The partial support of this research by DARPA through BBN/DARPA Award HR0011-08-C-0004 under subcontract 9500009235, the US Government through NSF Award IIS-0812111 is gratefully acknowledged.

Table 3. MEDIAN ROOT-MEAN-SQUARED-ERROR (DMOS VERSUS PREDICTED DMOS)

	JP2K	JPEG	WN	BLUR	FF	ALL
CB	9.0317	8.2160	6.3086	8.0237	10.7731	10.3627
BIQI	14.8427	13.7552	8.4094	10.2347	19.2911	15.6223
PSNR	7.2999	7.7177	6.2736	9.5471	8.3803	8.9976

Table 4. VARIANCE OF THREE PERFORMANCE METRICS ON ALL DATA

Method	Spearman	Pearson	Rmse
	$(\times 10^{-4})$	$(\times 10^{-4})$	
CB	0.27	0.25	0.1342
PSNR	0.51	0.27	0.2210

6. REFERENCES

[1] "Perceptual blur and ringing metrics: application to JPEG2000," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 163 – 172, 2004.

[2] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *IEEE International Conference on Image Processing*, Rochester, New York, Sep. 22-25 2002.

[3] H. Tong, M. Li, H. Zhang, C. Zhang, J. He, and W. Ma, "Learning no-reference quality metric by examples," in *Proceedings of the 11th International Multimedia Modelling Conference*, Jan 2005, pp. 247 – 254.

[4] M. A. Saad, A. C. Bovik, and C. Charrier, "A DCT statistics-based blind image quality index," *IEEE Signal Processing Letters*, vol. 17, no. 6, pp. 583 –586, Jun. 2010.

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

[6] M. Tuceryan and A. K. Jain, "Texture analysis," in *Handbook Pattern Recognition and Computer Vision*, C.H. Chen, L.F. Pau, P.S.P. Wang, and eds., Eds., chapter 2, pp. 235–276. World Scientific, Singapore, 1993.

[7] Hideyuki Tamura, Shunji Mori, and Takashi Yamawaki, "Textural features corresponding to visual perception," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 8, no. 6, pp. 460 –473, 1978.

[8] F. Jurie and B. Triggs, "Creating efficient codebooks for visual recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, 2005.

[9] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE image quality assessment database release 2," Online, http://live.ece.utexas.edu/research/quality.

[10] V. Kyrki and J. Kamarainen, "Simple Gabor feature space for invariant object recognition," *Pattern Recognition Letters*, vol. 25, pp. 311–318, 2004.

[11] D. Gabor, "Theory of communications," *J. Inst. Electr. Engrs.*, vol. 93, no. 429-457, 1946.

[12] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," Mar 2000, Online, http://www.vqeg.org/.

[13] D. Mount and S. Arya, "Ann: Approximate nearest neighbors," Online, http://www.cs.umd.edu/ mount/ANN.