Unsupervised Feature Learning Framework for No-reference Image Quality Assessment

Peng Ye, Jayant Kumar, Le Kang, David Doermann
Institute for Advanced Computer Studies
University of Maryland, College Park, MD, USA
{pengye,jayant,lekang,doermann}@umiacs.umd.edu

Abstract

In this paper, we present an efficient general-purpose objective no-reference (NR) image quality assessment (IQA) framework based on unsupervised feature learning. The goal is to build a computational model to automatically predict human perceived image quality without a reference image and without knowing the distortion present in the image. Previous approaches for this problem typically rely on hand-crafted features which are carefully designed based on prior knowledge.

In contrast, we use raw-image-patches extracted from a set of unlabeled images to learn a dictionary in an unsupervised manner. We use soft-assignment coding with max pooling to obtain effective image representations for quality estimation. The proposed algorithm is very computationally appealing, using raw image patches as local descriptors and using soft-assignment for encoding. Furthermore, unlike previous methods, our unsupervised feature learning strategy enables our method to adapt to different domains. CORNIA (Codebook Representation for No-Reference Image Assessment) is tested on LIVE database and shown to perform statistically better than the full-reference quality measure, structural similarity index (SSIM) and is shown to be comparable to state-of-the-art general purpose NR-IQA algorithms.

1. Introduction

Our work addresses the problem of no-reference (NR) objective image quality assessment (IQA) on natural scene images. The goal is to build a computational model to predict human perceived image quality, accurately and automatically without access to reference images [1, 11, 13, 14, 17, 18, 22, 23, 26]. In the past several decades, there has been an increased interest in objective IQA due to the tremendous growth in the use of digital images for representing and communicating information. Objective image quality measures have been used in a wide range of computer vision and image processing applications. For example, image processing and transmission systems may have their parameters be adjusted according to the image quality [9]; image retrieval systems can use quality as an attribute to rank images and image processing algorithms may use image quality measures for evaluation.

Based on the availability of non-distorted reference images, objective image quality measures are typically classified into three categories: full-reference (FR) IQA, reduced-reference (RR) IQA and no-reference (NR) IQA. FR-IQA requires reference images to evaluate the quality of degraded images. Examples of state-of-the-art FR-IQA algorithms include VSNR [2], VIF [19], SSIM [25] and MSSIM [27]. For RR-IQA, partial information of reference image is necessary for computing the quality measure [8]. In many practical applications, however, information of reference images is not available, thus it is desirable to develop NR-IQA methods [1, 11, 13, 14, 17, 18, 22, 23, 26] where quality estimation are performed without using any information extracted from reference images.

1.1. Background

Most of the existing NR-IQA algorithms [1, 11, 26] limit themselves to one or more specific types of distortions such as blur, blockiness from JPEG compression [26], or ringing arising from JPEG2k compression [11], and thus have very limited application domains. In contrast, general purpose non-distortion-specific (NDS) NR-IQA methods do not examine the exact prior knowledge of distortion for quality estimation, thus are more practical in real-world applications.

Existing general-purpose NR-IQA algorithms can be broadly classified as (1) Natural scene statistics (NSS) based approaches [13, 14, 17, 18] and (2) Training-based approaches [7, 22, 29]. NSS based approaches are based on the hypothesis that natural scenes possess certain statistical properties which will be affected by the presence of distortion. Complex statistical models for wavelet coefficients [13, 14] or cosine coefficients [17, 18] were developed to characterize these natural scene properties. Esti-

978-1-4673-1228-8/12/$31.00 ©2012 IEEE

mated model parameters are used as features to perform regression. Then, based on the learned regression model, a final quality score is obtained for a given test image. The second approach relies on a large number of features which are designed to capture relevant factors affecting image quality, but these features may not be easily interpreted, and it is not clear what features are best for this problem.

1.2. Our approach

Our approach follows the latter trend. However, unlike all previous approaches which rely on features that are designed based on prior knowledge on the differences between non-distorted images and distorted images, our unsupervised feature learning strategy offers the following advantages. First, we use raw-image-patches local descriptors in our learning framework instead of hand-crafted features, which are more efficient and easily computable. Unlabeled data available on internet can be easily used in this framework. In contrast, current state-of-the-art general purpose NR-IQA algorithms [14, 18, 22, 29] use off-the-shelf image transformation and filtering techniques such as wavelet transform, cosine transform and Gabor filtering for extracting features, which can be very time consuming. Computation efficiency of NR-IQA method is very important when image quality measures are embedded in real-time imaging systems or image transmission systems. Second, we use a codebook based approach which allows to learn highly effective features automatically. Third, we use soft-assignment coding with max pooling for encoding. This process is computationally efficient and parameter-free. Fourth, non-linearities introduced in the features during pooling stage allows our method to use more efficient, linear-SVM for obtaining quality scores. Moreover, if the domain of the problem changes, say from natural scene images to document images, the performance of previous techniques is questionable, while the proposed method does not embed any prior knowledge about natural scene statistics, making it more general and giving it the potential to adapt to different domains.

The remainder of this paper describes CORNIA (COdebook Representation for No-reference Image quality Assessment) in detail and is organized as follows. In Section 2, previous work on general-purpose NR-IQA and unsupervised feature learning that is most relevant to our work is briefly reviewed. Section 3 describes details about the proposed framework. Experimental results and a thorough analysis of our results are presented in Section 4. Finally, Section 5 concludes with a summary of our work.

2. Related Work

2.1. No-reference Image Quality Assessment

The use of a visual codebook for NR-IQA problem was first proposed by Ye and Doermann [29]. They approach the NR-IQA problem from the perspective of texture-analysis, and use Gabor-filter based visual codewords for representing images. Quality measure of a new image is obtained by computing the average of quality scores of codewords, weighted by their "distances" to visual words in the image. Hard assignment based encoding and average pooling are adopted in their method. They show promising performance, but the simple averaging strategy does not allow the use of raw-image-patches as features. A very large codebook with approximately $300,000$ codewords is required to achieve good performance. The use of Gabor-filter based features and a large codebook also makes this method computationally expensive. Furthermore, they use a subset of labeled data to construct the codebook. In comparison, CORNIA does not require labels for codebook construction and a relatively small codebook is sufficient to obtain accurate quality estimation.

Natural scene statistics (NSS) based approaches [13, 14, 17, 18] have been extensively studied and applied to the NR-IQA problem. Statistics of wavelet transform coefficients are explored in [13, 14], while Saad and Bovik [17, 18] use DCT coefficient statistics for representing images. These methods require deep domain knowledge and may not be well adapted to quality assessment problem in other domains. Instead of designing and discovering features that are highly correlated with image quality, we concentrate on designing suitable feature extraction architecture and show that with a proper architecture, highly efficient features can be learned automatically and state-of-the-art NR-IQA can be achieved using raw image patches as local image descriptors.

2.2. Unsupervised Feature Learning

With the increasing availability of computational resources in recent years, there has been an emphasis on unsupervised feature learning. The goal of unsupervised feature learning is to automatically learn a good representation of the input from unlabeled data instead of hand-engineering feature representation. In recent years, there have been many attempts to use unsupervised feature learning in computer vision. Raina *et al.* [16] used sparse coding (SC) to construct high-level features from raw image patches and showed that the resulting sparse representations perform much better than conventional representations for image classification. Coates *et al.* [3] showed the success of applying unsupervised feature learning in detecting text and recognizing characters in scene images. Most previous work has focused on applying unsupervised feature learning to classification problem. This paper applies it to NR-IQA, a regression problem and serves as a case study for applying unsupervised feature learning to regression problems in general.

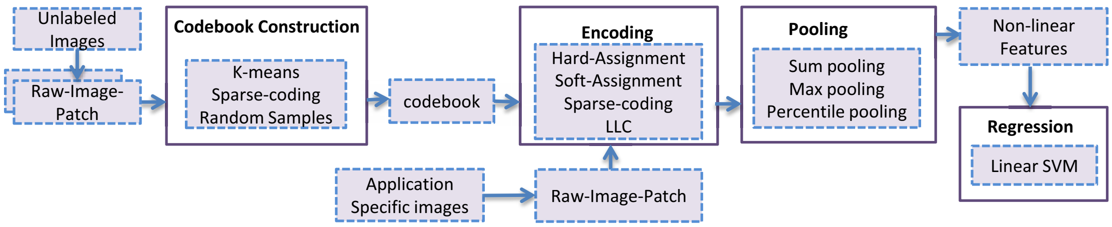

Figure 1. Codebook based framework

3. Learning Framework for NR-IQA

The learning framework adopted in this work is illustrated in Fig.1. Key components in this framework include (1) Local feature extraction, (2) Codebook construction, (3) Local feature encoding and (4) Feature pooling. We describe these components in more detail in the following sections.

3.1. Local feature extraction

Given an image I, local descriptors $X = [x_1, x_2, ..., x_N]$ are extracted from a set of $B \times B$ image patches, where $x_k \in R^d, d = B \times B$. To obtain X, we uniformly and randomly sample N different $B \times B$ image patches from I. Each patch is normalized by subtracting the mean and dividing by the standard deviation of its elements. Additionally, we perform Zero Components Analysis (ZCA) whitening to the normalized patches [4].

3.2. Codebook Construction

The visual codebook is constructed by performing K-means clustering on local features extracted from unlabeled training images. A matrix $D_{d\times K} = [D_1, D_2, ..., D_K]$ denotes a visual codebook, where $D_{i(i=1,...K)}$ are centroids of clusters learned by K-means clustering. More complex training methods such as sparse coding (SC) have been proposed to perform codebook construction (or dictionary learning) to improve system performance. But the use of K-means clustering in our work is motivated by the recent work of Coates et al. [5], where they found that a good encoding scheme is more critical than dictionary learning.

The learned codebook is normalized so that each of the bases has unit length. The normalized codebook is denoted as $\tilde{D}_{d\times K} = [\tilde{D}_1, \tilde{D}_2, ..., \tilde{D}_K]$. Examples of learned codewords are shown in Fig. 2. Codewords with the "dot" patterns come from patches with salt-pepper noise, while "smooth" codewords correspond to blurred patches and codewords with horizontal and vertical line patterns correspond to patches with "blockiness". As shown in Fig. 2, some codewords learned in this way resemble Gabor filters.

3.3. Local Feature Encoding

Soft-assignment coding: We perform soft-assignment coding using a normalized codebook $\tilde{D}$. Distances between

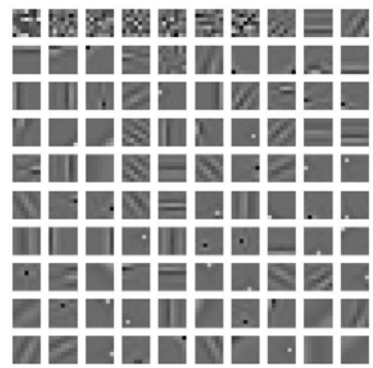

Figure 2. Randomly selected centroids trained on CSIQ database using K-means.

local descriptors and visual codewords in $\tilde{D}$ are computed using dot-products. Let s_{ij} being the similarity metric between the i^{th} local descriptor x_i and the j^{th} base $\tilde{D}_j$, then $s_{ij} = x_i \cdot \tilde{D}_j$. The code for local descriptor x_i can be written as follows:

$$\begin{aligned} c_i = [&\max(s_{i1}, 0), ..., \max(s_{id}, 0), \\ &\max(-s_{i1}, 0), ..., \max(-s_{id}, 0)]^T \end{aligned} \tag{1}$$

The soft-assignment function used here can be considered as a special case of the soft threshold function in [5] where we set the adjustable threshold to zero. The positive and negative components of s are split to obtain code c. This kind of rectification has been shown to improve the discriminative power of features [5].

3.4. Feature Pooling

The encoding step provides us with a coefficient matrix $C_{K\times N} = [c_1, c_2, ..., c_N]$, where $c_i = [c_{i1}, c_{i2}, ..., c_{iK}]^T (K = 2 \times d)$. In order to learn a regression model we need a fixed-length feature vector. In many image quality assessment algorithms [12, 18], "percentile pooling" has been adopted and it is motivated by the observation that humans tend to perceive "poor" regions in an image and these "poor" regions heavily affect the subjective impression. In image classification problems, however, max-pooling has demonstrated higher performance in image classification than average-pooling (or sum-pooling) [10, 28]. We empirically observed that max-pooling performs consistently better than average-pooling in our problem. It can be considered as a special case of "percentile pooling", where we always take responses from the "worst" regions.

Max-pooling: The image-level feature using max-pooling is given as:

$$\hat{\beta} = \Psi_{max}(C) \tag{2}$$

where Ψ_{max} is defined on each row of C to obtain $\hat{\beta} \in \mathbb{R}^K$ whose i^{th} element is given by:

$$\hat{\beta}_i = \max\{c_{1i}, c_{2i}, ..., c_{Ni}\} \tag{3}$$

$\hat{\beta}_i$ is the input to regression program.

4. Experimental Results

4.1. Protocol

Database for evaluation: To test the proposed framework, the following two IQA databases were used.

(1) *LIVE IQA database:* LIVE IQA database [20, 21] has been widely used for evaluating the performance of IQA systems. It consists of 29 reference images each with five different types of distortions - JPEG2k, JPEG, white Gaussian noise (WN), Gaussian blurring (BLUR) and fast fading channel distortion (FF), at 5 to 6 different levels. *differential mean opinion score* (DMOS) associated with distorted images are provided. DMOS is generally in the range $[0, 100]$, where lower DMOS indicates higher quality.

(2)*TID2008 database:* TID2008 [15] consists of 25 reference images and 1700 distorted images with 17 different distortions at 4 levels. The 17 types of distortions include: Additive Gaussian noise (WN), Additive noise in color components (WNC), Spatially correlated noise (SCN), Masked noise (MN), High frequency noise (HFN), Impulse noise (IN), Quantization noise (QN), Gaussian blur (BLUR), Image denoising (IDN), JPEG compression (JPEG), JPEG2000 compression (JPEG2K), JPEG transmission errors (JPEGTE), JPEG2000 transmission errors (JP2KTE), Non eccentricity pattern noise (NEPN), Local block-wise distortions of different intensity (LBD), Intensity shift (IS) and Contrast change (CC). Mean Opinion Score (MOS) is provided for each distorted images in this database. Higher value of MOS (0 - minimal, 9 - maximal) corresponds to higher visual quality of the image. It is worth noting that one of the 25 reference images in the TID2008 database is not natural scene image and it is included in our experiments. We will report results on the first 13 distortions in the TID2008 database. We do not evaluate the last four types of distortions, since they are not dealt with this work. For example, intensity shift and contrast change can be inherent properties of an image and it's a highly subjective task for people to tell their preference, especially when we want to compare images with different content.

Codebook Construction: For codebook construction, we used the CSIQ database [6]. CSIQ database consists of 30 reference images and their degraded versions with 6 different types of distortions at 4 to 5 different levels. There is no overlap between images in the CSIQ database and images in the LIVE and the TID2008 databases. We used images with four types of distortions present in the CSIQ database including JPEG compression, JPEG-2000 compression, additive pink Gaussian noise, and Gaussian blurring. Additionally, we produced images with some new types of distortions which include speckle noise, Poisson noise, Salt-pepper noise and zero-mean Gaussian white noise with an intensity-dependent variance [1]. Normalized and whitened B-by-B raw image patches are extracted from these images and then clustered to construct the codebook. ZCA whitening parameters are computed using randomly sampled patches from the CSIQ database.

Regression: In this work, we intend to use a simple regression method for quality estimation and we employed the Support-vector regression with linear kernel. Other regression methods may be explored to further improve the performance.

Evaluation: We evaluate our system performance using Linear Correlation Coefficient (LCC) and Spearman Rank Order Correlation Coefficient (SROCC). LCC can be considered as a measure of prediction accuracy of a model. SROCC is used to evaluate how well the relationship between the predicted quality score and true quality score can be described using a monotonic function. By default, all results reported in this section are obtained by 1000 train-test iterations with randomly selected 80% of the reference images and their associated distorted versions as training set and the remaining 20% of the reference images and their associated distorted versions as testing set. For the LIVE database, we use the realigned DMOS scores [20] and report results only on the distorted images. For the TID2008 database, MOS scores are used as a regression target and for both training and testing we use only distorted images.

4.2. Analysis

To show the effectiveness of the proposed feature extraction strategy, we compute SROCC and LCC between each one of the features and the DMOS in the LIVE database. The top five correlation coefficients for each distortion subset and the entire LIVE database are tabulated in Table 1. The high correlations shown in Table 1 demonstrates that our method can capture highly effective features. To give some intuitive idea about how these features are related to different distortions, examples of the most informative codewords for the five different types of distortions in the LIVE IQA database are shown in Fig. 3.

[1] We found that by adding these new distortions for constructing codebook, quality estimation performance can be further improved.

LIVE subset	Top five SROCC	Top five LCC
JPEG2K	0.8096, 0.8061, 0.7970, 0.7917, 0.7899	0.7951, 0.7867, 0.7823, 0.7812, 0.7810
JPEG	0.7826, 0.7646, 0.7516, 0.7463, 0.7423	0.7842, 0.7739, 0.7450 , 0.7309, 0.7281
WN	0.9358, 0.9269, 0.9261, 0.9248, 0.9245	0.9421, 0.9286, 0.9282, 0.9279, 0.9255
BLUR	0.9508, 0.9456, 0.9447, 0.9444, 0.9440	0.9416, 0.9371, 0.9361, 0.9350, 0.9342
FASTFADING	0.8435, 0.8420, 0.8415, 0.8388, 0.8386	0.8529, 0.8405, 0.8393, 0.8342, 0.8332
ALL	0.6962, 0.6889, 0.6839, 0.6805, 0.6791	0.6736, 0.6664, 0.6641, 0.6604, 0.6601

Table 1. Top five largest SROCC and LCC correlations between DMOS and each one of the feature values. (Parameters: patch size = 7×7, codebook size = $10,000$).

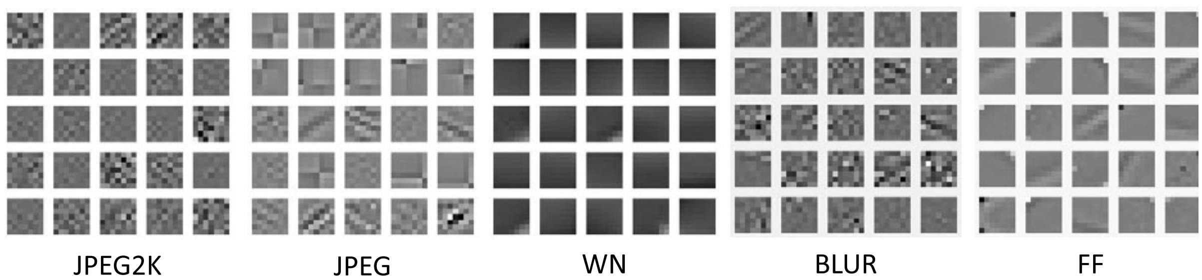

Figure 3. Codewords corresponding to features with the highest SROCC for five different types of distortions in the LIVE database.

4.3. Impact of algorithm parameters

The proposed framework include a number of parameters that can be changed: (1) the number of patches extracted from each image; (2) the number of codewords in the codebook; (3) the size of raw image patch and (4) the encoding method. In this section, we focus on the effect of choosing different codebook sizes and encoders. Number of patches extracted from each image is fixed at $10,000$ and patch size is fixed to 7-by-7. Results reported in this section are obtained on the LIVE database.

Effect of codebook size: We considered codebook size of 200, 400, 800, 1200, 2500, 5000 and 10000. As shown in Fig. 4, performance of this system improves as we increase the number of codewords in codebook.

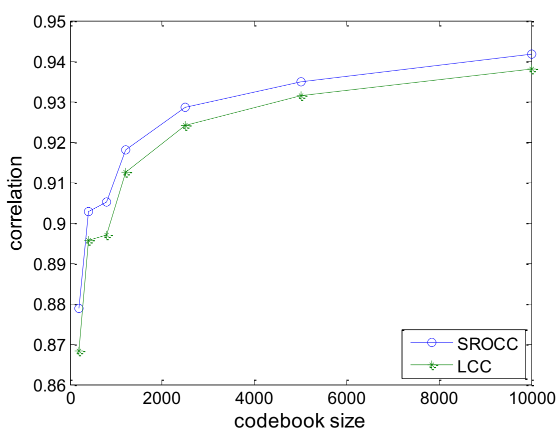

Figure 4. Effect of codebook size (tested on the LIVE database).

Effect of encoding methods: In addition to soft-assignment coding, we also tried using sparse coding (SC) [28], locality-constrained linear coding (LLC) [24], "Localized" soft-assignment coding (LSA) [10] and conventional hard-assignment coding (HA) for encoding in our experiment. SC, soft-assignment (SA) and LSA was used with rectification and max pooling; LLC was used with max pooling but no rectification [2] and HA was used with average pooling and no rectification [3]. In LLC and LSA, we need to specify the number of nearest neighbors used for encoding, we used five for both. Results are shown in Fig. 5. Surprisingly, we found that the simple SA encoding slightly outperforms the other four encoding methods, especially that LLC, LSA and SC have shown better performance than conventional SA in image classification problem. Similar results were presented in [5], where a variation of soft-assignment coding strategy was consistently able to compete with sparse coding in image classification.

4.4. Comparison with Full-Reference and No-Reference IQA Algorithms

Two FR-IQA measures: peak-signal-to-noise-ratio (PSNR) and the structural similarity index (SSIM) are tested for comparison. These results of FR-IQA measures are obtained as described above for obtaining results of our method. Specifically, in the training stage, parameters of a logistic function are learned and then applied to testing set to find the final quality score. We also report the performance of two recent state-of-the-art NR-IQA measures: DIIVINE [14] and BLIINDS-II [18]. Results for DIIVINE and BLIINDS-II were taken from the original paper, where

[2] We found that using LLC without rectification performs better than with rectification.

[3] The way we compute HA code gives a vector with elements non-negative, thus no rectification can be performed

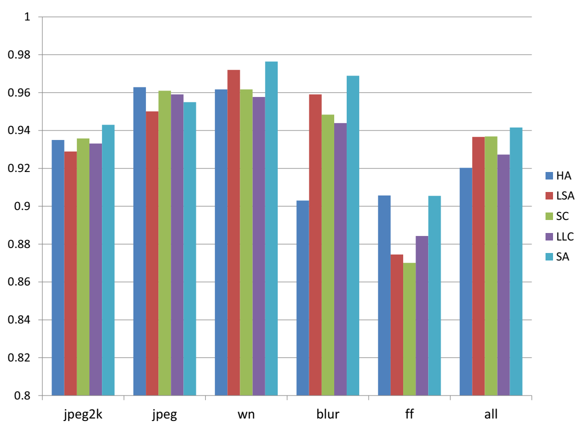

Figure 5. Effect of different encoders, evaluated using SROCC on the LIVE database.

experiments were performed the same way as we did here.

Parameters: In the following experiments, for the LIVE database, we used patch size of 7-by-7 and for the TID2008 database, we used patch size of 5-by-5. For both databases, the number of patches extracted from each image is $10,000$ and codebook size is fixed as $10,000$.

Distortion-specific experiment: First, we performed a distortion-specific (DS) experiment on different distortion subsets in the LIVE database and the TID2008 database. The objective of DS experiment is to see how the algorithm will perform if we only have images with one particular type of distortion. Results on the LIVE database are shown in Table 2 and 3 and results on the TID2008 database are shown in Table 4 and Table 5.

Non-distortion-specific experiment: In the non-distortion-specific (NDS) experiment, each train-test run is performed on images with all types of distortions under consideration. Results on the LIVE database are shown in the last column in Table 2 and 3. In these tables, you may also find results of PSNR, SSIM, BLIINDS-II and DIIVINE. And results on the TID2008 database are shown in Table 4 and Table 5. DIVINE and BLIINDS-II were not tested on the TID2008. The box plot of SROCC distributions of different quality measures from 1000 runs of experiments on the LIVE database are shown in Fig. 6. It is clear that on the LIVE database, CORNIA performs the best among the five measures under consideration. Although on the TID2008 database, CORNIA does not outperform SSIM, it still significantly outperforms PSNR. It is worth noting that PSNR and SSIM use reference image for quality estimation, while CORNIA predicts quality score given only the degraded image. We are unaware of any other general purpose NR-IQA algorithms that have reported NDS experimental results on so many different types of distortions in TID2008.

The standard deviation of the SROCC and LCC obtained

	JP2K	JPEG	WN	BLUR	FF	ALL
PSNR	0.872	0.885	0.941	0.764	0.875	0.867
SSIM	0.939	0.946	0.965	0.909	**0.941**	0.914
DIIVINE	0.913	0.910	**0.984**	0.921	0.863	0.916
BLIINDS-II	0.929	0.942	0.969	0.923	0.889	0.931
CORNIA	**0.943**	**0.955**	0.976	**0.969**	0.906	**0.942**

Table 2. Median SROCC with 1000 iterations of experiments on the LIVE database. (*Italicized* algorithms are NR-IQA algorithms, others are FR-IQA algorithms.)

	JP2K	JPEG	WN	BLUR	FF	ALL
PSNR	0.873	0.874	0.928	0.774	0.869	0.855
SSIM	0.920	0.955	0.982	0.891	**0.939**	0.906
DIIVINE	0.922	0.921	**0.988**	0.923	0.888	0.917
BLIINDS-II	0.935	**0.968**	0.980	0.938	0.896	0.930
CORNIA	**0.951**	0.965	0.987	**0.968**	0.917	**0.935**

Table 3. Median LCC with 1000 iterations of experiments on the LIVE database. (*Italicized* algorithms are NR-IQA algorithms, others are FR-IQA algorithms.)

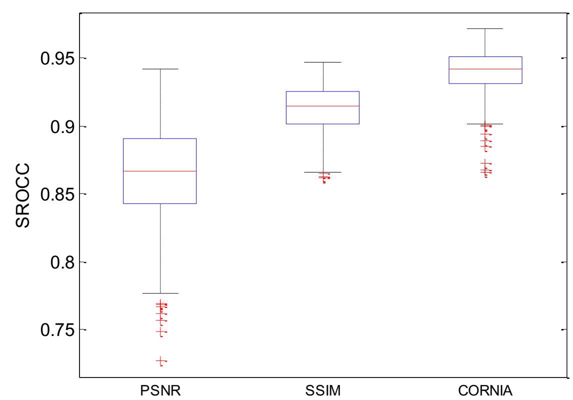

Figure 6. Box plot of SROCC distributions of algorithms from 1000 runs of experiments on the LIVE database.

	PSNR	SSIM	BLIINDS-II	CORNIA
SROCC	0.0348	0.0159	0.0277	**0.0150**
LCC	0.0326	**0.0167**	0.0252	0.0174

Table 6. Standard deviation of SROCC and LCC for 1000 iterations of experiments on the LIVE database.

from the 1000 runs of experiments on the LIVE database are reported in Table 6. This shows the consistency of the performance of CORNIA.

Statistical significance testing: We performed a two sample T-test with 95% confidence level between SROCC generated by PSNR, SSIM, single-scale BLIIND-II[4] and our algorithms in 1000 iterations of experiments on the LIVE database. Test results are shown in Table 7. From this table, we can see that CORNIA is statistically superior to PSNR,

[4] A multi-scale extension of the single scale measure was also reported in [18], however, we were not be able to obtain code for that part and similar multi-scale extension can also be done for our work, so t-test was only performed on single-scale DIIVINE-II.

	WN	WNC	SCN	MN	HFN	IN	QN	BLUR	IDN	JPEG	JPEG2K	JPEGTE	JP2KTE	ALL
PSNR	0.920	0.914	0.931	0.868	0.935	0.928	0.898	0.928	0.941	0.922	0.884	0.808	0.804	0.669
SSIM	0.854	0.837	0.847	0.813	0.901	0.727	0.878	0.958	0.959	0.934	0.961	0.872	0.878	0.878
CORNIA	0.913	0.928	0.868	0.879	0.921	0.924	0.886	0.932	0.887	0.929	0.919	0.726	0.785	0.813

Table 4. Median SROCC with 1000 iterations of experiments on the TID2008 database. (*Italicized* algorithms are NR-IQA algorithms, others are FR-IQA algorithms.)

	WN	WNC	SCN	MN	HFN	IN	QN	BLUR	IDN	JPEG	JPEG2K	JPEGTE	JP2KTE	ALL
PSNR	0.944	0.933	0.962	0.889	0.975	0.920	0.911	0.912	0.951	0.928	0.907	0.798	0.798	0.652
SSIM	0.827	0.852	0.827	0.845	0.871	0.699	0.837	0.953	0.967	0.963	0.971	0.886	0.846	0.857
CORNIA	0.911	0.932	0.852	0.886	0.933	0.922	0.892	0.932	0.908	0.963	0.929	0.732	0.762	0.837

Table 5. Median LCC with 1000 iterations of experiments on the TID2008 database. (*Italicized* algorithms are NR-IQA algorithms, others are FR-IQA algorithms.)

	PSNR	SSIM	BLIINDS-II	CORNIA
PSNR	0	-1	-1	-1
SSIM	1	0	1	-1
BLIINDS-II	1	-1	0	-1
CORNIA	1	1	1	0

Table 7. Results of the two sample T-test performed between SROCC values obtained by different measures. **1** (**-1**) indicates the algorithm in the row is statistically superior (inferior) than the algorithm in the column. 0 indicates the algorithm in the row is statistically equivalent to the algorithm in the column.

	PSNR	SSIM	CORNIA
SROCC	0.817	0.903	0.880
LCC	0.776	0.901	0.890

Table 8. Database independence test: CORNIA was trained on LIVE and tested on TID2008.

SSIM and single-scale BLIINDS-II.

Database Independence: Additionally, we tested CORNIA by performing training on the LIVE database and testing on the TID2008 database. With trained model on LIVE, the output quality score is in the range from around 0 to 100. But MOS in TID2008 is in the range from 0 to 9. We perform similar nonlinear mapping using a logistic function which is usually applied to FR-measure to obtain a consistent quality measure in certain range. Specifically, we estimate parameters in logistic function using 80% of TID2008 data and then test on the rest 20% of data. (Results of PSNR and SSIM were also obtained in this way.) Since only four types of distortions in the TID2008 database have examples in the LIVE database, we only report results on this four distortions - JPEG2k, JPEG, WN and BLUR. SROCC and LCC are reported in Table 8.

Failure modes: CORNIA works well on distortions which are spatially independent since codewords extracted from a distorted image are treated equally in our framework. Most commonly occurring distortions have such property. However, this is not true for JPEGTE and JP2KTE in TID2008 where distortions may be localized spatially appearing, for example as transmission artifacts or ghosting effects.

	BLIINDS-II (1-scale)	DIIVINE	CORNIA
Time	81.08	29.88	1.59

Table 9. Feature extraction time (in seconds).

4.5. Algorithm Complexity

In this section, we present an informal analysis on the speed of CORNIA. Suppose the number of patches extracted from each image is N, patch size is d and the size of codebook is K. Then the computational complexity of CORNIA is at the order of $O(Nd^2K)$

Time taken for extracting feature of a 512×768 image is about 1.59 seconds. This result is obtained by running an un-optimized matlab program on SunFire x4170 with 2.80GH processor. Speed test has also been performed for DIIVINE and single-scale BLIINDS-II on the same machine, where codes for these two algorithms are downloaded from the author's web site. For all three methods, we only consider time used for feature extraction, time for SVM training and prediction are negligible. Time taken for loading codebook and SVM model to memory is also negligible, since if we have a large number of images to process, we only need to load these data once. As is shown in Table 9, CORNIA is much faster than the other two NR-IQA algorithms.

5. Conclusion

We have presented a simple, efficient and effective algorithm for NR-IQA problem, which advances the state-of-the-art. The proposed algorithm is computationally appealing since we use raw-image-patches as local descriptors and soft-assignment coding for encoding. The proposed algorithm has shown good performance on both LIVE database and TID2008 database. On LIVE database, it is outperforming PSNR and SSIM and is comparative to state-of-the-art general-purpose NR-IQA algorithms in predicting human perceived quality. Furthermore, it is much faster than current state-of-the-art general-purpose NR-IQA algorithms.

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.

References

[1] T. Brandão and M. P. Queluz. No-reference image quality assessment based on DCT domain statistics. *Signal Processing*, 88:822–833, Apr. 2008.

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

[3] A. Coates, B. Carpenter, C. Case, S. Satheesh, B. Suresh, T. Wang, D. J. Wu, and A. Y. Ng. Text detection and character recognition in scene images with unsupervised feature learning. In *Intl. Conf. on Document Analysis and Recognition (ICDAR)*, 2011.

[4] A. Coates, H. Lee, and A. Y. Ng. An analysis of single-layer networks in unsupervised feature learning. In *In Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2011.

[5] A. Coates and A. Y. Ng. The importance of encoding versus training with sparse coding and vector quantization. In *International Conference on Machine Learning (ICML)*, pages 921–928, June 2011.

[6] 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, 2010.

[7] C. Li, A. C. Bovik, and X. Wu. Blind image quality assessment using a general regression neural network. *IEEE Transactions on Neural Networks*, 22(5):793 –799, may 2011.

[8] Q. Li and Z. Wang. Reduced-reference image quality assessment using divisive normalization-based image representation. *IEEE Journal of Selected Topics in Signal Processing*, 3(2):202 –211, april 2009.

[9] C. Liu, W. Freeman, R. Szeliski, and S. B. Kang. Noise estimation from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 901 – 908, 2006.

[10] L. Liu, L. Wang, and X. Liu. In defense of soft-assignment coding. In *IEEE International Conference on Computer Vision (ICCV)*, 2011.

[11] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi. Perceptual blur and ringing metrics: application to JPEG2000. *Signal Processing: Image Communication*, 19(2):163–172, 2004.

[12] A. K. Moorthy and A. C. Bovik. Visual importance pooling for image quality assessment. *IEEE Journal of Selected Topics in Signal Processing*, 3(2):193 –201, April 2009.

[13] A. K. Moorthy and A. C. Bovik. A two-step framework for constructing blind image quality indices. *IEEE Signal Processing Letters*, 17(5):513–516, May 2010.

[14] A. K. Moorthy and A. C. Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE Transactions on Image Processing*, 20(12):3350–3364, Dec. 2011.

[15] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti. Tid2008 - a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radio Electronics*, 10:30–45, 2009.

[16] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng. Self-taught learning: transfer learning from unlabeled data. In *International conference on Machine learning (ICML)*, pages 759–766, 2007.

[17] M. A. Saad, A. C. Bovik, and C. Charrier. A DCT statistics-based blind image quality index. *IEEE Signal Processing Letters*, 17(6):583–586, Jun. 2010.

[18] M. A. Saad, A. C. Bovik, and C. Charrier. Model-based blind image quality assessment using natural DCT statistics. *IEEE Transactions on Image Processing*, 2011 (to appear).

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

[20] 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.

[21] 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.

[22] H. Tang, N. Joshi, and A. Kapoor. Learning a blind measure of perceptual image quality. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 305 – 312, 2011.

[23] 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*, pages 247 – 254, Jan 2005.

[24] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3360 –3367, june 2010.

[25] 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.

[26] 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 (ICIP)*, Rochester, New York, Sep. 22-25 2002.

[27] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multi-scale structural similarity for image quality assessment. *Invited Paper, IEEE Asilomar Conference on Signals, Systems and Computers*, 2003.

[28] J. Yang, K. Yu, Y. Gong, and T. Huang. Linear spatial pyramid matching using sparse coding for image classification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1794 –1801, 2009.

[29] P. Ye and D. Doermann. No-reference image quality assessment based on visual codebook. In *IEEE International Conference on Image Processing (ICIP)*, 2011.