NATURAL DCT STATISTICS APPROACH TO NO-REFERENCE IMAGE QUALITY ASSESSMENT

Michele A. Saad, Alan C. Bovik

The University of Texas at Austin

Christophe Charrier

The University of Caen, France

ABSTRACT

General-purpose no-reference image quality assessment approaches still lag the advances in full-reference methods. Most no-reference methods are either distortion specific (i.e. they quantify one or more distortions such as blur, blockiness, or ringing), or they train a learning machine based on a large number of features. In this approach, we propose a discrete cosine transform (DCT) statistics-based support vector machine (SVM) approach based on only 3 features in the DCT domain. The approach extracts a very small number of features and is entirely in the DCT domain, making it computationally convenient. The results are shown to correlate highly with human visual perception of quality.

***Index Terms*—** No-reference image quality assessment, discrete cosine transform, anisotropy, entropy, support vector machine.

1. INTRODUCTION

The massive dissemination of digital visual information (in the form of images and video) and the broad range of applications that rely on it, such as PDAs, high definition televisions, internet video streaming, and video on demand, to name a few, necessitates means to evaluate the visual quality of this information. Only recently did full-reference image quality assessment (FR-IQA) methods reach a satisfactory level of performance by correlating highly with human visual perception of quality. SSIM [1], MS-SSIM [2], and the VIF index [3] are examples of FR-IQA algorithms, to name a few. These methods, however, rely on a reference signal against which to compare the test signal. This strictly limits their application domain and calls for reliable blind/no-reference image quality assessment (NR-IQA) algorithms. However, no current NR-IQA algorithm exists that provides consistently reliable generic performance.

NR-IQA algorithms generally follow one of three trends: 1) Distortion-specific approaches: these quantify one or more distortions such as blockiness [4], blur [5], [6], or ringing [7] and score the image accordingly. 2) Learning-based approaches: these train a model to predict the image quality score based on a number of features extracted from the image [8], [9]. 3) Natural scene statistics (NSS) approaches: these rely on the hypothesis that images of the natural word (i.e. distortion free images) occupy a small subspace in the space of all possible images and seek to find a *distance* between the test image and the subspace of natural images [10]. The first approach is distortion-specific and hence obviously application specific. The second approach is only as reliable as the representativeness of the features used to train the learning model. Moreover, most existing algorithms following this trend rely on a large number of features. The third approach is a promising one but relies on extensive statistical modeling and reliable generalization of the models.

In this paper, we propose an SVM regression-based learning model that relies on only 3 features extracted entirely from the DCT domain, making it computationally convenient. The 3 DCT features are chosen based on the observation that their statistics change as the image quality changes. This makes our method a hybrid of trends 2 and 3. We show that the method correlates highly with human visual perception. We also report how each feature alone correlates with human visual perception.

The rest of the paper is organized as follows. In Section 2, we describe the 3 DCT-domain features and the motivation behind the choice of the features. In Section 3, we show how each feature alone correlates with subjective *differential-mean-opinion-score* (DMOS). In Section 4, we describe the learning model. We present the results in Section 5, and we conclude in Section 6.

2. DCT DOMAIN FEATURES

The performance of a learning model is a function of the representativeness of the features (that are used for prediction) of image quality. In other words, the prediction is only as good as the choice of features extracted. It has been hypothesized that the human visual system (HVS) is adapted to the statistics of images in its natural surrounding, and that natural images exhibit strong structural dependencies between their pixel intensity levels [1]. Consequently, we choose to extract features representative of image structure, and whose statistics are observed to change with image distortions. We elect to extract structure features locally from local DCT coefficients, (in particular DCT transform of 17×17 image patches). We ignore the DC coefficient whose magnitude is usually much higher than the higher-frequency DCT coefficients in a local image patch. We illustrate how the statistics of the higher frequency DCT coefficients change as an image becomes distorted in Fig.1 and Fig.2, which show the DCT coefficient histograms of a distortion free and a Gaussian blur distorted image, respectively. Similar trends in the histogram statistics are observed throughout the LIVE IQA database of images [11], on which we perform our study. Among the observed differences in the histograms is the peakedness at zero, (the distorted images are observed to have a higher histogram peak at zero), and the variance along the support of the histogram. We seek to make use of statistical differences, such as the ones demonstrated above, to develop a NR-IQA index.

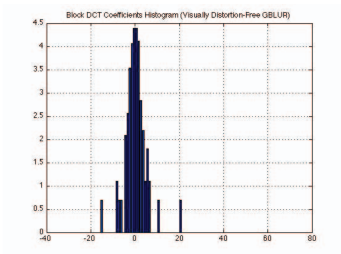

Fig. 1. Original image DCT histogram. Horizontal axis is non-DC DCT coefficient magnitudes

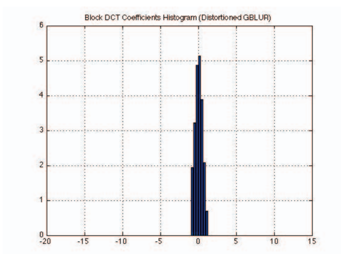

Fig. 2. Gaussian blur distorted image DCT histogram. Horizontal axis is non-DC DCT coefficient magnitudes

2.1. Kurtosis

To capture the statistical traits of the DCT histogram we compute its *kurtosis*, which quantifies the degree of its *peakedness* and tail weight, and is given by:

$$\kappa(x) = \frac{E(x-\mu)^4}{\sigma^4} \tag{1}$$

where μ is the mean of x, and σ is its standard deviation.

The kurtosis of each 17×17 DCT image patch is computed, and the resulting values are pooled together by averaging the lowest tenth percentile of the obtained values to obtain a global image kurtosis value.

2.2. Anisotropy Features

It has been hypothesized that degradation processes damage a scene's directional information. Consequently, anisotropy, which is a directionally dependent quality of images, was shown by Gabarda *et al.* in [12] to decrease as more degradation is added to the image. In [12] anisotropy is computed via the generalized Renyi entropy and a windowed pseudo-Wigner distribution (PWD). In this work, we compute a modified version of the anisotropy measure described in [12]. The anisotropy measure we compute is derived from the DCT coefficients of 17×1 oriented image patches. Our anisotropy computation proceeds as follows: DCT image patches are computed along four different orientations $(0°, 45°, 90°$ and $135°)$ at each pixel in the image. Each patch consists of the DCT coefficients of 17 oriented pixel intensities. (We discard the DC coefficient, since the focus is on directional information). Let the DCT coefficients of a certain patch be denoted by $P[n, k]$, where k is the frequency index of the DCT coefficient $(1 < k \leq 17)$, and n is the spatial index where the DCT patch was computed. Each DCT patch is then subjected to a normalization of the form:

$$\tilde{P}_\theta[n, k] =] \frac{P_\theta[n, k]^2}{\sum_k P_\theta[n, k]^2}, \tag{2}$$

where θ is one of the four orientations. The Renyi entropy for that particular image patch is then computed as

$$R_\theta[n] = -\frac{1}{2} log\Big(\sum_k \tilde{P}_\theta[n, k]^3 \Big). \tag{3}$$

Let M_θ be the number of image patches for orientation θ, then the average per orientation for all patches of orientation θ is obtained. This is denoted as $E[R_\theta]$. The variance across all four orientations (denoted as $var(E[R_\theta])$) along with the maximum $E[R_\theta]$ across the four orientations (denoted as $max(E[R_\theta])$) are chosen as measures of anisotropy.

3. CORRELATION WITH SUBJECTIVE DMOS

We report how each feature alone correlates with subjective DMOS provided with the LIVE IQA database. The LIVE IQA database consists of 5 subsets of images according to the type of distortion introduced to the image. These are JPEG2000 distortions, JPEG distortions, white noise, Gaussian blur, and fast fading channel distortions (these are simulated by JPEG2000 errors followed by channel bit errors). We compute the Spearman correlation for each of the 3 extracted features on each of the 5 LIVE IQA database subsets separately (prior to using the features for prediction in the learning model). These are reported in Table 1. While some correlations are low in some subsets (and high in others), others are significantly high, such as the kurtosis feature that consistently results in a high correlation value across all

LIVE Subset	Kurtosis	Max entropy	Entropy variance
JPEG2000	0.8378	0.6120	0.6000
JPEG	0.7756	0.5901	0.7874
White Noise	0.9322	0.9351	0.9617
Gaussian Blur	0.9481	0.7756	0.2111
Fast Fading	0.6999	0.6526	0.0686

Table 1. Spearman correlation (subjective DMOS versus each DCT-based feature)

5 subsets. These numbers motivate us to employ these features in a more elaborate model for quality score prediction. Towards this end we use these features to train a polynomial kernel regression SVM.

4. PREDICTION MODEL

Previous work in IQA has shown that extracting features and performing analysis at multiple scales can improve the quality assessment method. This is due to the fact that the perception of image details depends on the image resolution, the distance from the image plane to the observer, and the acuity of the observer's visual system. A multiscale evaluation accounts for these variable factors. One example is the multi-scale structural similarity index (MS-SSIM) [2] which outperforms the single scale SSIM index. We thus extract the same features described above at two scales. The features at the second scale are extracted, in the same manner as explained in the previous sections, after performing a down-sampling operation (by a factor of two in each spatial dimension) on the image in the spatial domain.
We then use the features at the 2 scales to train and test a 3^{rd} order polynomial kernel regression SVM [13]. The training and test sets are completely content independent, in the sense that no two images of the same scene are present in both sets. The LIVE database is derived from 29 reference images. The training set contains images derived from 15 reference images, and the test set contains the images derived from the other 14.

5. RESULTS

To evaluate the method, the linear Pearson correlation as well as the Spearman correlation were computed between the reported subjective DMOS and the DMOS predicted by our method. These are computed for each of the 5 LIVE IQA database subsets as well as on the entire database. The results are displayed in Table 2. A plot of the predicted DMOS versus the subjective one for each of the data set subsets is shown in Figs 3-8.

6. CONCLUSION

Our proposed method is a general (non-distortion specific) approach to NR-IQA using a minimal number of features extracted entirely from the DCT-domain (which is computationally convenient). The method is shown to correlate highly with human visual perception of quality.

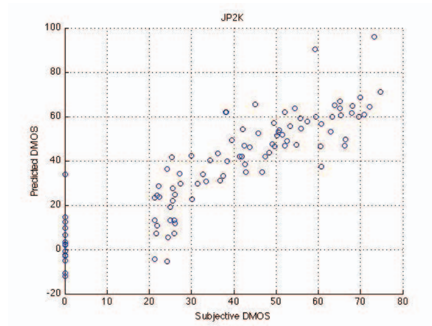

Fig. 3. Predicted DMOS versus Subjective DMOS (JPEG2000)

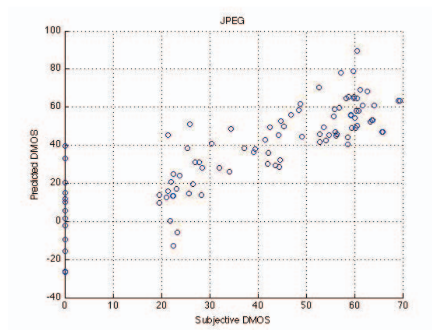

Fig. 4. Predicted DMOS versus Subjective DMOS (JPEG)

7. ACKNOWLEDGEMENTS

This research was supported by the CSOSG ANR-08-SECU-007-04 grant.

8. REFERENCES

[1] 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*, vol. 13, no. 4, pp. 600–612, April 2004.

LIVE Subset	Pearson Corr	Spearman Corr
JPEG2000	0.8892	0.9091
JPEG	0.8495	0.8647
White Noise	0.9740	0.9747
Gaussian Blur	0.9655	0.9652
Fast Fading	0.8296	0.8134
All	0.8001	0.8248

Table 2. Spearman and Pearson correlations (subjective DMOS versus predicted DMOS)

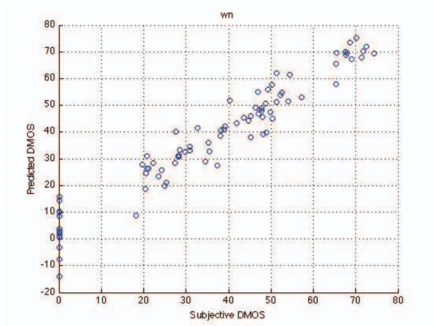

Fig. 5. Predicted DMOS versus Subjective DMOS (White Noise)

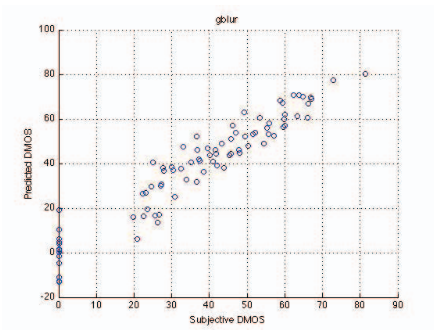

Fig. 6. Predicted DMOS versus Subjective DMOS (Gaussian Blur)

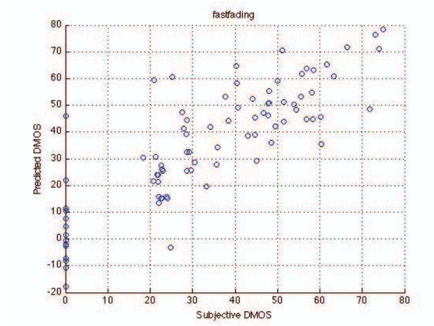

Fig. 7. Predicted DMOS versus Subjective DMOS (Fast Fading)

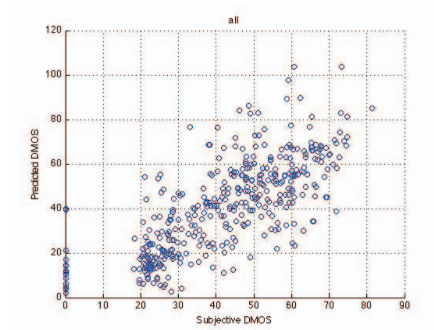

Fig. 8. Predicted DMOS versus Subjective DMOS (Entire Database)

[2] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity image quality assessment," in *37th Asilomar Conference on Signals, Systems, and Computers*, November 2003, vol. 2, pp. 1398–1402.

[3] 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*, vol. 14, no. 12, pp. 2117–2128, December 2005.

[4] Z. Wang, A. C. Bovik, and B. L. Evans, "Blind measurement of blocking artifacts in images," in *IEEE International Conference on Image Processing*. IEEE, September 2000, vol. 3, p. 981984.

[5] Z. M. Parvez Sazzad, Y. Kawayoke, and Y. Horita, "No-reference image quality assessment for jpeg2000 based on spatial features," *Signal Processing: Image Communication*, vol. 23, no. 4, pp. 257–268, April 2008.

[6] X. Zhu and P. Milanfar, "A no-reference sharpness metric sensitive to blur and noise," in *QoMEX*, July 2009.

[7] X. Feng and J. P. Allebach, "Measurement of ringing artifacts in jpeg images," in *Proceedings of the SPIE*, January 2006, vol. 6076, pp. 74–83.

[8] M. Jung, D. Lger, and M. Gazalet, "Univariant assessment of the quality of images," *Journal of Electronic Imaging*, vol. 11, no. 3, July 2002.

[9] C. Charrier, G. Lebrun, and O. Lezoray, "A machine learning-based color image quality metric," in *Third European Conference on Color Graphics, Imaging, and Vision*, June 2006, pp. 251–256.

[10] T. Brandao and M. P. Queluz, "No-reference image quality assessment based on DCT-domain statistics," *Signal Processing*, vol. 88, no. 4, pp. 822–833, April 2008.

[11] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE image quality assessment database release 2," *http://live.ece.utexas.edu/research/quality*, 2006.

[12] S. Gabarda and G. Cristobal, "Blind image quality assessment through anisotroy," *Journal of Optical Society of America*, vol. 24, no. 12, pp. B42–B51, December 2007.

[13] C-C. Chang and C-J. Lin, "LIBSVM: a library for support vector machines," 2001, http://www.csie.ntu.edu.tw/ cjlin/libsvm.