IMAGE QUALITY ASSESSMENT BASED ON DCT SUBBAND SIMILARITY

Amnon Balanov, Arik Schwartz, Yair Moshe, and Nimrod Peleg

Signal and Image Processing Laboratory (SIPL)
Department of Electrical Engineering, Technion – Israel Institute of Technology
yair@ee.technion.ac.il, http://sipl.technion.ac.il/

ABSTRACT

Measuring image quality becomes increasingly important due to the many applications involving digital imaging and communication. Image quality assessment aims to develop a visual quality metric that correlates well with human visual perception. In this paper, we present a full-reference image quality assessment technique based on DCT Subbands Similarity (DSS). The proposed technique exploits important characteristics of human visual perception by measuring change in structural information in subbands in the discrete cosine transform (DCT) domain and weighting the quality estimates for these subbands. The proposed technique was tested with public image datasets and shows higher correlation with subjective results than state-of-the-art techniques. Another advantage of the proposed technique is its low computational cost.

Index Terms—Image quality assessment (IQA), objective quality measure, Discrete Cosine Transform (DCT).

1. INTRODUCTION

Image quality assessment (IQA) becomes increasingly important with the ongoing demand of media in a growing number of applications such as mobile phones, tablets, live streaming, and social networks. Most of this media undergoes processing and transmission that introduce distortions and that degrade the perceived image quality for the end human observer. It is desirable to evaluate the quality degradation introduced by these distortions in order to benchmark image processing systems and algorithms, for quality control systems or in order to optimize parameter values of image processing algorithms. The most reliable evaluation method is by subjective human interpretation and evaluation. However, this method is costly and time-consuming. Therefore, it is desired to look for an automatic tool that will produce a quantitative assessment that correlates well with human judgment of the quality.

There are three approaches for evaluating image quality: full-reference (FR), no-reference (NR) and reduced-reference (RR). In this paper, we will concentrate on FR methods, where the original reference image is available. We will now give a short review of recent full-reference image quality assessment techniques. For more detailed information, the reader is referred to [1, 2].

Conventional metrics, such as the peak signal-to-noise ratio (PSNR), are commonly used but can be poor predictors of perceived visual quality [3]. Due to the limitations of PSNR, much effort has been spent on designing better visual quality metrics. The approaches in metric design can be classified into two groups - a vision modeling approach and an engineering approach [4]. The vision modeling approach is based on modeling various components of the human visual system (HVS). HVS-based metrics try to incorporate aspects of human vision deemed relevant to picture quality, such as luminance adaptation, contrast sensitivity, color perception and pattern masking, using models developed on the basis of psychophysical experiments. An early metric in this category is the digital video quality (DVQ) metric [5] that operates in the DCT domain. The visual SNR (VSNR) metric [6] operates in the wavelet domain. Contrast thresholds for detection of distortions and Euclidean distances are computed via wavelet-based models of visual masking and visual summation. The most apparent distortion (MAD) metric [7] uses two separate strategies for measuring near-threshold distortions and clearly visible distortion. Then, the two measures are combined to provide a single quality value.

The engineering approach is based primarily on the extraction and analysis of certain features or artifacts of the signal. These can be statistical features, structural similarity, luminance or color distortion, or specific distortions that are introduced by common image processing steps, such as blockiness and blurring. These metrics look at how pronounced these features are in the image to estimate overall quality. This does not mean that such metrics disregard human vision, as they often consider psychophysical effects as well, but they concentrate on signal analysis rather than vision modeling. The structural similarity (SSIM) index [8] computes the mean, standard deviation and covariance of small patches inside an image and combines the measurements into a distortion map. These three terms measure luminance distortion, contrast distortion and the loss of correlation, respectively. A more advanced IQA technique based on the SSIM is the multi-scale SSIM (MS-SSIM) [9]. The MS-SSIM index considers the viewing-distance and accounts for the multi-scale

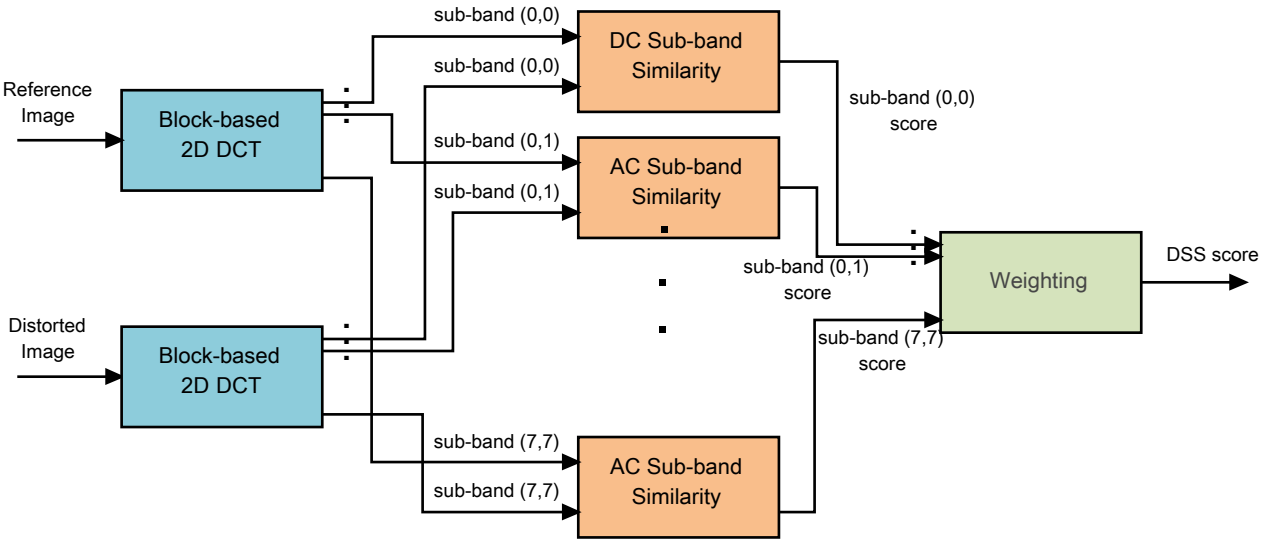

Fig. 1. DCT Sub-bands Similarity (DSS) image quality assessment scheme.

nature of the HVS. This index calculates multiple SSIM values at multiple image scales and aggregates them. CW-SSIM [10] is a wavelet domain version of SSIM that is insensitive to image translation, scaling and rotation. SSIM can also be expressed in the frequency domain using a DCT formulation [11]. Another image quality measure based on the SSIM is the feature similarity (FSIM) index [12]. In this measure, the phase congruency and the gradient magnitude are used to weight the SSIM index. A different approach for IQA, based on information theory, is taken by the visual information fidelity (VIF) index [13]. VIF tries to quantify the amount of information that can ideally be extracted from an image. In particular, it models the statistical image source, the image distortion channel and the human visual distortion channel, and exploits these models to assess image quality. Although MS-SSIM and VIF based on structural similarity and information fidelity, respectively, they have been shown to be equivalent [14].

2. DCT SUBBAND SIMILARITY

A well-known property of natural images is that their image power spectra tend to fall off with increasing spatial frequency. The distribution of DCT image coefficients is often approximated by a Laplacian distribution or by a generalized Gaussian distribution (which includes the Gaussian and Laplacian distributions as special cases) [15, 16]. Recently, the influence of different distortions on the distribution of these coefficients was analyzed [15]. An insight of this analysis is that one of the statistical features that differentiates the DCT coefficient distributions of distorted and nondistorted images is their variance. This property is exploited in the quality metric we propose.

The response of the HVS to changes in intensity of the illumination is known to be nonlinear. Due to the frequency and texture masking phenomena (reduction in the visibility of one image component by the presence of another), the HVS is more sensitive to changes in low spatial frequency regions than to changes in high spatial frequency regions [17]. Therefore, when evaluating image quality, it is beneficial to give more weight to distortions in low frequencies. In addition, many artifacts in images are introduced by coding. Modern image and video coders process each spatial frequency separately and may introduce artifacts with different strength into each subband. Measuring image quality in each spatial frequency subband fits naturally the scheme of these coders so it has the potential to estimate accurately the effect of artifacts introduced by such coders on the HVS. In the proposed IQA technique, we estimate quality in the DCT domain. The main motivation behind quality assessment in the DCT domain is the observation that the statistics of DCT coefficients change with the degree and type of image distortion [15]. The 2D DCT transform is a natural choice in this case since many image and video coding techniques are based on block-based DCT transforms (JPEG, MPEG-2, and H.263, which use the DCT; H.264 and HEVC, which use a variation of the DCT).

A high-level diagram of the proposed DCT subbands similarity (DSS) image quality assessment scheme is depicted in Fig. 1. First, channel decomposition is performed for both the reference and distorted images by separating each image into subbands that correspond to different 2D DCT spatial frequencies. After channel decomposition, the similarity between each subband in the reference image and its counterpart subband in the distorted image is computed, resulting in a subband similarity score. Finally, subband similarity scores are weighted resulting in a scalar DSS quality score.

In the first stage of DSS computation, both images are decomposed into spectral subbands. Each image is divided into small nonoverlapping rectangular blocks of size 8x8 and a

2D DCT transform is applied to each block. The resulting DCT coefficients are decomposed into 64 subbands. Coefficient extraction is performed locally in the spatial domain in accordance with the fact that the HVS processes the visual space locally. We denote by $X_{m,n}(p,q)$ the coefficient at location (p,q) in the subband (m,n), $0 \le m,n \le 7$, of the 2D DCT transform of the image x. Notice that each subband matrix $X_{m,n}$ is 64 times smaller than the image x.

After subband decomposition, the similarity between the coefficients of the reference image and the distorted image is computed for each subband. The aim of this stage is to measure the amount of local change in the statistics of coefficients due to various distortions. We compute the local variance $\sigma_{m,n}(p,q)^2$ for patches of size $k \times k$ in each subband (related to pixel patches of size $8k \times 8k$). We use $k = 3$ and measure the change in these variances as follows:

$$DSS_{m,n}(p,q) = \frac{2\sigma^X_{m,n}(p,q)\sigma^Y_{m,n}(p,q) + C}{\sigma^X_{m,n}(p,q)^2 + \sigma^Y_{m,n}(p,q)^2 + C} \quad , \qquad (1)$$

where $\sigma^X_{m,n}(p,q)^2$ is the local variance at location (p,q) in subband (m,n) of the reference image, $\sigma^Y_{m,n}(p,q)^2$ is the local variance at location (p,q) in subband (m,n) of the distorted image, and C is a constant added for numerical stability. Subband similarity score computed according to (1) is normalized so it is less or equal to 1. The power spectra of natural images tend to fall off with increasing spatial frequency [15], thus subband variances are smaller for higher spatial frequencies. Normalization allows weighting subband scores while ignoring this fall off. Typical values for the constant C are $100 \le C \le 1000$. The results are robust to the selection of C.

For a natural image, there is a high correlation between adjacent pixels. The correlation between adjacent pixels generates image structure that is important for the perception of image quality. The distortion in image structure can be measured by crosscorrelation between pixels of the reference and distorted image [8]. In the proposed IQA metric, we measure the change in image structure by the change in the auto-correlation function of counterpart patches of these images. The autocorrelation function has been suggested as the basis of a texture measure [18] since the locations of peaks in the autocorrelation function involve the shape and arrangement of texture primitives in the image. Thus, the autocorrelation function captures local image structure. According to the Wiener–Khinchin theorem, the autocorrelation function equals to the inverse Fourier transform of the power spectrum of the image. A similar result can be shown to exist also for the DCT transform. As we decompose the image into spectral subbands, we would like to implicitly consider the autocorrelation of neighboring pixels in these subbands. It can be shown that computing the local variance of a subband of 2D-DCT coefficients is related to computing the auto-correlation of an image patch in the pixel domain. Thus, (1) measures the change in the local auto-correlation function of the image. This is equivalent to measuring image structure of the distorted image relative to the structure of the reference image.

A meticulous reader may notice the similarity between (1) and the equation for computing the contrast component in the SSIM metric [8]. In SSIM, the computation is performed on pixels and accounts for the contrast distortion of an image. An important feature of this function is that with the same amount of contrast change $\Delta\sigma = \sigma_x - \sigma_y$, this measure is less sensitive to high base contrast σ_x than to low base contrast. This is consistent with the contrast masking feature of the HVS. In DSS however, the computation is performed on DCT coefficients in a specific subband and accounts for the change of variance of these coefficients. The measure is less sensitive to the same amount of change in variance for high base variance. The base variance is high when there are large changes in the local pixel region. This is related to high contrast so the masking feature of the HVS is considered here as well.

While AC coefficients give an indication of the amplitude and orientation of edges within an image, the DC coefficient is proportional to the average image amplitude. Hence, it is required to compute the similarity between DC subbands in a way that takes into account also other properties in the DC subbands. We found that the crosscorrelation between the two DC subband images (of the reference and distorted images) plays an important role in estimating the perceived image quality. Thus, we compute the similarity score between the reference and distorted DC subbands as follows:

$$DSS_{0,0}(p,q) = \frac{2\sigma^X_{0,0}(p,q)\sigma^Y_{0,0}(p,q) + C}{\sigma^X_{0,0}(p,q)^2 + \sigma^Y_{0,0}(p,q)^2 + C} \cdot \frac{\sigma^{XY}_{0,0}(p,q) + C}{\sigma^X_{0,0}(p,q)\sigma^Y_{0,0}(p,q) + C} \quad , \qquad (2)$$

where $\sigma^{XY}_{0,0}(p,q)$ is the crosscorrelation at location (p,q) in the DC subband between the reference and distorted image. The first term in the equation is similar to (1) and measures the change is the local variance while the second term is the local Pearson crosscorrelation metric. Like in (1), C is a constant added for numerical stability.

The similarity scores in (1) and (2) are computed pointwise resulting in a similarity score for each coefficient. We would like to use these scores in order to obtain a scalar similarity score for each subband. Humans tend to perceive "poor" regions in an image with more severity than "good" ones, and hence penalize heavily images with even a small number of "poor" regions. Namely, worst quality regions in an image dominate human perception of image quality. This effect can be taken into account by weighting more heavily quality scores from lower scoring regions. One of the simplest and best performing strategies for achieving this goal is percentile scoring [19]. In this pooling technique, the scalar similarity score is the mean of only the lowest w% percentiles

Table 1. Performance of DSS and other image quality assessment techniques on images from the LIVE, CSIQ, and TID databases. The best results are in bold.

		PSNR	SSIM	MSSSIM	VSNR	VIF	MAD	FSIM	DSS
SROCC	LIVE	0.876	0.910	0.944	0.928	0.963	0.968	0.963	**0.970**
	CSIQ	0.806	0.837	0.914	0.811	0.919	0.947	0.924	**0.953**
	TID	0.525	0.645	0.853	0.705	0.750	0.834	0.881	**0.891**
	Average	0.736	0.797	0.904	0.815	0.877	0.916	0.923	**0.938**
LCC	LIVE	0.870	0.938	0.933	0.923	0.960	**0.968**	0.960	**0.968**
	CSIQ	0.800	0.815	0.897	0.800	0.925	**0.950**	0.912	**0.950**
	TID	0.536	0.652	0.839	0.682	0.806	0.831	0.874	**0.897**
	Average	0.735	0.802	0.890	0.802	0.897	0.916	0.915	**0.938**
RMSE	LIVE	13.368	11.790	8.946	10.506	7.673	6.900	7.678	**6.830**
	CSIQ	0.158	0.133	0.115	0.158	0.098	0.082	0.108	**0.078**
	TID	1.137	0.851	0.730	0.982	0.789	0.750	0.653	**0.593**
running time [ms]		**5**	95	270	60	2000	2800	1350	200

of the scores. A typical value of w is 5. We apply percentile scoring to each subband separately, getting 64 subband similarity scores $DSS_{m,n}, 0 \le m,n \le 7$.

After evaluating the distortion in each DCT subband, the 64 similarity scores are weighted to obtain a scalar score:

$$DSS = \sum_{m,n=0}^{7} w_{m,n} DSS_{m,n} \quad , \qquad (3)$$

where $w_{m,n}$ is the weight of the score of subband (m,n). As previously mentioned, an important and well-modeled characteristic of the HVS is its higher sensitivity to distortions in low spatial frequencies [17]. It is desired to select a weighting function $f(m,n) = w_{m,n}$ that takes this characteristic into account by giving higher weights to subbands with lower spatial frequencies. We tried several weighting functions that satisfy this property and empirically selected a Gaussian weighting function. The standard deviation of the Gaussian determines the proportion between the weight given to low spatial frequencies and to high spatial frequencies. A typical value used for the standard deviation is $\sigma = \sqrt{6}$.

3. RESULTS

In order to examine the performance of the proposed image quality assessment technique, we have evaluated it against subjective results on three public image datasets [20] – the LIVE Image Quality Assessment Database [21], the Tampere Image Database (TID) [22], and the Categorical Subjective Image Quality (CSIQ) Database [7]. Each database contains hundreds of distorted images and subjective results of few dozen subjects. We use the Spearman rank-order correlation coefficient (SROCC), the Pearson linear correlation coefficient (LCC) and the Root Mean Square Error (RMSE), between the subjective scores and the quality indices to compare the relative performances between the proposed technique and other state-of-the-art full-reference image quality assessment techniques. The SROCC serves as a measure of prediction monotonicity, while the LCC and RMSE serve as measures of prediction accuracy. A better objective IQA measure should have higher SROCC, higher LCC, and lower RMSE. In order to evaluate the performance of a particular quality assessment technique, we have applied a 5-parameter nonlinear mapping recommended in [23]. In an attempt to obtain a linear relationship between the predicted and subjective quality scores, this mapping function transforms the predicted quality indices to the same scale as the subjective results before applying SROCC, LCC or RMSE.

Table 1 lists the SROCC, LCC and RMSE performance on the three image databases. Eight FR IQA metrics were examined: PSNR, SSIM [8], MS-SSIM [9], VSNR [6], VIF [13], MAD [7], FSIM [12] and DSS. For the three databases and for the three evaluation criteria, the proposed DSS metric yields, on average, the highest correlation with subjective results compared with all other image quality assessment metrics. The table also lists execution time of each of the metrics in MATLAB on Intel Core i5 running at 2.67GHz. DSS has higher computational complexity than metrics such as PSNR and SSIM that have substantially lower correlation with subjective results, but is an order of magnitude faster than techniques such as VIF, MAD and FSIM that have high correlation with subjective results.

4. CONCLUSION

In this paper, we propose a novel technique for full-reference image quality assessment that measures change in structural information in subbands in the discrete cosine transform (DCT) domain and weighting the quality estimates for these subband scores. The proposed technique achieves, on average, higher correlation with subjective results and has a lower computational complexity compared with competing state-of-the-art image quality assessment techniques.

5. ACKNOWLEDGMENT

The authors would like to thank Prof. David Malah, head of SIPL, for his support, advice, and helpful comments.

5. REFERENCES

[1] W. Lin and C. C. Jay Kuo, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation,* vol. 22, pp. 297-312, 2011.

[2] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *Broadcasting, IEEE Transactions on,* vol. 57, pp. 165-182, 2011.

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

[4] S. Winkler and P. Mohandas, "The evolution of video quality measurement: from PSNR to hybrid metrics," *Broadcasting, IEEE Transactions on,* vol. 54, pp. 660-668, 2008.

[5] A. B. Watson, J. Hu, and J. F. McGowan, "Digital video quality metric based on human vision," *Journal of Electronic imaging,* vol. 10, pp. 20-29, 2001.

[6] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *Image Processing, IEEE Transactions on,* vol. 16, pp. 2284-2298, 2007.

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

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

[9] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," presented at the Signals, Systems and Computers, 37th Asilomar Conference on, 2003.

[10] Z. Wang and E. P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," in *In Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, 2005.

[11] S. S. Channappayya, A. C. Bovik, and R. W. Heath, "Rate bounds on SSIM index of quantized images," *Image Processing, IEEE Transactions on,* vol. 17, pp. 1624-1639, 2008.

[12] L. Zhang, D. Zhang, and X. Mou, "FSIM: A Feature Similarity Index for Image Quality Assessment," *Image Processing, IEEE Transactions on,* vol. 20, pp. 2378-2386, 2011.

[13] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *Image Processing, IEEE Transactions on,* vol. 15, pp. 430-444, 2006.

[14] K. Seshadrinathan and A. C. Bovik, "Unifying analysis of full reference image quality assessment," presented at the Image Processing (ICIP'08)), 15th IEEE International Conference on, 2008.

[15] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *Image Processing, IEEE Transactions on,* vol. 21, pp. 3339-3352, 2012.

[16] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *Image Processing, IEEE Transactions on,* vol. 9, pp. 1661-1666, 2000.

[17] W. K. Pratt, "Digital Image Processing: PIKS Scientific Inside," ed: Wiley-Interscience, 2007.

[18] H.-C. Lin, L.-L. Wang, and S.-N. Yang, "Extracting periodicity of a regular texture based on autocorrelation functions," *Pattern Recognition Letters,* vol. 18, pp. 433-443, 1997.

[19] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *Selected Topics in Signal Processing, IEEE Journal of,* vol. 3, pp. 193-201, 2009.

[20] S. Winkler, "Analysis of public image and video databases for quality assessment," *Selected Topics in Signal Processing, IEEE Journal of* vol. 6, pp. 616 - 625, 2012.

[21] 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, 2005.

[22] 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 Radioelectronics,* vol. 10, pp. 30-45, 2009.

[23] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms," *Image Processing, IEEE Transactions on,* vol. 15, pp. 3440-3451, 2006.