BLIND PERCEPTUAL QUALITY ASSESSMENT METHOD FOR DCT-BASED ENCODED IMAGES

Tomás Brandão[1,3] *and Maria Paula Queluz*[2,3]

[1] Department of Sciences and Information Technologies, ISCTE/UI-Lisbon, Portugal;
[2] Department of Electrical and Computer Engineering, IST/TU-Lisbon, Portugal;
[3] IT-Lisbon, Torre Norte, Piso 10, Av. Rovisco Pais, 1, 1049-001 Lisbon, Portugal;
phone/fax: +(351)218418454/218418472; emails: tomas.brandao, paula.queluz@lx.it.pt; www.it.pt

ABSTRACT

This paper describes a technique whose purpose is to estimate the perceptual quality of DCT-based encoded images, without requiring the original data. To achieve this objective, a watermark is embedded in the DCT domain using a non-uniform quantization scheme. At the receiver side, the original DCT coefficients data distribution is estimated using a maximum likelihood approach. These distributions and the extracted watermark are then combined to estimate the error between reference and distorted DCT coefficients. This error is perceptually weighted, using a DCT domain perceptual model, allowing to blindly score the quality of the received media. Results have shown the effectiveness of the proposed algorithm when scoring the quality of images subject to lossy compression.

1. INTRODUCTION

Quality monitoring of multimedia data is becoming an important matter, especially due to the increasing transmission of multimedia contents over the internet and mobile networks. The most reliable scores for the perceived quality of multimedia data are achieved by means of *subjective metrics*, which result from an evaluation performed directly by human viewers. The score that results from this evaluation is frequently referred as *mean opinion score* (MOS). Since these scores require multiple viewers under controlled conditions, they are time consuming and thus useless in real-time environments. An alternative is to use *objective metrics*, allowing to automatically compute a quality score that should resemble the one that results from subjective evaluation.

Most of the research performed on objective metrics has been focused on the development of so-called *full reference* (*FR*) quality metrics, which are computed using both the original and distorted media. *FR* metrics are typically used as benchmarks for image processing algorithms (e.g. lossy coding, watermarking, image restoration, etc.). However, *FR* metrics can not be used in the context of media distribution, since the original media is not available at the receiver. In this kind of scenario, it is expected that content providers will be able to track the quality of the perceived media at the reception. This will enable new services, such as users paying proportionally to the quality they get at the reception and new server options, such as adjustment of streaming parameters as a function of the perceived quality. It is thus necessary to increase the effort in the development of the so called *reduced* (*RR*) and *no-reference* (*NR*) quality metrics: the first uses an additional channel to transmit side information about the reference media, while the later provide quality scores based only on received media.

Recent publications [1, 2, 3, 4] suggest the use of watermarking in order to provide additional help when attaining for *RR* and *NR* metrics scores. The use of watermark-based approaches in quality assessment problems is motivated by the fact that, once the host media is watermarked, both the host and the watermark will follow the same path, being subject to the same distortion. At the receiver, it should be possible to conclude about distortion of the host signal by analysis of the received watermark signal. However, quality metrics retrieved by current algorithms do not correlate well with the human perception of quality. We propose to overcome this limitation by considering the perceptual characteristics of the human eye, in terms of *just noticeable differences*, when computing the quality score at the reception. The objective is to estimate perceptual weights and distortion errors at the reception, in such a way that quality scores given to the distorted image resemble the perceptual metric proposed by A. B. Watson in [5].

In this paper, the watermark is embedded in the 8×8 block-based DCT domain of the image, using the non-uniform quantization-based technique described in [4]. However, at the receiver, instead of computing distortion distances based in empirically derived weighting functions [4], the original DCT coefficient distributions are estimated by considering natural scene statistics corrupted by quantization noise (which is valid for both JPEG and MPEG encoding). These statistics are then used together with the watermark in order to compute a perceptual weight map and to estimate the error between reference and received media. Related work is presented in [6], where blind quality scores are computed by measuring and combining specific compression artifacts. The algorithm proposed in this paper can also be used in other applications, since it is capable of providing local error estimates (which may be used in different perceptual weight-based quality metrics or even in artifact reduction, for instance). The expense is the presence of a watermark and a slight increase in complexity. Results have shown that the proposed scheme provides quality scores that correlate well with the human perception of quality. Furthermore, it is also able to compute very accurate estimates of rougher objective metrics such as the PSNR.

This document is organized as follows: section 2 gives a brief summary of the model proposed by Watson; watermarking scheme is depicted in section 3; section 4 describes how to get quality scores from the watermark and from the received DCT coefficient data; results are presented in the section 5; finally, conclusions and suggestions for fu-

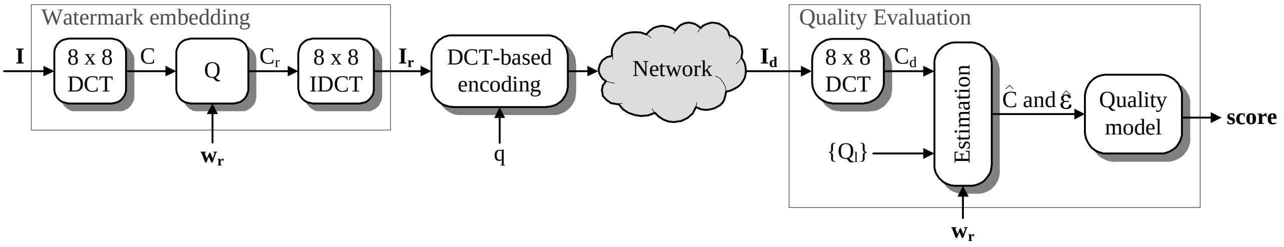


Figure 1: Watermark-based quality evaluation framework.

ture work are given in section 6.

2. WATSON'S MODEL IN BRIEF

In [5], Watson estimates the perceptibility of modifications in individual DCT coefficients in terms of *just noticeable differences* (JNDs), whose threshold values are called *slacks*. The proposed model comprises two components accounting for luminance and contrast masking effects. Let $C(i,j,k)$ represent the original block-based DCT coefficient at position (i,j) of the k-th block. The correspondent luminance masking threshold, $T_{lum}(i,j,k)$, is given by [5]:

$$T_{lum}(i,j,k) = T(i,j)\left(\frac{C(i,j,k)}{C_{00}}\right)^{\alpha_T},$$

where $T(i,j)$ is the frequency sensitivity for position (i,j), C_{00} is the average of the DC coefficients in the image, and α_T is a constant with a suggested value of 0.649. Slack values, $s(i,j,k)$, are computed by also considering the effect of contrast masking, through:

$$s(i,j,k) = \begin{cases} T_{lum}(i,j,k), & \text{if } |C(i,j,k)| \le T_{lum}(i,j,k); \\ \\ |C(i,j,k)|^{\beta(i,j)} T_{lum}(i,j,k)^{1-\beta(i,j)}, & \text{otherwise,} \end{cases} \tag{1}$$

where $\beta(i,j) = 0$ for $(i,j) = (0,0)$ and $\beta(i,j) = 0.7$, otherwise. The local perceptual error, $\varepsilon_p(i,j,k)$, is computed dividing the error between original and distorted coefficient values, $C_d(i,j,k)$, by the corresponding slack value:

$$\varepsilon_p(i,j,k) = \frac{\varepsilon(i,j,k)}{s(i,j,k)}, \tag{2}$$

with

$$\varepsilon(i,j,k) = |C(i,j,k) - C_d(i,j,k)|.$$

A global metric, d_{Watson}, is computed by combining all perceptual errors. Watson suggests the use of $L4$ error polling, i.e.:

$$d_{Watson} = \sqrt[4]{\frac{1}{M}\sum \varepsilon_p(i,j,k)^4}, \tag{3}$$

where M is the number of DCT coefficients under analysis.

Throughout this paper, Watson's model will be used in two distinct processes: at watermark embedding, with the purpose of maximizing the embedding strength without compromising watermark imperceptibility; and at the reception, for quality scoring of the received media.

3. WATERMARKING SCHEME

Consider that a binary watermark message, w_r, is embedded into the luminance component of a host image I. Watermark embedding and extraction are performed in the 8×8 block-based DCT domain of I, using the quantization-based approach proposed in [4].

The left side of figure 1 depicts the embedding scheme. Let Q_l represent the quantizer's output value at level l. Each coefficient used for embedding is modified to the nearest quantization level whose least significant bit (*LSB*) is equal to the watermark bit to be embedded. Formally, assuming that Q_n is the quantization value nearest to $C(i,j,k)$, the watermarked coefficient, $C_r(i,j,k)$, is obtained by:

$$C_r(i,j,k) = \begin{cases} Q_n, & \text{if } mod(n,2) = w_r(i,j,k); \\ Q_{n+t}, & \text{otherwise.} \end{cases}$$

where $w_r(i,j,k)$ is the watermark bit to embed, $mod(x,y)$ is the remainder of the integer division of x by y and t is defined as:

$$t = sgn(C(i,j,k) - Q_n),$$

with $sgn(x) = -1$ if $x < 0$ and $sgn(x) = 1$, otherwise. Concerning the quantization function, a non-uniform quantization scheme [4], that incorporates the main features of Watson's perceptual model, has been applied. The quantizer's output value at level l can be defined recursively as:

$$Q_l = \begin{cases} \alpha\frac{T(i,j)}{2}, & \text{if } l = 0; \\ Q_{l-1} + \alpha Q_{l-1}^{\beta(i,j)} T(i,j)^{1-\beta(i,j)}, & \text{otherwise.} \end{cases}$$

where α is a constant that regulates the embedding strength and the remaining parameters are those described in the previous section. The goal is to assign larger quantization steps to coefficients that allow greater modifications, while keeping the watermark's imperceptibility.

To complete the process, the inverse DCT transform is computed, resulting in the watermarked image I_r, which will be considered as the reference image.

At the receiver, estimates for the local error absolute value, $\hat{\varepsilon}(i,j,k)$ and for the original coefficient values, $\hat{C}(i,j,k)$ are computed based on the received coefficient data, $C_d(i,j,k)$, on w_r bits and on the values of Q_l. Quality scores result from $\hat{\varepsilon}(i,j,k)$ and $\hat{C}(i,j,k)$. Note that the reference watermark, w_r, must be known (or generated) by the receiver. The quality evaluation system will be discussed in the following section.

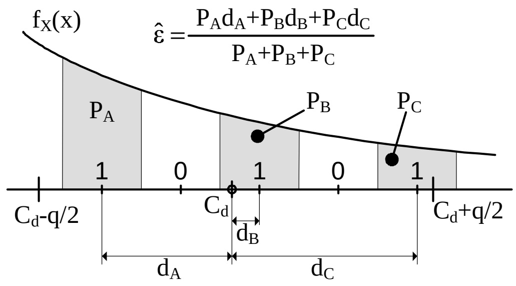

$f_X(x)$
$\hat{\varepsilon} = \frac{P_A d_A + P_B d_B + P_C d_C}{P_A + P_B + P_C}$
P_A
P_B
P_C
1
0
1
0
1
C_d-q/2
C_d
C_d+q/2
d_B
d_A
d_C

Figure 2: Error estimation.

4. QUALITY ESTIMATION

In this section it will be assumed that distortion on the reference (watermarked) image is due to linear quantization of the DCT coefficients, which is realistic in the presence of JPEG and MPEG encoding. For simplicity purposes, the notation that is being used for DCT block position and indexing will be dropped (e.g., $C(i,j,k)$ will be simply referred as C). Under these conditions, the expected value of the local absolute error, $\hat{\varepsilon}$, between reference and distorted coefficients in a given position, can be estimated by:

$$\hat{\varepsilon} = \frac{\sum_l P(Q_l)|C_d - Q_l|}{\sum_l P(Q_l)}, \text{ for } \begin{cases} Q_l \in [C_d - \frac{q}{2}, C_d + \frac{q}{2}] \\ LSB(l) = w_r, \end{cases} \qquad (4)$$

where $P(Q_l)$ is the probability of the reference coefficient value (after watermark embedding) to be Q_l and q is the quantization step used for image encoding at the corresponding coefficient's position. To illustrate (4), consider figure 2, which represents a set of reference coefficient values Q_l (small ticks labeled with '1's and '0's). Let's admit that the embedded reference watermark bit at a given coefficient's position is '1'. $\hat{\varepsilon}$ is estimated by first computing the distances from the received coefficient, C_d, to the reference points laying inside the interval $[C_d - \frac{q}{2}, C_d + \frac{q}{2}]$ and assigned to watermark bit '1'. In the figure, those distances are represented by d_A, d_B and d_C. Each distance is weighted by P_A, P_B and P_C, respectively, which are the probabilities of the reference coefficient value to be in each of the points with $w_r = 1$.

However, knowledge about the probability $P(Q_l)$ is not available at the receiver, thus it must be estimated based on the received coefficient data. In order to do so, the original coefficient data is modeled using a *Laplace* probability density function (*pdf*) with parameter λ, which represents a reasonable trade-off between model accuracy and simplicity. According to this model, the *pdf* for the original coefficient values, $f_X(x)$, is given by:

$$f_X(x) = \frac{\lambda}{2} e^{(-\lambda|x|)},$$

where λ is the distribution's parameter at the corresponding DCT frequency.

For algebraic simplicity purposes, it will be considered that further quantization of the reference signal gives approximately the same results as direct quantization of the original signal. This is a reasonable approach, since distortion due to watermark embedding is much lower than the expectable distortion caused by compression. Thus, assuming that lossy encoding results from linear quantization with step q, the probability for the original coefficient, x, to be quantized to value X_l is:

$$P(X_l)) = \int_{X_l - \frac{q}{2}}^{X_l + \frac{q}{2}} \frac{\lambda}{2} e^{-\lambda|x|} dx. \qquad (5)$$

If the quantization function is symmetric and includes the zero value, which is the case for JPEG and MPEG-2 encoding, (5) can be rewritten as:

$$P(X_l) = \begin{cases} 1 - e^{-\frac{\lambda q}{2}}, & \text{if } X_l = 0; \\ \frac{1}{2} e^{-\lambda|X_l| + \frac{\lambda q}{2}} (1 - e^{-\lambda q}), & \text{otherwise.} \end{cases} \qquad (6)$$

In order to estimate the parameter λ of the original *pdf* using the quantized coefficient values, the maximum-likelihood (ML) method is used, following the approach presented in [7]:

$$\hat{\lambda} = \arg\max_{\lambda} \{\log \prod P(X_l)\}. \qquad (7)$$

Substituting (6) in (7) leads to:

$$\begin{aligned} \hat{\lambda} = \arg\max_{\lambda} \{ & \sum_{X_l = 0} \log(1 - e^{-\frac{\lambda q}{2}}) \\ & + \sum_{X_l \neq 0} \log \frac{1}{2} (e^{-\lambda|X_l| + \frac{\lambda q}{2}})(1 - e^{-\lambda q})\} \end{aligned} \qquad (8)$$

Representing by N_0 and N_1 the number of points quantized to zero and non-zero values, respectively, and considering $S = \sum_{X_l \neq 0} |X_l|$, (8) can be rewritten as:

$$\begin{aligned} \hat{\lambda} = \arg\max_{\lambda} \{ & N_0 \log(1 - e^{-\frac{\lambda q}{2}}) \\ & - \lambda S + \frac{N_1 \lambda q}{2} + N_1 \log(1 - e^{-\lambda q})\}. \end{aligned} \qquad (9)$$

The derivative of (9) with respect to λ leads to:

$$(Nq + 2S) e^{-\lambda q} + N_0 q e^{-\frac{\lambda q}{2}} + N_1 q - 2S = 0, \qquad (10)$$

where N represents the number of coefficients at a given frequency. Equation (10) resembles a second order polynomial in $e^{-\frac{\lambda q}{2}}$, with solution:

$$\hat{\lambda} = -\frac{2}{q} \log \frac{-N_0 q + \sqrt{N_0^2 q^2 - 4(Nq + 2S)(N_1 q - 2S)}}{2Nq + 4S}. \qquad (11)$$

The parameter $\hat{\lambda}$ retrieved by (11) can then be used to compute the values of $P(Q_l)$:

$$P(Q_l) = \int_{\frac{Q_{l-1} + Q_l}{2}}^{\frac{Q_l + Q_{l+1}}{2}} \frac{\hat{\lambda}}{2} \exp(-\hat{\lambda}|x|) dx. \qquad (12)$$

The absolute value for the local error can then be estimated by using (12) in (4). The resulting error estimates can be used for quality estimation purposes. For instance, an estimate for the distorted image PSNR can be computed according to:

$$PSNR_{est[dB]} = 10 \log_{10} \frac{255^2}{\frac{1}{M} \sum \hat{\varepsilon}^2}, \qquad (13)$$

where M is the number of DCT coefficients.

However, it is far more interesting to use the estimated error with the purpose of scoring the perceptual quality of the received images. A "no-reference" approach for Watson's model is achieved by computing estimates for slack values, $\hat{s}$, based on the received coefficient values. Attending to (1):

$$\hat{s} = \begin{cases} \hat{T}_{lum}, & \text{if } |\hat{C}| \leq \hat{T}_{lum}; \\ |\hat{C}|^{\beta}\hat{T}_{lum}^{1-\beta}, & \text{otherwise,} \end{cases} \tag{14}$$

where $\hat{T}_{lum}$ is an estimate for Watson's luminance threshold and $\hat{C}$ is an estimate for the original coefficient value, which are given by:

$$\hat{T}_{lum} = T\left(\frac{\hat{C}}{C_{00}}\right)^{\alpha_T}; \qquad \hat{C} = C_d + \hat{\varepsilon}',$$

where $\hat{\varepsilon}'$ is an estimate for the local error value. It can be computed similarly to (4), using the difference $(C_d - Q_l)$ inside the summation, instead of its absolute value. This error is added to the received coefficient value, thus obtaining an estimate for the original coefficient value. This value is then used for computing slack values and, through (2), to blindly compute the local perceptual error:

$$\hat{\varepsilon}_p = \frac{\hat{\varepsilon}}{\hat{s}}, \tag{15}$$

To conclude, a global perceptual distortion measure is obtained using (15) in (3).

5. RESULTS

5.1 PSNR estimation

The proposed scheme has been evaluated using the LIVE image set database [8]. The images have been watermarked and JPEG encoded using quality factors in the range of $10-100$, using steps of 10. The watermark embedding strength α was set to 0.5, which guarantees watermark imperceptibility. All AC coefficients have been used for watermark embedding. After lossy encoding, the no-reference PSNR estimates given by (13) have been compared with the true PSNR values.

Figure 3(a) depicts the results attained for an image randomly chosen from the LIVE database and figure 3(b) depicts the PSNR estimates versus their true values, for all database images. Statistics regarding the PSNR estimation accuracy have been synthesized in table 1.

As can be observed from both the table and the figures, the proposed algorithm is quite accurate for the purpose of PSNR estimation.

5.2 Quality scores

The results for quality assessment have been evaluated by comparing the quality scores retrieved by the algorithm with the ones that result from a subjective test. LIVE database contains subjective scores for images subject to JPEG compression using different quality factors. Subjective scores are expressed by their *differential mean opinion scores* (DMOS), which is the quality score difference between the reference and the distorted image (quality decreases with increasing values of DMOS).

Table 1: PSNR estimation accuracy.

Average estimation error	0.703 dB
Error standard deviation	0.543 dB
Correlation (estimated and true PSNR)	0.984

Table 2: Evaluation of the proposed blind metric.

Root mean square (RMS)	6.245
Pearson's correlation coefficient (CC)	0.979
Spearman's rank order coefficient (RC)	0.967

Figure 4(a) depicts the estimated Watson's distance, using (3) and the perceptual error estimates given by (15), versus the corresponding DMOS values. Following a procedure similar to what is suggest by the *Video Quality Experts Group* (VQEG) in [9], a logistic function was used in order to normalize the values retrieved by (3), using the perceptual error estimates given by (15), into a linear quality scale from $0-100$. The logistic function has the form:

$$\text{Estimated DMOS} = \theta_0 + \frac{\theta_1}{1+\exp(\theta_2 d_{Watson} + \theta_3)}, \tag{16}$$

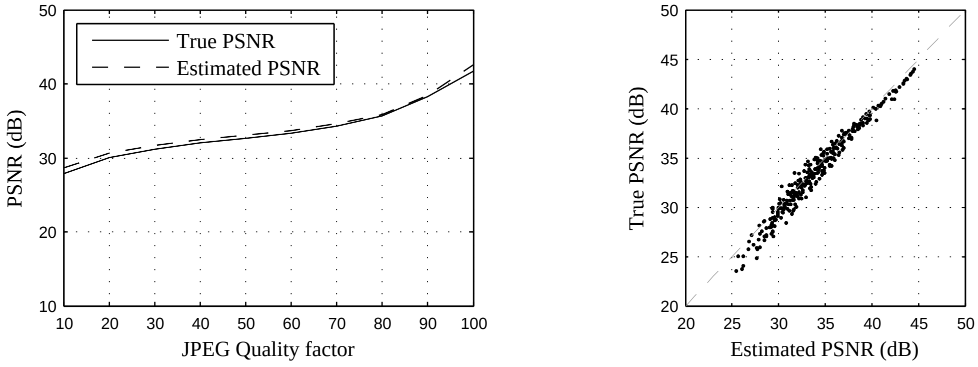
where θ_0 to θ_3 are parameters to estimate. These parameters have been computed in order to minimize the square differences between the estimated DMOS scores given by (16) and the true DMOS values in a given training set. The training set consists of DMOS scores given to the JPEG encoded versions of 15 reference images randomly chosen from LIVE database. The θ parameters have been computed using the *Levenberg-Marquardt* method for non-linear least squares minimization problems. The resulting logistic function can be observed in figure 4(a).

Figure 4(b) depicts the normalized "no-reference" quality scores versus their DMOS values. As can be observed, objective quality scores resulting from the proposed algorithm are well correlated with the subjective quality ranks. In [9], VQEG also suggests the use of several statistical measurements in order to evaluate the performance of an objective metric. These measurements have been synthesized in table 2.

The results confirm the good performance of the proposed algorithm. When compared with [6], the proposed scheme provides better results for RMS. Concerning the other measures, *CC* and *RC*, there is not sufficient data available to perform a fair comparison. Remember that [6] follows a completely different approach, where the proposed quality metrics result from compression artifacts measurements.

6. CONCLUSIONS AND DIRECTIONS

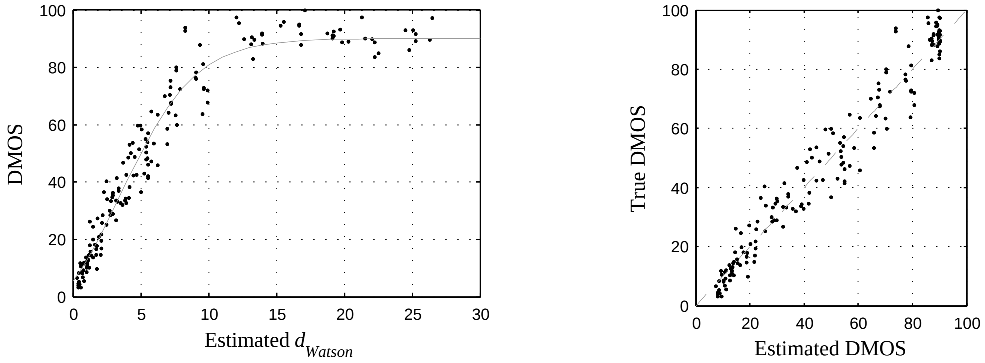
In this paper, a watermark-based algorithm that blindly scores the quality of DCT-based encoded images has been proposed. The main paper contribution is the achievement of a JND-based quality metric that resembles the perceptual model derived by A. B. Watson, without requiring the original data. Furthermore, it has also been shown that the proposed approach may be used in order to accurately estimate the PSNR metric.



(a) An example (*Lighthouse 2*).

(b) Global results.

Figure 3: PSNR estimation results.



(a) Estimated Watson distance vs. DMOS.

(b) Global results.

Figure 4: DMOS estimation results.

Further investigation is already undergoing in order to extend the developed work to DCT-based encoded digital video. It may also be worth to investigate the use of the proposed algorithm in other applications involving error estimation problems, such as coding artifacts reduction.

7. ACKNOWLEDGEMENTS

The authors would like to thank the *Laboratory for Image and Video Engineering* of Texas University for providing them access to the LIVE image database (special thanks to Dr. Hamid Sheikh).

REFERENCES

[1] P. Campisi, M. Carli, G. Giunta, and A. Neri, “Blind quality assessment system for multimedia communications using tracing watermarking,” *IEEE Transactions on Signal Processing*, vol. 51, no. 4, April 2003.

[2] S. Saviotti, F. Mapelli, and R. Lancini, “Video quality analysis using a watermarking technique,” in *Proc. of WIAMIS*, Lisbon, Portugal, April 2004.

[3] M. Holliman and M. Yeung, “Watermarking for automatic quality monitoring,” in *Proc. of SPIE Security and Watermarking of Multimedia Contents IV*, S. Jose, USA, January 2002.

[4] T. Brandão and M. P. Queluz, “Towards objective metrics for blind assessment of image quality,” in *Proc. of International Conference on Image Processing*, Atlanta, USA, October 2006.

[5] A. B. Watson, “DCT quantization matrices optimized for individual images,” in *Proc. of SPIE Human Vision, Visual Processing, and Digital Display IV*, S. Jose, USA, 1993.

[6] Z. Wang, H. Sheikh, and A. Bovik, “No-reference perceptual quality assessment of JPEG compressed images,” in *Proc. of International Conference on Image Processing*, Rochester, USA, 2002.

[7] J. Price and M. Rabbani, “Biased reconstruction for JPEG decoding,” *IEEE Signal Processing Letters*, vol. 6, no. 12, December 1999.

[8] H. Sheikh, Z. Wang, L. Cormack, and A. Bovik, “LIVE image quality assessment database release 2,” available at http://live.ece.utexas.edu/research/quality.

[9] VQEG, “Final report from the video quality experts group on the validation of objective models of video quality assessment, phase II,” August 2003, available at http://www.vqeg.org.