Available online at www.sciencedirect.com
ScienceDirect
Signal Processing 88 (2008) 822–833
SIGNAL PROCESSING
www.elsevier.com/locate/sigpro

No-reference image quality assessment based on DCT domain statistics

Tomás Brandão[a,b,*], Maria Paula Queluz[a,c]

[a]IT - Instituto de Telecomunicações, Torre Norte, Piso 10, Av. Rovisco Pais 1, 1049-001 Lisbon, Portugal
[b]ISCTE - University Institute of Lisbon, Av. Forças Armadas, 1649-026 Lisbon, Portugal
[c]Instituto Superior Técnico - Technical University of Lisbon, Av. Rovisco Pais 1, 1049-001 Lisbon, Portugal

Received 23 April 2007; received in revised form 25 September 2007; accepted 26 September 2007
Available online 5 October 2007

Abstract

This paper proposes a no-reference quality assessment metric for images subject to quantization noise in block-based DCT (discrete cosine transform) domain, as those resulting from JPEG or MPEG encoding. The proposed method is based on natural scene statistics of the DCT coefficients, whose distribution is usually modeled by a Laplace probability density function, with parameter λ. A new method for λ estimation from quantized coefficient data is presented; it combines maximum-likelihood with linear prediction estimates, exploring the correlation between λ values at adjacent DCT frequencies. The resulting coefficient distributions are then used for estimating the local error due to lossy encoding. Local error estimates are also perceptually weighted, using a well-known perceptual model by Watson. When confronted with subjective quality evaluation data, results show that the quality scores that result from the proposed algorithm are well correlated with the human perception of quality. Since no knowledge about the original (reference) images is required, the proposed method resembles a no-reference quality metric for image evaluation.

Keywords: Image quality; No-reference quality metric; DCT coefficient statistics; Parameter estimation

1. Introduction

Quality monitoring of multimedia data is becoming an important issue, especially due to the increasing transmission of multimedia contents over the internet and mobile networks. In the context of multimedia distribution scenarios, it is expected that content providers will be able to track media quality at the reception. This will enable new services, such as users paying proportionally to the quality they get, and new server possibilities, such as adjustment of streaming parameters as a function of the perceived quality.

The most reliable scores for evaluating the quality of multimedia data are those resulting directly from human viewers. The results from this evaluation are often referred to as *subjective scores* or *mean opinion scores* (MOS). To obtain these scores, several human viewers under controlled test conditions are required. Thus, subjective scores are not suitable for real-time environments.

*Corresponding author. IT - Instituto de Telecomunicações, Torre Norte, Piso 10, Av. Rovisco Pais 1, 1049-001 Lisbon, Portugal. Tel.: +351 218 412 184; fax: +351 218 418 472.
E-mail addresses: tomas.brandao@lx.it.pt (T. Brandão), paula.queluz@lx.it.pt (M.P. Queluz).

0165-1684/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2007.09.017

An alternative is to automatically score image quality using *objective metrics*. Most of the research performed on this area has been focused on the development of so-called *full reference* (FR) metrics [1–6], which are computed using both the original and the distorted media. FR metrics are typically used for benchmarking image processing algorithms (e.g. lossy coding, watermarking, image restoration) or for benchmarking media distribution systems during the testing phases. However, FR metrics are not suitable for practical media distribution scenarios, since the original media is generally not available at the receiver. It is thus desirable to have a quality measurement system that is able to provide perceived quality scores without requiring the reference (original) signals. This had lead to an increased effort in the development of the so-called *reduced* (RR) [7,8] and *no-reference* (NR) [9–15] quality metrics: the first uses a side channel to send additional information about the reference media, while the later provides quality scores based only on the received media.

This paper proposes a method that can be used for image quality evaluation without requiring knowledge about the original signal, thus belonging to the NR image quality metrics category. Quality scores rely on statistical properties of the original, block-based, DCT (discrete cosine transform) coefficient data. To compute these scores, it is thus necessary to accurately estimate the distribution of the original DCT coefficients.

In [13], Turaga et al. were probably the first authors to propose an algorithm that estimates image quality based on the statistical properties of DCT coefficients. On their work, image quality is measured by estimating the peak signal-to-noise ratio (PSNR), which is the most commonly used measure for quantifying image degradation due to lossy encoding. Although criticized by many authors due to its lack of correlation with the human perception of quality (e.g. [1,3,16]), it still is a very popular metric among the image coding community. In [13] it is reported that PSNR estimation accuracy decreases as coding rate decreases. This is due to the increasing number of DCT coefficients quantized to zero values, which lead to inaccurate statistical parameter estimation. Aware of this problem, Ichigaya et al. have proposed a method for PSNR estimation [15] that tries to overcome this issue by estimating a weighted mixture of Laplacian probability density functions (pdf): one is computed by considering all quantized coefficient values and the other is computed by considering the non-zero quantized values only. Although a considerable improvement is reported in [15], the proposed method still fails when all (or almost all) DCT coefficients at the same frequency are quantized to zero. Since this situation is quite common for high-frequency coefficients subject to DCT-based encoding, even at average compression rates, this topic deserves further investigation.

One of the main goals of the work described in this paper is to improve the estimation of the original DCT coefficient data distribution, from their quantized values available at the reception. For that, two new ideas are explored:

- explore the correlation between distribution parameters at adjacent DCT frequencies using a linear prediction scheme;
- combine the prediction results with maximum-likelihood (ML) parameter estimates.

Results show that the proposed parameter estimation technique is accurate for the situations falling in the cases described above (almost all of the DCT coefficients quantized to zero values). When applied to the NR image quality evaluation problem, the algorithm provides PSNR estimates that are more accurate than those provided by [15].

The proposed method is also used for perceptually based image quality scoring by considering perceptual characteristics of the human eye, in terms of *just noticeable differences* (JNDs). 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 Watson in [17]. The proposed algorithm is compared with [10], where quality scores are computed by measuring and combining specific JPEG compression artifacts, and with [14], where quality scores result from combining the outputs of neural networks, whose input data consists of block-based features taken from the image under evaluation.

The work described in this paper has been evaluated using JPEG encoded images. Nevertheless, it intends to present new ideas which can be extended to the blind quality evaluation of video sequences, subject to DCT-based lossy encoding, like in MPEG-2 and MPEG-4. Since the proposed algorithm provides local error estimates, it can be potentially used in other applications, such as image restoration or artifact reduction.

This document is organized as follows: after the Introduction, Section 2 gives a brief overview about ML estimation of the Laplace distribution parameter. Section 3 describes the proposed approach for estimating the distribution's parameter, exploring the correlation between adjacent coefficients distributions. The procedures to obtain quality scores from the estimated DCT distributions are explained in Section 4, and Section 5 depicts the results, which are compared with the methods proposed in [10,14,15]. Finally, conclusions and suggestions for future work are drawn in Section 6.

2. ML parameter estimation

Block-based DCT coefficient data distribution of natural images is well modeled by a *Laplace pdf* [18]. Using 8×8 blocks, for each horizontal/vertical frequency pair, $(i,j) \in \{0,\ldots,7\} \times \{0,\ldots,7\}$ and $(i,j) \neq (0,0)$, the coefficient's distribution is thus modeled by

$$f_X(x) = \frac{\lambda(i,j)}{2} \exp(-\lambda(i,j)|x|), \tag{1}$$

where $\lambda(i,j)$ is the distribution's parameter for frequency pair (i,j) (for simplicity, these indexes will be dropped along the text) and x is the coefficient value.

Other *pdf* models have been suggest in the literature, such as generalized Gaussian [19], Gaussian mixtures [20], or generalized gamma [21]. Nevertheless, the Laplacian model represents a good trade-off between model accuracy and mathematical simplicity. As an example, Fig. 1 shows a comparison between the histogram of coefficient values (for a given frequency) and the corresponding *pdf* model.

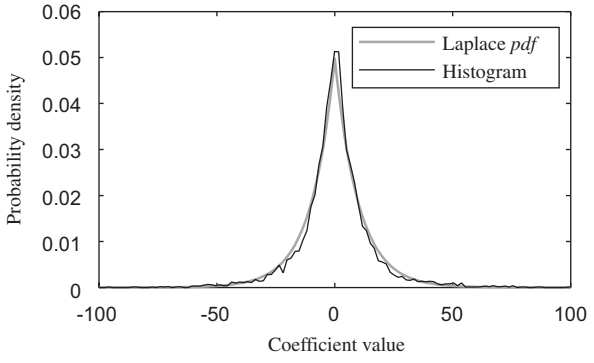

Fig. 1. Histogram for coefficient frequency (3,3) of image *Rapids*.

2.1. λ estimation using the original coefficient data

An estimate for λ, using the original coefficient data, is generally computed using the ML method [22]. Representing by x_k the kth original coefficient value at a given frequency, an ML estimate for λ is therefore given by

$$\lambda_{\mathrm{ML}} = \arg\max_{\lambda} \left\{ \log \prod_{k=1}^{N} f_X(x_k) \right\}, \tag{2}$$

where N represents the number of DCT coefficients at that frequency (which is the same as the number of blocks). The value of λ_{ML} is computed by finding the zeros of the derivative with respect to λ, which leads to

$$\lambda_{\mathrm{ML}} = \frac{N}{\sum_{k=1}^{N} |x_k|} = \frac{1}{E[|x|]}, \tag{3}$$

where $E[\cdot]$ represents the expected value. In the sequel, λ_{ML} will be used as benchmark for parameter estimation based on quantized data. Therefore, it will be referred to as *original* λ.

2.2. λ estimation using quantized coefficient data

Let us now suppose that the original coefficient values are subject to quantization noise and that the quantizer, $Q(.)$, is linear with constant quantization step q. Considering (1), the probability of having value X_k at the quantizer's output is given by

$$P(X_k = Q(x_k)) = \int_{X_k - q/2}^{X_k + q/2} \frac{\lambda}{2} \mathrm{e}^{-\lambda|x|} \, \mathrm{d}x. \tag{4}$$

Assuming that the quantization function is symmetric and includes the zero value (midtread quantizer), which is true for JPEG and MPEG-2 encoding, (4) can be rewritten as

$$P(X_k) = \begin{cases} 1 - \mathrm{e}^{-\lambda q/2} & \text{if } X_k = 0, \\ \frac{1}{2}\mathrm{e}^{-\lambda|X_k| + \lambda q/2}(1 - \mathrm{e}^{-\lambda q}) & \text{otherwise.} \end{cases} \tag{5}$$

For simplicity, $P(X_k = Q(x_k))$ has been denoted by $P(X_k)$. In order to estimate the parameter λ of the original *pdf* using the quantized coefficient values, the ML method can be used, following an approach similar to what is presented in [23], i.e.:

$$\hat{\lambda}_{\mathrm{ML}} = \arg\max_{\lambda} \left\{ \log \prod_{k=1}^{N} P(X_k) \right\}. \tag{6}$$

Substituting (5) in (6) leads to

$$\hat{\lambda}_{\mathrm{ML}} = \arg\max_{\lambda}\left\{\sum_{k_0=1}^{N_0}\log(1-\mathrm{e}^{-\lambda q/2}) + \sum_{k_1=1}^{N_1}\log\frac{1}{2}(\mathrm{e}^{-\lambda|X_{k_1}|+\lambda q/2})(1-\mathrm{e}^{-\lambda q})\right\}, \quad (7)$$

where N_0 and N_1 represent the number of coefficients at a given frequency, quantized to zero and non-zero values, respectively. The two summation terms in (7) correspond to the two possible cases in (5) and the quantized coefficient set, $\{X_k\}$, has been split into sets $\{X_{k_0}\}$ and $\{X_{k_1}\}$, according to those cases. Using the substitution

$$S = \sum_{k_1=1}^{N_1}|X_{k_1}|, \quad (8)$$

and after simple algebraic manipulations, (7) can be rewritten as

$$\hat{\lambda}_{\mathrm{ML}} = \arg\max_{\lambda}\left\{N_0\log(1-\mathrm{e}^{-\lambda q/2}) - \lambda S + \frac{N_1\lambda q}{2} + N_1\log(1-\mathrm{e}^{-\lambda q})\right\}. \quad (9)$$

Once again, the solution is given by finding the zeros of the derivative of (9) with respect to λ, leading to

$$(Nq+2S)e^{-\lambda q} + N_0 q\mathrm{e}^{-\lambda q/2} + N_1 q - 2S = 0. \quad (10)$$

Eq. (10) can be viewed as a second order polynomial in $\mathrm{e}^{-\lambda q/2}$, whose solution is

$$\hat{\lambda}_{\mathrm{ML}} = -\frac{2}{q}\log\frac{-N_0 q + \sqrt{N_0^2q^2 - 4(Nq+2S)(N_1q-2S)}}{2Nq+4S}. \quad (11)$$

Let us now take a deeper insight into Eq. (11), for the case where most of the DCT coefficients have been quantized to value 0. As the number of coefficients quantized to zero increases, $N_0 \to N$, and it is easy to conclude that $N_1 \to 0$ and $S \to 0$. Under these conditions, the argument of the logarithm in (11) will tend to zero and therefore $\hat{\lambda}_{\mathrm{ML}}$ will tend to infinity, meaning that the estimated distribution will approach a *Dirac's delta* function. This situation will be common in the presence of DCT-based compression, since high-frequency DCT coefficients are typically quantized to zero, even at average compression rates. As consequence, ML estimates for λ based on the quantized data will be inaccurate for these cases.

This phenomena has already been noticed in [15], where the authors propose an "extension" to the Laplace distribution. Their idea is to compute the original DCT distribution as a mixture of two Laplace *pdf*s: one of them is estimated by considering all the quantized coefficient values, while the other is estimated by considering quantized non-zero values only. Although a considerable improvement is achieved by using the proposed method, it still fails if all (or almost all) DCT coefficients at a given frequency are quantized to zero. Due to its characteristics, the algorithm proposed in [15] will be addressed to as *Laplacian compensation* method, for the remainder of the paper.

3. λ estimation using prediction

In order to tackle the problem described at the end of the previous section, this paper proposes a new approach to improve the estimation of the Laplacian parameter: to explore the correlation between λ values at neighboring DCT frequencies.

Consider Figs. 2(a) and (b), which represent the original λ values, computed using Eq. (3), for all 8×8 frequencies in two test images. The figures show a strong correlation between λ values at adjacent frequencies. In fact, using the reference (original) images in LIVE database [24], the correlation between the value of λ at a given frequency and the values in adjacent frequencies,

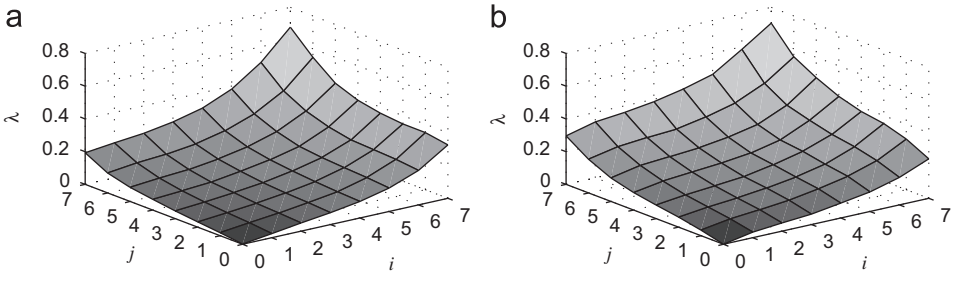

Fig. 2. Original λ values for each frequency. (a) *Rapids*. (b) *House*.

using a 4-connected neighborhood, is 0.971. One possible way to explore this correlation is to use a linear predictor. Representing the prediction result by $\hat{\lambda}_{\mathrm{p}}$, we can write

$$\hat{\lambda}_{\mathrm{p}} = \beta_0 + \sum_{k=1}^{K} \lambda_{\upsilon_k}\beta_k, \tag{12}$$

where K is the neighborhood size, λ_{υ_k} is the kth neighbor λ value and β_k is the corresponding linear weight, whose value will be found by a training procedure. Using matrix notation, Eq. (12) can also be written as

$$\hat{\lambda}_{\mathrm{p}} = \boldsymbol{\lambda}_{\upsilon}^{\mathrm{T}}\boldsymbol{\beta} \quad \text{with } \boldsymbol{\lambda}_{\upsilon} = \begin{bmatrix} 1 \\ \lambda_{\upsilon_1} \\ \vdots \\ \lambda_{\upsilon_K} \end{bmatrix} \text{ and } \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}. \tag{13}$$

The prediction $\hat{\lambda}_{\mathrm{p}}$ that results from (13) can then be used together with $\hat{\lambda}_{\mathrm{ML}}$ in order to improve the estimation accuracy for the original DCT distribution parameter.

Since ML estimates become more inaccurate as the rate of coefficients quantized to zero increases, more thrust should be given to the predictor in these kind of situations. On the other hand, if the number of coefficients quantized to zero is low, the ML estimator will most likely get accurate results, so there is no real need for the predicted value. Based on these premises, a simple criterion for combining $\hat{\lambda}_{\mathrm{p}}$ with $\hat{\lambda}_{\mathrm{ML}}$ is to weight them proportionally to the rate of DCT coefficients quantized to zero:

$$\hat{\lambda}_{\mathrm{f}} = r_0\hat{\lambda}_{\mathrm{p}} + (1 - r_0)\hat{\lambda}_{\mathrm{ML}}, \tag{14}$$

where $r_0 = N_0/N$ represents the rate of coefficients quantized to zero and $\hat{\lambda}_{\mathrm{f}}$ is the final estimation for the distribution's parameter.

3.1. Predictor training procedure

The objective of the training procedure is to find a weight vector $\boldsymbol{\beta}$ that is suitable for the proposed linear prediction scheme, given by Eq. (12). One possible way to compute $\boldsymbol{\beta}$ is by minimizing the square error between the original λ and $\hat{\lambda}_{\mathrm{p}}$, for all images in the training set. Admitting that L images are available for training, then we will have L vectors $\boldsymbol{\lambda}_{\upsilon}$ and L values of λ per DCT coefficient frequency. Therefore, it can be written as

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{l=1}^{L} (\lambda^{(l)} - \hat{\lambda}_{\mathrm{p}}^{(l)})^2 \right\} \\ &= \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{l=1}^{L} (\lambda^{(l)} - \boldsymbol{\lambda}_{\upsilon}^{\mathbf{T}(l)}\boldsymbol{\beta})^2 \right\}. \end{aligned} \tag{15}$$

Using matrix notation, the equation above can be rewritten as

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \{[\boldsymbol{\lambda} - \boldsymbol{\Lambda}\boldsymbol{\beta}]^{\mathrm{T}}[\boldsymbol{\lambda} - \boldsymbol{\Lambda}\boldsymbol{\beta}]\}, \tag{16}$$

where $\boldsymbol{\Lambda}$ is an $L \times K$ matrix, with the neighborhood samples for image l, $\lambda_{\upsilon_k}^{(l)}$, organized in rows and $\boldsymbol{\lambda}$ is a vector with the original λ values at the position to predict, i.e.:

$$\boldsymbol{\Lambda} = \begin{bmatrix} 1 & \lambda_{\upsilon_1}^{(1)} & \cdots & \lambda_{\upsilon_K}^{(1)} \\ 1 & \lambda_{\upsilon_1}^{(2)} & \cdots & \lambda_{\upsilon_K}^{(2)} \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_{\upsilon_1}^{(L)} & \cdots & \lambda_{\upsilon_K}^{(L)} \end{bmatrix}, \quad \boldsymbol{\lambda} = \begin{bmatrix} \lambda^{(1)} \\ \lambda^{(2)} \\ \vdots \\ \lambda^{(L)} \end{bmatrix}.$$

The solution of Eq. (16) can be found by differentiating with respect to $\boldsymbol{\beta}$:

$$\nabla_{\boldsymbol{\beta}} = 0 \;\Leftrightarrow\; \boldsymbol{\Lambda}^{\mathrm{T}}(\boldsymbol{\lambda} - \boldsymbol{\Lambda}\boldsymbol{\beta}) = 0. \tag{17}$$

Finally, if $\boldsymbol{\Lambda}^{\mathbf{T}}\boldsymbol{\Lambda}$ is non-singular, the unique solution for $\hat{\boldsymbol{\beta}}$ is given by

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{\Lambda}^{\mathbf{T}}\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\lambda}. \tag{18}$$

The original image set in LIVE database [24] has been randomly divided into two subsets, one for training (15 images) and another for validation (14 images). The neighborhood configuration used in the experiments is illustrated in Fig. 3. This structure has been chosen with the purpose of recursively predict values for λ, starting from the

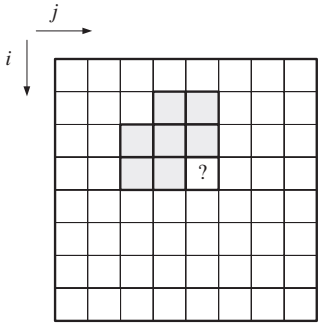

Fig. 3. Neighborhood configuration used for prediction.

low frequency positions. The procedure to obtain vector $\boldsymbol{\beta}$ can thus be described in the following steps:

(1) for each image in the training set, the original values of λ are computed for all frequencies, using (3) and the original coefficient data;
(2) for each frequency, the neighborhood matrix $\boldsymbol{\Lambda}$ and vector $\boldsymbol{\lambda}$ are constructed using the values that result from the previous step;
(3) predictor weights are computed for each frequency, using Eq. (18).

3.2. Prediction accuracy

To evaluate the effectiveness of the prediction scheme, two experiments have been performed using the validation image set. In a first experiment, $\hat{\lambda}_{\mathrm{p}}$ was computed using the original λ values in the neighborhood. The main purposes of this experiment were to evaluate the lowest prediction error that could be expect and to validate the training procedure. The average relative prediction error (in percentage) is depicted in Table 1(a). It can be observed that the relative error is generally low, which confirms that, due to high correlation of λ values in adjacent frequencies, prediction results are quite accurate. It can also be observed that the error is greater for the cases where less neighborhood values are available for the prediction (e.g. first row and first column).

Table 1
Average relative prediction error (%)

$\overset{j}{\rightarrow}$							
$i\downarrow$							
(a) *Normal prediction*							
–	–	11.7	9.9	4.9	4.5	4.7	7.7
–	11.9	6.2	5.1	2.7	2.3	3.3	4.1
10.9	5.2	2.7	2.5	2.8	2.4	4.0	2.0
7.5	4.0	2.8	4.4	2.6	3.1	2.7	4.6
6.1	3.7	4.3	5.4	2.1	4.3	4.2	2.7
6.5	3.0	2.6	3.3	3.4	2.2	1.6	2.8
5.8	2.7	1.8	2.6	2.4	1.8	3.4	2.7
3.3	5.0	2.9	3.0	2.5	2.5	2.3	1.9

$\overset{j}{\rightarrow}$							
$i\downarrow$							
(b) *Recursive prediction*							
–	–	–	–	–	4.5	6.0	8.8
–	–	–	–	2.7	3.6	7.4	7.3
–	–	–	2.5	4.2	4.2	8.5	8.2
–	–	2.8	4.5	5.5	6.5	11.0	18.3
–	3.7	4.7	6.1	5.7	6.6	8.5	13.9
6.5	6.0	7.3	7.5	7.8	8.7	10.2	13.9
12.2	10.0	10.6	8.5	11.4	10.0	12.0	14.8
19.1	12.8	10.9	11.2	13.0	14.0	16.5	18.3

A second experiment tried to make an approximation to what happens when images are subject to lossy compression. Remember that, in the presence of lossy compression, most high-frequency coefficients are quantized to zero, resulting in unreliable ML parameter estimates. On the other hand, low frequency coefficients are less affected by distortion, and thus more reliable parameter estimates can be performed. To approximate this situation, the original values of λ have been computed only for the low frequency range (frequency pairs (i,j) with $i+j<5$ and $(i,j)\neq(0,0)$, which correspond to 14 frequencies, marked with symbol '–' in Table 1(b)). The remaining values of λ have been computed recursively from the values at those frequencies, using Eq. (12). The average relative prediction error that has result from this test is shown in Table 1(b). It can be observed that prediction error increases with increasing frequency. This is due to error propagation in the recursive algorithm: for lower frequencies, predictions are performed using the original values of λ, but with increasing frequency, predictions are based on previously predicted values. Nevertheless, the results are quite satisfactory.

Additionally, Fig. 4 illustrates λ estimation results in the presence of lossy DCT-based encoding. For this example, a test image (*Womanhat*) has been subject to JPEG encoding, with the quality factor set to 50. Fig. 4(a) shows the original values of λ, computed using Eq. (3). These values should be seen as an estimation benchmark. Fig. 4(b) depicts the estimated values for λ using the quantized coefficient data, computed through (11). Due to all data quantized to zero at the medium–high DCT frequencies, the resulting estimates at those frequencies are infinite (for better visualization, λ values in these conditions have been represented by zeros—the darker areas in the corresponding plot). By combining ML estimates with the proposed prediction scheme, using Eq. (14), it can be observed in Fig. 4(c) that the resulting λ estimates are stable (no infinite values are computed) and quite close to the estimates based on the original coefficient data. For a better comparison, Fig. 4(d) joins together on a 1-D plot the results presented in Figs. 4(a)–(c).

4. Quality estimation

4.1. Blind PSNR estimation

The Laplacian parameter estimation method described throughout this paper can be used as a

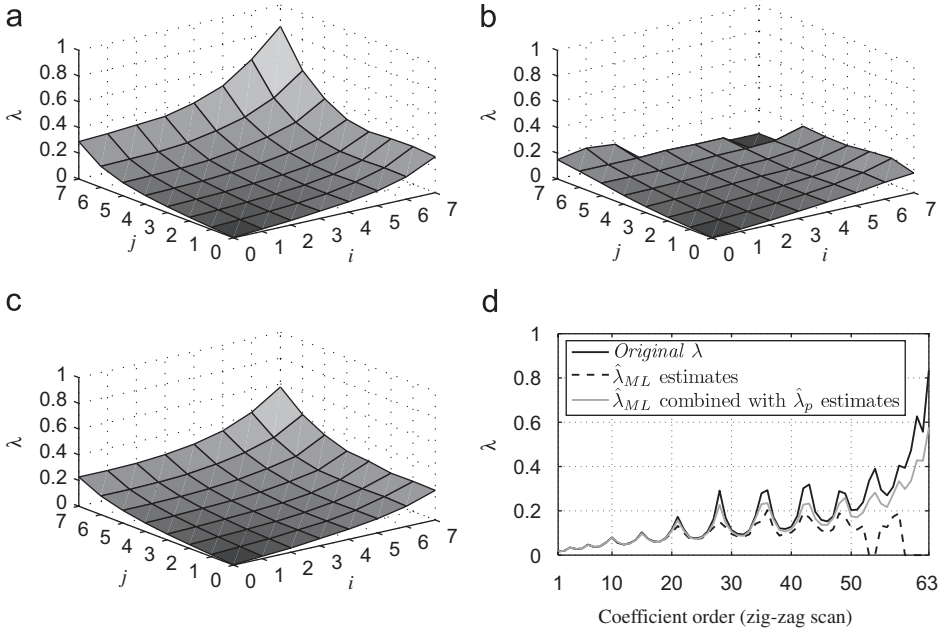

Fig. 4. λ estimation results for a JPEG encoded test image. (a) Original λ values. (b) $\hat{\lambda}_{\mathrm{ML}}$ estimates. (c) $\hat{\lambda}_{\mathrm{ML}}$ combined with $\hat{\lambda}_{\mathrm{p}}$ estimates. (d) Comparison.

tool to estimate the PSNR of an image subject to DCT-based lossy encoding, without requiring the original image data. For images with pixel values in the range of $[0;255]$, the PSNR is usually computed by

$$\mathrm{PSNR}_{[\mathrm{dB}]} = 10\log_{10}\frac{255^2}{(1/M)\sum_{k=1}^{M}\varepsilon_k^2}, \tag{19}$$

where M is the number of pixels and ε_k^2 is the squared error between the kth reference and the kth distorted pixel. In the context of this paper, M will be the number of DCT coefficients under analysis and $\varepsilon_k^2 = (X_k - x_k)^2$ will be the squared difference between original and quantized coefficients. Note that, in accordance with Parceval's theorem (and since DCT is an unitary transform), it is indifferent to measure image PSNR in pixel or in DCT domains.

The knowledge of the original DCT data distributions can be used for estimating the local mean square error $\hat{\varepsilon}_k^2$ at the kth coefficient:

$$\hat{\varepsilon}_k^2 = \int_{-\infty}^{+\infty} f_X(x|X_k)(X_k - x)^2\,\mathrm{d}x, \tag{20}$$

where $f_X(x|X_k)$ represents *pdf* of the original coefficient value to be x, given that it has been quantized to value X_k. According to *Bayes rule*, it can be written as

$$f_X(x|X_k) = \frac{P(X_k|x)f_X(x)}{P(X_k)}, \tag{21}$$

where $P(X_k|x)$ is the probability of having quantizer's output X_k given the coefficient's value x. This probability is 1 if x lies in the quantization interval centered in X_k and 0 otherwise. Formally

$$P(X_k|x) = \begin{cases} 1 & \text{if } x \in \left[X_k - \dfrac{q}{2}; X_k + \dfrac{q}{2}\right], \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

Substituting (21) in (20) leads to

$$\hat{\varepsilon}_k^2 = \frac{1}{P(X_k)}\int_{X_k - q/2}^{X_k + q/2} f_X(x)(X_k - x)^2\,\mathrm{d}x, \tag{23}$$

with $f_X(x)$ and $P(X_k)$ given by Eqs. (1) and (5), respectively. Both equations use the estimated value of $\hat{\lambda}_{\mathrm{f}}$ that results from (14). These estimates have been computed recursively starting from the lower frequencies (zig-zag scan order [25]), according to the following procedure:

(1) $\hat{\lambda}_{\mathrm{ML}}$ and r_0 have been computed for each frequency;
(2) $\hat{\lambda}_{\mathrm{p}}$ is computed by using previously estimated values of $\hat{\lambda}_{\mathrm{f}}$ in the neighborhood (or by using the values of $\hat{\lambda}_{\mathrm{ML}}$ at the start of the recurrence).

4.2. Perceptual quality

In order to score the perceptual quality of the received images, we propose a "NR" approach for the 8×8 block-based DCT-domain perceptual model introduced by Watson in [17].

4.2.1. Watson's model overview

In [17], Watson proposes to estimate the perceptibility of modifications in 8×8 block-based DCT coefficients in terms of JNDs, whose threshold values are called *slacks*. The proposed model comprises two masking components which account for luminance and contrast effects.

Let $x_k(i,j)$ represent the kth original (reference) DCT coefficient at frequency (i,j) of the 8×8 block. The correspondent luminance masking threshold, $T'_k(i,j)$, is given by

$$T'_k(i,j) = T(i,j)\left(\frac{x_k(0,0)}{\overline{x}_{00}}\right)^{\alpha_T}, \tag{24}$$

where $T(i,j)$ is the sensitivity threshold for frequency (i,j) (given in [17]), $\overline{x}_{00}$ is the average of the DC coefficients in the image, and α_T is a constant with a suggested value of 0.649.

Considering the effect of contrast masking, slack values, $s_k(i,j)$, are computed by

$$s_k(i,j) = \begin{cases} T'_k(i,j) & \text{if } |x_k(i,j)| \leqslant T'_k(i,j), \\ |x_k(i,j)|^{\omega(i,j)} & \\ \quad \times T'_k(i,j)^{1-\omega(i,j)} & \text{otherwise,} \end{cases} \tag{25}$$

where $\omega(i,j) = 0$ for $(i,j) = (0,0)$ and $\omega(i,j) = 0.7$, otherwise. The local perceptual error, $\varepsilon_{\mathrm{p}_k}(i,j)$, is obtained dividing the error between original and distorted coefficient values, $\varepsilon_k(i,j)$, by the corresponding slack value:

$$\varepsilon_{\mathrm{p}_k}(i,j) = \frac{\varepsilon_k(i,j)}{s_k(i,j)}, \tag{26}$$

with

$$\varepsilon_k(i,j) = |X_k(i,j) - x_k(i,j)|. \tag{27}$$

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

$$D_{\mathrm{W}} = \sqrt[4]{\frac{1}{M}\sum_{k=1}^{N}\sum_{i=0}^{7}\sum_{j=0}^{7}\varepsilon_{\mathrm{p}_k}(i,j)^4}, \tag{28}$$

where M is the total number of coefficients under analysis and N is the number of coefficients per frequency (or the number of 8×8 blocks).

4.2.2. Perceptual quality estimation

Let us consider Watson's model, namely the equation for the perceptual error given by (26). If the objective is to compute a "NR" estimate for the perceptual error, it will be necessary to estimate the local error $\varepsilon_k(i,j)$ and the associated slack value $s_k(i,j)$. For simplicity purposes, the indexes (i,j) will be dropped once again.

An estimate for the absolute error, $\overline{\varepsilon_k}$, can be computed in the same way as for the PSNR case, using Eq. (23), but replacing the squared term inside the integral with the absolute value of the error, i.e.:

$$\overline{\varepsilon_k} = \frac{1}{P(X_k)}\int_{X_k-q/2}^{X_k+q/2} f_X(x)|X_k - x|\,\mathrm{d}x. \tag{29}$$

The estimate for the slack value can be obtained based on Eq. (25). It can be written as

$$\hat{s}_k = \begin{cases} \hat{T}'_k & \text{if } |\hat{x}_k| \leqslant \hat{T}'_k, \\ |\hat{x}_k|^{\omega}\hat{T}'^{1-\omega}_k & \text{otherwise,} \end{cases} \tag{30}$$

with

$$\hat{T}'_k = T\left(\frac{\hat{x}_k}{\overline{x}_{00}}\right)^{\alpha_T} \quad \text{and} \quad \hat{x}_k = X_k + \hat{\varepsilon}'_k. \tag{31}$$

$\hat{T}'_k$ is an estimate for Watson's luminance threshold and $\hat{x}_k$ is an estimate for the original coefficient value. The latter can be obtained by adding the received coefficient value to an estimate for the error, $\hat{\varepsilon}'_k$. This error can be computed similarly to (29), using the difference $(X_k - x)$ inside the summation, instead of its absolute value.

After estimating slack values, (26) is used to blindly compute the local perceptual error:

$$\hat{\varepsilon}_{\mathrm{p}_k} = \frac{\hat{\varepsilon}_k}{\hat{s}_k}. \tag{32}$$

To conclude, a global perceptual distortion measure is obtained using (32) in (28).

5. Results

5.1. PSNR scores

In order to evaluate the accuracy of the blind PSNR estimates, all reference images in LIVE database have been subject to JPEG compression, with quality factors in the range from 5 to 90 with

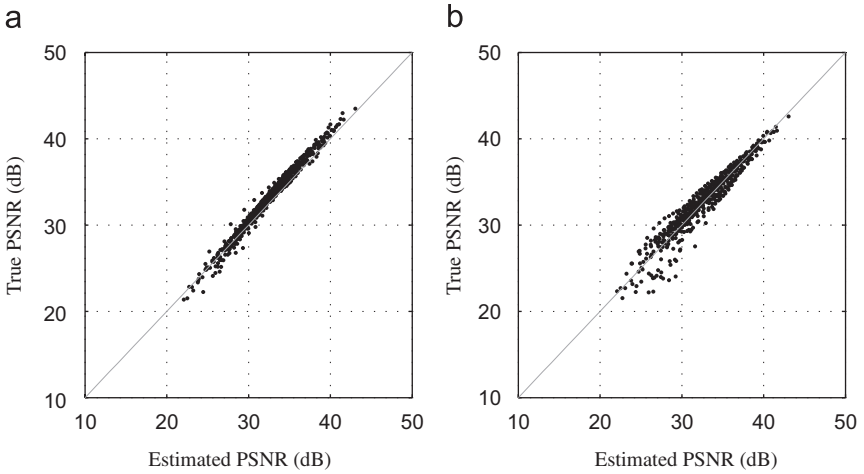

Fig. 5. Global PSNR estimation results. (a) λ prediction. (b) Laplacian compensation.

increments of 5. After image encoding, the true and estimated values for the PSNR have been computed and plotted in Fig. 5. Fig. 5(a) shows the results using our proposed algorithm, while Fig. 5(b) depicts the results achieved by the algorithm in [15]. Table 2 depicts global statistics of the error between true and estimated PSNR, that result from this experiment, for both algorithms. As can be observed from both figures and table, PSNR estimates based on the proposed algorithm for λ estimation are quite accurate, and better than the ones resulting from the *Laplacian compensation* method in [15].

Fig. 6 depicts PSNR estimation results attained for two images, subject to JPEG compression, with quality factors in the range 10–90. These examples represent the best and the worst PSNR estimates for the images used in the experiments.

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 (this means that quality decreases with increasing values of DMOS).

Fig. 7(a) depicts the estimated Watson's distance, using (28) and the perceptual error estimates given by (32), versus the corresponding DMOS values.

Table 2
PSNR estimation results

	λ prediction	Laplacian compensation
Average absolute error (dB)	0.660	1.047
Root mean square error (dB)	0.789	1.324
Correlation coefficient	0.992	0.957

Following a procedure similar to what is suggest by the *video quality experts group* (VQEG) in [26], a logistic function has been used in order to map the values retrieved by the algorithm, into the interval $[0;100]$, where 0 is the best quality. The logistic function has the form:

$$\text{Estimated DMOS} = \theta_0 + \frac{\theta_1}{1+\exp(\theta_2 D_W + \theta_3)}, \tag{33}$$

where θ_0 to θ_3 are parameters to estimate. In our case, they have been computed using the *Levenberg–Marquardt* method seeking the minimization of the square difference between true and estimated DMOS scores. The training set used for this procedure consists of DMOS values associated to the JPEG encoded images in LIVE database (125 out of 175). The remaining 50 DMOS values have been assigned to the validation set. The resulting logistic function can be observed in Fig. 7(a).

Fig. 7(b) depicts the normalized NR quality scores versus their DMOS values. As can be concluded, the resulting objective quality scores using the proposed algorithm are well correlated with the subjective quality scores.

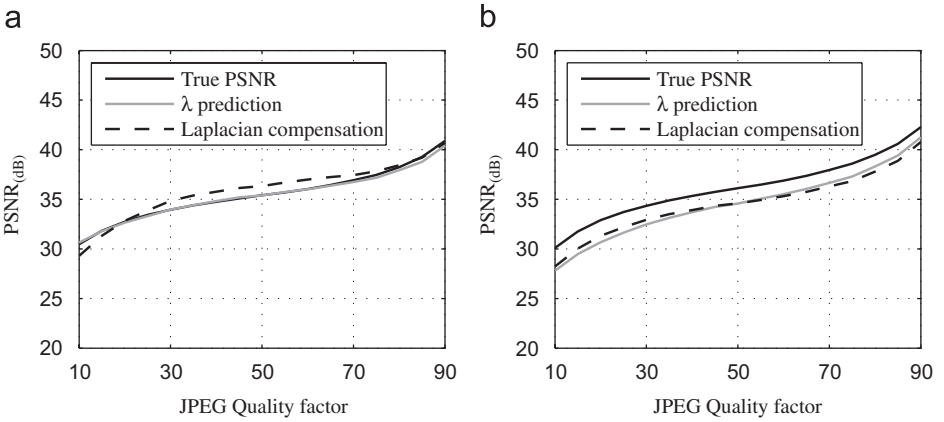

Fig. 6. Best and worst case PSNR estimates. (a) Best case (image *Womanhat*). (b) Worst case (image *Monarch*).

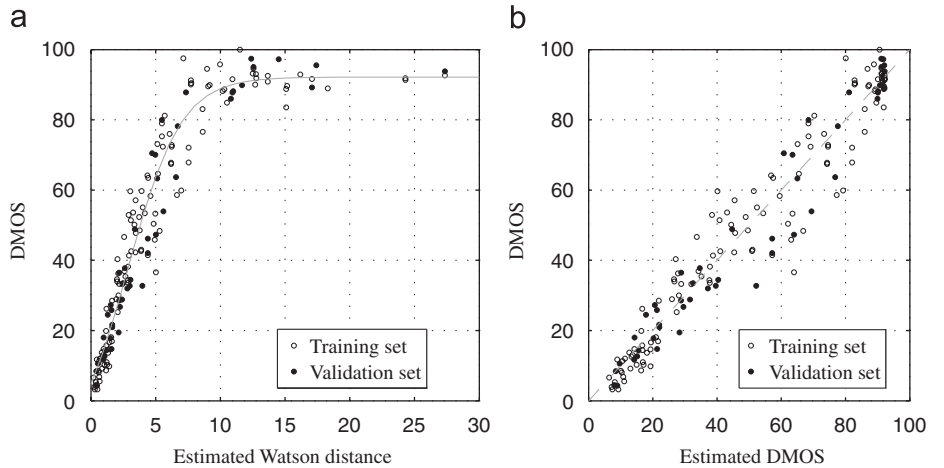

Fig. 7. DMOS estimation results. (a) Watson distance vs. DMOS. (b) DMOS estimation results.

In [26], 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 3, where a comparison with the algorithms proposed in [10] and in [14] is also depicted.

Remember that the algorithm proposed in [10] estimates quality scores based on artifact measurements (more specifically, blurriness and blockiness effects). In order to perform a fair comparison with the algorithm proposed in this paper, it has been implemented following the description given in [10]. As for the algorithm proposed in [14], quality scores result from combining the outputs of neural networks, whose input values are block-based features extracted from the image under evaluation. The comparison with this algorithm has been performed using the results given in the paper, which were also obtained through training and validation image sets taken from LIVE database (with the same sizes as described in this paper). The results depicted in [14] have also been scaled from the range $[-1;1]$ to the range $[0;100]$. The measurements affected by this scaling have been signaled with '*' in Table 3.

Table 3
Evaluation of the proposed blind metric

	Wang et al. [10]	Gastaldo and Zunino [14]	Proposed
Average absolute error	5.814	5.5*	5.661
Root mean square error (RMS)	7.760	7.0*	7.439
Pearson correlation coefficient (CC)	0.970	0.95	0.973
Spearman rank order coefficient (RC)	0.960	N/A	0.978

These results confirm the good performance of the algorithm proposed in this paper. When compared with the performance of [10], the proposed scheme shows better results for all the measurements, with more emphasis on the RC measurement. When compared with the results depicted in [14], it shows slightly worse results for the RMS and the average absolute error measurements, but results for the CC measurement are better. Considering that the ideal target value for CC and RC measurements is 1.00, the improvement on these measurements is noticeable.

6. Conclusions

A new approach for estimating original DCT coefficient distribution parameters from their quantized values has been proposed in this paper. It explores the correlation between coefficients distribution at adjacent DCT frequencies. The resulting distribution estimates are then used for computing a "NR" quality score of images subject to quantization noise. In this paper, two different approaches for quality scoring have been considered: PSNR estimation and perceptual quality estimation, based on a JND model by Watson.

For the blind PSNR estimates, results have shown that the proposed algorithm provides estimates that are more accurate than the ones provided by a state-of-the-art algorithm [15]. The results concerning perceptual quality scores have also been quite satisfactory, showing a high correlation with the human perception of quality. They have also been compared with other NR metrics for evaluating the quality of JPEG encoded images [10,14], generally exhibiting better results.

Further investigation is already undergoing in order to extend the ideas presented in this paper for DCT-based encoded video. In this case, the procedure for PSNR estimation can be used directly in the reference frames (e.g. I-frames in MPEG-2 and H-264/AVC). For perceptual quality estimation, a JND-based perceptual model for video, such as the one presented in [6], is necessary. Since the proposed method allows to compute local error estimates, it can also be potentially used in other applications such as image reconstruction, block-artifact removal or to blindly estimate other image quality metrics.

Acknowledgments

The authors would like to thank *Laboratory of Image and Video Engineering* at the University of Texas at Austin for providing them access to the LIVE database of images.

References

[1] S. Winkler, A perceptual distortion metric for digital color images, in: Proceedings of IEEE International Conference on Image Processing, vol. 3, Chicago, USA, 1998, pp. 399–403.
[2] S. Winkler, A perceptual distortion metric for digital color video, in: Proceedings of SPIE, vol. 3644, S. Jose, USA, 1999, pp. 175–184.
[3] Z. Wang, A. Bovik, A universal image quality index, IEEE Signal Process. Lett. 9 (3) (2002) 81–84.
[4] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[5] E. Ong, X. Yang, W. Lin, Z. Lu, S. Yao, Perceptual quality metric for compressed videos, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 2, Philadelphia, USA, 2005, pp. 581–584.
[6] A.B. Watson, J. Hu, J.F. McGowan, DVQ: a digital video quality metric based on human vision, J. Electron. Imaging 10 (1) (2001) 20–29.
[7] Z. Wang, G. Wu, H. Sheikh, E. Simoncelli, E.-H. Yang, A. Bovik, Quality-aware images, IEEE Trans. Image Process. 15 (6) (2006) 1680–1689.
[8] M. Masry, S. Hemami, Y. Sermadevi, A scalable wavelet-based video distortion metric and applications, IEEE Trans. Circuits Systems Video Technol. 16 (2) (2006) 260–273.
[9] H. Wu, M. Yuen, A generalized block-edge impairment metric for video coding, IEEE Signal Process. Lett. 4 (11) (1997) 317–320.
[10] Z. Wang, H. Sheikh, A. Bovik, No-reference perceptual quality assessment of JPEG compressed images, in: Proceedings of IEEE International Conference on Image Processing, vol. 1, Rochester, USA, 2002, pp. 477–480.
[11] L. Meesters, J.-B. Martens, A single-ended blockiness measure for JPEG-coded images, Signal Processing 82 (3) (2002) 369–387.
[12] F. Pan, X. Lin, S. Rahardja, W. Lin, E. Ong, S. Yao, Z. Lu, X. Yang, A locally adaptive algorithm for measuring blocking artifacts in images and videos, Signal Processing: Image Communication 19 (6) (2004) 499–506.
[13] D.S. Turaga, Y. Chen, J. Caviedes, No-reference PSNR estimation for compressed pictures, Image Commun. (special issue on Objective Video Quality Metrics) 19 (2) (2004) 173–184.
[14] P. Gastaldo, R. Zunino, No-reference quality assessment of JPEG images by using CBP neural networks, in: Proceedings of International Symposium on Circuits and Systems, vol. 5, Vancouver, Canada, 2004, pp. 772–775.
[15] A. Ichigaya, M. Kurozumi, N. Hara, Y. Nishida, E. Nakasu, A method of estimating coding PSNR using quantized DCT coefficients, IEEE Trans. Circuits Systems Video Technol. 16 (2) (2006) 251–259.
[16] B. Girod, What's wrong with mean-square error?, in: A.B. Watson (Ed.), Digital Images and Human Vision, MIT Press, Cambridge, MA, 1993.

[17] A.B. Watson, DCT quantization matrices optimized for individual images, in: Proceedings of IV SPIE Human Vision, Visual Processing, and Digital Display IV, S. Jose, USA, 1993.
[18] E. Lam, J. Goodman, A mathematical analysis of the DCT coefficient distributions for images, IEEE Trans. Image Process. 9 (10) (2000) 1661–1666.
[19] F. Muller, Distribution shape of two-dimensional DCT coefficients of natural images, Electron. Lett. 29 (22) (1993) 1935–1936.
[20] T. Eude, R. Grisel, H. Cherifi, R. Debrie, On the distribution of the DCT coefficients, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 5, Adelaide, Australia, 1994, pp. 365–368.
[21] J.-H. Chang, J.W. Shin, N.S. Kim, S. Mitra, Image probability distribution based on generalized gamma function, IEEE Signal Process. Lett. 12 (4) (2005) 325–328.
[22] R. Duda, P. Hart, D. Stork, Pattern Classification, second ed., Wiley-Interscience, New York, 2000.
[23] J. Price, M. Rabbani, Biased reconstruction for JPEG decoding, IEEE Signal Process. Lett. 6(12).
[24] H. Sheikh, Z. Wang, L. Cormack, A. Bovik, LIVE image quality assessment database, release 2, ⟨http://live.ece.utexas.edu/research/quality⟩, 2006.
[25] ITU-T, Recommendation t.81—information technology, digital compression and coding of continuous-tone till images, requirements and guidelines, JPEG encoding standard.
[26] VQEG, Final report from the video quality experts group on the validation of objective models of video quality assessment, phase ii, available at: ⟨http://www.vqeg.org⟩, August 2003.