Contents lists available at SciVerse ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

A hybrid system for distortion classification and image quality evaluation

Aladine Chetouani [a], Azeddine Beghdadi [a], Mohamed Deriche [b],*

[a] *L2TI, University Paris 13, France*
[b] *EE, KFUPM, Saudi Arabia*

ABSTRACT

Numerous Image Quality Measures (IQMs) have been proposed in the literature with different degrees of success. While some IQMs are more efficient for particular artifacts, they are inefficient for others. The researchers in this field agree that there is no universal IQM which can efficiently estimate image quality across all degradations. In this paper, we overcome this limitation by proposing a new approach based on a degradation classification scheme allowing the selection of the "most appropriate" IQM for each type of degradation. To achieve this, each degradation type is considered here as a particular class and the problem is then formulated as a pattern recognition task. The classification of different degradations is performed using simple Linear Discriminant Analysis (LDA). The proposed system is developed to cover a very large set of possible degradations commonly found in practical applications. The proposed method is evaluated in terms of recognition accuracy of degradation type and overall image quality assessment with excellent results compared to traditional approaches. An improvement of around 15% (in terms of correlation with subjective measures) is achieved across different databases.

© 2012 Published by Elsevier B.V.

1. Introduction

During the last decade, we have witnessed an increasing demand for quality multimedia material. This is essentially due to the development of advanced image/video production technologies. Indeed, the progress achieved in these domains is unprecedented. Despite such a progress, quantifying and reducing image degradation continues to be a challenging problem. A typical example is that of image degradation due to blocking effects in JPEG compression and ringing effect in JPEG2000 compression [1]. These artifacts are among the most limiting factors in compression.

In recent years, substantial research efforts in image quality have led to the development of a number of Image Quality Measures (IQMs) [2,3]. These quality assessment methods are broadly classified into three categories, Full Reference (FR), Reduced-Reference (RR) and No-Reference (NR) metrics. In the first class, both the original image and its distorted version are available. In the case of RR methods, the image quality is estimated using some features extracted from the original and the degraded image. When neither the reference image nor any of its features is known, NR or blind methods are used. It is worth noting that most of the currently known methods, especially the universal ones, use the whole original image (FR methods). Furthermore, most NR techniques are not universal and respond effectively only to one or two types of distortions (e.g. block effects in JPEG and/or image blurring). To test the proposed algorithm across the largest range of possible distortions, we opted to focus on FR methods.

The most common FR measure is the traditional Peak Signal to Noise Ratio (PSNR). Unfortunately, PSNR provides poor results in terms of correlation with subjective measures such as the Mean Opinion Score (MOS) [4].

* Corresponding author.
E-mail address: mderiche@kfupm.edu.sa (M. Deriche).

Some methods, such as the PSNR-HVS, have been proposed to improve the PSNR by taking into account some Human Visual System (HVS) characteristics [5]. Such measures exploit the limitations of the HVS in discerning fine details and textures in an image. This limitation is expressed as a filtering process modeled though a band-pass filter called Contrast Sensitivity Function (CSF) [6]. A more recent version of the PSNR-HVS has been developed by taking into account masking effects in the DCT domain [7,8].

Many others HVS-inspired image quality measures have been proposed in the literature (see Refs. [9,10] for an overview of IQMs). The Visible Differences Predictor, proposed by Daly is one of the earliest and most popular IQMs based on HVS characteristics discussed in the literature [11]. However, the computational complexity of this IQM makes it less attractive than other simpler and more efficient measures. Other image quality measures, such as the SSIM [12], use local structural characteristics, or mutual information concepts [13] to quantify image quality.

Despite all these available IQMs, there is no universal or unique IQM that can predict or measure image quality across all degradations. Indeed, the efficiency of a given IQM may be very high for a given type of degradation but inefficient for others. This is essentially due to the fact that generally Full Reference (FR) IQMs cannot take into account the particular type of visual distortion contained in a given image. Moreover, it is worth noting that images, in real setups, may be subjected, simultaneously, to a multitude of degradations.

In this work, we adopt a new strategy for quantifying image quality. We start from the point of view that image quality is rather a multidimensional problem as already noted in [14,15]. To overcome the limitations of different IQMs, we propose to identity first the degradation type contained in an image then measure the quality of that image, using the most appropriate IQM for that specific degradation. Here, we do not focus on the particular artifacts such as blocking effects or ringing effects, but use a statistical framework that covers a large set of common degradations.

The paper is organized as follows: In Section 2, we first discuss the importance of the a priori knowledge of the degradation type, we then describe the image database used for our experimental setup, and the features used for distortion classification. The experimental results are discussed in Section 3 followed by some concluding remarks in Section 4.

2. The proposed method

2.1. The distortion-IQM correlation

The efficiency of existing IQMs depends highly on the type of degradation contained in a given image. For a given degradation, some IQMs are more adapted to subjective judgments than others. The idea developed here is to exploit this limitation to better estimate the quality of a given image. We propose to detect the type of degradation contained in a given degraded image before quantifying the quality of such an image using an appropriate distortion-based IQM (D-IQM).

Before describing the proposed method, we will discuss the importance of knowing the type of distortion contained in an image through the following simple experiment. In this example, we analyze the performance of some FR-IQMs across three common degradations, namely: JPEG, JPEG2000 lossy compression artifacts, and common blur distortion.

In JPEG compression, blocking effects appear quite significantly at low bit rates. Such annoying artifacts appear generally at block edges as artificial horizontal and vertical contours. This is mainly due to the fact that the blocks are transformed and quantized independently. The effect of such blocking effects depends highly upon the spatial intensity distribution in the image and its frequency content. Moreover, the Human Visual System (HVS) enhances the perceived contrast between adjacent regions. Furthermore, these artificial block transitions are accentuated by the Mach phenomena [16]. An example of an image and its JPEG compressed version exhibiting compression distortion is presented in Fig. 1.

Another type of distortion commonly encountered in real applications is blur. This distortion mostly affects salient features such as contours which correspond to high frequencies components in the image. The lossy compression operation acts as a low pass filter on these components leading to contrast attenuation around region transitions such as object contours (see Fig. 2).

The more recent JPEG 2000 compression standard has been shown to offer better performance than JPEG in many aspects. However, still, other annoying artifacts such as ringing effects, which commonly appear around edges, limit the performance of this compression method at very low bit rates (see Fig. 3). These artefacts result from the decimation and the quantization processes [1]. This phenomenon has traditionally been known as the Gibbs phenomenon.

Fig. 1. JPEG compression distortions: (a) original image, (b) JPEG blocking effect, (c) zoomed version of image (b).
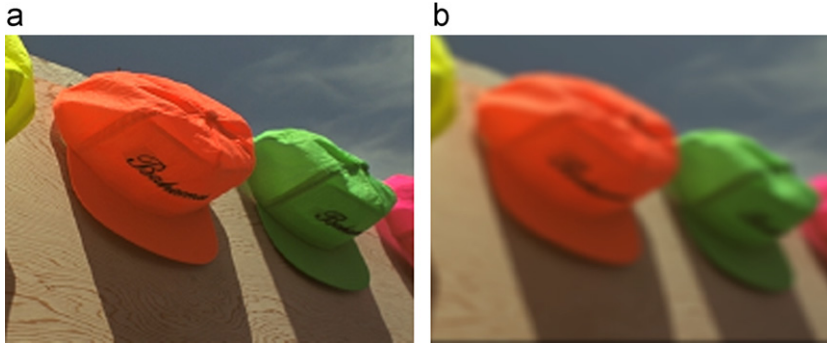
Fig. 2. Blur distortion: (a) original image and (b) its blurred version.

Fig. 3. JPEG2000 compression distortions: (a) original image, (b) ringing and blur effects, (c) zoomed version of (b).

Many of the experiments we carried revealed that for a given image subjected to various degradations with the same level of image quality (as quantified by objective measures), the perceptual appearance of the image can clearly differ from one distortion to another. This observation confirms that the use of a single IQM for quantifying different distortions cannot yield consistent results. Therefore, it would be inappropriate to use the same IQM for all types of distortions. For this reason, we started our experiments by considering different classes of distortion separately. For each type of distortion, we ranked the different IQMs using the Pearson's Correlation Coefficient (PCC) between the IQM indices and the Mean Opinion Score (MOS). For our experiments, we selected the most commonly used IQMs, namely: VIF, VIFP [13], PSNR-HVS (PSNRH) [5], PSNRHVS-M (PSNRM) [8], SSIM [12], UQI [17], IFC [18], WSNR [16], VSNR [20], XYZ [21] and PSNR (these measures will be briefly discussed in Section 2.3).

For the sake of completeness, the expression of the PCC is given below:

$$CORR_{ij} = \frac{\sum_{k=1}^{K}(IQM(k)_{ij}-\overline{IQM_{ij}}).(MOS(k)_i-\overline{MOS_i})}{\sqrt{\sum_{k=1}^{K}(IQM(k)_{ij}-\overline{IQM_{ij}})^2}\cdot\sqrt{\sum_{k=1}^{K}(MOS(k)_i-\overline{MOS_i})^2}} \quad (1)$$

where i and j stand for the ith degradation and the jth IQM, respectively. The index k stands for the kth image, and K is the total number of images considered in the experiment.

In our experiments, we used the Tampere Image Database [22]. We first started by ranking the different FR-IQMs for the 3 considered distortions. Table 1 summarizes the

Table 1
IQM ranking for blur, JPEG and JPEG2000 distortions.

IQM ranking	Degradation type		
	Blur (Class 8)	JPEG (Class 10)	JPEG2000 (Class 11)
1	VIFP	PSNRH	PSNRH
2	VIF	PSNRM	PSNRM
3	WSNR	VIF	NQM
4	VSNR	WSNR	WSNR
5	PSNRM	NQM	VIFP
6	PSNRH	VIFP	VSNR
7	SSIM	VSNR	UQI
8	UQI	SSIM	VIF
9	NQM	PSNR	PSNR
10	IFC	XYZ	SSIM
11	PSNR	UQI	IFC
12	XYZ	IFC	XYZ

obtained results. It is important to note that the ranking changes across the three types of degradations. Indeed, the best IQM for blur is the VIFP, while PSNRH appears as the most appropriate IQM for JPEG and JPEG 2000 compression artifacts.

The estimated PCCs obtained for VIFP and PSNRH under each of the degradations above, are given in Table 2.

The results above prompted us to investigate the effects of different distortions on quantifying image quality in more details. More specifically, we postulate here that one should first identify the distortion type before selecting the most appropriate IQM for quantifying image quality [23]. A relatively similar scenario for no reference

IQM estimation was independently proposed by Moorthy and Bovik in [24].

In summary, the flowchart of the proposed algorithm is presented in Fig. 4. First, some characterizing features are extracted from the original image and its degraded version. Then, after projecting these features onto a new space, the type of degradation is determined by using a simple minimum distance criterion. The new space is obtained using Linear Discriminant Analysis (LDA) projection.

Once the specific type of distortion is identified, the most appropriate IQM will then be used to quantify the quality of the distorted image. This will be discussed in more details.

Table 2
PCC for blur, JPEG and JPEG2000 distortions.

Degradation type	Pearson Correlation Coefficient	
	VIFP	PSNRHVS
Blur	**0.94**	0.91
JPEG	0.91	**0.95**
Ringing	0.94	**0.95**

Next, we briefly present the image database used in our experiments. Then, we introduce the feature extraction process and distortion classification followed by the final IQM estimation stage.

2.2. The Tampere Image Database (TID 2008)

In order to evaluate the performance of the proposed approach, we need a comprehensive database that covers the widest range of possible distortions. A number of image databases are available in the literature for testing IQM algorithms including the LIVE database [25], the IVC database [26], the Cornel database [27], and so on. In this work, we opted to use the TID 2008 image database [22]. This database consists of 17 types of degradations with 100 images per distortion from 25 reference images (i.e. 4 distortion levels per image and per degradation). Fig. 5 shows some reference images taken from the TID 2008 database. Table 3 lists the degradation types available in the database. We note that the database covers a wide range of possible distortions: compression artifacts such as JPEG and JPEG2000, blur, noise, and others. The MOS values for all the observed images are also available for this database.
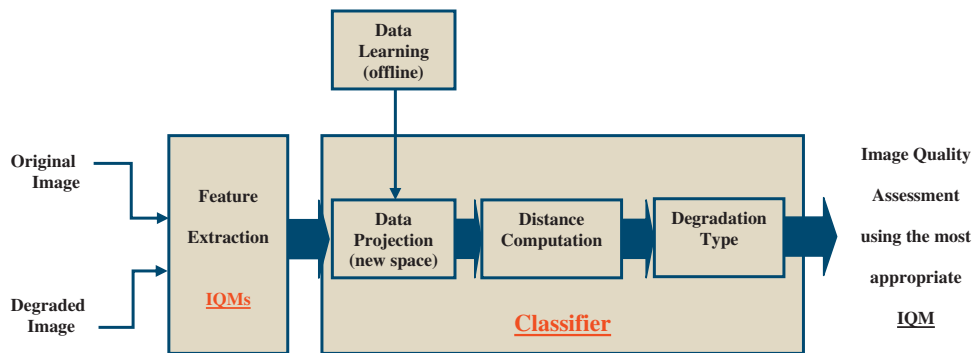

Fig. 4. Flowchart of the proposed system.

Fig. 5. Sample images from the TID2008 database.

Table 3
Types of degradation in the TID 2008 image database.

Degradation	Type
1	Additive Gaussian noise
2	Additive noise in color components
3	Spatially correlated noise
4	Masked noise
5	High frequency noise
6	Impulse noise
7	Quantization noise
8	Gaussian blur
9	Image denoising
10	JPEG compression
11	JPEG2000 compression
12	JPEG transmission errors
13	JPEG2000 transmission errors
14	Non eccentricity pattern noise
15	Local block-wise distortions of different intensity
16	Mean shift (intensity shift)
17	Contrast change

2.3. Features extraction for distortion identification

In order to identify the type of distortion affecting a certain image, some characterizing features need to be extracted from both the original image and its degraded version. Numerous features can be used for this purpose such as contrast, homogeneity and textured descriptors. Here, we propose to use directly the different IQMs as features. We consider that for a given degradation type, each FR-IQM exhibits a specific response. In other words, the FR-IQMs are used to discriminate between different degradations.

In the following, and for the sake of completeness, we describe the IQMs cited previously and considered here as features to be associated with the appropriate distortions. Note that in this study, we assume that we have access to both the original and the degraded images (i.e. Full Reference Measures are thus considered). We selected a range of IQMs commonly used in this research area. While some of the selected IQMs are HVS-based (such as the PSNRHVS), others are based on pixel-wise differences (MSE-based metrics), or local structural information such as the SSIM.

2.3.1. Peak Signal to Noise Rate (PSNR)

The Peak Signal to Noise Rate (*PSNR*) is the most commonly used measure for quantifying signal distortion [4]. In the absence of a well-accepted IQM, it is still in use in many applications and especially for real-time performance evaluation of some video and image processing systems.

2.3.2. Universal Image Quality Index

The Universal image Quality Index (*UQI*) [17] is based on a local analysis of the image. This measure represents structural information obtained by extracting some local statistical parameters from non-overlapping blocks.

This IQM has been shown to adapt well for measuring artifacts due to blur. Indeed, blur tends to decrease the gray-levels scatter around the mean of pixels of a given block in an image. However, this measure is unstable over homogeneous regions where the standard deviation is close to zero. Furthermore, it has recently been shown that this measure is directly related to the conventional, and often unreliable, mean squared error [28].

2.3.3. Structural SIMilarity and its multi-scale version

Another improved version of the UQI, called the Structural SIMilarity image quality index (*SSIM*), is a measure based on local structural information [39]. The SSIM measure is a function of three terms: luminance (*l*), contrast (*c*), and structure (*s*) factors. In [12], a multi-scale version of the *SSIM* was proposed. The same features (*l*, *c* and *s*) are extracted. However, *l* is derived from the last level of decomposition, whereas, *c* and *s* are computed at each level of decomposition. The overall image quality index is finally obtained by multiplying *l*, the sum of *c*, and the sum of *s*. Similarly to the UQI, blur degradation is well quantified by this measure. Note that both the SSIM and the MSSIM can be seen as improved versions of the UQI [17].

2.3.4. PSNRHVS and PSNRHVS-M

In [5], a perceptually motivated PSNR called *PSNR-HVS* index quality was proposed. The main rationale was to improve the performance of the PNSR by integrating some characteristics of the HVS. To achieve this, the authors incorporated the *CSF* model in the DCT domain.

A more recent version, named *PSNRHVS-M*, has also been proposed in [8] where a masking model was incorporated. It is worth noting that blocking effect can be well estimated using these measures for DCT-based compressed images. This is mainly due to the fact that these measures themselves are based on a block analysis and the DCT transform.

2.3.5. Weighted Signal to Noise Ratio (WSNR)

Contrary to the previous measures, the *WSNR* index is based on a frequency domain analysis [19]. The measure is basically a perceptually weighted signal-to-noise ratio. It is expressed as the ratio of the CSF-weighted Fourier spectrum of the original image over that of the distorted image.

2.3.6. Information Fidelity Criterion (IFC)

In [18], an IQM, based on some concepts from information theory, was proposed. This measure, called Information Fidelity Criterion (*IFC*) is computed using a source (*C*) and a distortion (*D*) model for some selected subbands in the wavelet domain. The wavelet coefficients of the different subbands are modeled using a Gaussian Mixture model.

2.3.7. Visual Information Fidelity (VIF)

An extension of the IFC measure, called Visual Information Fidelity (*VIF*), was proposed in [13]. The main difference is the incorporation of some characteristics of the HVS. Another version of VIF in spatial domain, called *VIFP*, was also proposed.

2.4. Distortion classification using Linear Discriminant Analysis (LDA)

To identify the type of distortion affecting a given image, we propose here to use a simple classifier based on Linear Discriminant Analysis (LDA). Under this framework, we consider the different IQMs as features extracted from the test image [29]. Each degradation is considered as a class among a set of M classes (here $M=17$). The IQMs estimated from a given image are concatenated as feature vectors of dimension n (here $n=12$).

Instead of dealing with all the extracted IQMs individually, and since many IQMs may exhibit large correlations as noted in [24], we propose to project these "feature" vectors onto an orthogonal space. The concept of data projection is not new and has been used in mathematics quite frequently. Linear subspace projection, in particular, has been used in numerous signal processing applications. Three popular approaches generally used under this class are: Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Independent Component Analysis. The PCA's basic concept is to find a set of the most representative projections such that the projected samples retain most information about the original samples and account for the most variance in the data. LDA, on the other hand, uses class information and finds as set of projection vectors that maximize the between-class scatter while minimizing the within-class scatter. Finally, ICA captures higher order statistics and projects data onto the basis vectors that are as statistically independent as possible.

While there were numerous attempts to compare the three approaches (ICA, PC, LDA), there has never been a consent of which one of the three performs best. One of the earliest attempts was discussed by Martinez and Kak [31] in which they say that even though LDA is seen a more appropriate approach for classification, PCA may outperform LDA when the number samples/class is small or when training data non-uniformly sample the underlying distribution. Moghaddam [33], on the other hand, showed that there was no significant difference between PCA and LDA. Eleyan and Demirel [34] carried an extensive analysis in which they analyze PCA and LDA both as classifiers and as preprocessing stages for a Neural Network (NN) classifier. They showed that LDA and NN-LDA consistently outperform PCA and NN-PCA for a medium to large number of images per class. Given that, in our application, we have a reasonable number of images for each class (distortion type), we opted to use the LDA approach.

As introduced above, Linear Discriminant Analysis (LDA) is a popular method for dimensionality reduction and classification that projects high-dimensional data onto a low dimensional space where the data achieves maximum class separability. The projection matrix or transformation is obtained by minimizing the within-class variability and maximizing the between-class distance simultaneously, hence achieving maximum class discrimination. It has been used successfully in many applications including face recognition, microarray gene expression data analysis [36–38]. The optimal transformation is readily computed by solving a generalized eigenvalue problem.

The original LDA formulation, known as the Fisher Linear Discriminant Analysis dealt with binary-class classifications. The key idea was to find a projection direction that separates the class means efficiently (when projected onto that direction) while achieving a small variance around these means. Discriminant Analysis is generally used to find a subspace with $M-1$ dimensions for multi-class problems, where M is the number of classes in the training dataset.

Contrary to Principal Component Analysis (PCA) which considers each observation vector as a class on its own, LDA achieves dimensionality reduction while preserving as much of the class discriminatory information as possible [30,31] and takes into account the fact that several observations may come from the same class. Linear Discriminant Analysis searches for those vectors in the underlying space that best discriminate between classes (rather than those that best describe the data as in PCA).

Mathematically speaking, for all the samples of all classes we define two measures: (i) one called within-class scatter matrix, as given by

$$S_w = \sum_{J=1}^{M} \sum_{i=1}^{N_i} (\boldsymbol{x}_i^j - \boldsymbol{\mu}_j)(\boldsymbol{x}_i^j - \boldsymbol{\mu}_j)^T \tag{11}$$

where $\boldsymbol{x}_i^j$ (dimension $n \times 1$) is the ith sample vector of class j, $\boldsymbol{\mu}_j$ is the mean of class j, M is the number of classes, and N_i is the number of samples in class j. The second measure (ii) is called the between-class scatter matrix defined as

$$S_b = \sum_{j=1}^{M} (\boldsymbol{\mu}_j - \boldsymbol{\mu})(\boldsymbol{\mu}_j - \boldsymbol{\mu})^T \tag{12}$$

where $\boldsymbol{\mu}$ is mean vector of all classes.

The goal is to find a linear transformation expressed through the matrix $\boldsymbol{W}$, that maximizes the between-class measure while minimizing the within-class measure. One way to do this is to maximize the ratio $\det(\boldsymbol{S}_b)/\det(\boldsymbol{S}_w)$. The advantage of using this ratio is that it has been proven that if $\boldsymbol{S}_w$ is a non-singular matrix then this ratio is maximized when the column vectors of the projection matrix, $\boldsymbol{W}$, are the eigenvectors of $\boldsymbol{S}_w^{-1}.\boldsymbol{S}_b$. It should be noted that: (i) there are at most $M-1$ nonzero generalized eigenvectors, and so an upper bound on reduced dimension is $M-1$, and (ii) we require at least n(size of original feature vectors)$+M$ samples to guarantee that S_w does not become singular.

Note however that the dimension in the projection space does not have to be $M-1$ as the number of important eigenvalues (in the energy sense) may be much smaller than $M-1$.

3. Experimental results

To evaluate the performance of the proposed method for degradation classification and quality evaluation, several experiments were carried covering over 400 natural images (different from those used during the learning stage). The experimental procedure is quite simple and requires the original and distorted images. Fig. 6 shows the Global Image Quality Assessment System for a JPEG
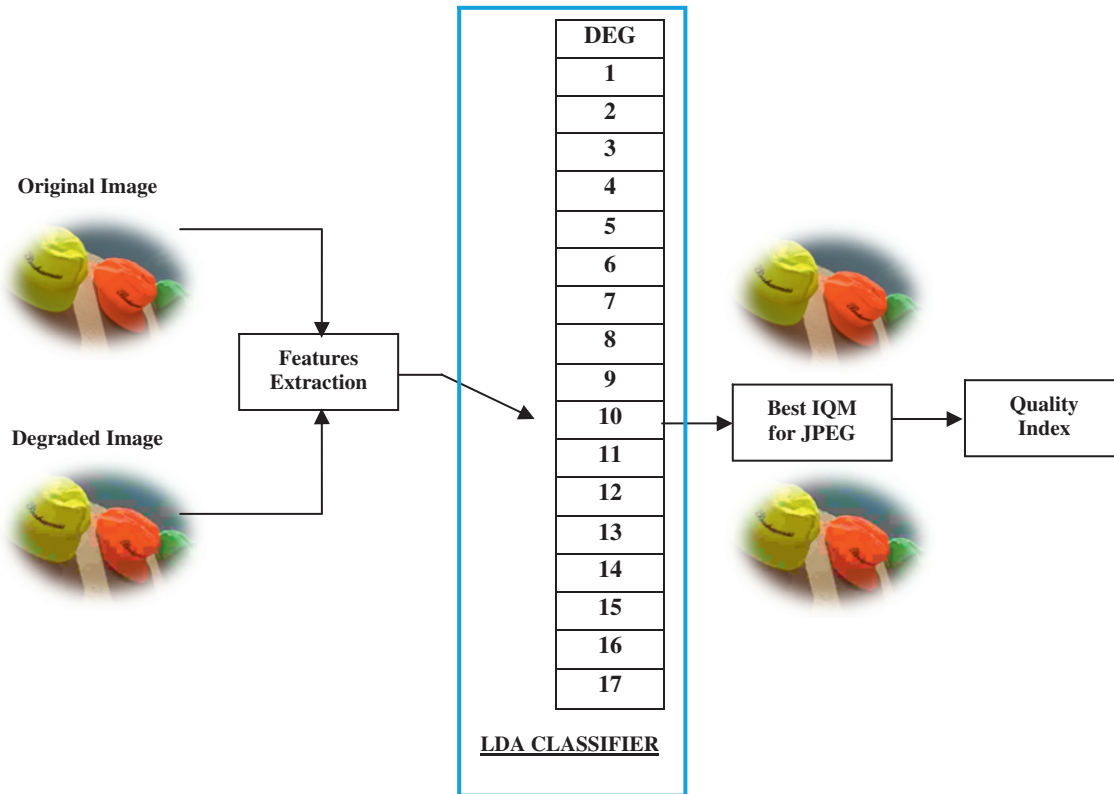
Fig. 6. Example of the global Image Quality Assessment System for JPEG compression degradation.

compressed image. We list below the main steps of the algorithm:

1. All features vectors (of dimension 12×1) are extracted from the different images in the training set.
2. The features vectors from the training stage are used to find the LDA transformation.
3. Using the transformation from (2), the feature vectors in step 1 are projected into the new subspace. These vectors will be called Reduced Feature Vectors from Training (RFVTR vectors).
4. For a given pair of images (original and its degraded version), the 12×1 feature vector is obtained. This vector is then projected onto the new space to get the Reduced Feature Vector from TeSting (RFVTS vector).
5. Using the Euclidean or Mahalanobis distance, find the RFVTR vector that is close to the RFVTS vector.
6. The class to which belongs the selected RFVTR vector is declared as the unknown degradation type corresponding to the test image.

For the sake of simplicity, the Euclidian distance was used in our implementation. The results using the Euclidian and the Mahanalobis distances were very comparable. Note that some of the references comparing LDA, PCA, and ICA did discuss the issue of most appropriate metrics for classification and showed that L_1, L_2, and Cosine distances all provide very comparable results [31].

We first evaluated the performance of our method in terms of accuracy in identifying the type of degradation contained in different test images. Fig. 7 displays the classification accuracy for each type of degradation. Note that for all types of distortions, a classification accuracy of more than 90%, is achieved with an average classification accuracy of **98.11%**.

To better visualize our results, the confusion matrix for different classes was computed and is shown in Fig. 8. Note that the lowest performance was obtained for classes 9, 12 and 13 corresponding to 92% classification accuracy.

We noticed from the experimental results that misclassification occurred only for some types of degradations with similar visual appearance or when the image contains a mixture of distortions such as blur and ringing (this may occur in JPEG2000 compression). An example of such confusion is illustrated in Fig. 9, denoising (degradation number 9) is identified as JPEG and JPEG2000 compression distortions.

To further analyze the performance of the proposed system, we used some images from the IVC database (see Fig. 10). With this database, we achieved an overall classification accuracy of 81.05% for JPEG2000 and JPEG distortions. Furthermore, we also tested our algorithm using selected images from the LIVE image database. The percentages in classification accuracy obtained for these types of degradations were 82% and 76%, respectively.

Finally, we tested our system with distortions of unknown nature. For this purpose, some LAR (Locally Adaptive Resolution) [32] compressed images were used
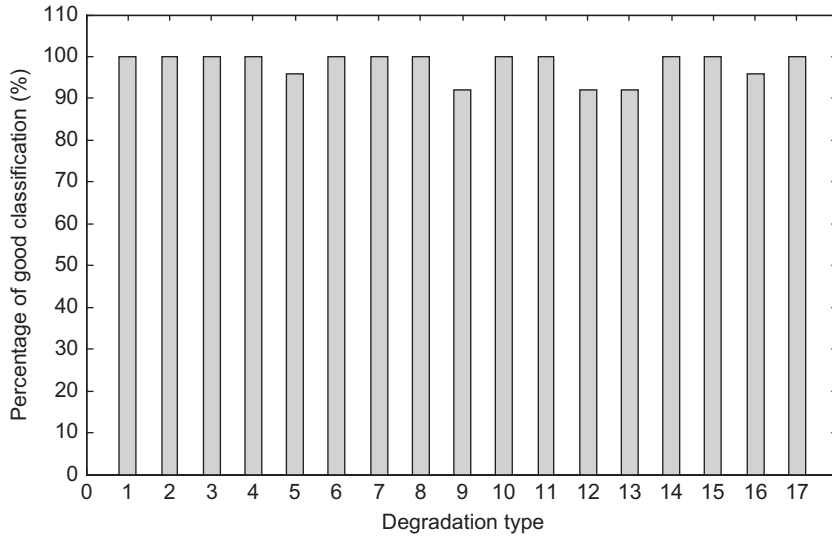
Fig. 7. Degradation classification accuracy.

ESTIMATED CLASS

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
T	**1**	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	**2**	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U	**3**	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	**4**	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	4	96	0	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0
C	**8**	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0
L	**9**	0	0	0	0	0	0	0	0	92	4	4	0	0	0	0	0	0
A	**10**	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0
S	**11**	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0
S	**12**	0	0	0	0	0	0	0	0	0	0	0	92	8	0	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	8	92	0	0	0	0
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100		0
	16	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	96	0
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100

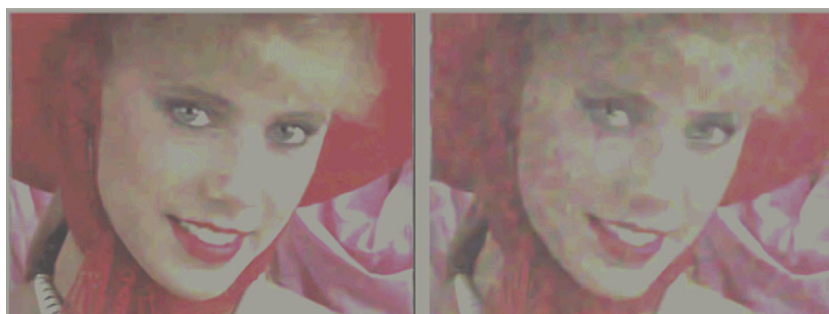**Fig. 8.** Confusion matrix for degradation classification (values are expressed in %).

Fig. 9. Classification errors: confusion between degradations 9 and 11 (left), and degradations 9 and 10 (right).

Fig. 10. A sample of reference images taken from the IVC image database.
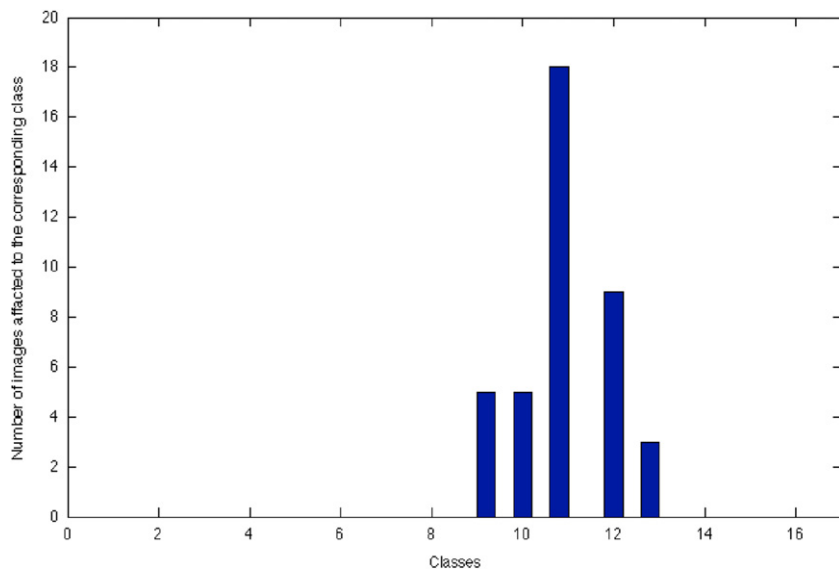

Fig. 11. LAR compressed images.

Fig. 12. Classification results for LAR compressed images.

Table 4
Criteria commonly used to compare IQMs.

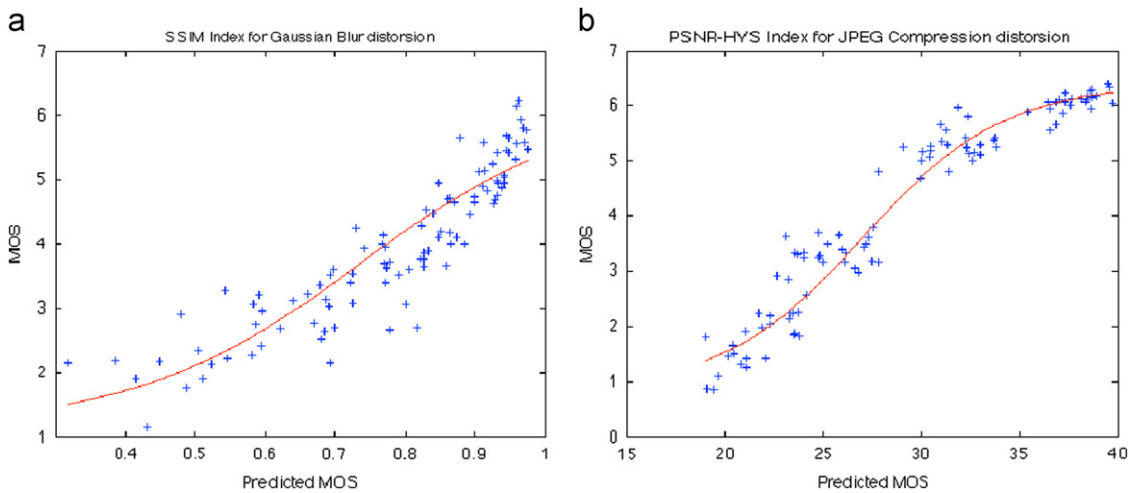| Criterion	Computation	Description
Logistic function	$MOSp = \dfrac{A}{(1+\exp^{(G*Index\ Quality - Dm)})} + B$	Where A, G, Dm and B are adjusted during the fitting process
RMSE	$RMSE = \sqrt{\dfrac{1}{N}\sum_{i=1}^{N} Qerror[i]}$	With $Qerror[i] = MOS[i] - MOSp[i]$ and N is the number of images
Outlier ratio	$OR = \dfrac{(number\ of\ outliers)}{N}$	With $number\ of\ outliers = \lvert Qerror[i] \rvert > 2*\sigma error_{MOS}[i]$

Fig. 13. MOS vs. predicted MOS for Gaussian blur: (a) SSIM index, (b) VIFP index (Best IQM for this degradation).
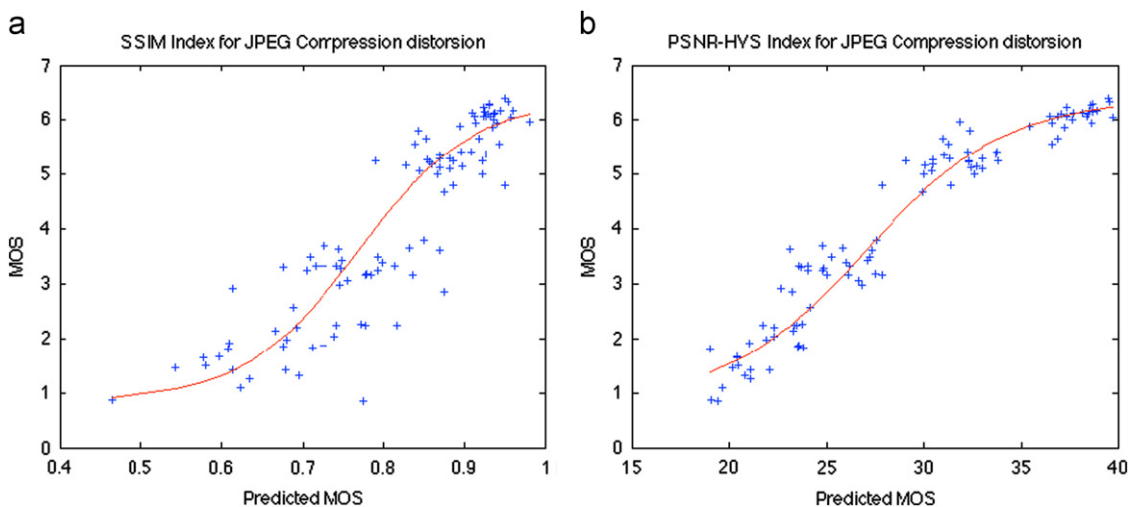
Fig. 14. MOS vs. predicted MOS for JPEG compression: (a) SSIM index, (b) PSNR-HVS index (Best IQM for this degradation).

(see Fig. 11). The distortions generated using LAR are generally very similar to common conventional blurring effect and blocking artifacts.

With the 40 distorted images used in the test, most were classified as JPEG2000 compressed images (class 11), see Fig. 12.

The proposed method has been successfully evaluated on a large number of images taken from a range of applications. As explained earlier, once the degradation type is identified, the quality of a given image can be better measured using the most appropriate IQM (Section 2). Table 7 summarizes the results obtained for 17 types of degradations. The best IQM for each of the considered degradations (using only the selected IQMs) and its corresponding PCC are given. Here, we limited the study to a selected set of IQMs but other objective image quality measures can be considered. Note that, the consistency of these IQMs with subjective evaluation can also be analyzed using some objective functions.

Indeed, in [35], numerous criteria have been used including curve fitting using the logistic function and the RMSE between the MOS and the predicted MOS (MOSp), to mention few. Some of these measures are listed in Table 4. Figs. 13 and 14 show that the distribution of the SSIM measure exhibits a more scattered behavior than the best metrics (VIFP for Gaussian blur, and PSNR-HVS for JPEG compression degradations). These results are further confirmed in Tables 5 and 6. Across all correlation coefficients,

the best selected IQM for the given degradation type consistently outperforms the SSIM measure.

In this study, the best metric for each degradation type is selected according to the PCC, computed using the logistic function. This criterion is generally used to evaluate the performance of IQMs. Table 7 provides a comprehensive summary showing that the proposed degradation-based IQM consistently outperforms one of the most commonly used IQM, namely the SSIM. For example, for class 11 (JPEG2000 compression), while the SSIM achieves only 0.84 in correlation (with the MOS), using PSNRHSV, we reach a PCC value of 0.96. The results show that using the "optimal" IQM for each type of degradation leads to an "optimal" quality assessment measure. The average correlation gain is around **14%** for the considered distortions. To consider the variations in types of distortions, we also carried a number of experiments by varying the number of considered distortions between 8 and 17. The overall improvement in correlation was consistently above **10%.**

To further analyze the performance of the proposed approach, we also tested it on a set of selected images from the LIVE image database. Table 8 summarizes our results. An improvement in correlation was achieved across 5 typical degradation types. The best result was obtained for Gaussian blur.

All of the results across different databases and different sets of distortions showed a consistent improvement in terms of PCC when the proposed algorithm is used. We display in Fig. 15 our overall hybrid system for

Table 5
Correlation coefficients obtained for Gaussian blur degration.

Metric	Pearson correlation	Spearman correlation	Kendall correlation
SSIM	0.90	0.94	0.78
VIFP	0.94	0.94	0.79

Table 6
Correlation coefficients obtained for JPEG compression degration.

Metric	Pearson correlation	Spearman correlation	Kendall correlation
SSIM	0.90	0.90	0.71
PSNR-HVS	0.97	0.96	0.82

Table 8
LIVE image database: best IQM for each considered degradation using Pearson Correlation Coefficient (PCC).

Degradation type	PCC for SSIM	Gain (%)
1: JPEG2000 compression	0.91	2
2: JPEG compression	0.84	10
3: White noise	0.95	3
4: Gaussian blur	0.84	14
5: Fast fading	0.90	0

Table 7
TID 2008 image database: best IQM for each considered degradation using Pearson Correlation Coefficient (PCC).

| Degradation type	PCC for SSIM	PCC for Best IQM	Gain (%)
1: Additive Gaussian noise	0.78	PSNRHVS (0.94)	21
2: Additive noise in color components	0.80	PSNR (0.92)	15
3: Spatially correlated noise	0.80	PSNR (0.95)	19
4: Masked noise	0.81	VIF (0.89)	10
5: High frequency noise	0.86	PSNRHVS (0.97)	13
6: Impulse noise	0.73	PSNR (0.90)	24
7: Quantization noise	0.78	PSNR (0.89)	15
8: Gaussian blur	0.90	VIFP (0.94)	5
9: Image denoising	0.89	PSNRHVS-M (0.96)	8
10: JPEG compression	0.90	PSNRHVS (0.97)	8
11: JPEG2000 compression	0.84	PSNRHVS (0.96)	15
12: JPEG transmission errors	0.81	VIF (0.87)	8
13: JPEG2000 transmission errors	0.81	PSNRHVS (0.92)	14
14: Non eccentricity pattern noise	0.65	IFC (0.84)	30
15: Local blockwise distortions of different intensity	0.89	SSIM (0.89)	0
16: Mean shift (intensity shift)	0.72	WSNR (0.73)	2
17: Contrast change	0.67	VIF (0.88)	32
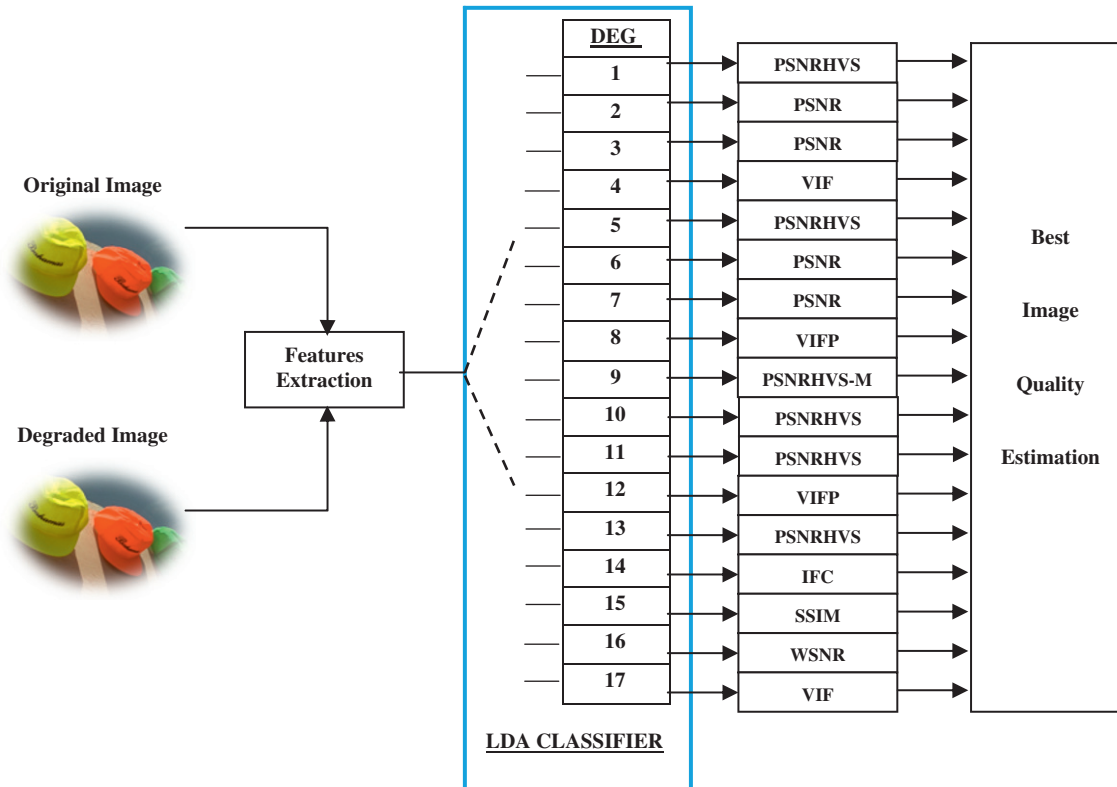
Fig. 15. Proposed global IQM system.

degradation classification and image quality evaluation. For a given degraded image, we get at the output: the degradation type, the best corresponding IQM, and its value. To the best knowledge of the authors, the proposed system presented here is the only one able to provide such a comprehensive description of image distortion identification and estimation.

4. Conclusions

We propose in this paper a new framework for estimating and predicting image quality. In particular, we present an LDA-based technique for classifying degradations before estimating image quality. The classification stage uses the different IQMs estimated from the given image (original and its degraded version) as features. Our experimental results show that the type of degradation can be estimated with more than 90% accuracy. Such knowledge is crucial in determining the types of IQMs that need to be used for evaluating quality. The proposed system not only evaluates quality but also identifies the type of distortion the test image was subjected to.

In future works, we will complete the system by taking into account additionally IQMs and by introducing a feature selection method for choosing the more appropriate features (IQMs) for the classification step. Other degradation types will be also considered, and particularly color degradations, such as color bleeding and false color artifacts.

References

[1] D. Taubman, M. Marcellin, JPEG2000: Image Compression Fundamentals, Standards and Practice, Kluwer Academic Publishers, Boston, 2001.
[2] L.J. Karam, T. Ebrahimi, S. Hemami, T. Pappas, R. Safranek, Z. Wang, A.B. Watson, Introduction to the special issue on visual media quality assessment, IEEE Journal on Special Topics in Signal Processing, Special Issue on Visual Media Quality Assessment 3 (2) (2009) 189–192.
[3] W. Lin, C.-C. Jay Kuo, Perceptual visual quality metrics: a survey, Journal of Visual Communication and Image Representation 22 (4) (2011) 297–312.
[4] Z. Wang, A.C. Bovik, Mean squared error: love it or leave it?—A new look at signal fidelity measures, IEEE Signal Processing Magazine 26 (2009) 98–117.
[5] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, M. Carli, New full-reference quality metrics based on HVS, in: Proceedings of the Second International Workshop on Video Processing and Quality Metrics, USA, 2006.
[6] A.B. Watson, H.B. Barlow, J.G. Robson, What does the eye see best? Nature 302 (1983) 419–422.
[7] G. Legge, J. Foley, Contrast masking in human vision, Journal of Optical Society of America 70 (1980) 1458–4471.
[8] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, V. Lukin, On between-coefficient contrast masking of DCT basis functions, in: Proceedings of the International Workshop on Video Processing and Quality Metrics, 2007.
[9] A.B. Watson, Digital Images and Human Vision, MIT Press, 1993.

[10] M. Pedersen, J.Y. Hardeberg, Survey of Full-Reference Image Quality Metrics, Høgskolen i Gjøviks rapportserie ISSN: 1890-19520X, 2009.

[11] S. Daly, The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity, Digital Images and Human Vision, MIT Press, 1993 (Chapter14, pp. 179–206).

[12] Z. Wang, E.P. Simoncelli, A.C. Bovik, Multi-scale structural similarity for image quality assessment, in: Proceedings of the Asilomar Conference on Signals, Systems and Computers, vol. 2, 2003, pp. 1398–1402.

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

[14] J. Ahumada, J. Albert, C.H. Null, Image quality: a multidimensional problem, digital images and human vision, MIT Press, 1993 (pp. 141–148).

[15] J.B. Martens, Multidimensional modeling of image quality, in Proceedings of the IEEE 2002, vol. 90, pp. 133–153.

[16] T.N. Cornsweet, Visual Perception, Academic Press, NY, 1970.

[17] Z. Wang, A.C. Bovik, A universal image quality index, IEEE Signal Processing Letters 9 (2002) 81–84.

[18] H.R. Sheikh, A.C. Bovik, G. de Veciana, An information fidelity criterion for image quality assessment using natural scene statistics, IEEE Transactions on Image Processing 14 (2005) 2117–2128.

[19] T. Mitsa, K. Varkur, Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms, in; Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, 1993, pp. 301–304.

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

[21] B.W. Kolpatzik, C.A. Bouman, Optimized universal color palette design for error diffusion, Journal Electronic Imaging 4 (1995) 131–143.

[22] N. Ponomarenko, M. Carli, V. Lukin, K. Egiazarian, J. Astola, F. Battisti, Color image database for evaluation of image quality metrics, in: Proceedings of the International Workshop on Multimedia Signal Processing, Australia, 2008, pp. 403–408.

[23] A. Chetouani, A. Beghdadid, M. Deriche, Classification of image degradation using multiple image quality metrics and linear discriminant analysis, in: Proceedings of the European Signal Processing Conference, 2010, pp. 319–322.

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

[25] H.R. Sheikh, Z. Wang, L. Cormack, A.C. Bovik, Live Image Quality Assessment Database, ⟨http://live.ece.utexas.eduesearch/quality⟩, 2006.

[26] P. Le Callet, F. Autrusseau, Subjective Quality Assessment IRCCyN/IVC Database, ⟨http://www.irccyn.ec-nantes.fr/ivcdb/⟩, 2005.

[27] D. Chandler, S. Hemami, Subjective Image database, ⟨http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html⟩, 2007.

[28] R. Dosselman, X.D. Yang, A comprehensive assessment of the structural similarity index, Signal, Image and Video Processing 5 (1) (2011) 81–91.

[29] G.J. Mclachlan, E. Kuh, R.E. Welsch, Discriminant Analysis and Statistical Pattern Recognition, Wiley Series in Probability and Statistics, 2004.

[30] N. Peter, J. Belhumeur, P. Hespanha, D. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.

[31] A.M. Martinez, A.C. Kak, PCA versus LDA, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2001) 228–233.

[32] O. Deforges, J. Ronsin, Locally adaptive resolution method for progressive still image coding, IEEE International Symposium on Signal Processing and Its Applications 2 (1999) 825–829.

[33] B. Moghaddam, Principal manifolds and probabilistic subspaces for visual recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (6) (2002) 780–788.

[34] A. Eleyan, H. Demirel, PCA and LDA Based Neural Networks for Human Face Recognition, Face Recognition, Kresimir Delac and Mislav Grgic (Eds.), 978-3-902613-03-5, InTech, 2007.

[35] N. Seshadrinathan, R. Soundararajan, A. Bovik, L. Cormack, Study of subjective and objective quality assessment of video, IEEE Transactions on Image Processing 19 (N6) (2010).

[36] H. Yu, J. Yang, A direct LDA algorithm for high-dimensional data with application to face recognition, Pattern Recognition 34 (10) (2001) 2067–2069.

[37] F. Yue, K. Wang, W. Zuo. Informative gene selection and tumor classification by null space LDA for microarray data, in: Proceedings of ESCAPE'07, Lecture Notes in Computer Science, vol. 4614, 2007, pp. 435–446.

[38] K.K. Paliwal, A. Sharma, Improved direct LDA and its application to DNA microarray gene expression data, Pattern Recognition Letters 31 (2010) 2489–2492.

[39] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing 13 (4) (2004) 600–612.