A REDUCED REFERENCE IMAGE QUALITY METRIC BASED ON FEATURE FUSION AND NEURAL NETWORKS

Aladine Chetouani[1], Azeddine Beghdadi[1], Mohamed Deriche[2], and Abdesselam Bouzerdoum[3]

[1] L2TI, Institut Galilée, Université Paris13
99, av. J-B Clément, 93430, Villetaneuse, France
[2] EE. Dept., King Fahd University
P.O. BOX 1427, KFUPM, Dhahran, 31261, Saudi Arabia
[3] School of Electrical, Computer & Telecommunication Engineering,
University of Wollongong, Wollongong, NSW 2522, Australia
email: aladine.chetouani@univ-paris13.fr, azeddine.beghdadi@univ-paris13.fr,
mderiche@kfupm.edu.sa, a.bouzerdoum@uow.edu.au

ABSTRACT

A Global Reduced Reference Image Quality Metric (IQM) based on feature fusion using neural networks is proposed. The main idea is the introduction of a Reduced Reference degradation-dependent IQM (RRIQM/D) across a set of common distortions. The first stage consists of extracting a set of features from the wavelet-based edge map. Such features are then used to identify the type of degradation using Linear Discriminant Analysis (LDA). The second stage consists of fusing the extracted features into a single measure using Artificial Neural Networks (ANN). The result is a degradation-dependent IQM measure called the RRIQM/D. The performance of the proposed method is evaluated using the TID 2008 database and compared to some existing IQMs. The experimental results obtained using the proposed method demonstrate an improved performance even when compared to some Full Reference IQMs.

1. INTRODUCTION

In practical applications, images undergo different types of processing including acquisition, transmission or compression, which often generate some annoying distortions. The most annoying impairments are blocking, ringing, and blur artifacts. Blocking manifests as artificial horizontal and vertical discontinuities in some block-based compression methods. Blur is also a common artifact which affects the details of the image due to several phenomena such as defocusing or filtering. Ringing is another annoying degradation which is essentially due to quantization, and is generally defined as noise around edge points or in contrasted transitions.

To quantify the visual impact of these annoying degradations, a number of subjective and objective measures have been proposed [1]. Subjective evaluation is regarded as the most reliable approach for assessing image quality. Unfortunately, subjective methods are complex, time consuming and impractical for real-time applications. Understanding and applying the knowledge of human visual perception is recognized as the most promising approach for developing objective methods consistent with human judgement. Three categories of image quality assessment are commonly used: Full-Reference (FR), No-Reference (NR) and Reduced Reference (RR) methods. Full Reference methods need both the original image and its degraded version. Most of the existing Image Quality Metrics (IQM), such as PSNR, SSIM [2], SNRWAV [3], VIF [4] and so on, belong to this family. However, in real applications the original image is not always available. Hence, NR methods are most appropriate as they require only the degraded image. During the last decade, a number of NR metrics have been proposed [5]-[8]. However, NR-IQMs are limited by the type of the degradation contained in the image. Indeed, NR-IQMs are developed for particular artifacts, hence limiting their use to certain applications only.

Reduced Reference approaches present a good compromise between FR and NR approaches. Only some features, such as edges or some visual descriptors, are extracted from the original and the degraded images. From the structural information conveyed by these descriptors, a metric is then derived and used as an IQM. There are relatively few successful RR-IQMs that have been discussed in the literature, including the popular RRIQA [9]-[10].

In this paper, we focus our work on RR-IQMs approaches. In particular, we propose a new metric based on a multiscale feature fusion scheme using an Artificial Neural Network (ANN) model. The main idea developed here is to compute an RR-IQM which is degradation-dependent (RR-IQM/D). The type of the degradation contained in the image is initially detected using simple Linear Discriminant Analysis, (LDA) based on the extracted features from the edge map of the image and its distorted version.

The remainder of the paper is organized as follows. Section 2 discusses the feature extraction and distortion classification stages, and presents the proposed neural network fusion method. The experimental results are presented in Section 3, followed by concluding remarks and some perspectives in Section 4.

2. THE PROPOSED APPROACH

The flowchart of the proposed method is presented in Fig.1. For a given type of degradation, a number of features are extracted from the original image and its degraded version. These features are then combined using a neural network scheme in order to estimate a local image quality index, depending on the degradation type. This degradation-dependent IQM is called RR-IQM/D. The type of the distortion contained in the degraded image is determined through an initial classification step.

In what follows, the features extraction process, the fusion and classification steps are discussed in more details.

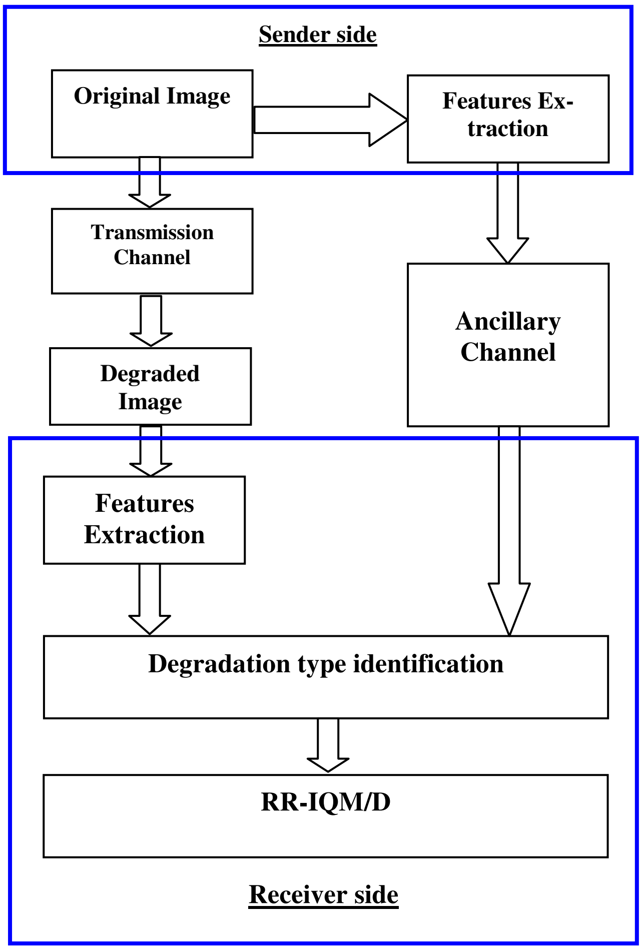

Figure 1 – Flowchart of the proposed RR-IQM/D.

2.1 Feature Extraction Stage

The features are extracted from the edge map image derived from the wavelet domain as illustrated in Fig.2.

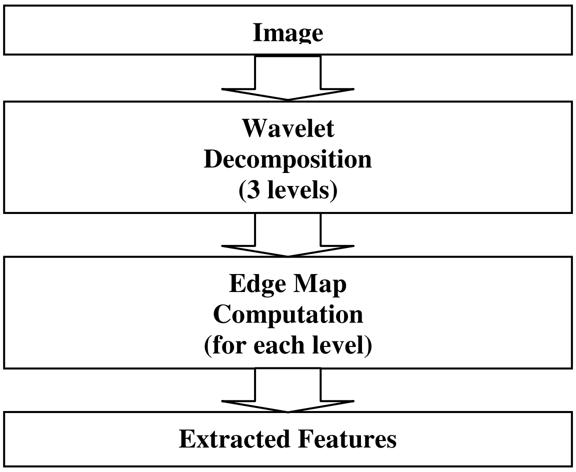

Figure 2 – The features extraction stage.

For a given image, a simple wavelet decomposition is performed. Here, the number of decomposition levels is fixed to 3 (additional layers did not improve the results). Then, an edge map is derived at each decomposition level k [11] using:

$$EMap(k) = \sqrt{CH(k)^2 + CV(k)^2 + CD(k)^2} \quad (1)$$

where CH, CV and CD denote the horizontal, vertical, and diagonal details, respectively.

Fig. 3 illustrates the wavelet decomposition and the associated edge map for an image containing a multitude of spatial frequency components. Note that the approximation coefficients are not considered in the edge map, see Eq. (1).

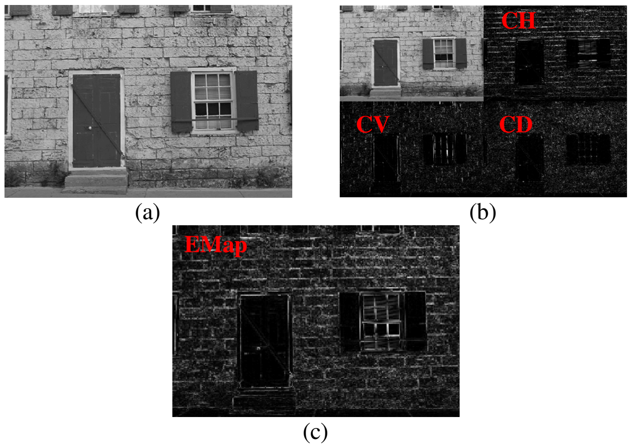

Figure 3 – a) Test image, b) Wavelet decomposition, c) Edge map (EMAP for k=1)

Once the different edge maps are obtained, a set of key feature is derived from it. Here, we use the mean and the standard deviation as primary descriptors. Six features are selected from each image: 2 features per decomposition level. Overall, 12 features are extracted, 6 from the original image at the transmitter side, and 6 from the degraded image at the receiver end. For a given pair of images (original and degraded images), the extracted features are concatenated in a single feature vector as follows:

$$Input = \begin{bmatrix} \mu_{EMAP_s}(1) \\ \mu_{EMAP_d}(1) \\ \mu_{EMAP_s}(2) \\ \mu_{EMAP_d}(2) \\ \mu_{EMAP_s}(3) \\ \mu_{EMAP_d}(3) \\ \sigma_{EMAP_s}(1) \\ \sigma_{EMAP_d}(1) \\ \sigma_{EMAP_s}(2) \\ \sigma_{EMAP_d}(2) \\ \sigma_{EMAP_s}(3) \\ \sigma_{EMAP_d}(3) \end{bmatrix} \qquad (2)$$

where $\mu_{EMAP_s}(k)$ and $\mu_{EMAP_d}(k)$ denote the means of the k[th] edge map of the original and the degraded images, respectively, and $\sigma_{EMAP_s}(k)$ and $\sigma_{EMAP_d}(k)$ represent the standard deviations of k[th] edge map of the original image and its degraded version, respectively.

2.2 A Reduced Reference Image Quality Metric

For a given distortion, the RR-IQM/D is obtained by combining the extracted features using an Artificial Neural Networks (ANN). Here, a simple Multi Layer Perceptron (MLP) is used (see Fig. 4).

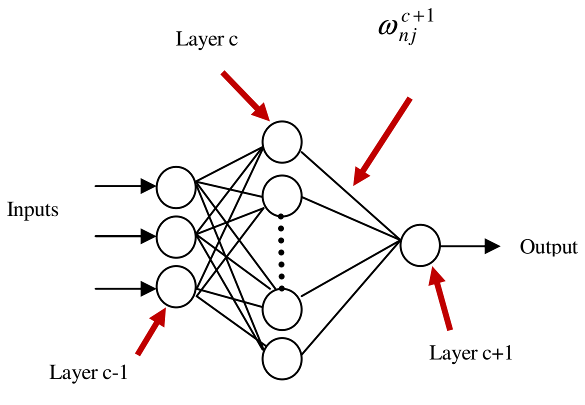

Figure 4 – Basic structure of the MLP ANN.

The output value x_n^c of a given neuron n at Layer c is given by:

$$x_n^c = \sigma\left(\sum_{u=0}^{N_{c-1}} \omega_{nu}^c \cdot x_u^{c-1}\right) \qquad (3)$$

with

$$\sigma(V) = \frac{e^{2V} - 1}{e^{2V} + 1} \qquad (4)$$

where N_c denotes the number of neurons in Layer c, ω_{nu}^c is the weight of the neuron n, x_u^{c-1} is the output of the neuron u in Layer c-1, and σ(.) is the activation function.

For a given distortion, the ANN model consists of 12 inputs and 1 output. The inputs correspond to the extracted features and the output is the predicted subjective score (Mean Opinion Score (MOS)). The number of hidden layers is fixed here to 1. The input and output values are first scaled to the range [-1, +1]. For the output, values -1 and +1 denoting worst and best quality respectively.

The back propagation algorithm is used to train the MLP. The different parameters of the MLP model are displayed in Table 1. Note that for each type of degradation, we design a different ANN model.

TABLE 1. ANN model for each considered distortion

Inputs	12 (i.e. number of extracted features).
Hidden layer	1 (the number with).
Output	1 (i.e. MOS).
Activation function	Sigmoid
Learning step	Back propagation with cross validation.

2.3 Identifying the Degradation Type

Before using the proposed IQM discussed above, we need to identify the type of degradation affecting the test image. Different classifiers can be used. In this study, a simple Linear Discriminant Analysis (LDA) classifier is used. The extracted features above are used as input patterns in the training stage. After projecting the patterns over the LDA orthogonal basis, a simple minimum distance classifier is used for distortion identification.

3. RESULTS AND DISCUSSION

To evaluate the performance of the new proposed image quality index, we used the popular Tampere Image Database (TID 2008) [14]. This database uses 25 reference images and contains 17 types of degradations with 100 images per distortion (with their associated MOS). Here, we focus only on some common distortions, namely Gaussian noise (GN), Gaussian Blur (GB), JPEG (i.e. blocking) and JPEG2000 (i.e. Ringing) artifacts as shown in Table 2. All images from the TID 2008 database are used through a cross-validation approach. Here, at each fold, 80% of the image database is used for the training step and 20% for the test step (5-fold cross validation). All images are then used to test the efficiency of the proposed method.

Table 2. Degradations considered in this work

Degradation	Type
1	Additive Gaussian noise (WN)
2	Gaussian blur (GB)
3	JPEG compression (JPEG)
4	JPEG2000 compression (JP2K)

To evaluate the performance of the proposed approach, we compared our results to those of RRIQA (Reduced Reference Image Quality Assessment) metric proposed in [9]. This method is based on a statistical model in the wavelet domain. The histograms of some selected wavelet sub-bands are first modelled (Gaussian models). The Kullback-Leiber distance is then used to evaluate the similarity index between the original and the degraded images. The proposed method is also compared to 2 full reference measures, namely the SSIM [2] (structural-based) and the VIF [4] (mutual information-based). These metrics are the most used and are available in [15].

Table 3. TID 2008 database: Pearson and Spearman correlations for each considered distortion, Gaussian Noise (GN), Gaussian Blur (GB), JPEG and JPEG2000.

IQM	Pearson Correlation			
	GN	*GB*	*JPEG*	*JPEG2000*
SSIM [2]	0.96	0.58	0.75	0.85
VIF [4]	0.95	0.83	0.89	0.93
RRIQA [9]	0.59	0.89	0.84	0.92
RR-IQM/D	0.93	0. 89	0.97	0.98
	Spearman Correlation			
SSIM [2]	0.97	0.61	0.76	0.86
VIF [4]	0.95	0.79	0.92	0.97
RRIQA [9]	0.59	0.90	0.83	0.93
RR-IQM/D	0.93	0.91	0.93	0.96

Table 3 presents the Pearson and Spearman correlation coefficients obtained for four different IQMs, including the proposed method RR-IQM/D. The proposed IQM index clearly outperforms the reduced reference RRIQA [9] in all considered degradations, except for blur where both methods give similar results. Even though the proposed method does not exploit the full reference data, it outperforms some traditional FR-IQMs. Indeed, better results are obtained for Gaussian blur, JPEG and JPEG2000 degradations compared to those obtained with the full reference metrics SSIM and VIF.

Once the efficiency of the proposed IQM scheme is validated, the classification step is then assessed in terms of classification accuracy. The overall classification system is composed of 4 classes (i.e. 4 degradation types) and 12 inputs (see equation 2). Table 4 shows the confusion matrix obtained for each considered degradation. The mean percentage is equal to 93.5% with some confusions between classes, particularly between GB and JPPEG2000 distortions. This is essentially due to the fact that blur appears also in JP2K compressed images.

True Distortion		Predicted Distortion			
		WN	GB	JPEG	JP2K
	WN	**94**	0	**1**	**5**
	GB	0	**92**	**1**	**7**
	JPEG	**1**	0	**97**	**2**
	JP2K	**3**	**6**	0	**91**

4. CONCLUSION

In this study, a new approach for Reduced Reference Image Quality evaluation is introduced. We show that the multiscale feature extraction stage based on the wavelet decomposition leads to robust visual descriptors of the image structural information. The fusion of this visual information through a neural network process offers an efficient image quality metric which can successfully be used for real-time applications such as image quality monitoring. In the future, we will analyze the impact of wrong distortion type classification. Also, the proposed method will be extended to other artifacts such as color distortions.

REFERENCES

[1] H.R. Wu and K.R. Rao, "Digital Video Image Quality and Perceptual Coding," CRC Press, 2005.

[2] Z. Wang, E.P. Simoncelli and A.C. Bovik, "Multi-scale structural similarity for image quality assessment," IEEE Asilomar Conference on Signals, Systems and Computers, 2003, Vol. 2, pp. 1398-1402.

[3] A. Beghdadi and B. Pesquet-Popescu, "A New Image Distortion Measure Based Wavelet Decomposition," 2003, IEEE International Symposium on Signal Processing and Its Applications, 2003, pp. 485-488.

[4] H.R. Sheikh and A.C. Bovik, "Image information and visual quality," IEEE Transactions on Image Processing, 2006, Vol.15, no.2, pp. 430-444.

[5] Z. Wang, H.R. Sheikh and A.C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," IEEE International Conference on Image Processing, 2002, pp. 477-480.

[6] F. Crête, "Estimer, mesurer et corriger les artefacts de compression pour la télévision," Université Joseph Fourier, 2007.

[7] A. Chetouani, G. Mostafaoui and A. Beghdadi, "A New Free Reference Image Quality Index Based on Perceptual Blur Estimation," IEEE Pacific-Rim Conference on Multimedia, 2009, pp.1185-1196.

[8] A. Chetouani, A. Beghdadi and M. Deriche, "A new free reference image quality index for blur estimation in the frequency domain," IEEE International Symposium on Signal Processing and Information Technology, 2009, pp. 155-159.

[9] Z. Wang and E.P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," Human Vision and Electronic Imaging X, Proc. SPIE, 2005, Vol. 5666, pp 149-159.

[10] M. Carnec, P. Le Callet and B. Dominique, "Objective quality assessment of color images based on a generic perceptual reduced reference," Signal Processing: Image Communication, 2008, Vol. 23, Issue 4, pp. 239-256.

[11] H. Tong, M. Li, H. Zhang and C. Zhang, "Blur detection for digital images using wavelet transform," IEEE International Conference on Multimedia and Exposition, 2004, Vol. 1, pp. 17-20.

[12] A. Chetouani, A. Beghdadi and M. Deriche, "Statistical Modeling of Image Degradation Based on Quality Metrics," International Conference on Pattern Recognition, 2010, pp.714-717.

[13] N. Peter, J. Belhumeur, P. Hespanha and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", IEEE Transactionson Pattern Analysis and Machine Intelligence, 1997, pp. 711-720.

[14] N. Ponomarenko, M. Carli, V. Lukin, K. Egiazarian, J. Astola and F. Battisti, "Color Image Database for Evaluation of Image Quality Metrics," International Workshop on Multimedia Signal Processing, 2008, pp. 403-408. http://www.ponomarenko.info/tid2008.htm

[15] http://live.ece.utexas.edu/research/Quality/index.htm