IMAGE QUALITY ASSESSMENT BASED ON HARMONICS GAIN/LOSS INFORMATION

Irwan Prasetya Gunawan and Mohammed Ghanbari

Department of Electronic Systems Engineering, University of Essex, Colchester, UK
Email: {ipguna,ghan}@essex.ac.uk

ABSTRACT

We present an objective reduced-reference image quality assessment method based on harmonic gain/loss information through a discriminative analysis of local harmonic strength (LHS). The LHS is computed from the gradient of images, and its value represents a relative degree of the appearance of blockiness on images when it is related to energy $gain$ within an image. Furthermore, comparison between local harmonic strength values from an original, distortion-free image and a degraded, processed, or compressed version of the image shows that the LHS can also be used to indicate other types of degradations, such as blurriness that corresponds with energy $loss$. Our simulations show that we can develop a single metric based on this gain/loss information and use it to rate the quality of images encoded by various encoders such as DCT-based JPEG, wavelet-based JPEG 2000, or various processed images. We show that our method can overcome some limitations of the traditional PSNR.

1. INTRODUCTION

The perceived quality of digital images and video has become an important issue in the rapid growth of multimedia applications. Although it is believed that a reliable method to judge the quality of such applications relies on the user through subjective tests, the results of such tests may not be tractable. In addition, subjective evaluation is time-consuming, laborious, expensive, and non-repeatable. Therefore, an objective automatic prediction of the perceived quality is in great demand. Research on the objective methodology has usually been aimed at replacing the widely used metrics such as peak signal-to-noise ratio (PSNR) or mean squared error (MSE) with a more realistic quality measure.

This paper concentrates on the reduced-reference (RR) image quality assessment. In RR framework, a set of side information is utilised to help the assessment. This extra information usually comprises some of the important features which have been extracted from the original/reference image.

In this work, we have developed a method based on a discriminative harmonics analysis of the spatial gradient computed from the image. A non-discriminative version of harmonics analysis has been used in a full-reference (FR) framework [1, 2] to measure the amount of blockiness distortions typically found in block-based encoded images. However, since the harmonics analysis method in the FR model was optimised for measuring blocking artefacts only, it may not be suitable for other types of distortions and consequently for the quality assessment of images encoded by other than the block-based and DCT-based codecs. Our proposed method is able to mitigate these limitations.

The contribution of our work in this paper is manifold: $1)$ the use of a single tool to quantify different types of distortions; $2)$ design simplicity without having to resort to complex model such as the human visual systems (HVS); $3)$ improved capability of harmonic analysis based quality assessment method such that it works not only on block-based DCT coded images for which it was originally designed, but also on wavelet-based coded images. We have evaluated the performance of our quality meter with images encoded by the two standard codecs: JPEG (which is based on DCT) and JPEG 2000 (based on wavelet). In addition, we have found that the proposed method in this paper is also useful for quality assessment of pictures contaminated by various types of distortions, such as additive, multiplicative, and impulsive noise, as well as low pass filtered images. More importantly, our method outperforms the shortcomings of the traditional, widely-used PSNR measure.

This paper is organized as follows. Section 2 gives the description of our proposed technique. Experimental results are given in Section 3. Finally, we conclude this paper in Section 4.

2. METHOD

We follow the *downstream* model [3] of the reduced-reference method. We also focus on the two most known compression artefacts namely blocking and blurring [4]. Both feature extraction stages of the original and the processed images use the same algorithm (Fig. 1).

Firstly, an image has to go through an edge-detection stage. At this stage, we calculate the gradient of image by applying a 3×3 Sobel operator. The gradient image is then subjected to a non-overlap block segmentation, with a sufficiently large enough blocksize to account for any vertical and horizontal activity within each block. Optionally, we would like to have a perfect alignment between these blocks with the DCT block boundary; hence inclusion of the DCT block boundary detection is useful. However, even when such perfection can not be achieved (for example, in a non-DCT coded images), the subsequent process in the model is not affected too much owing to the frequency domain operation which is insensitive to spatial shift.

Harmonics analysis is then applied to each segmented block in the intended parts of the gradient picture. The analysis is originally devised to detect the presence of blockiness on DCT-based compressed images [1]. It is based on detecting the appearance of blockiness tiling pattern that creates a pseudo-periodic signal on the gradient image and generates outstanding harmonics in the frequency-domain. Transformation to the frequency domain is done via 2-D Fast Fourier Transform (FFT) to each block. The harmonic analysis isolates and accumulates the harmonics components of the resulting FFT spectrum. The accumulated magnitudes of these harmonics components within each FFT block are chosen as the *local harmonic strength* (LHS) feature for the reduced-

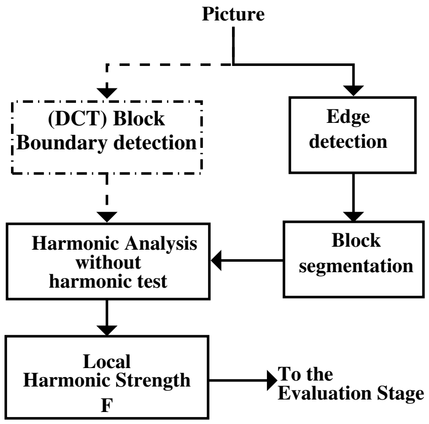

Fig. 1. Feature extraction stage of the proposed reduced-reference image quality model. Dotted lines imply optional.

reference information.

Each of the LHS values corresponds with their respective FFT blocks and can be identified by the location of the block within an image. Therefore, all the local harmonic strengths in a picture can be collected as a matrix,

$$\mathbf{F} = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{m1} & f_{m2} & \cdots & f_{mn} \end{pmatrix} \tag{1}$$

with its element, f_{ij}, represents the local harmonic strength value with $1 \leq i \leq m = \lfloor H/B \rfloor$, $1 \leq j \leq n = \lfloor W/B \rfloor$, and (i, j) represents the coordinate of each FFT block, while $W, H,$ and$B$ are the picture width, height, and the size of the FFT window, respectively. For the reference and the decoded picture we have $\mathbf{F}^{(r)}$ and $\mathbf{F}^{(d)}$. We may not need all the elements of each of these matrices, because we can restrict our attention to some important parts or regions of the image. If R represents a set of FFT blocks coordinates of these regions, one may use a more compact representation of the reduced reference information as

$$\hat{\mathbf{F}} = \{(i, j, f_{ij}) | (i, j) \in R\}\,. \tag{2}$$

The discriminative analysis (evaluation stage) is performed once all the features from the reference and the degraded pictures, $\hat{\mathbf{F}}^{(r)}$ and $\hat{\mathbf{F}}^{(d)}$, are collected. First, we calculate a local difference as $\Delta_{ij} = f_{ij}^{(d)} - f_{ij}^{(r)}$. Using Δ_{ij}, one can define the *local harmonic gain*, e_{ij}^{+}, and the *local harmonic loss*, e_{ij}^{-}, of the ij-th block in a frame/image as:

$$e_{ij}^{+} = \begin{cases} \Delta_{ij} & \text{when } \Delta_{ij} > \delta_h, \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

$$\text{and } e_{ij}^{-} = \begin{cases} |\Delta_{ij}| & \text{when } \Delta_{ij} < 0 \text{ and } |\Delta_{ij}| > \delta_h, \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

respectively. In these equations, δ_h is the harmonic threshold value below which the perceived harmonic differences is considered insignificant. In the areas where $\Delta_{ij} > \delta_h$, spatial 'activity gain' is indicated, and it is likely that this gain is proportional to the appearance of the blocking artifacts. On the other hand, regions with $\Delta_{ij} < -\delta_h$ signify spatial 'activity loss' that correspond to blurring and/or disappearance of the contextual details.

The local harmonics gain/loss values give insight on how degradations are distributed across the picture frame. To produce a single quality index for an image from these localised information, a spatial collapsing functions can be used. For example, in this work a simple arithmetic average produces good results (although different spatial pooling method is also possible). If n_e^{+} and n_e^{-} denote the number of blocks identified as $gain$ and $loss$, respectively, then the *mean harmonic gain*, e^{+}, and the *mean harmonic loss*, e^{-}, can be expressed as

$$\overline{e^{+}} = \frac{1}{n_e^{+}} \sum_i \sum_j e_{ij}^{+} \text{ and } \overline{e^{-}} = \frac{1}{n_e^{-}} \sum_i \sum_j e_{ij}^{-}. \tag{5}$$

To take into account the non-linearity behaviour of the objective metric, or saturation effect in the vicinity of extreme values, we may use the logarithmic of the mean harmonic gain, G, and loss, L, as follows:

$$G = \log_{10}(1 + \overline{e^{+}}) \quad \text{and} \quad L = \log_{10}(1 + \overline{e^{-}}) \tag{6}$$

where an offset of unity is added to ensure that the resulting values are always positive. The intermediate quality metric *LHS* based on the discriminative analysis of the local harmonic strength is defined as

$$LHS \;=\; |\alpha G + \beta L - \theta| \tag{7}$$

where α and β are the weighting coefficients for each quality factor, and θ is an offset value. According to [3], there is a potential to use non-linear mapping of the objective output defined in Eq. (7) to a subjective rating by using a non-linear logistic function, with the constraint that the function remains monotonic over the full range of data. The function to transform the set of model outputs of Eq. (7) to a set of predicted mean opinion scores (MOSp) is defined as

$$LHS^{*} = \frac{b_1 - b_2}{1 + \exp\left\{\dfrac{b_3 - LHS}{|b_4|}\right\}} + b_2 \tag{8}$$

where $b_1 - b_4$ are the coefficients of the 4-parameter logistic curve. A calibration process is then applied to the Eq. (8) to map its output to the subjective data. Note that lower LHS^{*} values correspond to lower picture quality that may be due to the appearance of blockiness/tiling degradations on the picture, the information loss as a result of blurring/smearing, or both. On the other hand, higher LHS^{*} implies that the picture in question is of higher quality because it contains a relatively small amount of degradations.

3. RESULTS

3.1. Test on compressed images

We tested the model against various images after calibrating the model to some subjective data. Calibration is conducted by training the model on a set of images annotated with subjective quality ratings. Subsequently, the validation of the model is conducted on a different set of annotated images not used in the training. The coefficients of Eq. (7) and Eq. (8) can be determined by training the model with any libraries of images. In this paper, we have used

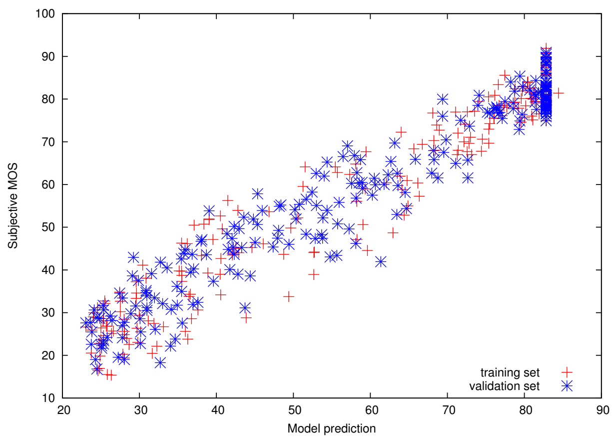

Fig. 2. Scatter plot between subjective MOS and the model's MOSp for JPEG and JPEG 2000 images from LIVE image quality database [5].

JPEG and JPEG-2000 images from LIVE image quality assessment database [5] for testing the model. In the test, the database which consists of 460 images is divided into two subsets: one for training, and the other for validation. For the given subjective scores in the database, we found that the set of coefficients suitable for assessment in this work are $\alpha = 0.968$, $\beta = 2.601$, $\theta = 0.838$ for Eq. (7) and $b1 = 0.853$, $b2 = 0.219$, $b3 = 2.538$, and $b4 = 0.534$ for Eq. (8). We also used a window size of 32×32 pixels for the 2-D FFT computation and $\delta_h = 2.00$ for the harmonic threshold in Eq. (3) and (4).

The results from using these parameters on LIVE database are shown in Fig. 2. The model performs remarkably well, both on the training and the validation sets. The apparent vertical lines on the right end of the graph are the scatter plot of the lossless versions of the original images; it shows that even for the exact copy of the original image, human judgement varies. The prediction performance of the proposed model, as well as the contribution of the two quality factors (harmonics $gain$ and $loss$) is summarised in Table 1. Along with our own metric, we have also included the assessment by the traditional PSNR and the scaled-version of full-reference blockiness detector output [1] for comparison.

From Table 1 we would like to emphasize the important role of combining $gain$ and $loss$ factor to increase the performance of the proposed technique. When using only $gain$ without $loss$ factor (or vice versa) the performance is far from remarkable; even the PSNR can outperform the $gain$ factor. On the other hand the $loss$ parameter demonstrates a better performance than $gain$ for all types of images; it even outperforms the FR blockiness detector. This means in image quality assessment, the $loss$ could play more important role than the $gain$. However, taking into account *both* the $gain$ and $loss$ factors gives the objective model LHS^* even better performance and better ability to differentiate the visual quality. Overall, our model outperforms the PSNR measure, despite the surprisingly well performance of the PSNR on LIVE dataset. In addition, later in the subsequent section of this paper we will show an example of the shortfall of the PSNR in discriminating the subjective quality. In contrast, throughout the datasets used in this work we can observe that the LHS method has given more stable performance than the PSNR.

Model	JPEG (A)		JPEG 2000 (B)		All (A+B)	
	PC	SC	PC	SC	PC	SC
LHS*	0.975	0.968	0.961	0.935	0.965	0.948
Gain	0.851	0.934	0.886	0.894	0.668	0.889
Loss	0.925	0.933	0.932	0.937	0.924	0.933
FR	0.911	0.902	0.602	0.598	0.737	0.739
PSNR	0.877	0.890	0.887	0.904	0.877	0.896

Table 1. Prediction performance in terms of Pearson correlation (PC) and Spearman rank correlation (SC) for LIVE image quality database [5]. LHS* = the proposed reduced-reference model using local harmonic strength; Gain/Loss = using only one factor (either $gain$ or $loss$ from Eq. (5), but not both); FR = full-reference blockiness detector of [1]; PSNR = using PSNR in dB.

Table 1 also shows a significant improvement our proposed model has over the FR blockiness detector. Although it is true that the FR model performs generally well on the JPEG coded images, this may not be the case for the JPEG 2000 image dataset. This is understandable since the FR model of [1] was designed primarily to detect blockiness on images. Unfortunately, this means that other types of distortions may still have gone unnoticed by the FR model. This is illustrated in Table 1 by the lowest performance of the FR model for JPEG 2000 images, where it fails to capture the degradation caused by the JPEG 2000 encoder.

So far, the benchmarks for our reduced-reference model have been the full-reference metric; i.e., the PSNR and FR blockiness detector. Although this may seem unfair, these comparisons are necessary to show that even with the reduced number of supporting information our model can still outperform them. For comparison with other reduced-reference model, we found that our model has better correlation coefficients than an HVS-based reduced-reference model in [6]; against the JPEG subset, for example, our LHS^* model can achieve a correlation of 0.975 (Table 1) while the model in [6] is reported to have a correlation of 0.961. It is also worth to note that the number of data in the reduced-reference used for our model (typically around 330 real number) is considerably less than those required in [6] (= 1056 real numbers). A typical size of LHS^* reduced-reference data file per image is about 3.0 kilobytes (and this can be reduced further by using data compression), while the HVS-based model in [6] requires around 18.0 kilobytes (with no compression). This shows the advantage of the LHS model which –despite its simple design– can still compete with more complex model such as those based on HVS.

By using the coefficients from the calibration process, we can even apply the proposed method to a wide range of corruptions on images. This is explained in the next section.

3.2. Application to images with different types of distortions

We also tested the proposed method against images distorted by various types of distortions such as those listed in Table 2. The test images along with their subjective test scores were taken from [7]. The subjective scores are expressed in terms of the mean subjective rank (MSR), which is the average of visual quality differences between the original and the processed images, rated by a group of subjects. Hence the low value of subjective rank indicates higher visual quality, and lower visual quality is represented by higher value of the subjective rank. Note that the data in Table 2 have been presented in a descending order of their MSR; i.e., from

Distortion Type	**MSR** from [7]	**UQI** from [7]	**PSNR** (dB)	**FR** [1]	**Proposed** **LHS***
Mean Shift	1.59	0.9894	24.61	1.0000	0.8283
Contrast Stretching	1.64	0.9372	24.61	0.8371	0.8471
Salt-Pepper Noise	3.32	0.6494	24.60	0.3516	0.5397
Speckle Noise	4.18	0.4408	24.61	0.3811	0.5610
Gaussian Noise	4.27	0.3891	24.61	0.3707	0.4922
Blurring	6.32	0.3461	24.63	0.7380	0.3915
JPEG Compression	6.68	0.2876	24.80	0.1969	0.3778
Pearson Correlation		-0.94	0.65	-0.61	-0.95
Spearman Correlation		-1.00	0.73	-0.68	-0.92

Table 2. Assessment of "Lena" image [7] distorted by various corruptions. Note: MSR = Mean Subjective Rank. Correlations values are calculated with respect to the subjective score of MSR

good quality picture to the worst one. The results of the assessment for sample images ("Lena" images, not presented here due to space limit) by the proposed method, as well as the output of the full-reference model Universal Quality Index (UQI) given in [7], are also tabulated in Table 2. All images are tuned to yield the same PSNR relative to the original image.

In these samples, despite the differences in subjective quality (in terms of the MSR), the table shows that the change in PSNR values across pictures with different quality is insignificant; i.e., the discriminative ability of the PSNR is very poor. In contrast, the LHS^* measure demonstrates a consistent ability to differentiate the quality, far better than the PSNR does.

Table 2 shows that the LHS^* measure agrees relatively well with the MSR compared with the other metrics; in fact, the Pearson correlation of the LHS^* is the highest amongst the others presented here, and it outperforms that of the UQI metrics which has the advantage of being the full-reference model. The only exception is the output of the LHS^* for the contrast-stretched image, which is rated slightly better than the mean-shifted image. This, however, is probably no surprise because many consider contrast-stretching as an image enhancement process, which increases the visual quality; actually, it was reported subjectively that the contrast-stretched image is better than the original [7].

4. CONCLUSIONS

This paper has presented a framework to develop an objective image quality assessment model using local harmonic strength as a reduced-reference information. We have devised a method to extract the harmonic gain and loss information from a picture and use them for quality assessment of images contaminated by various types of degradations, including additive, multiplicative, and impulsive noise as well as coding distortion in the block-based and the wavelet-based coded images; e.g., JPEG and JPEG 2000 images. Although the concept of different treatment for gain and loss distortions in our method may resemble those presented in [8], the features being used in the model and also the definition of each gain and loss are different. The model in [8] works in the spatial domain, uses features based on the statistics of the output of some edge enhancement filters, employs some relative measures of these outputs, and defines the gain and loss as some normalized ratios. Our method, on the other hand, uses some insights in the frequency domain and defines the harmonics gain and loss based on the actual differences between features in this domain. The results of our experiments indicate that the approach presented here is promising and shows an advantage over the traditional PSNR. Therefore, the proposed LHS^* method has a great potential to replace the PSNR for image quality evaluation technique in the absence of full original images. Due to the simple approach we have taken in designing this quality model, a variety of services and applications for image quality monitoring can benefit from the method proposed in this paper.

5. ACKNOWLEDGEMENTS

The work in this paper is supported by the BT Exact, UK.

6. REFERENCES

[1] K. T. Tan and M. Ghanbari, "Measuring blocking artefacts using Harmonic Analysis," *Electronic Letters*, vol. 35, no. 16, pp. 1322–1323, August 1999.

[2] ——, "Blockiness detection for MPEG2-coded video," *IEEE Signal Processing Letters*, vol. 7, no. 8, pp. 213–215, August 2000.

[3] Video Quality Expert Group, *RRNR-TV Group Test Plan Draft Version 1.6*, May 2004. [Online]. Available: http://www.vqeg.org

[4] M. Yuen and H. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal Processing*, vol. 70, pp. 247–278, 1998.

[5] H. R. Sheikh, A. C. Bovik, L. Cormack, and Z. Wang, "LIVE Image Quality Assessment Database," 2004. [Online]. Available: http://live.ece.utexas.edu/research/quality

[6] M. Carnec, P. L. Callet, and D. Barba, "An image quality assessment method based on perception of structural information," in *Proceedings of the ICIP*, vol. CD-ROMs, no. ISBN: 0-7803-7750-8. Barcelona: IEEE, September 2003.

[7] Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, March 2002. [Online]. Available: http://www.cns.nyu.edu/∼zwang/

[8] S. Wolf and M. H. Pinson, "Spatial-temporal distortion metrics for in-service quality monitoring of any digital video system," in *SPIE International Symposium on Voice, Video, and Data Communication*, Boston, MA, September 1999.