A Reduced-Reference Perceptual Quality Metric for In-Service Image Quality Assessment

Tubagus Maulana Kusuma and Hans-Jürgen Zepernick
Western Australian Telecommunications Research Institute
39 Fairway, Nedlands, WA 6907, Australia
E-mail: {mkusuma, hans}@watri.org.au

Abstract

User-oriented image quality assessment has become a key factor in multimedia communications. However, existing image quality metrics such as Peak Signal-to-Noise Ratio (PSNR) are inappropriate for in-service quality monitoring since they require the original image to be available at the receiver. Although PSNR and others are objective metrics, they are not based on human visual perception. In order to overcome these problems, we propose a novel reduced-reference objective hybrid image quality metric (HIQM) that accounts for the human visual perception and does not require a reference image at the receiver. This approach is based on a combination of various image artifact measures. The result is a single number, which represents an overall image quality. Experimental results indicate that HIQM outperforms PSNR.

1 Introduction

Transmission of multimedia services such as image and video over a wireless communication link can be expected to grow rapidly with the deployment of third and future generation mobile radio systems [1]. These mobile radio systems are also expected to offer higher data rates, improved reliability and spectrum efficiency. This will allow transmission of a variety of multimedia services over a hostile radio channel while maintaining a satisfactory quality of service. However, experiments have shown that the existing quality measures, such as Bit Error Rate (BER) are not adequate for image and video transmission. Therefore, user oriented quality metrics that incorporate human visual perception have become of great interest in image and video delivery services [2, 3]. Although the best and truest judge of quality is human (through subjective tests), continuous monitoring of communication systems quality by subjective methods is tedious, expensive and impossible in a real-time environment. Therefore, objective quality measurement methods which closely approximate the subjective test results are sought after.

Another incentive for the search after user oriented quality metrics is the fact that the commonly used image fidelity and quality metrics like Mean-Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) are inappropriate for in-service quality monitoring since the reference image is unavailable at the receiver. These existing metrics fall into the category of full-reference (FR) metrics [3]. In order to overcome this problem, we propose a novel reduced-reference (RR) objective perceptual image quality metric that does not require a reference image and which is based on the human visual system (HVS).

This paper is organized as follows. Different types of image artifacts and metrics are outlined in Section 2. The proposed Hybrid Image Quality Metric (HIQM) is explained in Section 3. In Section 4, application of HIQM for quality monitoring is discussed, followed by experimental results in Section 5. Finally, concluding remarks are given in Section 6.

2 Image Artifacts and Metrics

2.1 Image Artifacts

Image artifacts are caused by impairments such as transmission errors and depend on the image compression scheme used. A received data packet may have header information and/or data segments corrupted. In some image formats a single corrupted bit might lead to an incomplete or even undecodable image. In case of Joint Photographic Expert Group (JPEG) images, for example, the bit error location can have a significant impact on the level of image distortion. A bit error that occurs at a marker segment position can severely degrade the image quality. An image may be even completely lost because the decoder fails to recognize the compressed image.

In this paper, we consider five types of artifacts that have been observed during our simulations. These artifacts are smoothness, blocking, ringing, masking, and lost block and will be briefly described below [4][5][6]:

Smoothness, which appears as edge smoothness or texture blur, is due to the loss of high frequency components when compared with the original image. Blurring means that the received image is smoother than the original.

0-7803-7993-4/03/$17.00 © 2003 IEEE

Blocking appears in all block-based compression techniques and is due to coarse quantization of frequency components. It can be observed as surface discontinuity (edge) at block boundaries. These edges are perceived as abnormal high frequency components in the spectrum.

Ringing is observed as periodic pseudoedges around original edges. It is due to improper truncation of high frequency components.

Masking is the reduction in the visibility of one image component (the target) due to the presence of another (the masker). There are two kinds of masking effects. The first is called luminance masking, also known as light adaptation. The second is texture masking, which occurs when maskers are complex textures or masker and target have similar frequencies and orientations.

Lost block is an alteration of a pixel value, so that it does not match with its neighbours pixel value. In common operation of still image compression standards like JPEG, the encoder tiles the image into blocks of $n \times n$ pixels, calculates a 2-D transform, quantizes the transform coefficients and encodes them using Huffman coding. In common wireless scenario, the image is transmitted over wireless channel block-by-block. Due to severe fading, entire image blocks can be lost.

2.2 Image Metrics

Image metrics may be divided into two categories:

Image fidelity metrics indicate image differences by measuring pixel-by-pixel closeness between images. MSE and PSNR fall into this category.

Image quality metrics define quality based on individual image features; these metrics also incorporate HVS. Much research is being carried out on the image quality metrics, but they mostly concentrate on single artifacts [7, 8, 9, 10].

There are three approaches of measuring image fidelity and image quality as follows:

Full-Reference (FR) is a method that compares a distorted image with its undistorted original.

Reduced-Reference (RR) is a method that removes the need to store the entire original image by computing a few statistics from the distorted image and comparing them with corresponding stored statistics of the original image.

No-Reference (NR) techniques do not require any information about the original image but perform feature extraction from the image to search for the presence of known artifacts.

3 Hybrid Image Quality Metric

The proposed Hybrid Image Quality Metric (HIQM) employs several quality measurement techniques and is calculated as weighted sum of respective quality metrics. It is designed to detect and to measure different image artifacts. The result is a single number that correlates well with perceived image quality. HIQM does not require a reference image at the receiver to measure the quality of a target image.

The proposed RR approach considers five different artifact measurements relating to blocking, blur, image activity, and intensity masking detection.

The blocking measurement is based on the algorithm proposed by Wang et al. [7]. This algorithm extracts the average difference across block boundaries, averages absolute differences between in-block image samples, and calculates the zero-crossing rate. The system has been trained using subjective test results in order to comply with human visual perception. The final blocking measure is calculated using statistical non-linear curve fitting techniques. This metric is classified as an NR type because only the received image is needed to measure the blocking. In our approach, this metric is also used to detect and to measure lost blocks.

The blur measurement algorithm is based on the work of Marziliano et al. [9]. This algorithm accounts for the smoothing effect of blur on edges by measuring the distance between edges from local maximum and local minimum, the so-called local blur. A Sobel filter is used to detect the edges. Once the edges are detected, the distance between local maximum and local minimum can be measured for both horizontal and vertical directions. The final blur measure is obtained by averaging the local blur values over all edge locations in both directions.

Another important characteristic of an image relates to the activity measure that indicates the 'busyness' of the image. The active regions of an image are defined as those with strong edges and textures. Due to distortion, a received image normally has more activity compared to the original image. The technique used by HIQM is based on Saha and Vemuri's algorithm [10]. We use this metric to detect and to measure ringing and lost blocks. Especially, two types of Image Activity Measures (IAM) are deployed, edge and gradient-based IAM. For an $M \times N$ binary image, the edge activity measure is given by [10]:

$$IAM_{edge} = \left[\frac{1}{M \cdot N} \sum_{i=1}^{M \cdot N} B(i) \right] \cdot 100 \qquad (1)$$

where $B(i)$ denotes the value of the detected edge at pixel location i, M is the number of image rows and N is the number of image columns.

The gradient-based activity measure for an $M \times N$ image is given by [10]:

$$IAM_{grad} = \frac{1}{MN} \left[\sum_{i=1}^{M-1} \sum_{j=1}^{N} |I(i,j) - I(i+1,j)| - \right.$$
$$\left. - \sum_{i=1}^{M} \sum_{j=1}^{N-1} |I(i,j) - I(i,j+1)| \right] \qquad (2)$$

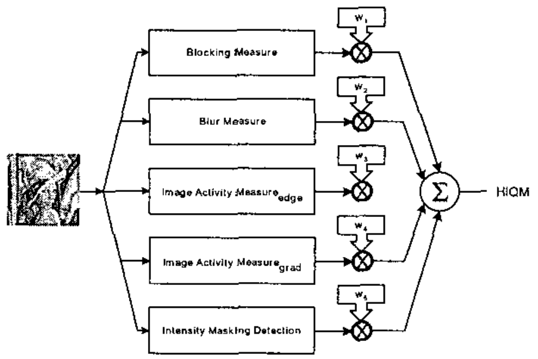

Figure 1: Formation of the combined metric HIQM.

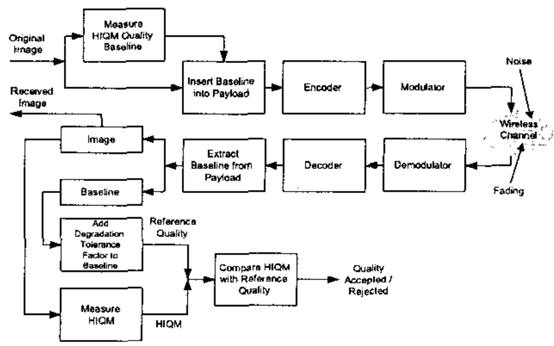

Figure 2: Reduced reference in-service quality monitoring using the proposed HIQM.

where $I(i,j)$ denotes the intensity value at pixel location (i,j) and $|\cdot|$ denotes absolute value.

Finally, the intensity masking detection is based on the standard deviation of the first-order image histogram which indicates the distribution of the image data values. The shape of the histogram provides many insights into the character of an image [11].

The proposed overall quality measure is a weighted sum of all the aforementioned metrics (see Figure 1). The weight allocation for individual metrics was based on the impact of the metric on the overall perceptibility of images by human vision. The fine-tuning of the weights was done empirically and was justified by requesting opinion from a group of unbiased test persons. Initially, all the metrics were given the same weight $w \in [0 \cdots 1]$ and were then adjusted based on the contribution of each metric to the perceptibility of image by human eye. In JPEG, for example, blocking is given a higher weight compared to other metrics because it is the most frequently observed artifact in this particular image format and can be easily perceived by human vision. The overall quality is given by:

$$\begin{aligned} HIQM &= w_1 \cdot blockMetric \\ &+ w_2 \cdot blurMetric \\ &+ w_3 \cdot IAM_{edge} + w_4 \cdot IAM_{grad} \\ &+ w_5 \cdot mask_{int} \end{aligned} \tag{3}$$

where w_i denotes the weight of a particular metric.

4 Quality Monitoring using HIQM

In image transmission systems, HIQM can be used for continuous in-service quality monitoring since it does not require a reference image to measure the quality of a received image. However, different original images differ in activity and other characteristics. Therefore, each image has its individual HIQM value, which we will refer to as quality baseline.

To obtain a proper measure, we need to normalize or calibrate the measurement system to the quality baseline. Therefore, the quality baseline of an original image needs to be communicated to the receiving end of the system under test. Obviously, this constitutes an RR approach and may replace conventional FR quality techniques. As correct reception of the quality baseline is vital for image quality assessment at the receiver, error control coding is recommended to protect this important parameter.

The basic steps of in-service quality monitoring using HIQM can be summarized as follows (Figure 2):

1. At the transmitter, measure the quality baseline of the original image in terms of its HIQM value.
2. Concatenate quality baseline and image file to form the overall payload before transmission of respective packets.
3. At the receiver, provide a reference quality by extracting quality baseline from received payload and adding a tolerable degradation value to it.
4. Measure HIQM of the received image and compare it with the reference quality.

5 Experimental Results

In this section, we provide experimental results for test images "Goldhill" (Quality baseline = 1.34) and "Tiffany" (Quality baseline = −0.57). The test scenario for these standard test images was chosen as a Rayleigh flat fading channel. A simple (31, 21) BCH code was applied for error protection purposes along with a soft-combining scheme using a maximum of two retransmissions. The average Signal-to-Noise Ratio (SNR) was set to $5dB$.

For this test scenarios, the weights of the various metrics were finally obtained from involving a group of test persons as: $w_1 = 1$, $w_2 = 0.5$, $w_3 = 1$, $w_4 = 0.5$, and $w_5 = 0.3$. From the experimental results, it can be concluded that HIQM inversely relates to PSNR (see Figures 3 and 4). In other words, the lower the HIQM value, the better the quality whereas the lower the PSNR value, the worse the quality (see Figure 5).

Figures 3 and 4 show some outliers which are due to disagreement of HIQM with PSNR. These disagreements are mostly due to the misjudgement of the PSNR in relation to the opinion of the human subjects. One sample of misjudgement is presented in Figure 6. For justification, we interviewed 10 persons to give their opinion about the quality when disagreements occurred between HIQM and PSNR. The average opinion of these people were that HIQM provides a better judgement than PSNR.

6 Summary

A new reduced-reference image quality measurement technique was presented. It was shown by way of experiment that the proposed HIQM outperforms PSNR with respect to quantifying user perceived quality. The introduced HIQM may be used for reduced-reference in-service image quality monitoring.

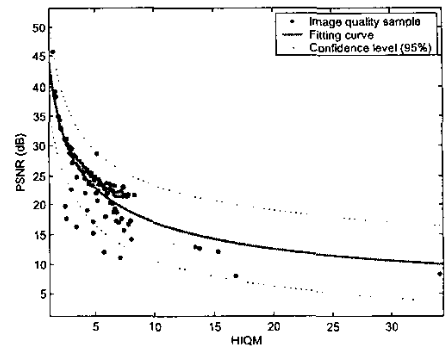

Figure 3: Metric comparison for image "Goldhill".

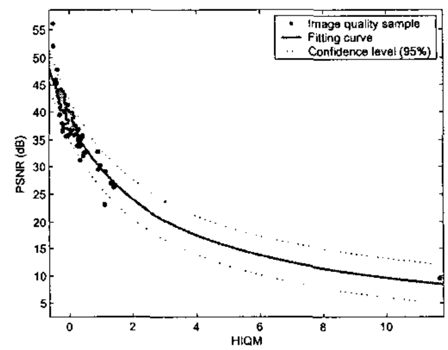

Figure 4: Metric comparison for image "Tiffany".

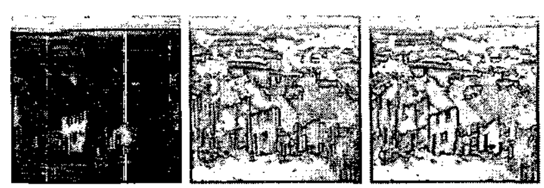

Figure 5: Quality samples for test image "Goldhill":
(L) $PSNR = 8.28dB$ and $HIQM = 34.09$;
(M) $PSNR = 22.05dB$ and $HIQM = 6.95$;
(R) $PSNR = 45.70dB$ and $HIQM = 1.43$.

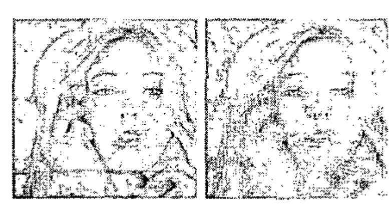

Figure 6: Quality samples for test image "Tiffany":
(L) $PSNR = 23.17dB$ and $HIQM = 3.30$;
(R) $PSNR = 24.15dB$ and $HIQM = 7.78$.

References

[1] V. Weerackody, C. Podilchuk, and A. Estrella, "Transmission of JPEG-Coded Images over Wireless Channels," *Bell Laboratories Technical Journal*, Autumn 1996.

[2] M. Knee, "The Picture Appraisal Rating (PAR) - A Single-ended Picture Quality Measure for MPEG-2," *Technical Report*, Snell and Wilcox, Jan. 2000.

[3] A. Webster, "Objective and Subjective Evaluation for Telecommunications Services and Video Quality," *National Telecommunications and Information Administration (NTIA)/Institute for Telecommunication Science (ITS), Rapporteur Q21/9, 2002.*

[4] S. Winkler, "Vision Models and Quality Metrics for Image Processing Applications," *Ph.D. Thesis, École Polytechnique Fédéralé De Laussane (EPFL)*, Dec. 2000.

[5] A. Jakulin, "Baseline JPEG and JPEG2000 Artifacts Illustrated," <http://ai.fri.uni-lj.si/~aleks/jpeg/artifacts .htm>, accessed on 8 May 2003.

[6] S.D. Rane, J. Remus, and G. Sapiro, "Wavelet-Domain Reconstruction of Lost Blocks in Wireless Image Transmission and Packet-Switched Networks," *IEEE Int. Conf. on Image Proc.*, pp. 309-312, 2002.

[7] Z. Wang, H.R. Sheikh, and A.C. Bovik, "No-Reference Perceptual Quality Assessment of JPEG Compressed Images," *IEEE Int. Conf. on Image Proc.*, pp. 477-480, Sept. 2002.

[8] Z. Wang, A.C. Bovik, and B.L. Evans, "Blind Measurement of Blocking Artifacts in Images," *IEEE Int. Conf. on Image Proc.*, pp. 981-984, Sept. 2000.

[9] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A No-Reference Perceptual Blur Metric," *IEEE Int. Conf. on Image Proc.*, pp. 57-60, Sept. 2002.

[10] S. Saha and R. Vemuri, "An Analysis on the Effect of Image Activity on Lossy Coding Performance," *IEEE Int. Symp. on Circuits and Systems*, pp. 295-298, May 2000.

[11] A.R. Weeks, *Fundamentals of Electronic Image Processing*, SPIE/IEEE Series on Imaging Science and Engineering, 1998.