EVALUATION OF FULL-REFERENCE IMAGE QUALITY ASSESSMENT METHODS

Ing. Martin SLANINA, Doctoral Degree Programme (1)
Dept. of Radio Electronics, FEEC, BUT
E-mail: xslani04@stud.feec.vutbr.cz

Supervised by: Prof. Václav Říčný

ABSTRACT

In this paper, different image quality assessment methods are compared. Their basic principles are introduced and evaluation results are given. To compare the overall performance of objective methods, correlation with subjective DSCQS method results is computed.

1 INTRODUCTION

Image and video quality measures play an important role in a variety of image and video processing applications. Very often the quality of an image needs to be quantified. This can be done by subjective testing sessions, or by objective – computational metrices. The aim of the objective metrices is to predict how much of the distortion will be observed by user. Many metrices based on different principles have been developed.

2 OBJECTIVE – DSCQS

Subjective methods for digital image (video) quality assessment are defined in ITU-R Rec. BT.500-11 [1]. More methods are still being developed, some of them are included in ITU-R Report BT.1082. The principle of subjective methods is that a group of assessors (or even a single assessor) judge the quality of an image or video being presented to them. Subjective methods are the most accurate in determining "how much" of image distortion can be perceived, and thus can be a measure of the performance of objective assessment methods. The disadvantage of these subjective methods is clear: they are expensive and impossible to be included in automatic systems (e.g. setting parameters of a system according to the instant output image / video quality).

In this contribution, only full-reference methods are compared. In other words, we always have a "perfect" image to compare the quality of a distorted one. Among the methods described in ITU-R BT.500-11, the "double stimulus" methods meet this prerequisite. In our experiments, the double stimulus continuous quality-scale method (DSCQS) is used. Two versions of each picture are presented. The observers are asked to assess the overall quality of each picture by inserting a mark on a continuous vertical scale (Fig. 1).

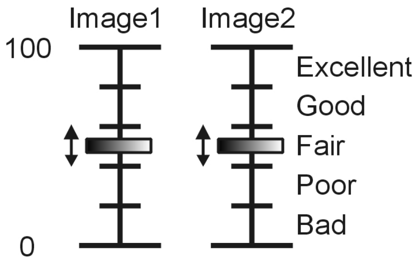

Fig. 1: *Continuous quality scale.*

The results are analysed as follows: Positions on the vertical scale are converted to normalized scores in the range 0 to 100. Each pair of scores is then converted to rating difference. The overall difference in quality is given as DMOS (differential mean opinion score), which is computed as the mean value the differences from all observers related to one image pair. The higher the DMOS, the more distortion in the image is visible.

3 MEAN SQUARED ERROR (MSE), PEAK SIGNAL TO NOISE RATIO (PSNR)

The simplest objective assessment methods are statistically defined mean squared error and peak signal to noise ratio (PSNR). These methods are pixel-based, i.e. the distorted picture and the reference are compared pixel-by-pixel. The MSE is computed according to [2] as follows:

$$MSE = \frac{1}{X \cdot Y} \sum_{i=1}^{X} \sum_{j=1}^{Y} \left[I(i,j) - \tilde{I}(i,j) \right]^2 , \qquad (1)$$

where $X \cdot Y$ is the size of the image and $I(i,j)$ and $\tilde{I}(i,j)$ are the luminance values of the reference and the distorted image, respectively. PSNR can easily be computed as

$$PSNR = 10 \cdot \log_{10} \frac{m^2}{MSE} \qquad \text{[dB]}, \qquad (2)$$

where m is the maximum pixel value (e.g. 255 for 8-bit images).

4 WEIGHTED SIGNAL TO NOISE RATIO (WSNR)

In [4], a different approach to PSNR was presented: As the human visual system (HVS) is not equally sensitive to all spatial frequencies, a contrast sensitivity function (CSF) is taken into account. The CSF is simulated by a lowpass or bandpass frequency filter.

First of all, the difference of the reference and the distorted image is computed. Then the difference is transformed into frequency domain using 2-dimensional fast Fourier transform. The obtained error spectrum is weighted by the CSF resulting in weighted error spectrum. The last thing to do is to compute the power of the weighted error spectrum and the power of the signal (also transformed into frequency domain).

5 STRUCTURAL SIMILARITY INDEX (SSIM)

This method, presented in [3], is based on comparing the structures of the reference and the distorted images.

Let *x* and *y* be two non-negative signals corresponding to the reference and distorted images, and let $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2$ and σ_{xy} be the mean of *x*, the mean of *y*, the variance of *x*, the variance of *y*, and the covariance of *x* and *y*, respectively. Here the mean and the standard deviation (square root of the variance) of a signal are roughly considered as estimates of the luminance and the contrast of the signal.

As in [3], the measure called Structural Similarity (SSIM) index is given by

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \tag{3}$$

In this formula, three different measures are involved: the luminance, contrast and structure comparison measures as follows:

$$l(x, y) = \frac{2\mu_x\mu_y}{\mu_x^2 + \mu_y^2}, \quad c(x, y) = \frac{2\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2}, \quad s(x, y) = \frac{\sigma_{xy}}{\sigma_x\sigma_y} \tag{4}$$

The two constants, C_1 and C_2 are added to prevent unstable measurement when $(\mu_x^2 + \mu_y^2)$ or $(\sigma_x^2 + \sigma_y^2)$ is close to zero. They are given by

$$C_1 = (K_1 L)^2 \text{ and } C_2 = (K_2 L)^2, \tag{5}$$

where L is the dynamic range of the pixel values (L = 255 for 8-bit gray scale images) and K_1 and K_2 are the same as in [3]: K_1 = 0.01 and K_2 = 0.03.

The SSIM indexing algorithm is applied for quality assessment of still images using a sliding window approach. The window size used in our experiments is 8x8. The SSIM indices are calculated within the sliding window, which moves pixel-by-pixel from the top-left to the bottom-right corner of the image. This results in a SSIM index max of an image, which is also considered as the quality map of the distorted image being evaluated. The overall quality value is defined as the average of the quality map – the mean SSIM (MSSIM) index.

6 TESTING MATERIAL

In our experiments, five different images were used. Each of these was subject to five different types of distortion (jpeg compression, jpeg 2000, gaussian blur, fast fading and additive white noise) with several levels of degradation, giving the total of 134 images. The quality of these pictures was evaluated using the MSE, PSNR, WSNR and SSIM metrices.

All the images were taken from the LIVE image quality assesment database from the Texas University [5]. The database provides subjective testing results (DMOS) for all images.

All the metrices were evaluated using MATLAB. The MSE and PSNR had to be implemented. The MATLAB code for SSIM and WSNR can be found on the Internet [6],[4].

Fig. 2: *Images used for evaluation.*

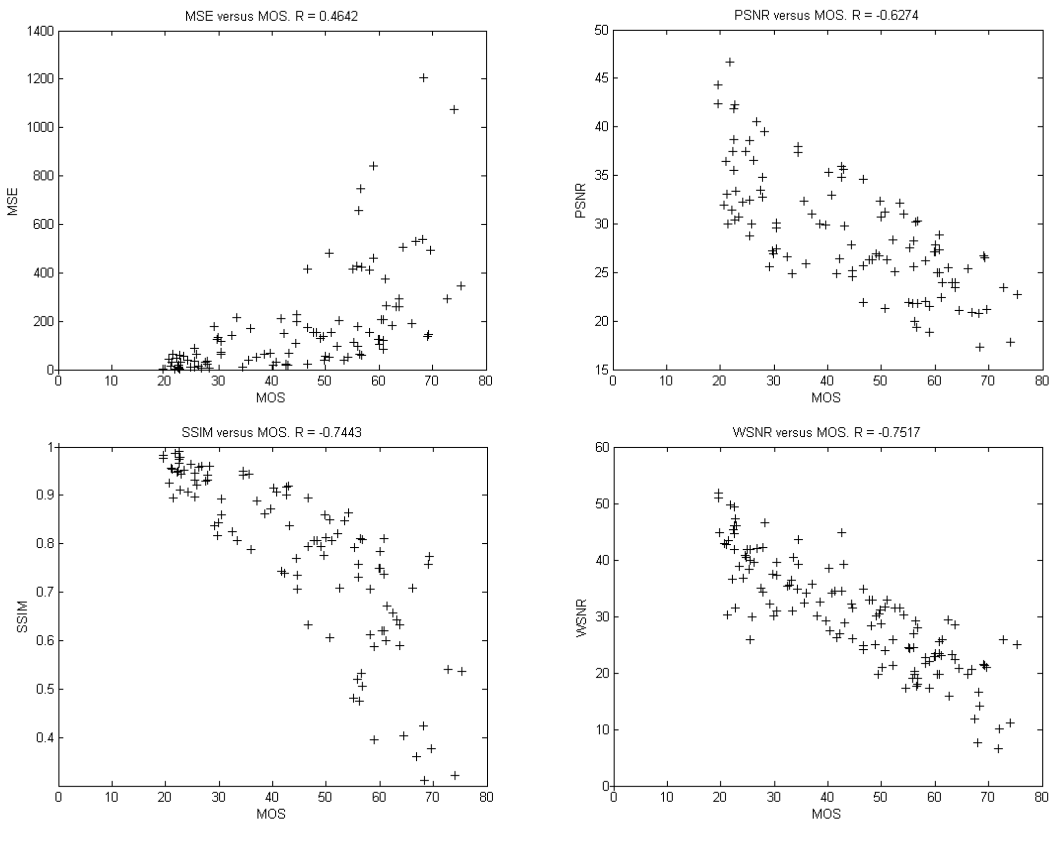

Fig. 3: *Scatter plot comparison of the evaluated metrices: MSE (R = 0.4642), PSNR (R = -0,6274), SSIM (R = -0,7443) and WSNR (R = -0,7517).*

7 CONCLUSIONS

The results of metric evaluation are shown on Fig. 3. The correspondence between DMOS and the respective metric is presented. For each graph, the overall correlation coefficient R is computed.

As seen on Fig. 3, MSE gives the worst results among the proposed metrices (R=0.4642). We get much better results by computing PSNR. The correlation coefficient increases to R=-0.6274. The higher the PSNR, the higher the image quality. That's why R is negative (in DMOS the higher the value, the worse the image quality). The performance of PSNR is highly increased by weighting the spacial frequencies with CSF. The correlation of WSNR is as high as R=-0.7517. This is comparable to the results of SSIM.

ACKNOWLEDGEMENTS

This paper is supported by the grant project of Research programme of Brno University of Technology MSM0021630513 “Electronic Communication Systems and New Generation Technology (ELKOM)”.

REFERENCES

[1] ITU-R BT.500-11. Methodology for the subjective assessment of the quality of television pictures, 2002

[2] Winkler, S.: Digital Video Quality. John Wiley & Sons, Ltd., Chichester 2005

[3] Wang, Z., Lu, L., Bovik, A. C.: Video Quality Assessment Based on Structural Distortion Measurement, Signal Processing: Image Communication, Vol. 19, No. 2, P. 121-132, 2004

[4] Damera-Venkata, N. et al.: Image Quality Assessment Based on Degradation Model http://www.ece.utexas.edu/~bevans/papers/2000/imageQuality/

[5] Sheikh, H. R., Wang, Z. et al.: LIVE Image Quality Assessment Database Release 2 http://live.ece.utexas.edu/research/quality

[6] Wang, Z.: The SSIM Index for Image Quality Assessment, http://www.cns.nyu.edu/~zwang/files/research/ssim/