A NEW IMAGE AND VIDEO QUALITY CRITERION

Ľudmila MACEKOVÁ, Stanislav MARCHEVSKÝ
Department of Electronics and Multimedia Communications, Faculty of Electrical Engineering and Informatics,
Technical University of Košice, Park Komenského 13, 041 20 Košice, Slovak Republic, tel. 055/602 2853,
E-mail: ludmila.macekova@tuke.sk, stanislav.marchevsky@tuke.sk

SUMMARY

The well known quality criteria of images and video as MSE or MAE are not corresponding sufficiently with the quality perceived by human visual system (HVS). HVS is mostly sensitive to the structural character of images and to structural errors, too. The new criterion of quality respects this aspect and can also be considered universal, because of its value not exceeding one, which is the best quality (identity actually) and the others values represent the worse quality.

Keywords: *image quality criterion, video quality perceived by human visual system (HVS)*

1. INTRODUCTION

In various areas of application, it is important to appreciate the quality of images or image sequences by mathematical criterion. The mean absolute error (MAE), mean squared error (MSE), signal-to-noise ratio (SNR) or its modification [e.g. 4] are already well known and often used. Their advantage is their independency of viewing conditions in opposite to subjective appraisal of quality. The subjective measurement of image or video quality can gain as many values as a lot of observing conditions there are. But, on the other side, the values of the numerical criteria mentioned above do not often correspond to quality perceived by *human visual system* (*HVS*).

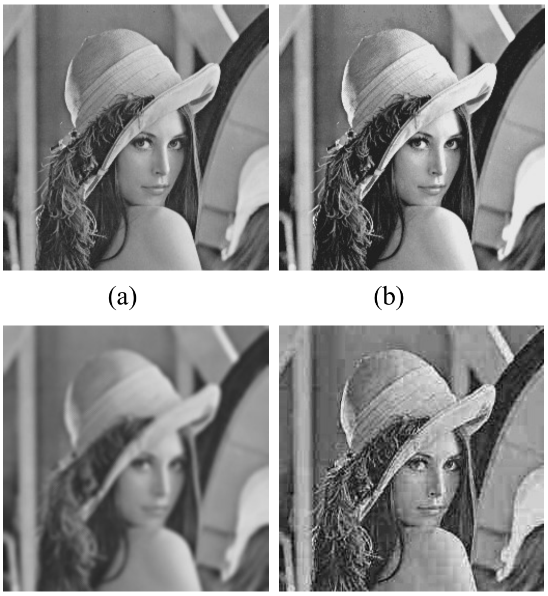

Fig. 1 The standard gray image Lena 256 x 256: a) original, b) image changed by increasing contrast, c) image distorted by blurring, d) image after JPEG compression. The b,c,d images have MSE about 225 in comparison with the original.

The good example illustrating this problem is presented in fig.1. There are noticeable differences between three images with approximately equal MSE value. The first one is original Lena and the others are an image with increased contrast and images degraded by blurring and by JPEG compression, respectively. The last three ones have MSE about 225. Therefore it is needed to find such numerical criterion, which better reflects the serious quality and which approximate to the quality perceived by HVS.

The photos in fig. 1 suggest, that our visual system is sensible to texture in image, which is for us the main carrier of image information. Therefore, we are mainly sensible to texture distortion, too. This fact is the base idea of derivation of new – structural criterion of image quality.

This article presents a new criteria of image quality and image sequence quality, based on structural features of image or video. The second part describes the mathematical derivation of criterion for static image, the third part contains derivation and application of new criteria for image sequences. The fourth part deals with experiments and their results and the last one is a conclusion.

2. THE DEFINITION OF STRUCTURAL SIMILARITY INDEX (SSIM)

If we have two digitized images x, y being compared (or just only their little parts corresponding to each other), we can describe them by values x_i, y_i; i = 1, ...n. Their statistical mean μ_x, μ_y, dispersions σ_x^2, σ_y^2 and covariance σ_{xy} are as follows:

$$\mu_x = \frac{1}{n}\sum_n x_i \ , \quad \mu_y = \frac{1}{n}\sum_n y_i \tag{1}$$

$$\sigma_x^2 = \frac{1}{n-1}\sum_n (x_i - \mu_x)^2$$
$$\sigma_y^2 = \frac{1}{n-1}\sum_n (y_i - \mu_y)^2 \tag{2}$$

$$\sigma_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y) \tag{3}$$

The mean and standard deviation (square root of the variance) roughly match to the luminance and the contrast of the signal, respectively. The covariance reflects the linear correlation between **x** and **y**.

Measures for luminance, contrast and structure comparisons (*l, c s)* of 2 image flats can be define [6]:

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

The value *s* is the different kind of similarity than luminance or contrast similarity. It reflects the structural similarity of two images, it equals one only if the structures of both compared image are exactly the same.

Then, the overall similarity index S(x, y) for comparing two similar image fragments can be expressed as the product of *l,c,s*.

$$S(\mathbf{x},\mathbf{y}) = l(\mathbf{x},\mathbf{y})c(\mathbf{x},\mathbf{y})s(\mathbf{x},\mathbf{y}) = \frac{4\mu_x\mu_y\sigma_{xy}}{(\mu_x^2+\mu_y^2)(\sigma_x^2+\sigma_y^2)} \tag{5}$$

When the member $(\mu_x^2+\mu_y^2)(\sigma_x^2+\sigma_y^2)$ is close to zero (in both too dark or too smooth-faced flats) the resulting term become unstable. This problem is eliminated by some modifications of (5) – i.e. by definition of new measure of image comparing named *Structural SIMilarity* (*SSIM*) index:

$$SSIM(\mathbf{x},\mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \tag{6}$$

where $C_1 = (K_1\, L)^2$, $C_2 = (K_2\, L)^2$ (7)

In (6), and (7) there are 3 constants established, which depend on the character of image or sequence. L is the dynamic range of pixel values - for 8 bits per pixel in gray scale images L=255. K_1, K_2 are set low enough, such that C_1, C_2 will take effect only when $(\mu_x^2+\mu_y^2)$ or $(\sigma_x^2+\sigma_y^2)$ is very low. In experiments K_1=0.01, K_2=0.03 were used.

The SSIM index has the following properties:

1. SSIM(x, y) = SSIM(y, x)

2. SSIM(x, y) ≤ 1

3. SSIM(x, y) = 1 if and only if x = y (in discrete signals there must be $x_i = y_i$ for i = 1, 2, …, N).

Thus, by the definition and by the properties of SSIM, it is simple to evaluate the quality of destroyed image, if it is compared with original image of perfect quality. The more the SSIM index value differs from 1, the worse image quality.

In practice, the application of SSIM criterion for image does not execute in one step for the whole image. First, the criterion values are evaluated in each position of 8x8 sample window (in comparison with window in original image). The sample window is sliding across the whole image, pixel by pixel. In that way, we gain so called *quality map* of image. Subsequently, the *mean SSIM (MSSIM) index* Q is evaluated as an overall image quality measure:

$$Q = \frac{\sum_{i=1}^{N} SSIM_i}{N} \tag{8}$$

where N is the number of image pixels (horizontal dimension multiplied by vertical one).

2.1 SSIM index for color images

In the case of color image, one must consider computation of local $SSIM_i$ index for all color components independently. For example for the Y, Cr, Cb components there will be $SSIM_i^Y$, $SSIM_i^{Cr}$, $SSIM_i^{Cb}$, respectively. Thus, the overall index, with respect to particular component weights, is [7]:

$$SSIM_i = W_Y\, SSIM_i^Y + W_{Cb}\, SSIM_i^{Cb} + W_{Cr}\, SSIM_i^{Cr} \tag{9}$$

In the experiments the weights were fixed: W_Y = 0.8, W_{Cb} = 0.1, W_{Cr} =0.1.

2.2 A Video quality assessment

It would be simple to calculate the video sequence quality by MSSIM index for each frame, and after this, by mean value for whole sequence. But it involves a huge volume of calculations. The next work therefore is to find possibilities of their elimination.

At first one can eliminate the calculation by restricting the number of sample windows. Only the fixed smaller count of local windows will be chosen from random positions in each frame.

The second problem is that the overall mean SSIM index is not optimal. It does not response to the quality perceived by HVS. Because of not equal importance of all particular areas of the frame for human eyes, these sample windows can not have the same weight in the term (8) for the frame quality index. The HVS perceives dark frame areas less then light ones. This phenomenon is crucial for specification of each *local weight* in this work. The darker areas the smaller are their weights. The ground for choice of threshold can be e.g. the mean local luminance about value of 40 (for 255 gray levels).

Likewise, the third reason to reform overall video quality criterion is that not all the frames in sequence have the same importance for HVS. In both cases of grate value of motion in the scene or of high speed moving camera a frame quality is not as such important as in the case of quiet frames or of a small moving. For example, some blurring is usually very disagreeable type of distortion here. Hence, in a process of quality assessment only the frames with both no and small motion will get the non-zero weights.

All the above mentioned aspects lead to video quality comparative assessment technique as follows.

- The *local windows* (e.g. 8x8) are randomly drawn from both original and inquired video frames (at the same position.). The $SSIM_{ij}$ of each local window is calculated using (6) and (9), where window index is i=1,...,R_S, R_S is the count of windows and j denotes the frame item.
- For each random i-th sampling window (in j-th distorted frame) the mean luminance μ_{ij}, is evaluated by term (1) (in the case of color frames it is the mean of Y component) and *local weighting* is differently adjusted by an outline introduced:

$$w_{ij} = \begin{cases} 0 & \mu_{ij} \le 40 \\ (\mu_{ij} - 40)/10 \quad \text{for} & 40 < \mu_{ij} \le 50 \\ 1 & \mu_{ij} > 50 \end{cases} \tag{10}$$

- Now, one can evaluate the SSIM index Q_j for each frame by weight summing of sample window quality index values:

$$Q_j = \frac{\sum_{i=1}^{Rs} w_{ij}\, SSIM_{ij}}{\sum_{i=1}^{Rs} w_{ij}} \tag{11}$$

- Assigning the weight W_j to each frame can be realized after studying the motion value. *Method of block-based motion estimation* can be employed for each i-th sample window, by comparing the actual and the next frame [7]. This step results in a frame set of *local motion vector lengths* m_{ij}. Afterwards the *frame motion level* M_j is:

$$M_j = \sum_{i=1}^{Rs} m_{ij}\,)/R_s \tag{12}$$

And, the weight W_j of the j-th frame is designated by comparing W_j with the motion level threshold t_M.

$$W_j = \begin{cases} \sum_{i=1}^{Rs} w_{ij} & M_j \le t_M \\ 0 & M_j > t_M \end{cases} \tag{13}$$

The threshold can be set to 16. As well as for sampling window weights, the frame weights can be more fine-tuned [7].

- Finally, the result step of algorithm is the calculating of *video quality* Q_v:

$$Q_v = \frac{\sum_{j=1}^{F} W_j\, Q_j}{\sum_{j=1}^{F} W_j} \tag{14}$$

3. EXPERIMENTS AND RESULTS

The first goal of our experiments was to compare the values of new quality index with subjective evaluations for several types of distortion of *Lena image*, which have around equal MSE. These observations have the most marked results and are therefore proposed in this paper.

The standard test image Lena was distorted by blurring, contrast stretching, impulsive salt-and-pepper, multiplicative noise and JPEG compression, respectively (see the fig. 1 or fig. 2).

(a) (b)

Fig. 2 The Lena image distorted by salt-and-pepper noise (a), and by multiplicative noise (b), respectively. The both noised images have the MSE value closed to 225 in comparing to original image.

All distortion types caused the MSE value around 225. The new numerical quality index *Q* was evaluated for each distorted image by means of method of sliding 8 x 8 window and by using of terms (6) - (8). On the other hand, in the subjective experiment, ten people who were not acquainted with image processing area, compared these 5 (and original one) images and designated the ranks of quality from value of 1 (original image) to 6.

The results of the above mentioned experiments are documented in table 1 and confirm our assumption: The subjective rank is similar to Q index rank; the best subjective rank was given to Contrast Stretching Lena and the worst one to Multiplicative noise image. The Contrast stretching image obtained the highest index again (near value 1), and the Multiplicative Noise Lena got the lowest one.

Distortion type	MSE	Mean subjective rank	Q
Contrast stretching	226.36	1,9	0.943
'Salt-and-pepper' noise	226.23	2,47	0.786
JPEG compression	225.92	5,14	0.745
Blurring	225.23	5,24	0.741
Multiplicative noise	224.62	5,43	0.59

Tab. 1 Comparison of subjective, MSE, and structural similarity index (Q) ranking of damaged Lena image versions

A lot of other calculations were performed to gain Q values of *black-and-white (BW) Lena and Bridge*, damaged by several types and values of noise, and filtered by several filters, as well. The results led to similar conclusions as above mentioned ones.

There were also evaluated the *color images* (the color Lena and Mandrill) disturbed by impulsive correlated noise of volumes both 10% and 20% and filtered by median filters with a few square window sizes. Some representative of this area are presented in table 2. The more detailed describing of these experiments can be found in work [2].

Image type	Noise volume	Filter window	MSE	Q
color Lena	10%	3 × 3	58.57	0.905
		5 × 5	109.13	0.838
		7 × 7	159.34	0.786
	20%	3 × 3	86.8	0.874
		5 × 5	124.36	0.826
		7 × 7	175.06	0.774
Mand rill	10%	3 × 3	322.76	0.600
		5 × 5	407.17	0.457
		7 × 7	467.18	0.381
	20%	3 × 3	358.99	0.576
		5 × 5	424.31	0.448
		7 × 7	481.56	0.374

Tab. 2 Demonstration of use of Q criterion for measurement of filtering efficiency

The new numerical *quality index of little gray standard video* Salesman (50 frames, 256x256 dimensions, 255 levels gray scale) was investigated ultimately [2]. The decomposed image sequence was artificial damaged by the BW spots (1% of all frame pixels) and subsequently filtered by several modifications of median filter [1]. One-step filtering and two-step filtering were realized, both with and without spots detection. The quality of result sequence was then calculated. In the table 3 there are a few of the results introduced. Though, the enumeration was made by simplified procedure - with all sample window weights and frame weights equal to 1.

It is known, from many previous works of various authors [e.g. 3 and 5], the noise or blotch filters work better with detectors of distortion, and the two-stage median filtering with blotch detector (MMF2+detector) shows the best visual results [1]. The highest quality rank of this filter type in the table 3 corresponds with this fact.

Filter type	MSE	Q_v
W7	101.12	0.594
W7+detektor	21.94	0.842
MMF2	23.90	0.851
MMF2+detektor	22.29	0.861

Tab. 3 Results of experiments in form of MSE and SSIM index for image sequence Salesman filtered by several filter types

4. CONCLUSION

The new image quality criterion recently proposed in [6] and improved in [7] seems very useful and comprehensible for purpose of quality assessment closed to human visual perceiving. It implies the change of structural properties of distorted image or video, because of their priority for human eyes.

We have proofed the new quality criterion for many standard noised and filtered images and image sequences, which were examined previously by MAE and MSE. Based on the all results of our objective and subjective experiments one can establish, that this criterion is really effective and it more correlates with the quality perceived by human visual system, than the criterion MSE or its derivative. Because of its value below one it is more practical for the purpose of image and video quality assessment, as well. Its use would be appropriate in the future image processing research. Of course, there are the areas for more improvements, like consideration both of motion, and of perfect video with a damaged partitions [7], etc.

REFERENCES

[1] Hrešo, M.: Restoration of Damaged Black-and-White Image Sequences (in Slovak), Diplomová práca, Technická univerzita v Košiciach, 2001

[2] Hvizdoš, J.: Image quality evaluation by univerzal criterion (in Slovak), Dipl.práca, Technická univerzita v Košiciach, 2004

[3] Kokaram, A.C., Morris, R.D., Fitzgerald, W.J., Rayner, P.J.W.: Detection of Missing Data in Image Sequences, IEEE Transactions on Image Processing, Vol.4, No.11, Nov. 1995, pp. 1496-1508

[4] Moucha, V., Marchevský, S., Lukáč, R., Stupák, C.: Digital Filtering of Image Signals (in Slovak), Edičné stred. VLA gen. M.R.Štefánika v Košiciach, 2000
[5] Stupák, C., Lukáč, R., Marchevský, S.: Utilization of the Impulse Detectors in Grayscale Image Filtering, Journal of Electrical Engineering, Vol. 51, 07-08 2000, pp. 173-181
[6] Wang, Z., Bovik, C.: A Universal Image Quality Index, IEEE Signal Processing Letters, Vol.9, No.3, March 2002
[7] Wang, Z., Lu, L., Bovik, A. C.: Video Quality Assessment Based on Structural Distortion Measurement, Signal Processing: Image Communication, Vol. 19, No. 1, Jan. 2004

BIOGRAPHY

Ľudmila Maceková graduated (MSc. equiv. degree) in radioelectronics from the Technical University of Košice in 1983. Since 1991 she was with Department of Electronics and Multimedia Communications of the Faculty of Electrical Engineering and Informatics of this university as assistant professor and nowadays as research assistant. She is working in projects in area of image processing. The image and image sequence filtering is a problem content of her PhD work, as well.

Stanislav Marchevský received the M.Sc. in electrical engineering at the Faculty of Electrical Engineering, Czech Technical University in Prague, in 1976 and PhD. degree in radioelectronics at the Technical University of Košice in 1985. Currently he is a Professor of Electronics and Multimedia Communication Department of Faculty of Electrical Engineering and Informatics of Technical University of Košice. His teaching interests include switching theory, digital television technology, and satellite communications. His research interests include image nonlinear filtering, neural networks, genetic algorithms, and multiuser detection, spacetime communication, diversity communications over fading channel, and power and bandwidth-efficient multiuser communications.