WEIGHTED MSE BASED METRICS FOR CHARACTERIZATION OF VISUAL QUALITY OF IMAGE DENOISING METHODS

Nikolay Ponomarenko (*), Sergey Krivenko (*), Karen Egiazarian (**), Jaakko Astola (**) and Vladimir Lukin (*)

(*) National Aerospace University, Kharkov, Ukraine
(**) Tampere University of Technology, Tampere, Finland

ABSTRACT

We propose modifications of standard MSE and some other metrics related to it for estimation of visual quality of images filtered by different methods. The modified metrics are based on assumption that it is important for a human observer that details and edges in a filtered image are not smeared. We propose to apply a weighted version of MSE as follows. For image fragments where differences of filtered image values with respect to noise-free image are larger than for the corresponding noisy image the weights are set larger than for other image fragments. An experiment on estimation of filtered image visual quality is carried out for a test set containing 72 images. The weight value equal to 5 is recommended. It is shown that such method of calculating MSE can be used for modifying the metrics PSNR, PSNR-HVS, PSNR-HVS-M. This allows increasing their correspondence to human perception and evaluation of filtered images visual quality.

1. INTRODUCTION

Filtering of noisy images is a typical task in image processing [1]. Numerous filters have been already proposed, tested and compared, mostly using mean squared error (MSE) or peak signal-to-noise ratio (PSNR) as filtering efficiency criteria. However, in many applications visual quality of processed images are of major importance [2, 3] and, as known, MSE and PSNR do not correlate well enough with visual quality of images. In particular, this relates to analysis of efficiency of image filtering methods for which the absence of metrics adequate to human perception is still a problem. However, recently a great attention has been paid to design of new visual quality metrics and to advancing existing quality indices [3, 4].

In such design, full reference metrics are commonly used. This means that their calculation is based on comparison of a distorted and the corresponding noise-free images without taking into account what are the most "important" fragments of an image subject to visual analysis by an observer.

Usually in analysis of filter efficiency, one has a reference (noise-free) image I^{et} that is then artificially distorted by adding noise of a given model with pre-determined parameters obtaining I^{n}. Then, image I^{n} is a subject to filtering resulting in a processed image I^{p}. After this, filtering quality metric can be calculated both for I^{p} and I^{n} and the results compared. A filtering method is considered to be more efficient if the quality of the image I^{p} with respect to the image I^{n} improves more according to the chosen quantitative measure. It often happens that image quality, according to PSNR, has increased by around 1-2 dB but visual quality of the filtered image has reduced due to smearing image details and edges.

It is known that observers, while analyzing images, mainly pay more attention to such image features as edges, fine details and textures [5, 6] than to the noise present in image homogeneous regions. Meanwhile, most filters might introduce distortions just in image locally active areas (neighborhoods of edges, fine details and textures). Therefore, aforementioned properties of human vision system and filters have to be taken into account by metrics that are intended for adequate characterization of image processing methods. This paper addresses experimental verification of this hypothesis and the corresponding practical steps.

Section 2 describes the main idea of the proposed method and it contains basic expressions how to calculate the proposed modified MSE. Test image set and the conducted experiments are described in Section 3. Weight optimization for several visual quality metrics based on the results of experiments is carried out in Section 4. Finally, the conclusions follow.

2. MAIN IDEA

Conventional MSE between a reference image I^{et} and a processed image I^{p} is defined as

$$MSE(I^{et},I^{p}) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(I^{et}_{ij} - I^{p}_{ij})^2 ,$$

where M, N define number of rows and columns in images. Respectively, PSNR can be easily calculated (for standard 8-bit image representation) as

$$PSNR(I^{et},I^{p}) = 10 \log_{10}(255^2/MSE(I^{et},I^{p})).$$

As it is seen, conventional MSE and PSNR need two images, I^{et} and I^{p}, for their calculation. We propose a new metric wMSE (weighted MSE) that requires availability of all three images (I^{et} , I^{n} , and I^{p}) of size MxN and is defined as

$$wMSE(I^{et},I^{n},I^{p}) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\delta_{ij}(I^{et}_{ij} - I^{p}_{ij})^2 / \sum_{i=1}^{M}\sum_{j=1}^{N}\delta_{ij}$$

$$\delta_{ij} = \begin{cases} 1, & \left|I^{et}_{ij} - I^{p}_{ij}\right| \le \left|I^{et}_{ij} - I^{n}_{ij}\right| \\ W_{dist}, & \left|I^{et}_{ij} - I^{p}_{ij}\right| > \left|I^{et}_{ij} - I^{n}_{ij}\right| \end{cases}, \qquad (1)$$

where W_{dist}>1 is a weight, i and j are image indices. Thus, we penalize a filter if it produces an absolute difference between I^{et} and I^{p} larger than an absolute difference between I^{et} and noisy image I^{n} at the same pixel position. Fig. 1 shows a map of δ for the test image Barbara corrupted by additive Gaussian noise with variance σ^2=400 and then processed by DCT filter [7]. As it is seen, filtering has led to distortions mainly introduced in edge/detail neighborhoods and textural regions.

Fig. 1. An example of the map of δ where filtering has led to distortions (dark color pixels).

Note that similarly it is possible to evaluate efficiency for not only methods of image filtering but also for other image processing methods. For example, this can be done for image reconstruction and impairment correction methods, i.e. for applications where original, distorted (I^{n}) and processed (I^{p}) images are available.

Recently, the visual quality metrics PSNR-HVS [7] and PSNR-HVS-M [8] have been proposed. MSE for them is calculated in spectral (DCT) domain. Then, expression (1) can be modified as

$$wMSE(I^{et},I^{n},I^{p}) = K\sum_{i=1}^{N-7}\sum_{j=1}^{M-7}\sum_{m=1}^{8}\sum_{n=1}^{8}\delta_{ijmn}(X^{et}_{ijmn} - X^{p}_{ijmn})^2 ,$$

$$K = 1/(N-7)/(M-7)/64/\sum_{i=1}^{N-7}\sum_{j=1}^{M-7}\sum_{m=1}^{8}\sum_{n=1}^{8}\delta_{ijmn} ,$$

$$\delta_{ijmn} = \begin{cases} 1, & \left|X^{et}_{ijmn} - X^{p}_{ijmn}\right| \le \left|X^{et}_{ijmn} - X^{n}_{ijmn}\right| \\ W_{dist}, & \left|X^{et}_{ijmn} - X^{p}_{ijmn}\right| > \left|X^{et}_{ijmn} - X^{n}_{ijmn}\right| \end{cases}, \qquad (2)$$

where X_{ijmn} denotes DCT coefficients of the DCT block of size 8x8 pixels with coordinates of its left upper corner defined by indices *ij*, and *m,n* are corresponding indices within the 8x8 block.

The proposed approach to weighted taking into account of distortions is quite universal. Modifications to derivation of MSE similar to (1) and (2) can be introduced in wavelet transform domain and, respectively, for the corresponding quality metrics.

Note that expression (2) can be easily modified taking into account human visual system (HVS) peculiarities as different sensitivity to distortions in different spatial frequencies, masking effects, etc. as this was done in the papers [8] and [9] for the metrics PSNR-HVS and PSNR-HVS-M.

3. TEST IMAGE SET AND EXPERIMENTS WITH HUMAN OBSERVERS

It is clear that properties of the proposed metrics depend upon W_{dist} and it is desirable to optimize it or, at least, to give some practical recommendations. To adjust a proper value of W_{dist}, an experiment with volunteers has been carried out to get a mean opinion score (MOS) for image visual quality.

Two-stimulus comparisons of images from a set have been performed. This set contained 24 images for each of three test images. For each of three 8-bit test images (Barbara, Baboon, Lenna), four noisy images corrupted by additive i.i.d. Gaussian noise with variances σ^2 equal to 50, 100, 200, and 400, have been generated.

Each noisy image has been processed by the standard mean, median [1], local statistic Lee [10], sigma [11], and DCT [7] filters. All noisy and processed images were subject to pairwise comparisons with ordering and, finally, obtaining mean opinion score (MOS). Thus, the total number of processed images in the test set was equal to 5x12 = 60.

The metric *wMSE* was calculated for both filtered images and noisy ones ($wMSE(I^{et},I^{n},I^{n})$), and analysis (assessment) of image visual quality has been carried out for all noisy and filtered images of the set.

Experiments have been performed for each reference image set separately according to the methodology described in [8]. 63 humans (students studying telecommunications and radio engineering, researchers and few people who do not deal with image processing at all) participated in experiments. Visualization and observation conditions varied in reasonable limits to be comfortable for each participant. Different monitors were used, both LCD and CRT, mainly 19” with the preset resolution 1152x864 pixels. Average time for one experiment was about 6 minutes. Totally more than 7000 elementary (pair-wise) comparisons have been performed. As the result, MOS estimates have been obtained and later used for optimization of W_{dist}.

Analysis has shown that among the considered filters the best visual quality has been usually provided by the DCT filter [7] except the case of highly textural test image Baboon corrupted by noise with variance 50 for which the local statistic Lee [10] filter was the “best”. Standard mean and median filters were always the worse according to MOS.

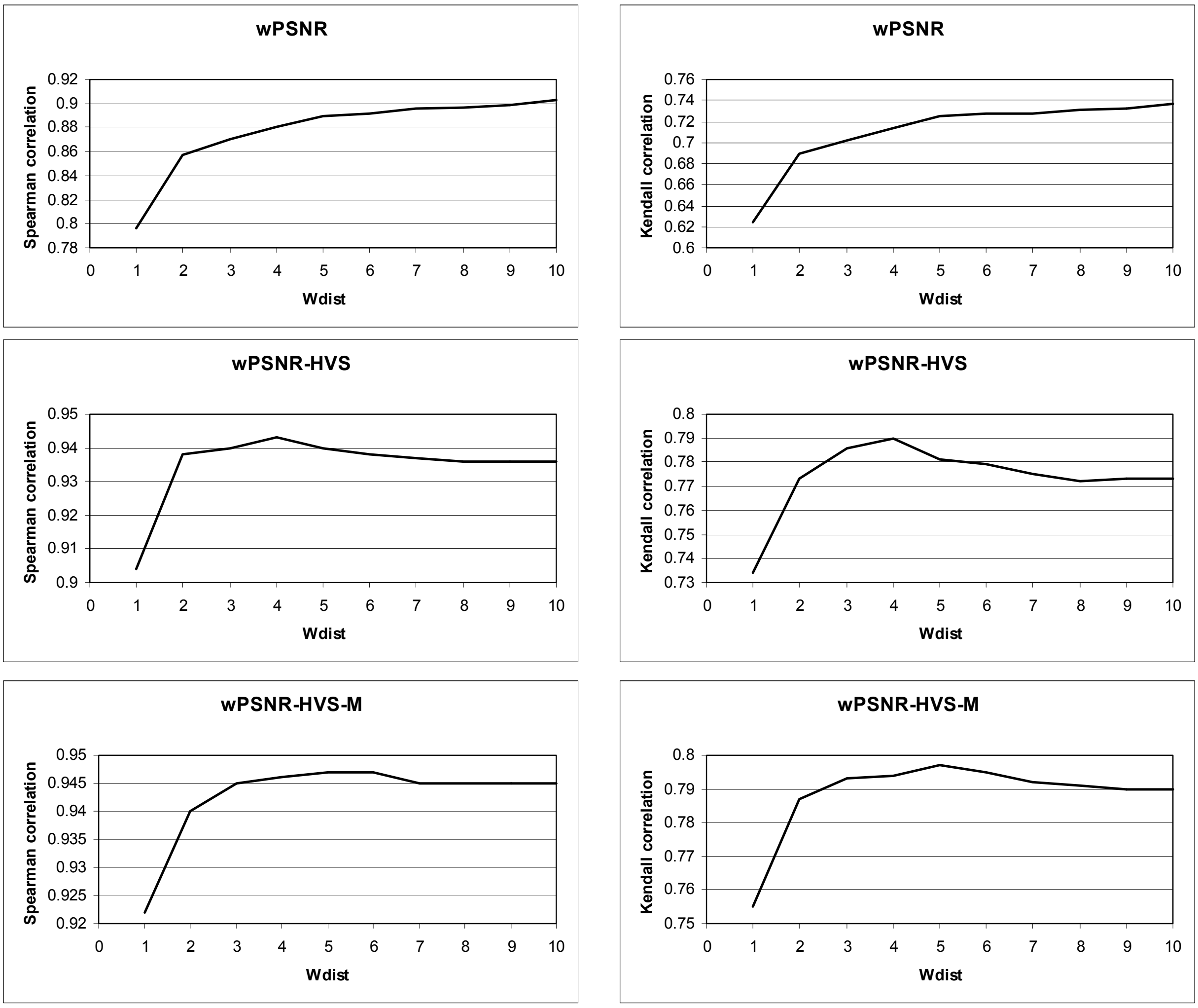

Fig. 2. Dependence of rank correlation with MOS on W_{dist} for metrics wPSNR, wPSNR-HVS, and wPSNR-HVS-M

Moreover, for small noise variance, the visual quality of images processed by standard mean and median filters was perceived as worse than the quality of the corresponding noisy images.

4. ANALYSIS OF RESULTS AND WEIGHT OPTIMIZATION

Spearman and Kendall rank correlation factors [12] for the considered metrics and MOS have been determined. Table 1 presents the results for some known quality metrics. As it is seen, the best results are produced by the metric VIF [16], but it is desirable to provide larger correlation factors.

Table 1. Rank correlation factors for the obtained MOS and some visual quality metrics

Metric	Spearman correlation	Kendall correlation
PSNR	0.796	0.624
PSNR-HVS [7]	0.904	0.734
PSNR-HVS-M [8]	0.922	0.755
MSSIM [13]	0.874	0.687
SSIM [14]	0.883	0.700
VSNR [15]	0.889	0.720
VIF [16]	0.937	0.768

As it was mentioned above, the metrics PSNR, PSNR-HVS, and PSNR-HVS-M have been modified accordingly in order to use the proposed *wMSE* as a basis. As the result, the metrics wPSNR, wPSNR-HVS, and wPSNR-HVS-M have been obtained. Varying W_{dist} from 1 till 10 and checking its influence to rank correlation factors for the modified metrics with MOS, we obtained dependences presented in Fig. 2.

As it is seen from analysis of these plots, the use of weighted MSE in calculation of these metrics allows considerable increase of rank correlation factors, both Spearman and Kendall. This especially relates to the metric wPSNR for which Spearman correlation, for W_{dist} approaching 10, becomes larger than 0.9, i.e. larger than for such well-known metrics as MSSIM, SSIM, and VSNR. Note that both dependences do not have optimums.

On the contrary, dependences of Spearman and Kendall correlation factors for the metrics wPSNR-HVS and wPSNR-HVS-M on W_{dist} have optimums. They are observed for W_{dist} within the limits from 4 to 6. Taking into account the number of carried out experiments, it is possible to recommend using $W_{dist}=5$ for practical use. In this case, the metrics wPSNR-HVS and wPSNR-HVS-M provide rank correlation factors larger (over 0.94) than the metric VIF [16] which was the best according to results in Table 1.

We have also tested the possibility to improve performance of the metric SSIM by using its weighted modification. Both Spearman and Kendall correlations have increased and their optima have been observed for W_{dist} about 3. However, improvements with respect to original version of SSIM were not as large as for PSNR-HVS-M metric, only about 0.01 for both Spearman and Kendall correlations.

Table 2 presents results of filtered image visual quality assessment using standard PSNR and wPSNR ($W_{dist}=5$) for some images from the set and several variances of noise.

Table 2. Filtering efficiency assessment, dB

Image	σ^2	Filter	PSNR	wPSNR
Barbara	400	-	22.18	22.18
Barbara	400	Median	22.83	19.05
Barbara	400	Mean	23.30	19.30
Barbara	400	DCT	30.31	27.74
Baboon	50	-	31.14	31.14
Baboon	50	Sigma	31.78	30.90
Baboon	50	Lee	32.22	31.86
Baboon	50	DCT	32.70	31.67

An example for the test image Barbara corrupted by noise with variance 400 shows the following. The standard mean and median filters improve image quality according to PSNR, but not considerably. At the same time, comparison of image fragments in Fig. 3 shows that the filtered image is smeared (blurred) and its visual quality is not better (actually, worse) than for the noisy image. This is in agreement with results for the proposed metric wPSNR that indicates that the processed image visual quality is worse than original for the mean and median filters.

a) b)

Fig. 3. A fragment of the test image Barbara: a) corrupted by additive noise, σ^2=400, PSNR=wPSNR = 22.18 dB, b) the median filter output, PSNR = 22.83 dB, wPSNR = 19.05 dB

Besides, the modified versions allow assessing what filter is more efficient in the sense of providing better visual quality. An example given in Table 2 for the image Baboon corrupted by noise with variance 50 shows that the DCT filter is better than the local statistic Lee and sigma filters in PSNR. However, the local statistic Lee filter provides larger wPSNR than the DCT filter and this is in agreement with MOS analysis. These results also show that there are such images (mostly, textural ones) and noise variance range (quite small values) for which even efficient filtering produces practically no improvement of image visual quality.

5. CONCLUSIONS

This paper presents a simple weighted modification of MSE and three metrics based on it that allows increasing adequateness of metrics to visual quality of filtered images. The obtained results ones more confirm that distortions in edge/detail neighborhoods are paid much attention by observers. Thus, this should be taken into account in metrics design. For a practical use, we recommend to set weight W_{dist}=5 for the considered metrics.

REFERENCES

[1] J. Astola, P. Kuosmanen, *Fundamentals of nonlinear digital filtering*, Boca Raton (USA): CRC Press LLC, 1997.

[2] B.W. Keelan, *Handbook of Image Quality*, Marcel Dekker, Inc.: New York, USA, 2002.

[3] Z. Wang, A.C. Bovik, “Mean squared error: love it or leave it: a new look at signal fidelity measures,” *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98-117, Jan. 2009.

[4] N. Ponomarenko, F. Battisti, K. Egiazarian, J. Astola, V. Lukin, "Metrics performance comparison for color image database", *Fourth international workshop on video processing and quality metrics for consumer electronics*, Scottsdale, Arizona, USA. Jan. 14-16, 6 p, 2009.

[5] M. Pedersen, J.Y.Hardeberg, “A new spatial hue angle metric for perceptual image difference”, *Springer Series on LNCS*, Vol. 5646, pp. 81-90, 2009.

[6] C.M. Privitera, L.W. Stark, “Algorithms for defining visual regions-of-interest: comparisons with eye fixations”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No 7, pp. 970-982, 2000.

[7] V.V. Lukin, R. Oktem, N. Ponomarenko, K. Egiazarian, “Image filtering based on discrete cosine transform”, *Telecommunications and Radio Engineering*, Vol. 66, No 18, pp. 1685-1701, 2007.

[8] K. Egiazarian, J.Astola, N.Ponomarenko, V.Lukin, F.Battisti, and M.Carli, "New full-reference quality metrics based on HVS", *CD-ROM Proceedings of the Second Int. Workshop on Video Proc. and Quality Metrics*, Scottsdale, USA, 4 p, 2006.

[9] N.Ponomarenko, F.Silvestri, K.Egiazarian, M.Carli, J.Astola, and V.Lukin, "On between-coefficient contrast masking of DCT basis functions", *CD-ROM Proc. of the Third Int. Workshop on Video Proc. and Quality Metrics*, USA, 4 p, 2007.

[10] J.S. Lee, "Digital Image Enhancement and Noise Filtering by Use of Local Statistics", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, March, p.165-168, 1980.

[11] J.S. Lee, “Digital Image Smoothing and the Sigma Filter”, *Computer Vision, Graphics, and Image Processing* 24, p. 255-269, 1983.

[12] M.G.Kendall, *The advanced theory of statistics*, Vol. 1, London, UK, Charles Griffin & Company limited, 457 p, 1945.

[13] Z.Wang, E.P.Simoncelli, A.C.Bovik, "Multi-scale structural similarity for image quality assessment," *IEEE Asilomar Conf. on Signals, Systems and Computers*, 2003.

[14] Z.Wang, A.Bovik, H.Sheikh, E.Simoncelli, “Image quality assessment: from error visibility to structural similarity”, *IEEE Transactions on Image Processing*, vol. 13, issue 4, April, pp. 600-612, 2004.

[15] D.M.Chandler, S.S.Hemami, “VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images”, *IEEE Transactions on Image Proc.*, Vol. 16 (9), pp. 2284-2298, 2007.

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