Picture quality measures in image compression systems

Conference Paper · October 2003

DOI: 10.1109/EURCON.2003.1248017 · Source: IEEE Xplore

3 authors, including:

Marta Mrak
BBC

Sonja Grgic
University of Zagreb

Picture Quality Measures in Image Compression Systems

Marta Mrak, Sonja Grgic, and Mislav Grgic
University of Zagreb, Faculty of Electrical Engineering and Computing
Unska 3 / XII, HR-10000 Zagreb
CROATIA
E-mail: marta.mrak@fer.hr

***Abstract*--A major problem in evaluating picture quality in image compression systems is the extreme difficulty in describing the type and amount of degradation in reconstructed image. Because of the inherent drawbacks associated with the subjective measures of picture quality there has been a great deal of interest in developing an objective measure that can be used as a substitute. The aim of this paper is to examine a set of objective picture quality measures for application in still image compression systems and to highlight the correlation of these measures with subjective picture quality measures. Picture quality is measured using nine different objective picture quality measures and subjectively using Mean Opinion Score (MOS) as measure of perceived picture quality. The correlation between each objective measure and MOS is found. The effects of different image compression algorithms, image contents and compression ratios are assessed and the best objective measures are proposed. Our results show that some objective measures correlate well with the perceived picture quality for a given compression algorithm but they are not reliable for an evaluation across different algorithms. So, we compared objective picture quality measures across different algorithms and we found measures, which serve well in all tested image compression systems.**

***Index Terms*--Correlation, JPEG, JPEG2000, Objective Assessment, Picture Quality Measures, SPIHT.**

I. INTRODUCTION

WITH the increasing use of multimedia technologies, image compression requires higher performance. To address needs and requirements of multimedia and Internet applications, many efficient image compression techniques, with considerably different features, have recently been developed. Image compression techniques exploit a common characteristic of most images that the neighbouring picture elements (pixels, pels) are highly correlated [1]. It means that a typical still image contains a large amount of spatial redundancy in plain areas where adjacent pixels have almost the same values. In addition, still image can contain subjective redundancy, which is determined by properties of human visual system (HVS). HVS presents some tolerance to distortion depending upon the image content and viewing conditions. Consequently, pixels must not always be reproduced exactly as originated and HVS will not detect the difference between original image and reproduced image [2].

The evaluation of lossless image compression techniques is a simple task where compression ratio and execution time are employed as standard criteria. The picture quality before and after compression is unchanged. Contrary, the evaluation of lossy techniques is difficult task because of inherent drawbacks associated with both objective and subjective measures of picture quality. Objective measures of picture quality do not correlate well with subjective quality measures [3], [4]. Subjective assessment of picture quality is time consuming process and results of measurements should be processed very carefully.

In this paper we attempt to evaluate and compare objective and subjective picture quality measures. As test images we used images with different spatial and frequency characteristics. Images are coded using three different compression algorithms. The paper is structured as follows. In section 2 we define picture quality measures. In section 3 we briefly present image compression systems used in our experiment. In Section 4 we evaluate statistical and frequency properties of test images. Section 5 contains numerical results of picture quality measures. In this section we analyse correlation of objective measures with subjective grades and we propose objective measures which should be used in relation to each image compression system and objective measures which are suitable for the comparison of picture quality between different compression systems.

II. PICTURE QUALITY MEASURES

Among many objective numerical measures of picture quality, that are based on computable distortion measures, we have chosen those listed in Table I.

In our analysis digital image is represented as MxN matrix, where M denotes number of columns and N number of rows. $x_{j,k}$ and $x'_{j,k}$ denote pixel values of original image before compression and degraded image after compression.

Mean squared error (*MSE*) and Peak Signal to Noise Ratio (*PSNR*) are the most common measures of picture quality, despite the fact that they are not adequate as perceptually meaningful measures [5]. In addition to objective measures listed in Table I, we chose to use perception based objective evaluation, quantified by Picture Quality Scale (PQS) [6] and a perception based subjective evaluation, quantified by Mean Opinion Score (MOS) [7].

For the set of distorted images, the MOS values were obtained from an experiment involving 20 non-expert viewers. The testing methodology was the double-stimulus impairment scale method with five-grade impairment scale described in ITU-R BT Rec. 500 [7].

TABLE I
PICTURE QUALITY MEASURES

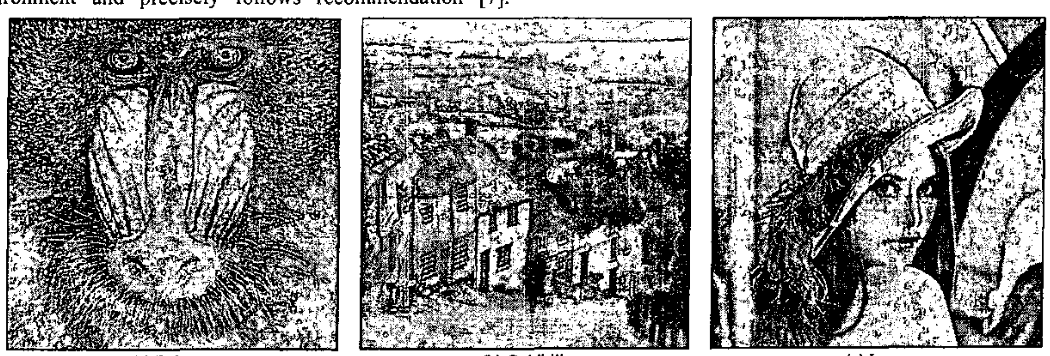
Mean Square Error	$MSE=\frac{1}{N}\sum_{j=1}^{M}\sum_{k=1}^{N}\left(x_{j,k}-x'_{j,k}\right)^2$
Peak Signal to Noise Ratio	$PSNR=10\log\frac{\left(2^n-1\right)^2}{MSE}=10\log\frac{255^2}{MSE}$
Normalised Cross-Correlation	$NK=\sum_{j=1}^{M}\sum_{k=1}^{N}x_{j,k}\cdot x'_{j,k}\Big/\sum_{j=1}^{M}\sum_{k=1}^{N}x_{j,k}^2$
Average Difference	$AD=\sum_{j=1}^{M}\sum_{k=1}^{N}\left(x_{j,k}-x'_{j,k}\right)/MN$
Structural Content	$SC=\sum_{j=1}^{M}\sum_{k=1}^{N}x_{j,k}^2\Big/\sum_{j=1}^{M}\sum_{k=1}^{N}x'^2_{j,k}$
Maximum Difference	$MD=Max\left(\left\lvert x_{j,k}-x'_{j,k}\right\rvert\right)$
Laplacian Mean Square Error	$LMSE=\sum_{j=1}^{M}\sum_{k=1}^{N}\left[O\left(x_{j,k}\right)-O\left(x'_{j,k}\right)\right]\Big/\sum_{j=1}^{M}\sum_{k=1}^{N}\left[O\left(x_{j,k}\right)\right]^2$ $O\left(x_{j,k}\right)=x_{j+1,k}+x_{j-1,k}+x_{j,k+1}+x_{j,k-1}-4x_{j,k}$
Normalised Absolute Error	$NAE=\sum_{j=1}^{M}\sum_{k=1}^{N}\left\lvert x_{j,k}-x'_{j,k}\right\rvert\Big/\sum_{j=1}^{M}\sum_{k=1}^{N}\left\lvert x_{j,k}\right\rvert$
Picture Quality Scale	$PQS=b_0+\sum_{i=1}^{3}b_iZ_i$

When the tests span the full range of impairments (as in our experiment) the double-stimulus impairment scale method should be used. The method uses the five-grade impairment scale with proper description for each grade: 5-imperceptible, 4-perceptible, but not annoying, 3-slightly annoying, 2-annoying and 1-very annoying. At the end of the series of sessions, MOS for each test condition and test image are calculated:

$$MOS=\sum_{i=1}^{5}i\cdot p(i) \tag{1}$$

where i is grade and $p(i)$ is grade probability.

To perform subjective assessment of picture quality we developed an application in Visual Basic, which enables equal viewing conditions for all viewers in our laboratory environment and precisely follows recommendation [7]. Viewing distance was $4\cdot H$, where H is image height displayed on monitor in full resolution. In addition to MOS, we used PQS methodology proposed in [6]. The PQS has been developed for evaluating the perceived quality of compressed images. It combines various perceived distortions into a single quantitative measure. To do so, PQS methodology uses some of the properties of HVS relevant to global image impairments, such as random errors, and emphasizes the perceptual importance of structured and localized errors. PQS is constructed by regressions with MOS, which is 5-level grading scale. PQS closely approximates the MOS in the middle of the quality range [8]. For very high quality images it is possible to obtain values of PQS larger than 5. At the low end of the image quality scale, PQS can obtain negative values (meaningless results).



(a) Baboon
$SFM=36.515$, $SAM=24.93$

(b) Goldhill
$SFM=16.167$, $SAM=126.77$

(c) Lena
$SFM=14.019$, $SAM=227.43$

Fig. 1. Test images

III. Compression Techniques

To produce test images for our objective and subjective picture quality assessments we used three different image compression systems: JPEG [9], [10], JPEG2000 [11], [12] and SPIHT [13]. JPEG (Joint Photographic Experts Group) corresponds to the ISO/IEC international standard 10928-1 for digital compression and coding of continuous-tone (multilevel) still images. Image compression scheme in JPEG is based on Discrete Cosine Transform (DCT) [14].

JPEG2000 should provide low bit-rate operation (below 0.25 bits/pixel) with subjective picture quality performance superior to existing standards. Image compression scheme in JPEG2000 Part I is based on discrete wavelet transform (DWT) [15], [16]. In our experiment JJ2000 implementation of JPEG2000 codec is used [17].

Set Partitioning in Hierarchical Trees (SPIHT) coding algorithm introduced by Said and Pearlman is a very efficient technique for wavelet image compression. SPIHT is improved and extended version of Embedded Zerotree Wavelet (EZW) coding algorithm introduced by J.M. Shapiro [18] and it is one of the best wavelet coder today.

IV. Test Images

The fundamental difficulty in testing image compression system is how to decide which test images to use for the evaluations. The image content being viewed influences the perception of quality irrespective of technical parameters of the system [19]. We have selected three test images (512×512, 8 bits/pixel) with different spatial and frequency characteristics: Baboon, Goldhill and Lena, Figure 1. The spatial frequency measure (*SFM*) indicates the overall activity level in an image [20]. *SFM* is defined as follows:

$$SFM = \sqrt{R^2 + C^2}\,,$$
$$R = \sqrt{\frac{1}{MN}\sum_{j=1}^{M}\sum_{k=2}^{N}\left(x_{j,k} - x_{j,k-1}\right)^2}\,, \qquad (2)$$
$$C = \sqrt{\frac{1}{MN}\sum_{k=1}^{N}\sum_{j=2}^{M}\left(x_{j,k} - x_{j-1,k}\right)^2}\,,$$

where R is row frequency, C is column frequency and $x_{j,k}$ denotes the samples of image; M and N are numbers of pixels in horizontal and vertical directions.

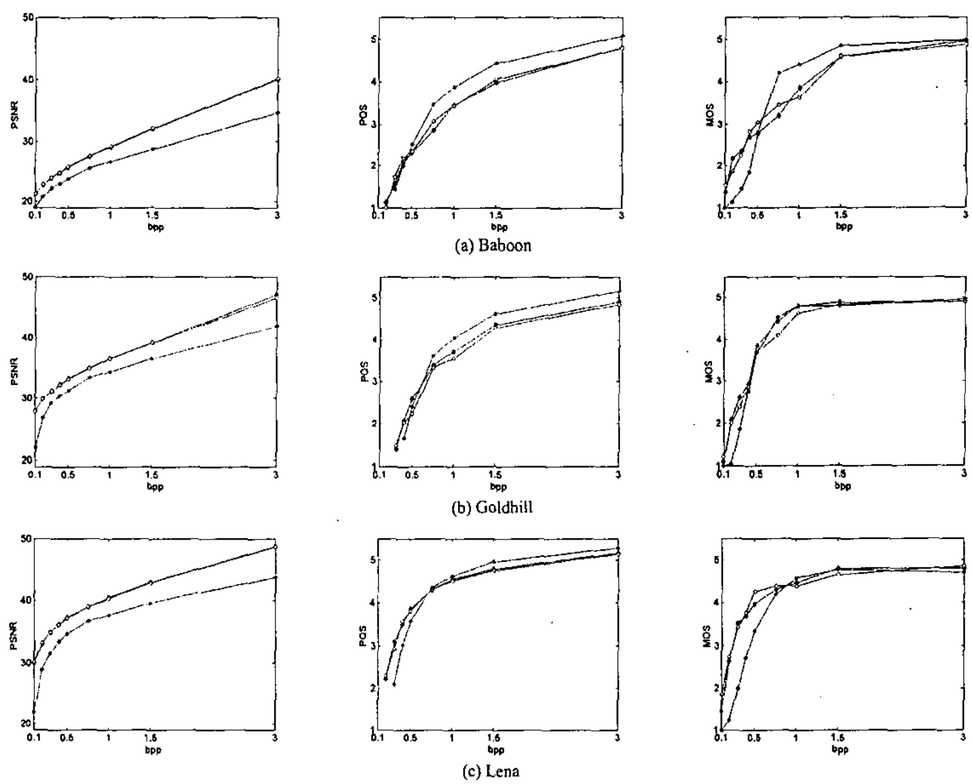


Fig. 2. PSNR (in dB), PQS and *MOS* results for test images (a) Baboon, (b) Goldhill, (c) Lena, and compression systems denoted as (♦ - JPEG2000; ○ - SPIHT; • - JPEG)

Spectral activity measure (*SAM*) is a measure of image predictability and it is evaluated in frequency domain [1]:

$$SAM = \frac{\frac{1}{M \cdot N}\sum_{j=0}^{M-1}\sum_{k=0}^{N-1}\left|F(j,k)\right|^2}{\left[\prod_{j=0}^{M-1}\prod_{k=0}^{N-1}\left|F(j,k)\right|^2\right]^{\frac{1}{M \cdot N}}}, \quad (3)$$

where $F(j,k)$ is (j,k)-th DFT coefficient of image. *SAM* has a dynamic range of $[1, \infty)$. Higher values of *SAM* imply higher predictability. Active images (*SAM* close to 1) are in general difficult to code.

Test image Baboon has large SFM and small SAM. Large value of SFM means that image contain components in high frequency area and small value of SAM means low predictability. For typical natural image largest value of SFM implies smaller value of SAM. Images Goldhill and Lena are images with less detail (smaller SFM) than Baboon.

V. Results

Test images Baboon, Lena and Goldhill are coded using JPEG, JPEG2000 and SPIHT. For each test image and compression method, nine different bit rates are selected: 0.1; 0.2; 0.3; 0.4; 0.5; 0.75; 1; 1.5 and 3 bits per pixel (bpp). Objective and subjective picture quality measures are calculated for all images. Results for *PSNR*, *PQS* and *MOS* are presented in Figure 2. For some very low quality images *PQS* is out of range.

PQS and MOS use the same quality scale so direct comparison between these two measures is possible. On the other hand, PSNR values depend very much on image content. For example, PSNR of image Lena is through all compression ratios for about 8-11 dB higher than PSNR for image Baboon. PSNR can not be used for quality comparison of different images. If we consider only *PSNR* values (Figure 2) we can conclude that JPEG2000 and SPIHT provides better picture quality than JPEG. If we take into account visual picture quality quantified by *MOS*, the conclusions are quite different. At high and moderate bitrates (above 0.75 bpp) for all test images JPEG produces better visual picture quality than wavelet-based techniques (JPEG2000 and SPIHT). At low bitrates (below 0.5 bpp) JPEG picture quality degrades below SPIHT and JPEG2000 picture quality, because of the artefacts introduced by block-based DCT scheme. It is clear example that PSNR can not be used as definitive picture quality measure. PQS grades follow the trend of MOS grades but MOS results show that human observers have more tolerance for moderately distorted images than PQS.

Table II shows the correlation between the numerical objective quality measures introduced in Table I and *MOS*. As a measure of the extent of the linear relationship, the Pearson product-moment (r) was used [20]. Correlation coefficient is defined as:

$$r = \frac{\sum(x_i - \bar{x})(x_i' - \bar{x}')}{\sqrt{\sum(x_i - \bar{x})^2 \sum(x_i' - \bar{x}')^2}} \quad (4)$$

where x and x' are two series between which correlation has to be found. The possible values of r are between -1 and +1; the closer r is to -1 or +1, the better the correlation is. The last row in the Table II contains average absolute values of correlation coefficients for each objective measure. The values of correlation coefficients indicate that commonly used measures of visual quality *PSNR* and MSE cannot be reliably used with all techniques, because they have poor correlation with *MOS*. *PQS* incorporates model of HVS and leads to the best correlation with *MOS* for all three compression systems and all test images but it needs too much time to be evaluated (approximately 15 sec per image in our test). Beside *PQS*, measures with good correlation with *MOS* are *MD*, *LMSE*, *NAE* and *NK* (see average absolute values of r in Table II). *MSE*, *PSNR* and *SC* cannot be reliably used with all techniques, because they have poor correlation with *MOS* for some of them. The poorest correlation has *AD*.

To evaluate usefulness of each quality measure in tested compression systems we found average absolute values of correlation coefficients for each compression system. Results are shown in Table III. Table III indicates that *PQS* is excellent measure of picture quality for all compression systems. In JPEG2000 and SPIHT compression systems *MSE* should be used instead of *PSNR* because of its better correlation with *MOS*. For JPEG2000 and SPIHT compression systems MD, MSE, LMSE and NAE measures demonstrate very good results. For JPEG compression system good results are achieved using PSNR, MD and LMSE.

In some image coding application, it is not appropriate to compute PQS because of its time expensiveness. Maximum difference (*MD*) has a good correlation with MOS for all tested compression techniques so we propose this very simple measure as a reference for measuring compressed picture quality in different compression systems. LMSE has also good correlation with MOS but this measure is not so simple as MD (see equations in Table I for MD and LMSE).

VI. Conclusion

The results of an evaluation concerning the usefulness of a number of objective quality measures in image compression systems have been presented. In addition, picture quality is measured subjectively using perceived picture quality. The correlation between each objective measure and subjective measure is found. We demonstrated that for a given compression system a group of numerical objective measures could reliably be used to specify the magnitude of degradation in reconstructed images. We also demonstrated that this group of objective measures is different for different compression systems. We proved that MSE and PSNR, as traditionally used objective measures of picture quality, are not adequate as perceptually meaningful measures in all tested compression systems. We found out that *PQS* is the most correlated measure with *MOS* for all compression techniques. In some image compression application, it is not possible to compute PQS because of its time expensiveness. So we considered other objective measures of picture quality for each compression technique and we found that maximum difference (*MD*) has a good correlation with MOS for all tested compression techniques. So, we propose use of *MD* for comparison of picture quality across different compression systems because of its good correlation with *MOS* and computing simplicity.

TABLE II
CORRELATION COEFFICIENTS FOR EACH COMPRESSION TECHNIQUE AND TEST IMAGE

Test Image	Codec	*MSE*	*PSNR*	*AD*	*SC*	*NK*	*MD*	*LMSE*	*NAE*	*PQS*
Baboon	JPEG2000	-0.95506	0.94071	0.84244	-0.92320	0.94264	-0.97625	-0.98814	-0.98912	0.99054
	SPIHT	-0.96117	0.93149	-0.50867	-0.95273	0.95803	-0.98784	-0.98947	-0.98905	0.98951
	JPEG	-0.89973	0.88491	0.38663	-0.94490	0.92001	-0.91574	-0.90406	-0.93453	0.97878
Goldhill	JPEG2000	-0.97097	0.83227	0.69434	-0.96565	0.97000	-0.96746	-0.93256	-0.95399	0.97033
	SPIHT	-0.96723	0.86626	-0.80212	-0.97618	0.97247	-0.96992	-0.96155	-0.96573	0.94765
	JPEG	-0.74839	0.89574	0.50969	0.42746	0.80153	-0.94067	-0.90946	-0.84936	0.97073
Lena	JPEG2000	-0.98327	0.88481	-0.71231	-0.95673	0.97326	-0.97585	-0.95612	-0.97600	0.99111
	SPIHT	-0.97636	0.88636	0.85532	-0.97525	0.97609	-0.97765	-0.95429	-0.97363	0.98234
	JPEG	-0.68045	0.93077	-0.43969	0.56470	0.80798	-0.85259	-0.90469	-0.78024	0.98867
Average absolute values of *r*		0.90	0.89	0.64	0.85	0.92	0.95	0.94	0.93	0.98

TABLE III
AVERAGE ABSOLUTE VALUES OF CORRELATION COEFFICIENTS FOR EACH COMPRESSION SYSTEM

Codec	*MSE*	*PSNR*	*AD*	*SC*	*NK*	*MD*	*LMSE*	*NAE*	*PQS*
JPEG2000	0.97	0.89	0.75	0.95	0.96	0.97	0.96	0.97	0.98
SPIHT	0.97	0.89	0.72	0.97	0.97	0.98	0.97	0.98	0.97
JPEG	0.78	0.9	0.45	0.65	0.84	0.9	0.91	0.85	0.98

REFERENCES

[1] N. Jayant, P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, Prentice Hall: Washington, 1984.

[2] N. Jayant, J. Johnston, R. Safranek, "Signal Compression Based on Models of Human Perception", *Proc. of IEEE*, vol. 81, pp. 1385-1422, Oct. 1993.

[3] S. Bauer, B. Zovko-Cihlar, and M. Grgic, "The Influence of Impairments from Digital Compression of Video Signal on Perceived Picture Quality", *Proceedings of the 3rd International Workshop on Image and Signal Processing*, IWISP'96, Manchester, pp. 245-248, Nov. 1996.

[4] S. Bauer, B. Zovko-Cihlar, M. Grgic, "Objective and Subjective Evaluations of Picture Quality in Digital Video Systems", *International Conference on Multimedia Technology and Digital Telecommunication Services*, ICOMT '96, Budapest, Hungary, pp. 145-150, Oct. 1996.

[5] S. Grgic, M. Grgic, B. Zovko-Cihlar, "Picture Quality Measurements in Wavelet Compression System", *Proceedings of the International Broadcasting Convention*, IBC'99, Amsterdam, The Netherlands, pp. 554-559, Sep. 1999.

[6] M. Miyahara, K. Kotani, and V. R. Algazi, "Objective Picture Quality Scale (PQS) for Image Coding", http://info.cipic.ucdavis.edu/scripts/reportPage?96-12

[7] ITU, "Methodology for the Subjective Assessment of the Quality of Television Pictures", *ITU-R Rec. BT. 500-9*, 1998.

[8] S. Grgic, M. Grgic, B. Zovko-Cihlar, "Performance Analysis of Image Compression Using Wavelets", *IEEE Transactions on Industrial Electronics*, vol. 48, issue 3, pp. 682-695, June 2001.

[9] ISO/IEC IS 10918, "Digital Compression and Coding of Continuous Tone Still Images", 1991.

[10] G. K. Wallace, "The JPEG Still Picture Compression Standard", *Communication of the ACM*, vol. 34, no. 4, pp. 30-44, 1991.

[11] ISO/IEC FDIS 15444-1, "JPEG2000 Part I Final Draft International Standard", Aug. 2000.

[12] N. Skodras, C. A. Christopoulos, and T. Ebrahimi, "JPEG2000: The Upcoming Still Image Compression Standard", *Proceedings of the 11th Portuguese Conference on Pattern Recognition*, Porto, Portugal, pp. 359-366, May 2000.

[13] A. Said and W. A. Pearlman, "A New, Fast, and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243-249, June 1996.

[14] K. R. Rao, P. Yip, *Discrete Cosine Transform: Algorithms, Advantages and Applications*, Academic Press: San Diego, 1990.

[15] M. Antonini, M. Barland, P. Mathieu, and I. Daubechies, "Image Coding Using the Wavelet Transform", *IEEE Trans. on Image Processing*, vol. 1, pp. 205-220, 1992.

[16] S. Grgic, K. Kers, and M. Grgic, "Image Compression Using Wavelets", *Proceedings of the IEEE International Symposium on Industrial Electronics*, ISIE'99, Bled, Slovenia, pp. 99-104, 1999.

[17] JJ2000 project by EPFL, Ericsson and CRF, World Wide Web: http://jj2000.epfl.ch/

[18] J. M. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients", *IEEE Transactions on Signal Processing*, vol. 41, pp. 3445-3462, Dec. 1993.

[19] S. Grgic, M. Mrak and M. Grgic, "Comparison of JPEG Image Coders", *Proceedings of the 3rd International Symposium on Video Processing and Multimedia Communications, VIPromCom-2001*, Zadar, Croatia, pp. 79-85, June 2001.

[20] A. M. Eskicioglu and P. S. Fisher, "Image Quality Measures and Their Performance", *IEEE Trans. on Communications*, vol. 43, no. 12, pp. 2959-2965, Dec. 1995.