QUALITY MEASUREMENT FOR MONOCHROME COMPRESSED IMAGES IN THE PAST 25 YEARS

Ahmet M. Eskicioglu

Thomson Consumer Electronics
101 West 103rd Street, Indianapolis, IN 46290, USA

ABSTRACT

While lossy image compression techniques are vital in reducing bandwidth and storage requirements, they result in distortions in compressed images. A reliable quality measure is a much needed tool for determining the type and amount of image distortion. The traditional subjective criteria, which involve human observers, are inconvenient, time-consuming, and influenced by environmental conditions. Widely used pixelwise measures such as the mean square error (MSE) cannot capture the artifacts like blurriness or blockiness, and do not correlate well with visual error perception. Attempts to improve quality measurement include incorporation of simple models of the human visual system (HVS) and multi-dimensional tool design. We review the criteria for monochrome compressed image quality from 1974 to 1999.

1. INTRODUCTION

In the past few decades, many compression techniques have been developed for efficient transmission and storage of images [1-5]. Most of these techniques make us of transform coding, vector quantization, fractals, or subband/wavelet coding for removing psychovisual and statistical image redundancies. As the bit rate is decreased, each compression technique introduces a known artifact, creating blocky, blurry, patchy or smudgy images.

Automated quantitative metrics are essential in testing and evaluating various image compression techniques or devices. Ideally, a useful metric should be easy to compute, independent of viewing distance, and able to quantify all types of image artifacts. Image quality measurement is still an unsolved problem today. New studies exploiting certain aspects of the HVS report reasonable success in quantifying certain types of distortion, providing higher correlation with subjective ranking.

2. IMAGE QUALITY MEASURES

In light of recent developments, it is possible to classify image quality criteria as shown in Table 1.

Subjective	**Quantitative**	
	Numerical	**Graphical**
Absolute Comparative	MSE Lp-norm Power spectrum Other	VDP Histograms Hosaka plots Eskicioglu charts

Table 1. Classification of image quality criteria

2.1 Subjective Criteria

As the ultimate assessment of image quality is made by human observers, the most reliable judgment is based on human perception. In subjective rating, a group of non-experts and experts may be used. Non-experts, representing average viewers, have little or no technical background in image compression, while experts are trained individuals having much familiarity with the relevant compression technologies.

Evaluation performed by the observers take two forms: Absolute and comparative. Absolute evaluation is a process whereby the observer assigns to an image a category in a given rating scale, whereas comparative evaluation is the ranking of a set of images from best to worst. The following rating scales have been used by researchers [6-11]:

5	Excellent	10, 9	Very good
4	Good	8, 7	Good
3	Fair	6, 5, 4	Fair
2	Poor	3, 2	Bad
1	Unsatisfactory	1, 0	Very bad
7	Best	1	Much better
6	Well above average	2	Better
5	Slightly above average	1	Slightly better
4	Average	0	Same
3	Slightly below average	-1	Slightly worse
2	Well below average	-2	Worse
1	Worst	-3	Much worse
1	Not noticeable		
2	Just noticeable		
3	Definitely noticeable but only slight improvement		
4	Impairment not objectionable		
5	Somewhat objectionable		
6	Definitely objectionable		
7	Extremely objectionable		

Bubble sort [6, 11, 12] is another technique used in rating images. The subject takes two images A and B from a group, and compares them. If his order is AB, he picks a third image to establish the order ABC or ACB. If the order is ACB, then a comparison is needed between A and C. The procedure ends with the best image at the top if no ties are allowed.

In many cases, subjective rating results may not be reproducible as they can be affected by a number of factors including:

a) type, size and range of images
b) observers' background and motivation,
c) experimental conditions (lighting, display quality, etc.).

2.2 Quantitative Criteria

Quantitative measures for image quality can be classified according to two criteria:

(1) the number of images used in the measurement;
(2) the nature or type of measurement.

According to the first criterion, the measures are divided into two classes: univariate and bivariate [10]. A univariate measure uses a single image, whereas a bivariate measure is a comparison between two images. According to the second criterion, there are again two classes: numerical and graphical. A numerical measure takes one (if the measure is univariate) or two (if the measure is bivariate) images as input, and processes the pixel values by an integration rule. The output of this processing is a single integer or real number. Graphical measures (univariate or bivariate) differ from numerical ones in that they do not reduce their output into a scalar value. They are multi-dimensional measures in the form of images, histograms, plots, or charts.

Quality measurements are usually made using the pixel elements of digitized images. For more accurate assessment, a continuous image field can be generated by two-dimensional interpolation of the pixel matrix.

Metrics for image quality have been defined in either spatial or frequency domain. Two-dimensional discrete Fourier transform is a common tool for frequency domain analysis.

2.2.1 Numerical (scalar) Measures

Denoting the samples on the NxM original image field as F(j,k), a spatial domain, univariate quality rating may be expressed in general as

$$\Sigma\Sigma\, O\{F(j,k)\},$$

where O {·} is some operator.

Bivariate measures are more frequently used in practice. A number of measures have been defined to determine the closeness of the degraded and original image fields. The alternatives are NMSE, PMSE, LMSE, L_p-norm, normalized absolute error, image fidelity, discrete cross-correlation, correlation quality, structural content, average difference and maximum absolute difference [6, 7, 10, 13].

2.2.2 Application of HVS Models

The human visual system (HVS) is enormously complex with optical, synaptic, photochemical, and electrical phenomena [14-16]. It has been the subject of study in several disciplines including biology, anatomy, psychology, and physiology. Two primary sources of information for developing a model of the HVS are psychophysical and physiological studies.

Image quality assessment can be improved by incorporating a model of the HVS into the evaluation process. Three distinct approaches have appeared in the literature:

- The MSE (or one of its variants) is modified be attaching a weight to the image samples in the frequency domain [11, 12, 17-20].
- The digital image power spectrum is weighted [21].
- The HVS is mimicked through several nonlinearities and filters [7, 22-34].

2.2.3 Graphical Measures

In a qualitative evaluation of distorted images, the human observer makes decisions concerning:

(1) the amount of distortion,
(2) the type of distortion,
(3) the distribution of error.

To be able to represent these decisions, it is clear that more than a single number is needed. Graphical measures compute summary statistics which, in some way, reflect the most important features in reconstructed versions of images. They provide an alternative means of mimicking the HVS.

The graphical measures that have general applicability in the field of image compression are:

- VDP (Visual Differences Predictor) [35],
- Histogram of the compression error [36, 37],
- Hosaka plots [38, 39],
- Eskicioglu charts [40-42].

2.3 Surveys on image quality measurement

We now mention four survey papers that discuss measures on image quality. Three of these also give comparative evaluations of selected measures.

Beaton [43] reports an evaluation of fourteen image quality metrics for hard-copy and soft-copy displays of digital images degraded by various levels of noise and blur. All the quality metrics were formulated to include the displayed modulation spectrum of the image. The digital image database used in this work consists of 250 digitally derived images from low altitude aerial photographs of domestic facilities. There are 10 unique scenes (4 naval ports, 4 airports, and 2 research and development facilities), each represented by 25 degradation levels. The degradation levels represent the effects of factorial combination of five static noise levels and five blur levels. Each digital image is a 4096 x 4096 pixel matrix with 256 levels of gray. For the hard-copy study, the digital images were presented on positive film transparencies. For the soft-copy study, the digital images were displayed on a CRT system. The subjective scaling task was carried out by military photo interpreters using a NATO-standardized scale.

A brief review of computational image quality metrics is given by Ahumada [44]. It starts with a framework for image quality measurements: For a given pair of images I_0 and I_1, the visual model P computes lists of visual system outputs $P(I_0)$, $P(I_1)$ from the images. The integration rule $Q(P(I_0), P(I_1))$ then gives the distance between pairs of perceptual outputs. A discussion of the features of the perceptual output and the integration rule is followed by a table of the models proposed by a number of researchers.

Eskicioglu, Fisher and Chen [37] present an extensive analysis of the performance of a number of image quality measures. Two of these measures are graphical (histograms and Hosaka plots), and the others are numerical. Four compression techniques were used in the study; JPEG and RLPQ induced blockiness, whereas EPIC and SLPQ led to blurriness. The 512 x 512 test images, Lenna (NITF6), Gilbert (US Navy) and Fingerprint (NITF5) had 256 gray levels. There were 84 images (3 test images x 4 compression codes x 7 degradation levels) in the image data set. The compression ratios ranged from 10:1 to 69:1, with an increment of about 10. Photographic samples of the degraded images were subjectively evaluated by ten observers chosen from the graduate students and faculty members.

Two experiments, one threshold and one suprathreshold, for evaluating psychophysical distortion metrics in JPEG encoded images are described by Fuhrmann, Baro and Cox, Jr. [45]. The threshold experiment was designed to determine the bit rate at which the observer is able to distinguish a distorted image from the original. In the suprathreshold experiment, image subblocks were ranked by observers according to perceived distortion. For these experiments, which involved eleven observers, gray-scale images were digitized at 8 bits/pixel. Using 512x512 original images, compressed images were generated at bit rates ranging from 0.1 to 4.0 with increments of 0.1. Subblocks of size 128x128 were chosen to mitigate the effects of attention shifting in larger images.

3. CONCLUSIONS

This survey covers image quality measurement in the past twenty-five years. Subjective criteria and the MSE, the least favorable of all tools, are indispensable in evaluating compression algorithms or real products because of a lack of a reliable quality metrics.

Recent research has focused on understanding the human vision through psychophysical and physiological experiments. There is enough evidence to show that simple HVS models incorporated into numerical quality measures result in higher correlation with subjective assessment. A typical approach is to obtain useful statistics (activity in blocks, errors on edges, block distortions, etc.) about the impairments in a compressed image. These statistics are then combined in a weighted sum to represent the error characteristics. Reduction of the output to a single value, however, is a major drawback because much of the useful information is lost. In addition, some of the numerical measures exploiting the HVS have limited use and scope as they are able to detect only a particular type of distortion (mostly blockiness).

Graphical measures attempt to simulate the HVS by introducing additional dimensions in quality assessment. Also referred to as multi-dimensional measures, they are shown to be more useful than scalar ones (with or without HVS weighting), since they provide information concerning the type of impairment and/or the locality of the error. Being in its infancy stage, this approach is believed to hold considerable promise, especially in optimizing image coders.

Review of color images and video sequences will be the subject of another study as they include color and temporal components.

4. REFERENCES

[1] T. C. Bell, J. G. Cleary and I. H. Witten, *Text Compression*, Prentice-Hall, Inc., USA, 1990.

[2] J. A. Storer (Editor), *Image and Text Compression*, Kluwer Academic Publishers, USA, 1992.

[3] R. J. Clarke, *Digital Compression of Still Images and Video*, Academic Press, England, 1996.

[4] V. Bhaskaran and K. Konstantinides, *Image and Video Compression Standards*, Kluwer Academic Publishers, USA, 1996.

[5] K. R. Rao and J. J. Hwang, Techniques and Standards for Image, Video and Audio Coding, Prentice Hall, Inc, USA, 1996.

[6] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, Inc., USA, 1989.

[7] J. O. Limb, “Distortion Criteria of the Human Viewer,” *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 9, No. 12, pp. 778-793, December 1979.

[8] A. N. Netravali and J. O. Limb, “Picture Coding: A Review,” *Proceedings of the IEEE*, Vol. 68, No. 3, pp. 366-406, March 1980.

[9] F. X. J. Lukas and Z. L. Budrikis, “Picture Quality Prediction Based on a Visual Model,” *IEEE Transactions on Communications*, Vol. 30, No. 7, pp. 1679-1692, July 1982.

[10] W. K. Pratt, *Digital Image Processing*, John Wiley and Sons, Inc., USA, 1978.

[11] J. L. Mannos and D. L. Sakrison, “The Effects of a Visual Fidelity Criterion on the Encoding of Images,” *IEEE Transactions on Information Theory*, Vol. 20, No. 4, pp. 525-536, July 1974.

[12] C. F. Hall, “Subjective Evaluation of a Perceptual Quality Metric,” *Proceedings of SPIE*, Vol. 310, pp. 200-204, 1981.

[13] H. L. Snyder, “Image Quality: Measures and Visual Performance,” *Flat-Panel Displays and CRTs*, L. E. Tannas, Jr., Ed., Van Nostrand Reinhold, New York, pp. 70-90, 1985.

[14] R. L. De Valois and K. K. Valois, *Spatial Vision*, Oxford University Press, Inc., New York, 1988.

[15] M. A. Ali and M. A. Klyne, *Vision in Vertebrates*, Plenum Press, New York, 1985.

[16] I. P. Howard, *Human Visual Orientation*, John Wiley and Sons Ltd., 1982.

[17] N. B. Nill, “A Visual Model Weighted Cosine Transform for Image Compression and Quality Assessment,” *IEEE Transactions on Communications*, Vol. 33, No. 6, pp. 551-557, June 1985.

[18] J. A. Saghri, P. S. Cheatham and A. Habibi, “Image Quality Measure Based on a Human Visual System Model,” *Optical Engineering*, Vol. 28, No. 7, pp. 813-818, July 1989.

[19] J. Farrell, H. Trontelj, C. Rosenberg and J. Wiseman, “Perceptual Metrics for Monochrome Image Compression,” *Society for Information Display International Symposium Digest of Technical Papers*, Vol. 22, pp. 631-634, 1991.

[20] G. G. Kuperman and B. L. Wilson, "Objective and Subjective Assessment of Image Compression algorithms," Ibid., pp. 627-630, 1991.

[21] N. B. Nill and B. H. Bouzas, "Objective Image Quality Measures Derived From Digital Image Power Spectra," *Optical Engineering*, Vol. 31, No. 4, pp. 813-825, April 1992.

[22] H. Marmolin, "Subjective MSE Measures," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 16, No. 3, pp. 486-489, May/June 1986.

[23] C. S. Stein, A. B. Watson and L. E. Hitchner, "Psychophysical Rating of Image Compression Techniques," *Proceedings of SPIE*, Vol. 1977, pp. 198-208, 1989.

[24] V. R. Algazi, Y. Kato, M. Miyahara and K. Kotani, "Compression of Image Coding Techniques with a Picture Quality Scale," *Proceedings of SPIE*, Vol. 1771, pp. 396-405, 1992.

[25] A. A. Webster, C. T. Jones, M. H. Pinson, S. D. Voran and S. Wolf, "An Objective Video Quality assessment System Based on Human Perception," *Proceedings of SPIE*, Vol. 1913, 1993.

[26] S. Wolf, M. Pinson, C. Jones and A. Webster, "A Summary of Methods of Measurement for Objective Video Quality Parameters Based on the Sobel Filtered Image and the Motion Difference Image," *U.S. Department of Commerce, National Telecommunications and Information Administration, Institute for Telecommunication Sciences, Boulder, CO, USA, Document No. T1A1.5/93-152*, November 8, 1993.

[27] A. M. Eskicioglu and P. S. Fisher, "The Variance of the Difference Image: An Alternative Quality Measure," *Proceedings of the International Picture Coding Symposium PCS '94*, Sacramento, CA, USA, September 21-23, pp. 88-91, 1994.

[28] W. Xu and G. Hauske, "Picture quality evaluation based on error segmentation," *Proceedings of SPIE*, Vol. 2308, pp. 1454-1465, 1994.

[29] H. Hamada and S. Namba, "A study on objective picture quality scales for pictures digitally encoded for broadcast," *IEICE Transactions on Communications*, Vol. E77-B, No. 12, December 1994.

[30] S. J. P. Westen, R. L. Lagendijk and J. Biemond, "Perceptual image quality based on a multiple channel HVS model," *Proceedings of the 1995 International Conference on Acoustics, Speech and Signal processing*, Vol. 4, Detroit, Michigan, USA, pp. 2351-2354, May 9-12, 1995.

[31] S. A. Karunasekera and N. G. Kingsbury, "A Distortion Measure for Blocking Artifacts in Images Based on Human Visual Sensitivity," *IEEE Transactions on Image Processing*, Vol. 4, No. 6, pp. 713-724, June 1995.

[32] M. Kazubek, A. Przelaskowski and T. Jamrógiewicz, "Quality measurement of compressed medical images: Block effect measures," *Medical & Biological Engineering & Computing, Vol. 34, Supplement I, Part I, Proceedings of the 10th Nordic-Baltic Conference on Biomedical Engineering, Medical and Biological Engineering & Computing*, pp. 235-236, 1996.

[33] P. Franti, "Blockwise distortion measure for statistical and structural errors in digital images," *Signal Processing: Image Communication*, Vol. 13, pp. 89-98, 1998.

[34] T. Eude, A. Mayache and C. Milan, ""A psychovisual quality metric based on multiscale texture anaysis," *Proceedings of SPIE*, Vol. 3644, pp. 235-244, 1999.

[35] S. Daly, "The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity," *Proceedings of SPIE*, Vol. 1616, pp. 2-15, 1992.

[36] W. D. Abbott III, R. T. Kay and R. J. Pieper, "Performance Considerations for the Application of the Lossless Browse and Residual Model," *Proceedings of 1994 Space and Earth Science Data Compression Workshop* (NASA Conference Publication 3258), University of Utah, Salt Lake City, UT, USA, pp. 43-54, April 2, 1994.

[37] A. M. Eskicioglu, P. S. Fisher and S. Chen, "Image Quality Measures and Their Performance," *IEEE Transactions on Communications*, Vol. 43, No. 12, pp. 2959-2965, December 1995.

[38] K. Hosaka, "A New Picture Quality Evaluation Method," *Proceedings of the International Picture Coding Symposium PCS '86*, Tokyo, Japan, pp. 17-18, April 1986.

[39] P. M. Farrelle, "*Recursive Block Coding for Image Data Compression*," Springer-Verlag New York, Inc., pp. 104-154, 1990.

[40] A. M. Eskicioglu, "A Multi-Dimensional Measure for Image Quality," *Proceedings of 1995 Space and Earth Science Data Compression Workshop* (JPL Publication 95-8), University of Utah, Salt Lake City, UT, USA, pp. 83-92, March 27, 1995.

[41] A. M. Eskicioglu, "An Improved Graphical Quality Measure for Monochrome Compressed Images," *Proceedings of Optical Engineering Midwest '95*, Illinois Institute of Technology, Chicago, IL, USA, pp. 692-701, May 18-19, 1995.

[42] A. M. Eskicioglu, "Application of Multi-Dimensional Measures to Reconstructed Medical Images," *Optical Engineering*, Vol. 35, No. 3, pp. 778-785, March 1996.

[43] R. J. Beaton, "Quantitative Models of Image Quality," *Proceedings of the Human Factors Society - 27th Annual Meeting*, pp. 41-45, 1983.

[44] A. Ahumada, "Computational Image Quality Metrics: A Review," *Society for Information Display Digest of Technical Papers*, Seattle, WA, USA, pp. 305-308, 1993.

[45] D. R. Fuhrmann, J. A. Baro and J. R. Cox, Jr., "Experimental evaluation of psychophysical distortion metrics for JPEG-encoded images," *Journal of Electronic Imaging*, Vol. 4, No. 4, pp. 397-406, October 1996.