

COMPUTATIONAL IMAGE QUALITY METRICS: A REVIEW

A. J. Ahumada, Jr.

NASA Ames Research Center, Moffett Field, CA

Abstract

A large number of computational methods for assessing the quality of static images have been proposed. They differ along many dimensions, especially their range of application areas and complexity. This review is intended to help the researcher in search of a suitable metric.

Introduction

Several recent projects associated with our laboratory have needed computational methods for assessing image quality. 1) A display modeling system needed a method for comparing proposed displays and evaluating the quality of its simulations (Martin, Ahumada, & Larimer, 1992). 2) A halftoning optimizing method needed a metric for evaluating the quality of its halftones (Mulligan & Ahumada, 1992a). 3) A DCT image compression project needed a method for measuring visibility of compression artifacts (Ahumada & Peterson, 1992; Watson, 1993). 4) A sensor fusion workstation project needed methods for evaluating the quality of radar and IR enhanced displays (Pavel, Larimer, & Ahumada, 1992). The goal of this paper will be to try to develop a framework for describing image quality metrics, summarize the methods that we have researched, describe some of the relationships between the methods, and provide recent entry references to the literature.

The scope of the paper is limited to static monochrome images and to methods that compute a distance function between two images, usually an original image and a corrupted version of it. This would still include all computational models of visual discrimination, so the scope is also restricted to vision models applied to image quality. Much image quality work has focussed instead on the properties of the image display system and provide what one might regard as estimates of display system channel capacity after the human visual system is included in the system (Barten, 1990; 1992). An alternative approach to system performance could be based on the approach of the metrics considered here. One could assess the corruption by a system to individual images and

then average over the class of expected images. This is more in the spirit of the noisy image quality measures, such as that of Barrett (1992).

Framework

The present framework provides for a pair of images, I_0, I_1 . The visual model P computes lists of visual system outputs $P(I_0), P(I_1)$ from the images. The integration rule $Q(P(I_0), P(I_1))$ gives the "distance" between pairs of perceptual outputs. For barely distinguishable perceptual outputs, Q should be monotonically related to the probability that an observer would see them as different.

For example, the optimum halftoning method mentioned above needs a very rapid method for evaluating the quality of its halftones. The perceptual output P is a lowpass (blurred) version of the image I and the integration rule $Q(P(I_0), P(I_1))$ is the square of the Euclidean or RMS distance between them,

$$\begin{aligned} Q &= |P(I_0) - P(I_1)|^2 \\ &= \sum_x \sum_y (p_{0,x,y} - p_{1,x,y})^2, \end{aligned}$$

where the $p_{i,x,y}$ are pixel values in the blurred images.

A useful simplification results when the perceptual output is a linear function of the input image, as in the simple blur case, and the integration rule is a function of the difference between the perceptual images. In this case the integration rule can be applied to the difference image $\Delta I = I_1 - I_0$, since

$$\begin{aligned} &Q(P(I_1), P(I_0)) \\ &= Q(P(I_1) - P(I_0)) \\ &= Q(P(I_1 - I_0)) \\ &= Q(P(\Delta I)) \end{aligned}$$

Features of the perceptual output

The perceptual output function incorporates the main features of image processing by the visual system:

- 1) Optical blur
- 2) Photoreceptor sampling and transduction
- 3) Retinal local contrast enhancement and gain control
- 4) Masking which is luminance, contrast, spatial frequency, orientation, and location specific.

These effects are all to some degree image specific. The blur function depends on the pupil size; the effective sampling depends on whether rods are playing a role. All the effects except optical blur are to some extent nonlinear, but are often approximated by linear computations. For example, the local contrast and gain control mechanisms are considered to be more divisive than subtractive, but are most frequently modeled by a linear filter whose response falls off for spatial frequencies below 1 cycle per degree. The spatial frequency and orientation specificity of masking has led to the development of multi-channel models, where each point in space is multiply represented.

The integration rule

Many researchers use the Euclidean distance metric described above or a generalization of that metric, the Minkowski metric or distance (de Ridder, 1992)

$$Q = [\sum_j |p_{0,j} - p_{1,j}|^E]^{1/E} .$$

It gives the Euclidean distance for $E = 2$, approximates probability summation (Watson, 1979) when $E = 4$, and finds the maximum absolute difference when $E \rightarrow \infty$.

Although a simple Euclidean distance function (RMS error) is most frequently used, some researchers have found it necessary to include a nonlinear rectification stage, like squaring, to compute unsigned errors and then follow this with a local integration stage to represent summation among errors that are close together in space. This precedes the general combination rule to represent the combination of errors that are far away in space or in some other channel altogether.

Model linearization

A useful technique for simplifying nonlinear models for evaluating the effects of small image distortions from a base image I_0 is to linearize the model in the region of I_0 . We find a linear function $P_{I_0}(I)$ such that

$$P(I_0 + \Delta I) \approx P(I_0) + P'_{I_0}(\Delta I).$$

If the integration rule Q is a function of the difference of the two images,

$$\begin{aligned} & Q(P(I_0 + \Delta I) - P(I_0)) \\ &= Q(P(I_0) + P'_{I_0}(\Delta I) - P(I_0)) \\ &= Q(P'_{I_0}(\Delta I)) \end{aligned}$$

Ahumada (1987) has discussed this technique for simplifying nonlinear vision models and Girod (1989) has illustrated its usefulness in the application of vision models to image quality assessment. Notice that the image distortions have to be so small that they do not significantly enter into the masking process, which is assumed to be only a function of the base image I_0 . Only very low contrast images can have their artifact masking adequately represented by a linear filter depending only on the luminance.

The Table

Table 1 summarizes the models of a number of researchers in the image quality field. The entries are ordered by generally by complexity, and, except for the Grogan and Keene (1992) metric, the models can be regarded as simplified versions of the Lubin (1993) model. The topics are coded according to the following key: HT = halftoning, IC = image compression, IQ = image quality.

Perceptual Model Properties

Several properties are used to categorize the perceptual models. First is the number of perceptual images generated. For the models that are both spatial frequency and orientation selective, a pair of numbers is given. The first is the number of spatial frequency channels, the second is the number of orientation channels. The second property is the filter shape, B for bandpass, L for lowpass. A multiple channel model with 5 spatial frequency channels and 4 orientations will have the specification (5×4 B) even though one of the channels is lowpass. The number of channels is given as a guide to complexity. Larger numbers of channels could be used with larger images, but little change in computation occurs if the low frequency channels are adequately subsampled.

The next three properties are local properties. The inclusion of point nonlinearities before the final metric stage is represented by N. These are most often concave downward functions that increase dynamic range, but can also be S-shaped functions

that mimic the thresholding and saturation properties of neurons. The use of a local intensity measure to compute local contrast in the manner of Peli (1990) and Duval-Destin (1991) is represented by C. The use of a local activity measure is denoted by A. Like the local intensity measure it is used to reduce the signal. It can be either a local variance measure or some type of edge energy measure obtained by high pass filtering and rectification. Both the local contrast and local energy measures could be regarded as generating additional channels in the models, although in the Zetzche and Hauske (1989) Ratio-of-Gaussian model, no additional channels are needed.

Metric Properties

The metric property S indicates the use of local error summation preceding the final error aggregation step. The value of the Minkowski exponent is then given, or the use of actual probability summation is indicated by P.

Discussion

The Table illustrates that a large range of metrics have been proposed. The obvious question is whether so many are needed. Clearly, to the extent that some are just computational approximations of more complex ones, the variety seems reasonable. The question which is most puzzling is the extent to which features that are specifically added by some models are automatically included by others. Many metrics include a term to specifically account for increased masking near edges. Zetzche and Hauske (1989) showed that this is an emergent feature of nonlinear output scaling in their oriented frequency channel model. If more studies compared more models, the apparent convergence towards these channel models could be seen more as the result of facts than fads.

It has sometimes been reported that using a filter to represent the visibility of artifacts does worse than just using the RMS distance between the original images (Farrell, et al., 1991). This result suggests not that visibility can be ignored, but rather that the work done in developing nonlinear, image-dependent metrics has probably not been done in vain.

Acknowledgments

Assistance was provided by Loretta Hidalgo, supported by NASA Cooperative Agreement NCC 2-307 with Stanford University. Funding was

provided by NASA RTOP 505-64-53. Helpful feedback was provided by A. Watson and R. Samadani. This paper is a slightly revised version of Ahumada (1993).

References

- A. Ahumada, Jr., (1987) Putting the noise of the visual system back in the picture. *JOSA A*, **4**, 2372-2378.
- A. Ahumada, Jr. and H. Peterson, (1992) Luminance-Model-Based DCT Quantization for Color Image Compression. *SPIE Proc.*, **1666**, 365-374.
- A. Ahumada, Jr., (1993) Computational Image Quality Metrics: A Review. *SID Digest*, **24**, Paper 21-1.
- M. Analoui and J. Allebach, (1992) Model-based halftoning by direct binary search. *SPIE Proc.*, **1666**, 96-108.
- H. Barrett, (1992) Evaluation of Image Quality through Linear Discriminant Models. *SID Digest*, **23**, 871-873.
- P. Barten, (1990) Evaluation of subjective image quality with the square foot integral method. *JOSA A*, **7**, 2024-2031.
- P. Barten, (1992) The SQRI as a measure for VDU image quality. *SID Digest*, **23**, 867-870.
- Z. Budrikis, (1972) Visual fidelity criterion and modeling. *Proc. IEEE*, **60**, 771-779.
- P. Burt and E. Adelson, (1983) The Laplacian Pyramid as a Compact Image Code. *IEEE Trans.*, **COM-31**, 532-540.
- C. Chu and W. Watunyuta, (1992) On Designing Dither Matrices Using a Human Visual Model. *SPIE Proc.*, **1666**, 134-143.
- S. Daly, (1992) Visible differences predictor: an algorithm for the assessment of image fidelity. *SPIE Proc.*, **1666**, 2-14.
- M. Duval-Destin, (1991) A spatio-temporal complete description of contrast. *SID Digest*, **22**, 615-618.
- J. Farrell, H. Trontelj, C. Rosenberg, and J. Wiseman, (1991) Perceptual metrics for monochrome image compression. *SID Digest*, **22**, 631-634.
- B. Girod, (1989) The Information Theoretical Significance of Spatial and Temporal Masking in Video Signals. *SPIE Proc.*, **1077**, 178-187.
- N. Griswold, (1980) Perceptual coding in the cosine transform domain. *Optical Eng.*, **19**, 306-311.
- T. Grogan and D. Keene, (1992) Image Quality

- Evaluation with a Contour-Based Perceptual Model. *SPIE Proc.*, **1666**, 188-197.
- S. Klein, A. Silverstein, and T. Carney, (1992) Relevance of human vision to JPEG-DCT compression. *SPIE Proc.*, **1666**, 200-215.
- B. Kolpatzik and C. Bouman, (1992) Optimized Error Diffusion Based on a Human Visual Model. *SPIE Proc.*, **1666**, 165-176.
- J. Limb, (1979) Distortion Criteria of the Human Viewer. *IEEE Trans.*, **SMC-9**, 778-793.
- J. Lubin, (1993) The use of psychophysical data and models in the analysis of display system performance. In A. Watson (Ed.), *Visual Factors in Electronic Image Communications*, Cambridge, MA: MIT Press.
- F. Lukas and Z. Budrikis, (1982) Picture Quality Prediction Based on a Visual Model. *IEEE Trans.*, **COM-30**, 1679-1692.
- J. Mannos and D. Sakrison, (1974) The effects of a visual fidelity criterion on the encoding of images. *IEEE Trans.*, **IT-20**, 525-536.
- R. Martin, A. Ahumada, Jr., and J. Larimer, (1992) Color matrix display simulation based upon luminance and chromatic contrast sensitivity of early vision. *SPIE Proc.*, **1666**, 336-342.
- T. Mitsa, (1992) Evaluation of halftone techniques using psychovisual testing and quantitative quality measures. *SPIE Proc.*, **1666**, 177-187.
- J. Mulligan and A. Ahumada, Jr., (1992) Principled Halftoning Based on Models of Human Vision. *SPIE Proc.*, **1666**, 109-121.
- A. Netravali and B. Prasada, (1977) Adaptive quantization of picture signals using spatial masking. *Proc. IEEE*, **65**, 536-548.
- K. Ngan, K. Leong, and H. Singh, (1986) Cosine transform coding incorporating human visual system model. *SPIE Proc.*, **707**, 165-171.
- N. Nill, (1985) A visual model weighted cosine transform for image compression and quality assessment. *IEEE Trans.*, **COM-33**, 551-557.
- T. Pappas and D. Neuhoff, (1992) Least-squares model-based halftoning. *SPIE Proc.*, **1666**, 165-176.
- T. Pappas, (1992) Perceptual Coding and Printing of Gray-Scale and Color Images. *SID Digest*, **23**, 689-692.
- M. Pavel, J. Larimer, and A. Ahumada, (1992) Sensor Fusion for Synthetic Vision. *SID Digest*, **23**, 475-478.
- E. Peli, (1990) Contrast in complex images. *JOSA A*, **7**, 2032-2040.
- H. de Ridder, (1992) Minkowski-metrics as a combination rule for digital-image-impairments. *SPIE Proc.*, **1666**, 16-26.
- D. Sakrison, (1977) On the role of the observer and a distortion measure in image transmission. *IEEE Trans.*, **COM-25**, 1251-1267.
- E. Shlomot, Y. Zeevi, and W. Pearlman, (1987) The Importance of Spatial Frequency and Orientation in Image Decomposition and Coding. *SPIE Proc.*, **845**, 152-158.
- T. Stockham, (1972) Image processing in the context of a visual model. *Proc. IEEE*, **60**, 828-842.
- J. Sullivan, L. Ray, and R. Miller, (1991) Design of minimum visual modulation halftone patterns. *IEEE Trans.*, **SMC-21**, 33-38.
- A. Watson, (1979) Probability summation over time. *Vision Res.*, **19**, 515-522.
- A. Watson, (1987) Efficiency of an image code based on human vision. *JOSA A*, **4**, 2401-2417.
- A. Watson and A. Ahumada, Jr., (1989) A hexagonal orthogonal-oriented pyramid as a model of image representation in visual cortex. *IEEE Trans.*, **BE-36**, 97-106.
- A. Watson, (1993) DCT quantization matrices visually optimized for individual images. *SPIE Proc.*, **1913**.
- A. Zakhor, S. Lin, and F. Eskafi, (1992) A New Class of B/W Halftoning Algorithms. *SPIE Proc.*, **1666**, 122-133.
- C. Zetsche and G. Hauske, (1989) Multiple channel model for the prediction of subjective image quality. *SPIE Proc.*, **1077**, 209-216.

Table 1

Researchers	Topic	Perceptual Transformation					Metric		Comments
		Images	Filter	N	C	A	S	Exp	
Lubin, 1993	IQ	7×4	B	N	C	A	S	2.4	peripheral summation ratio pyramid
Zetzsche & Hauske, 1989	IQ	5×6	B	N	C			2	
Watson, 1993	IC	8×8	B	N	C			4	DCT basis
Sakrison, 1977	IC	12×9	B	N		A		6	early theory
Watson, 1987	IC	6×4	B	N				4	efficient subsampling
Watson & Ahumada, 1989	IC	3×6	B	N					hexagonal array
Daly, 1992	IQ	6×6	B	N				P	
Shlomot et al., 1987	IC	6×6	B						
Martin et al., 1992	IQ	9×3	B					4	Haar pyramid
Ahumada & Peterson, 1992	IC	8×8	B					∞	DCT basis
Klein et al., 1992	IC	8×8	B					2	DCT basis
Grogan & Keene, 1992	HT	8	B	N				2	
Burt & Adelson, 1983	IC	5	B						Laplacian pyramid
Stockham, 1972	IQ	1	B	N				2	
Mannos & Sakrison, 1974	IQ	1	B	N				2	
Ngan et al., 1986	IC	1	B		C				DCT basis
Girod, 1989	IC	1	B		C		S	2	linearized
Budrikis, 1972	IC	1	B			A		1,2	TV quality meter
Limb, 1979	IQ	1	B			A	S	2	options compared
Griswold, 1980	IC	1	B					2	
Nill, 1985	IC	1	B					2	
Analoui & Allebach, 1992	HT	1	B					2	
Mitsa, 1992	HT	1	B					2	
Lukas & Budrikis, 1982	IQ	1	L		C	A		2,4	
Zakhor et al., 1992	HT	1	L					2	
Chu & Watunyuta, 1992	HT	1	L					2	
Sullivan et al., 1991	HT	1	L					2	
Pappas & Neuhoff 1992	HT	1	L					2	
Pappas, 1992	HT	1	L					2	
Kolpatzik & Bouman, 1992	HT	1	L					2	
Mulligan & Ahumada, 1992	HT	1	L					2	
Netravali & Prasada, 1977	IC	1				A		2	no CSF