

N94-28257

## IMAGE QUALITY MEASURES AND THEIR PERFORMANCE

Ahmet M. Eskicioglu, Paul S. Fisher, and Siyuan Chen

University of North Texas

Department of Computer Sciences

P.O. Box 13886

Denton, TX 76203, USA

Telephone: (817) 565-2767

Fax: (817) 565-2799

E-mail: [eskiciog@ponder.csci.unt.edu](mailto:eskiciog@ponder.csci.unt.edu), [fisher@gab.unt.edu](mailto:fisher@gab.unt.edu),  
[schen@ponder.csci.unt.edu](mailto:schen@ponder.csci.unt.edu)

56-61

460

P-13

### Abstract

A number of quality measures are evaluated for gray scale image compression. They are all bivariate, exploiting the differences between corresponding pixels in the original and degraded images. It is shown that although some numerical measures correlate well with the observers' response for a given compression technique, they are not reliable for an evaluation across different techniques. The two graphical measures (histograms and Hosaka plots), however, can be used to appropriately specify not only the amount, but also the type of degradation in reconstructed images.

### 1. Introduction

The need for storing and transmitting huge volumes of data in today's computer and communications systems necessitates data compression in many fields ranging from medicine to aerospace. Data compression is an encoding process to reduce the storage and transmission requirements in applications. Many efficient techniques with considerably different features have recently been developed for both lossless and lossy compression. The evaluation of lossless techniques is normally a simple and straightforward task, where a number of standard criteria (compression ratio, execution time, etc.) are employed. A major problem in evaluating lossy techniques is the extreme difficulty in describing the type and amount of degradation in reconstructed images. Because of the inherent drawbacks associated with the subjective measures of image quality, there has been a great deal of interest in developing a quantitative measure, either in numerical or graphical form, that can consistently be used as a substitute. We would like to have such a measure not only to judge the quality of images obtained by a particular algorithm, but also for quality judgment across various algorithms. The latter task is definitely more challenging since a wide range of image impairments is involved. An extensive survey and a classification of the quality measures that appeared in the relevant literature are given in [1].

It is known that the mean square error (MSE), the most common objective criterion, or its variants do not correlate well with subjective quality measures. A major emphasis in recent research has therefore been given to a deeper analysis of the human visual system (HVS). The HVS is too complex to fully understand with present psychophysical means, but the incorporation of even a simplified model into objective measures reportedly leads to a better correlation with the response of the human observers.

We attempt to evaluate the usefulness of some of the objective quality measures listed in [1] through a set of experiments.

### 2. Image Quality Measures, Compression Techniques, and Test Images

The quality measures included in our evaluation are listed in Table 1. They are all discrete and bivariate, i.e., they provide some measure of closeness between two digital images by exploiting

the differences in the statistical distributions of pixel values.  $F(j, k)$  and  $\hat{F}(j, k)$  denote the samples of original and degraded image fields.

Table 1. Image Quality Measures

Average Difference	$AD = \sum_{j=1}^M \sum_{k=1}^N [F(j, k) - \hat{F}(j, k)] / MN$
Structural Content	$AC = \sum_{j=1}^M \sum_{k=1}^N [F(j, k)]^2 / \sum_{j=1}^M \sum_{k=1}^N [\hat{F}(j, k)]^2$
N. Cross-Correlation	$NK = \sum_{j=1}^M \sum_{k=1}^N F(j, k)\hat{F}(j, k) / \sum_{j=1}^M \sum_{k=1}^N [F(j, k)]^2$
Correlation Quality	$CQ = \sum_{j=1}^M \sum_{k=1}^N F(j, k)\hat{F}(j, k) / \sum_{j=1}^M \sum_{k=1}^N F(j, k)$
Maximum Difference	$MD = \text{Max}\{ F(j, k) - \hat{F}(j, k) \}$
Image Fidelity	$IF = 1 - \left( \sum_{j=1}^M \sum_{k=1}^N [F(j, k) - \hat{F}(j, k)]^2 / \sum_{j=1}^M \sum_{k=1}^N [F(j, k)]^2 \right)$
Weighted Distance	WD: Every element of the difference matrix is normalized in some way and $L_1$ -norm is applied [1].
Laplacian Mean Square Error	$LMSE = \sum_{j=1}^{M-1} \sum_{k=2}^{N-1} [O\{F(j, k)\} - O\{\hat{F}(j, k)\}]^2 / \sum_{j=1}^{M-1} \sum_{k=2}^{N-1} [O\{F(j, k)\}]^2$
Peak Mean Square Error	$PMSE = \frac{1}{MN} \sum_{j=1}^M \sum_{k=1}^N [F(j, k) - \hat{F}(j, k)]^2 / [\text{Max}\{F(j, k)\}]^2$
N. Absolute Error	$NAE = \sum_{j=1}^M \sum_{k=1}^N  O\{F(j, k)\} - O\{\hat{F}(j, k)\}  / \sum_{j=1}^M \sum_{k=1}^N  O\{F(j, k)\} $
N. Mean Square Error	$NMSE = \sum_{j=1}^M \sum_{k=1}^N [O\{F(j, k)\} - O\{\hat{F}(j, k)\}]^2 / \sum_{j=1}^M \sum_{k=1}^N [O\{F(j, k)\}]^2$
$L_p$ -norm	$L_p = \left\{ \frac{1}{MN} \sum_{j=1}^M \sum_{k=1}^N  F(j, k) - \hat{F}(j, k) ^p \right\}^{1/p}, p = 1, 2, 3$
Hosaka plot	A graphical quality measure. The area and shape of the plot gives information about the type and amount of degradation [1,6].
Histogram	Another graphical quality measure. Gives the probability distribution of the pixel values in the difference image.
<p>Note: For LMSE, <math>O\{F(j, k)\} = F(j+1, k) + F(j-1, k) + F(j, k+1) + F(j, k-1) - 4F(j, k)</math>. For NAE, NMSE, and <math>L_2</math>-norm, <math>O\{F(j, k)\}</math> is defined in three ways: (1) <math>O\{F(j, k)\} = F(j, k)</math>, (2) <math>O\{F(j, k)\} = F(j, k)^{1/3}</math>, (3) <math>O\{F(u, v)\} = H\{(u^2 + v^2)^{1/2}\} F(u, v)</math> (in cosine transform domain).</p>	

Among the few models of the HVS that have been developed, we chose the one proposed by Nill for dealing with cosine transforms. The function for the model is defined as [2]

$$H(r) = \begin{cases} 0.05r^{0.554}, & \text{for } r < 7 \\ e^{-9[(\log_{10} r - \log_{10} 9)]^{2.3}}, & \text{for } r \geq 7, \end{cases}$$

where  $r = (u^2 + v^2)^{1/2}$ , and  $u, v$  are the coordinates in the transform domain. The subimage structure weighting factor  $W_i$  in the original model was not used in our computations because we wanted to investigate the effect of  $H(r)$  alone. Since  $W_i$  is proportional to the intensity level variance of subimage  $i$ , a separate analysis is needed to determine a suitable proportionality constant.

Table 2 Image Compression Techniques

JPEG	Fourth public release of the Independent JPEG Group's JPEG software
EPIC	Vision Science Group, The Media Laboratory, MIT
RLPQ	Department of Computer Sciences, University of North Texas
SLPQ	Department of Computer Sciences, University of North Texas

The implementations of the image compression techniques are given in Table 2. Both JPEG and EPIC belong to the class of transform coding techniques. The former performs the discrete cosine transform and the latter a wavelet transform. RLPQ and SLPQ contain several modifications to the Laplacian pyramidal decomposition and use a loose wavelet basis. After quantization, they employ arithmetic coding with a specifically tuned adaptive predictive model to compress the pyramid.

It should be noted that the choice of the compression techniques for an investigation of the performance of quality measures (especially those that are graphical) is important since it is desirable to include techniques which produce different types of impairments in the reconstructed images. Our purpose is to see how well the measures are able to describe image distortions of unsimilar nature. As we shall discuss later, the four codes in Table 2 serve this purpose.

The information about the three test images that we used can be seen in Table 3. Lenna and Fingerprint are in the set of the National Imagery Format Test Images. The third image, hurricane Gilbert, was obtained from the U.S. Navy.

Table 3 Test Images

Image	Source	Size(bytesxbytes)	Pixel Length(bits)	Spatial Frequency
Lenna	NITF	512x512	8	14.07
Gilbert	US Navy	512x512	8	31.25
Fingerprint	NITF	512x512	8	59.37

The spatial frequency for a given image is defined as follows [3]:

Consider an  $M \times N$  image, where  $M$  = number of rows and  $N$  = number of columns. The row and column frequencies are given by

$$\text{Row\_Freq} = \sqrt{\frac{1}{MN} \sum_{j=0}^{M-1} \sum_{k=1}^{N-1} [F(j,k) - F(j,k-1)]^2}$$

and

$$\text{Column\_Freq} = \sqrt{\frac{1}{MN} \sum_{k=0}^{N-1} \sum_{j=1}^{M-1} [F(j,k) - F(j-1,k)]^2}$$

The total frequency is then

$$\text{Spatial frequency} = \sqrt{(\text{Row\_Freq})^2 + (\text{Column\_Freq})^2}.$$

This definition of frequency in the spatial domain indicates the overall activity level in an image.

### 3. Performance Of Quality Measures

The gray scale image data set was obtained by coding and decoding the three test images with the compression codes listed in Table 2. For each test image, seven different compression ratios were selected for degradation. They range from 10:1 to 70:1 with an increment of about 10. (Our original intention was to use the ratios 10:1, 20:1, 30:1, 40:1, 50:1, 60:1, and 70:1, but because of the inflexibility in using the JPEG parameter, we ended up with some different ratios.)

The photographic samples of the degraded images were first subjectively evaluated in an office environment by ten observers who were chosen from the graduate students and faculty having some background in image compression. They were asked to rank the images in two ways: Within each technique and between the four techniques for a fixed compression ratio. The mean rating of the group for an evaluation was computed by

$$R = \left( \sum_{k=1}^{10} s_k n_k \right) / \left( \sum_{k=1}^{10} n_k \right),$$

where  $s_k$  = the score corresponding to the  $k$ th rating,  $n_k$  = the number of observers with this rating, and 10 = the number of grades in the scale. No limits were imposed on viewing time or distance for the observers.

Table 4 shows the correlation between the numerical objective quality measures and the subjective evaluation. As a measure of the extent of the linear relationship, the Pearson product-moment correlation coefficient ( $r$ ) was used. The possible values of  $r$  are between -1 and +1; the closer  $r$  is to -1 or +1, the better the correlation is.

The coefficient values in Part (a) of Table 4 indicate that the quality measures can be put into three groups according to their performance:

- Group I: AD, SC
- Group II: NK, CQ, LMSE, MD
- Group III: WD, PMSE, IF, NAE, NMSE,  $L_p$ .

Table 4. (a) Correlation coefficients for each technique

(1) Lenna

Measure/Code	JPEG	EPIC	RLPQ	SLPQ
AD	0.528	-0.154	0.864	0.984
SC	0.561	-0.117	-0.988	-0.971
NK	0.479	0.865	0.996	0.979
CQ	0.480	0.865	0.996	0.979
LMSE	-0.980	-0.794	-0.752	-0.803
MD	-0.964	-0.984	-0.883	-0.941
WD	-0.995	-0.993	-0.954	-0.970
PMSE	-0.999	-0.996	-0.991	-0.990
IF	0.999	0.996	0.991	0.990
NAE	-0.997	-0.996	-0.970	-0.973
NAE(1/3)	-0.996	-0.996	-0.969	-0.972
NAE(HVS)	-0.972	-0.977	-0.925	-0.940
NMSE	-0.999	-0.996	-0.991	-0.990
NMSE(1/3)	-0.999	-0.997	-0.989	-0.989
NMSE(HVS)	-1.000	-0.998	-0.995	-0.996
L1	-0.997	-0.996	-0.970	-0.973
L2	-0.994	-0.993	-0.966	-0.969
L2(1/3)	-0.995	-0.993	-0.965	-0.968
L2(HVS)	-0.988	-0.990	-0.969	-0.975
L3	-0.991	-0.991	-0.961	-0.964

(2) Gilbert

Measure/Code	JPEG	EPIC	RLPQ	SLPQ
AD	0.747	-0.527	0.820	0.969
SC	-0.243	-0.936	-0.987	-0.930
NK	0.768	0.981	0.984	0.936
CQ	0.768	0.981	0.984	0.936
LMSE	-0.869	-0.800	-0.809	-0.727
MD	-0.828	-0.929	-0.853	-0.687
WD	-0.960	-0.960	-0.958	-0.923
PMSE	-0.979	-0.986	-0.981	-0.943
IF	0.979	0.986	0.981	0.943
NAE	-0.967	-0.975	-0.975	-0.939
NAE(1/3)	-0.842	-0.987	-0.974	-0.945
NAE(HVS)	-0.941	-0.941	-0.961	-0.914
NMSE	-0.979	-0.986	-0.981	-0.943
NMSE(1/3)	-0.717	-0.992	-0.978	-0.958
NMSE(HVS)	-0.988	-0.989	-0.998	-0.967
L1	-0.967	-0.975	-0.975	-0.939
L2	-0.961	-0.965	-0.962	-0.917
L2(1/3)	-0.754	-0.985	-0.959	-0.934
L2(HVS)	-0.964	-0.968	-0.985	-0.941
L3	-0.948	-0.960	-0.946	-0.890

## (3) Fingerprint

Measure/Code	JPEG	EPIC	RLPQ	SLPQ
AD	0.803	-0.101	0.926	0.880
SC	0.325	-0.846	-0.955	-0.935
NK	0.895	0.975	0.958	0.944
CQ	0.895	0.975	0.958	0.944
LMSE	-0.906	-0.962	-0.737	-0.812
MD	-0.417	-0.956	-0.540	-0.402
WD	-0.962	-0.992	-0.938	-0.934
PMSE	-0.989	-0.999	-0.962	-0.953
IF	0.989	0.999	0.962	0.953
NAE	-0.975	-0.994	-0.956	-0.946
NAE(1/3)	-0.974	-0.993	-0.954	-0.939
NAE(HVS)	-0.948	-0.987	-0.936	-0.925
NMSE	-0.989	-0.999	-0.962	-0.953
NMSE(1/3)	-0.988	-0.995	-0.959	-0.934
NMSE(HVS)	-0.991	-0.996	-0.966	-0.954
L1	-0.975	-0.994	-0.956	-0.946
L2	-0.975	-0.995	-0.947	-0.937
L2(1/3)	-0.974	-0.993	-0.943	-0.920
L2(HVS)	-0.968	-0.997	-0.946	-0.930
L3	-0.975	-0.996	-0.934	-0.925

Table 4. (b) Correlation coefficients across techniques

## (1) Lenna

Measure/Ratio	69:1	59:1	52:1	42:1	30:1	20:1	10:1
AD	-0.470	-0.498	-0.051	-0.558	0.875	0.260	-0.656
SC	0.863	0.716	0.863	0.626	0.683	-0.780	0.364
NK	-0.834	-0.705	-0.834	-0.675	-0.582	0.858	-0.455
CQ	-0.834	-0.705	-0.834	-0.675	-0.582	0.858	-0.455
LMSE	0.231	0.163	-0.010	0.203	-0.720	-0.471	0.950
MD	0.033	0.564	0.332	0.541	-0.380	-0.958	0.681
WD	-0.914	-0.221	-0.097	0.519	-0.254	-0.792	0.941
PMSE	0.188	0.533	0.360	0.671	-0.085	-0.893	0.929
IF	-0.161	-0.520	-0.349	-0.666	0.087	0.892	-0.928
NAE	-0.805	-0.295	-0.133	0.534	-0.015	-0.862	0.915
NAE(1/3)	-0.790	-0.417	-0.302	0.434	-0.017	-0.858	0.915
NAE(HVS)	0.454	0.527	0.270	0.531	-0.272	-0.828	0.874
NMSE	0.161	0.520	0.349	0.666	-0.087	-0.892	0.928
NMSE(1/3)	-0.627	-0.342	-0.349	0.384	-0.119	-0.879	0.928
NMSE(HVS)	0.589	0.664	0.397	0.629	-0.202	-0.879	0.909
L1	-0.805	-0.295	-0.133	0.534	-0.015	-0.862	0.915
L2	0.164	0.503	0.332	0.651	-0.086	-0.884	0.932
L2(1/3)	-0.607	-0.313	-0.326	0.370	-0.123	-0.867	0.934
L2(HVS)	0.553	0.632	0.373	0.604	-0.187	-0.864	0.894
L3	0.461	0.627	0.401	0.670	-0.139	-0.893	0.938

**(2) Gilbert**

Measure/Ratio	69:1	59:1	52:1	42:1	30:1	20:1	10:1
AD	-0.015	0.968	0.664	0.913	0.835	0.896	0.661
SC	-0.883	0.466	-0.494	-0.641	-0.552	-0.697	-0.739
NK	0.871	-0.654	0.617	0.728	0.636	0.760	0.741
CQ	0.871	-0.654	0.617	0.728	0.636	0.760	0.741
LMSE	0.532	-0.600	0.171	-0.112	-0.403	0.125	0.673
MD	-0.762	0.881	-0.935	-0.891	-0.761	-0.255	0.458
WD	-0.048	0.871	0.132	-0.365	-0.480	-0.639	-0.616
PMSE	-0.517	0.953	-0.700	-0.688	-0.788	-0.866	-0.753
IF	0.517	-0.953	0.700	0.688	0.788	0.866	0.753
NAE	-0.140	0.947	-0.011	-0.318	-0.374	-0.628	-0.759
NAE(1/3)	0.772	0.990	0.952	0.087	0.977	-0.174	-0.007
NAE(HVS)	-0.941	-0.961	-0.962	-0.896	-0.834	-0.854	-0.835
NMSE	-0.517	0.953	-0.700	-0.688	-0.788	-0.866	-0.753
NMSE(1/3)	0.560	0.993	0.961	0.118	0.982	-0.076	0.071
NMSE(HVS)	-0.967	-0.952	-0.973	-0.908	-0.843	-0.885	-0.895
L1	-0.140	0.947	-0.011	-0.318	-0.373	-0.628	-0.759
L2	-0.539	0.954	-0.712	-0.693	-0.786	-0.868	-0.754
L2(1/3)	0.584	0.999	0.935	0.084	0.974	-0.110	0.057
L2(HVS)	-0.965	-0.950	-0.967	-0.896	-0.832	-0.878	-0.881
L3	-0.787	0.984	-0.918	-0.904	-0.941	-0.893	-0.391

**(3) Fingerprint**

Measure/Ratio	69:1	59:1	52:1	42:1	30:1	20:1	10:1
AD	-0.871	0.878	-0.930	0.135	0.345	-0.093	-0.656
SC	-0.946	-0.925	-0.975	-0.960	-0.903	-0.953	-0.887
NK	0.979	0.930	0.982	0.971	0.924	0.966	0.920
CQ	0.979	0.930	0.982	0.971	0.924	0.966	0.920
LMSE	0.804	-0.437	-0.592	0.208	0.014	0.002	0.232
MD	0.735	0.977	0.999	0.309	0.573	-0.412	0.574
WD	0.057	-0.126	-0.976	-0.881	-0.918	-0.993	-0.930
PMSE	-0.185	0.916	-0.920	-0.983	-0.981	-0.989	-0.966
IF	0.185	-0.916	0.920	0.983	0.981	0.989	0.966
NAE	-0.304	1.000	-0.970	-0.999	-0.992	-0.989	-0.964
NAE(1/3)	-0.553	-0.024	-0.913	-0.994	-0.982	-0.980	-0.974
NAE(HVS)	-0.888	-0.404	-0.959	-0.977	-0.986	-0.946	-0.866
NMSE	-0.185	0.916	-0.920	-0.983	-0.981	-0.989	-0.966
NMSE(1/3)	-0.826	-0.791	-0.923	-0.986	-0.969	-0.976	-0.968
NMSE(HVS)	-0.894	-0.442	-0.986	-0.983	-0.979	-0.961	-0.902
L1	-0.304	1.000	-0.970	-0.999	-0.992	-0.989	-0.964
L2	-0.192	0.914	-0.921	-0.984	-0.983	-0.990	-0.964
L2(1/3)	-0.830	-0.792	-0.926	-0.987	-0.972	-0.974	-0.967
L2(HVS)	-0.896	-0.440	-0.988	-0.985	-0.983	-0.962	-0.892
L3	-0.195	0.862	-0.544	-0.960	-0.960	-0.988	-0.974

The measures in Group I cannot be reliably used with all techniques as the sign of the correlation coefficient does not remain the same. Group II measures are consistent, but nevertheless have poor correlation with the observers' response for some of the techniques. Among the useful measures in Group III, NMSE(HVS) is the best one for all the test images. Except for a single

case, the incorporation of the HVS into NMSE makes the correlation slightly stronger. For the other two measures NAE and L2, however, there is no such improvement. (In fact, the visual model has an adverse effect on NAE.) The results reported in [4] and [5] support our conclusion that the HVS model does not always improve the correlation, and when it does, the gain is small. The nonlinear filter  $(\cdot)^{1/3}$ , on the other hand, seems to have a random behavior, but usually leads to a weaker correlation. As IF is defined in terms of NMSE, the results for these two measures are identical. It has been found that PMSE establishes the same relationship as well.

Part (b) of Table 4 is rather disappointing, and the information that can be extracted is limited. As the compression ratio is increased, the measures perform much poorer. This observation is not surprising because different techniques introduce different types of degradation into the reconstructed images. Since the metrics combine all the pixel differences between two given images into a single number, one cannot expect to know much about the annoyance experienced by the human observer. In our experiments, for instance, although JPEG was the code for which the errors were always the smallest, the observers found the tile effect very objectionable in Lenna, yet favored blockiness in the higher frequency images Gilbert and Fingerprint.

To the best of our knowledge, histograms and Hosaka plots are the only two image quality measures that are graphical. Before we evaluate their performance, a specification of the type of impairment caused by the techniques is needed. Because of space limitation, the results for only the first test image will be discussed here. Four degraded versions of Lenna for the highest compression ratio (69:1) are given in Figure 1. The original image is also included for a comparison. The major types of degradation in the images are blockiness with JPEG, blurriness with EPIC, both fuzziness and blockiness with RLPQ, and fuzziness with SLPQ (The term fuzziness is used in the sense of equal amount of blurriness over the entire image).

A histogram of the compression error is constructed by plotting the number of times a specific value occurs in the difference image versus the value itself. Typically, it looks like a Gaussian curve; the more it resembles a spike at  $x=0$ , the greater the fidelity of the reconstructed image. The seven histograms in Figure 2 were obtained using JPEG. They clearly depict the increase in the amount of blockiness as the compression ratio goes up. The concentration of low intensity pixels for the lowest ratio is gradually reduced and the distribution becomes more uniform. Our experience has shown that histograms may also be used to specify different types of degradation in images. In Figure 3, the histograms with low intensity pixel concentrations are associated with RLPQ and SLPQ, and they are in contrast with those corresponding to JPEG and EPIC. The uniform fuzziness over the entire image, it is understood, leads to a spiky histogram. Nevertheless, the similarity between the histograms in each pair makes it difficult to distinguish between the artifacts involved.

To construct a Hosaka plot, or an h-plot, we measure a number of features of the reconstructed image and compare these with the corresponding features in the original image [6]. The difference between the two feature vectors generates a vector error measure, which, unlike scalar quantities, allows for a description of not only the amount, but also the type of degradation. In the process, the original image is first segmented into blocks whose variance is less than some specified threshold. These blocks are then grouped together to form a number of classes which depend on the size of the blocks. Two features are computed for each class in both the original and the reconstructed images. One of them is related to the mean intensity values and the other is the mean standard deviation. The h-plot is constructed by plotting the errors in the corresponding features in polar coordinates. The radius denotes the feature error, the left and right half planes contain the vectors associated with standard deviations and means, respectively.

It is reported in [6] that when noise is added to an image, the area of the h-plot is proportional to the image quality, but the structure of the diagram depends on the type of distortion. If an image is





ORIGINAL



JPEG



EPIC



RLPQ



SLPQ

Figure 1

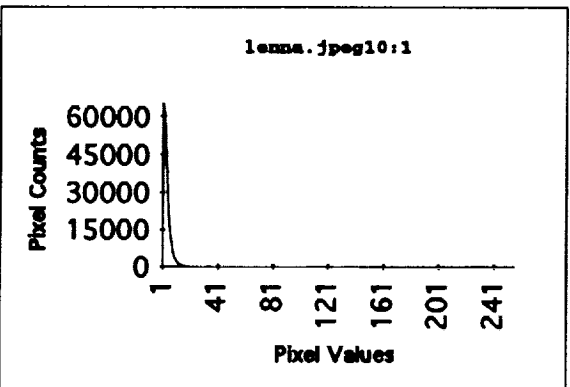
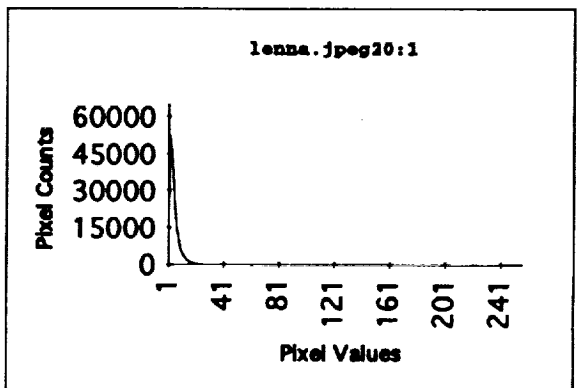
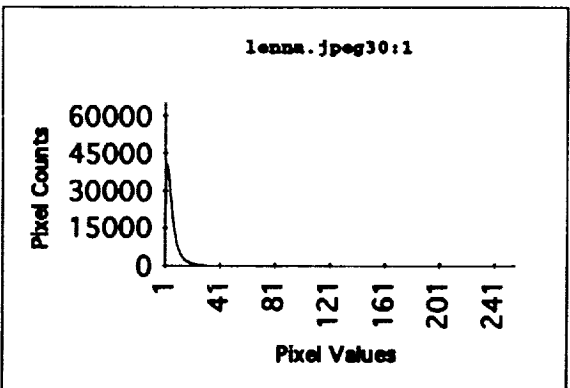
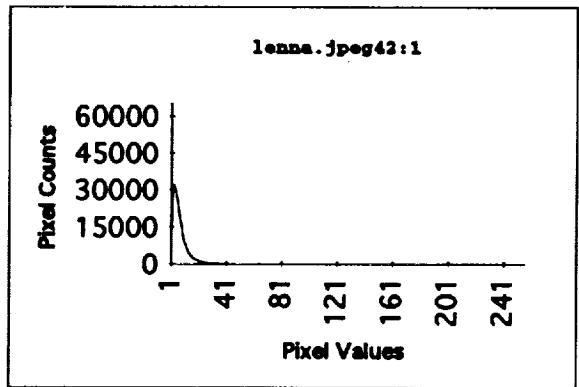
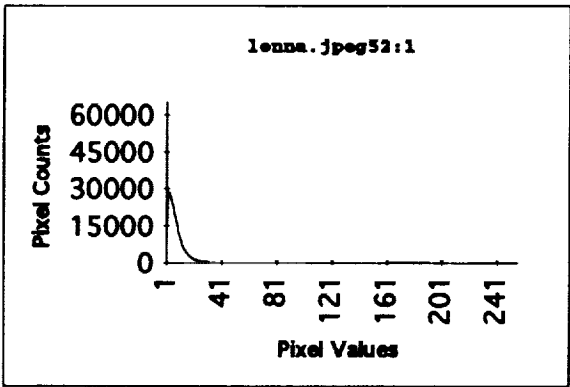
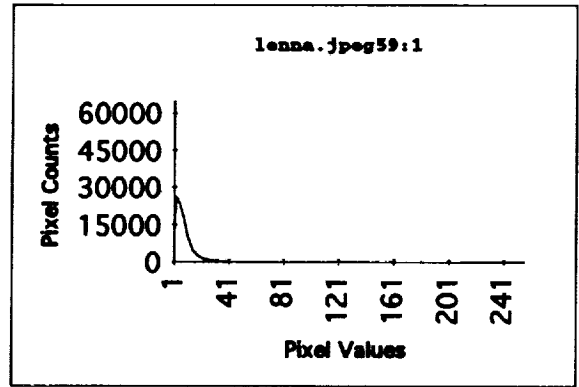
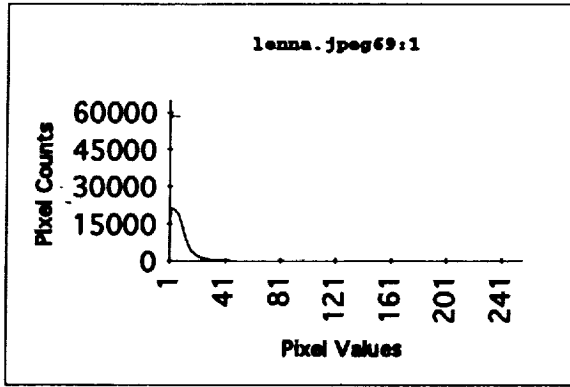


Figure 2

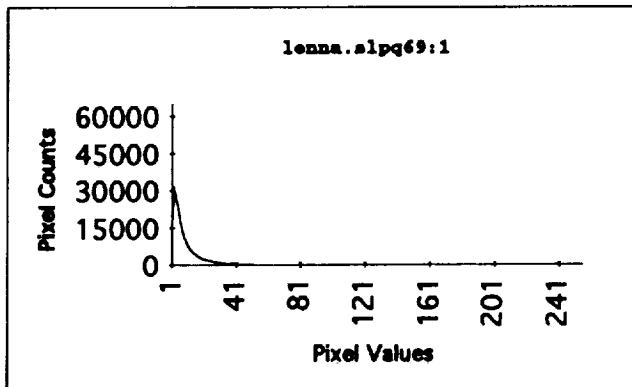
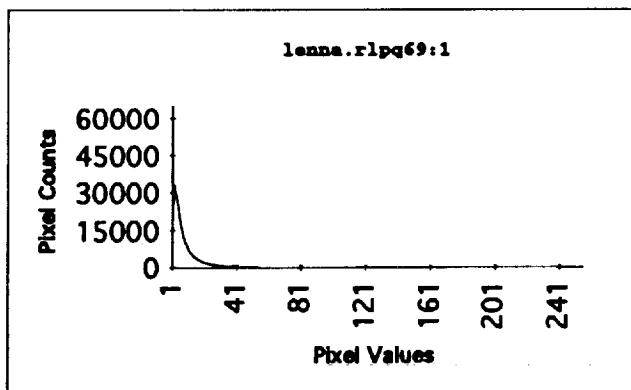
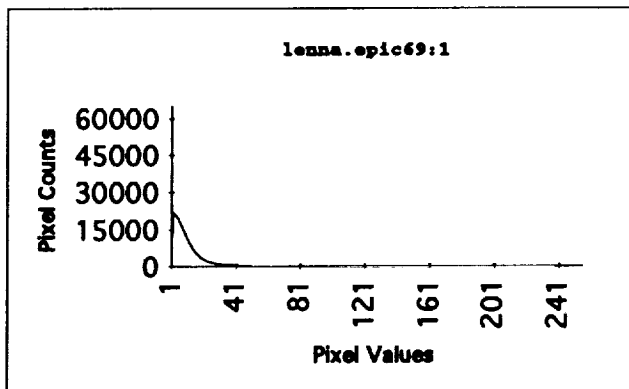
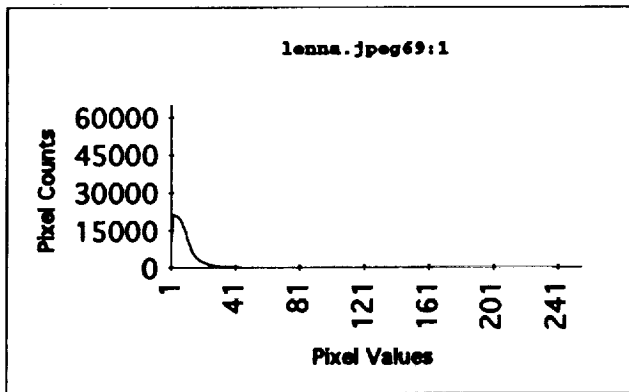


Figure 3

blurred, on the other hand, the pattern on the right side of the diagram remains fixed and increases in magnitude as the blurring increases while the left side is much less predictable.

The h-plots in Figure 4 were obtained using Lenna for all compression techniques and ratios. In each diagram, the length of a radius is 2.75 units. The blockiness is reflected on the right side of h-plots, whereas, the effect of blurriness can be traced on the left. By a simple comparison, we are able to see the way each code reduces the fidelity of the image. One can even learn how the distortion is distributed in the reconstructed images by looking at the relative lengths of the components along the axes. For example, it is evident that JPEG preserves the high frequency components (the feathers) of the image, whereas RLPQ induces uniform blockiness. Such information is extremely helpful considering the sensitivity of the human observer to the location of the image error. For the construction of the h-plots in Figure 4, the two parameters, the initial block size  $N$  and the variance threshold  $T$ , were chosen as 16 and 10, respectively, as in Hosaka's or Farrelle's work [6]. For high compression ratios, the h-plots for JPEG and RLPQ indicate that it may be worth trying larger values for these parameters.

#### 4. Conclusions

The results of an evaluation concerning the usefulness of a number of objective quality measures for grayscale image compression have been presented. It is understood that although a group of numerical measures can reliably be used to specify the magnitude of degradation in reconstructed images for a given compression technique, an evaluation across different techniques is not possible. This is because a single scalar value cannot be used to describe a variety of impairments. A simple analogy would be the futility in comparing apples with oranges. The two graphical measures, however, are fairly successful in specifying the type of degradation. Hosaka plots, in particular, provide a good indication of how images are degraded. A combination of numerical and graphical measures may prove more useful in judging image quality. There is also a need for the development of new graphical measures with superior judgment capabilities. Further research in these areas is now ongoing.

#### References

- [1] A.M. Eskicioglu and Paul S. Fisher, "A Survey of Quality Measures for Gray Scale Image Compression," Proceedings of 1993 Space and Earth Science Data Compression Workshop (NASA Conference Publication 3191), Snowbird, Utah, April 2, 1993.
- [2] N.B. Nill, "A Visual Model Weighted Cosine Transform for Image Compression and Quality Assessment," *IEEE Transactions on Communications*, Vol. COM-33, No. 6, pp. 551-557, June 1985.
- [3] Paul S. Fisher and Nassrin Tavakoli, Final Report for Compression of Geophysical Data, Space and Naval Warfare Systems Command Contract N00039-C-0063, Department of Computer Sciences, University of North Texas, Texas, May 1992.
- [4] G.G. Kuperman and D.L. Wilson, "Objective and Subjective Assessment of Image Compression Algorithms," Society for Information Display International Symposium Digest of Technical Papers, Vol. 22, pp. 627-630, May 1991.
- [5] 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.
- [6] P.M. Farrelle, *Recursive Block Coding for Image Data Compression*, Springer-Verlag New York Inc., USA, 1990.

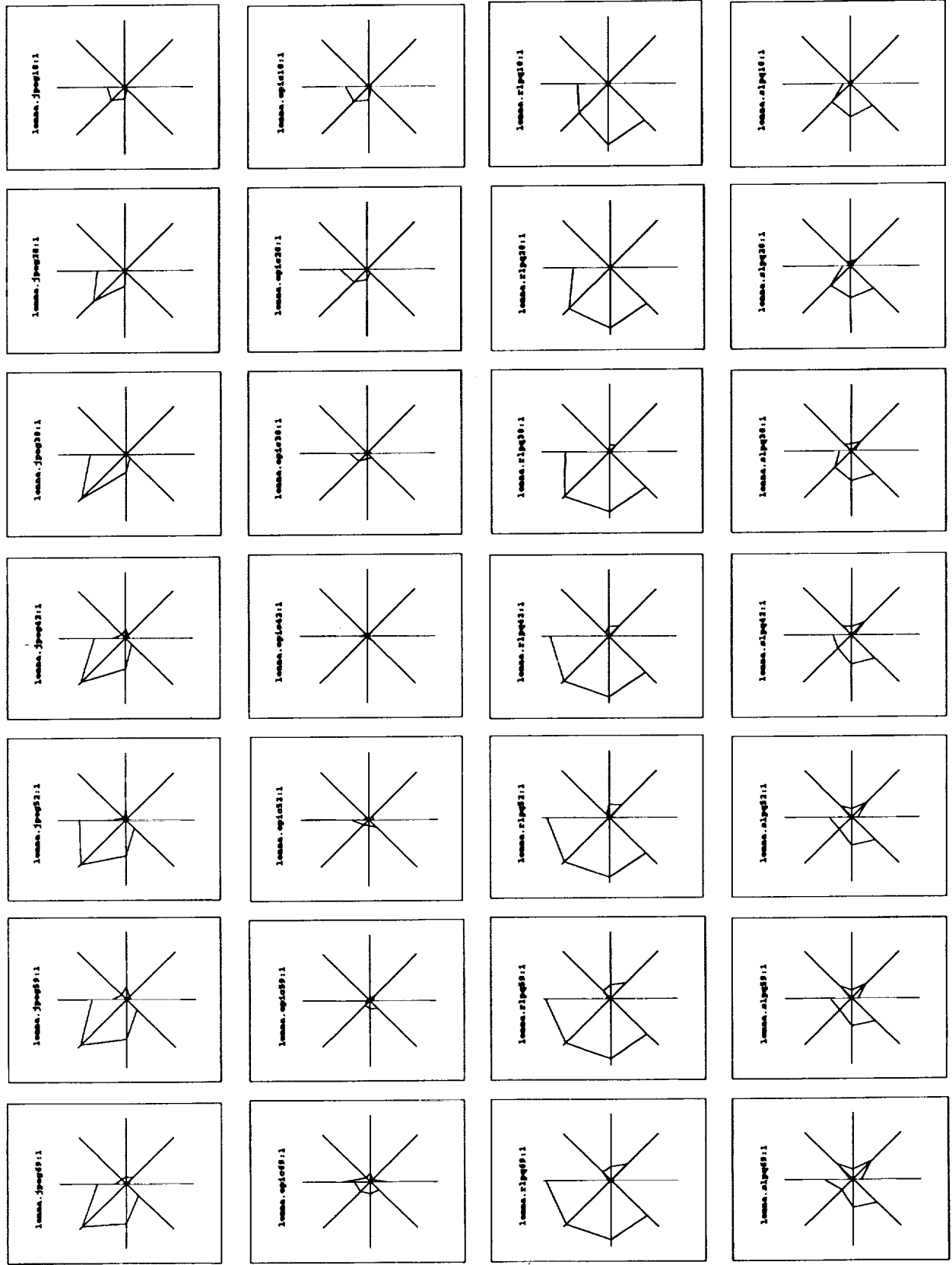


Figure 4

