Quality Measures for Gray Scale Image Compression

There is no single reliable objective criterion for measuring the quality of a compressed image.

One can not make a complete evaluation of various compression techniques

Subjective criteria: burdensome, inaccurate, inconsistent

Objective criteria: mean square error(MSE) do not have a good correlation with the viewer's response

To have a reliable visual quality measure:

a responsible model of the complex human visual system is required.

Such a measure is not only needed for comparing images produced by different techniques, but it is also instrumental in designing image processing/compression algorithms.

The efficiency of a compression algorithm is generally measured using three criteria:

1. Compression amount
2. Implementation complexity
3. Resulting distortion

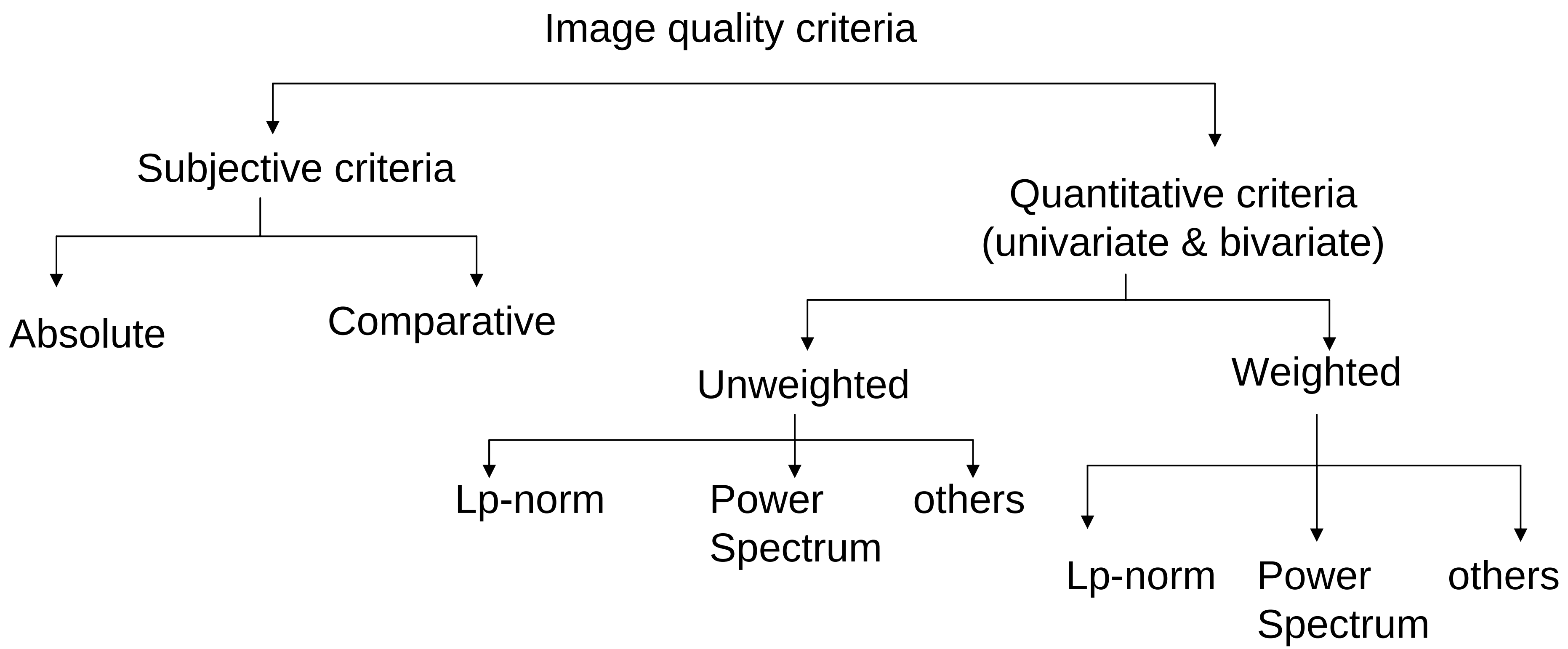
Amount of compression:

compression ratio, compression percentage

Algorithm complexity:

can be measured by considering the data structures as well as the type and number of operation required.

The difficulty in evaluating a lossy compression algorithm comes from the fact that there is no reliable and consistent measure for determining the magnitude of distortion resulting from the loss. – we lack a useful and practical measure for quality assessment.



Classification of Image Quality Criteria

Subjective Criteria:

As the final user of images are humans, the most reliable and commonly used assessment of image quality is the subjective rating by human observers.

Both expert and non expert observers are used in experiments; non expert represent the average viewer while experts are believed to be able to give better, more refined assessments of image quality since they have been trained and are familiar with images and their distortions.

In absolute evaluation, the observers view image and assess its quality by assigning to it a category in a given rating scale, whereas in comparative evaluation, a set of images are ranked from best to worst by the observers.

:Mean opinion Scale (MOS)

A

5 Excellent

4 Good

3 Fair

2 Poor

1 Unsatisfactory (bad)

Rating Scales Used in Subjective Evaluation

B

7 best

6 well above average

5 slightly above average

4 Average

3 slightly below average

2 well below average

1 worst

Rating Scales Used in Subjective Evaluation

C

1. Not noticeable (perceptible)
2. Just noticeable (perceptible)
3. Definitely noticeable but only slight impairment
4. Impairment not objectionable
5. Somewhat objectionable
6. Definitely objectionable
7. Extremely objectionable

Rating Scales Used in Subjective Evaluation

D

3 Much better

2 Better

1 Slightly better

0 Same

-1 Slightly worse

-2 Worse

-3 Much Worse

Rating Scales Used in Subjective Evaluation

E

5 Imperceptible

4 Perceptible

3 Slightly annoying

2 Annoying

1 Very annoying

Rating Scales Used in Subjective Evaluation

F

10,9	Very good
8,7	Good
6,5,4	Fair
3,2	Bad
1,0	Very bad

Rating Scales Used in Subjective Evaluation

The mean rating of a group of observers who join the evaluation is usually computed by

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

Where s_k = the score corresponding to the kth rating,

n_k = the number of observers with this rating, and

n = the number of grades in the scale.

Bubble sort is another technique used in image rating.

The subject compares two images A and B from a group and determines their order. Assuming that the order is AB, he/she takes a third image and compares it with B to establish the order ABC or ACB. If the order is ACB, then another comparison is made to determine the new order . The procedure continues until all the images have been used.

→ The best picture is bubbled to the top if no ties are accepted.

- The subjective rating results are affected by a number of factors including:

  (i) type and range of images

  (ii) level of expertise of the observer

  (iii) experimental conditions

Quantitative criteria: Univariate / Bivariate

A univariate measure assign to a single image a numerical value based upon measurements of the image field, and a bivariate measure is a numerical comparison between two images.

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

$$Q = \sum_{j=1}^{M} \sum_{k=1}^{N} O\{F(j,k)\}$$

For MxN samples, where O{ · } is some operator.

Bivariate measures are more frequently used in image quality measurement. If $\hat{F}$(j,k) denotes the samples on the degraded image field, a number of measures can be established to determine the closeness of the two images.

$$Q' = f\{ F(j,k) , \hat{F}(j,k) \}$$

(i) $$L_p = \{(1/MN)\sum_{j=1}^{M}\sum_{i=1}^{N} | F(j,k) - \hat{F}(j,k) |^P \}^{1/p}$$ Lp-norm

P determines the relative significance of errors of different magnitudes.

L1 = the average absolute error

L2 = the root mean square error (RMSE)

As the value of p is increased, a greater relative emphasis is given to large errors in the image

(ii) Low order moment of a power spectrum

(iii) $$K = \sum_{j=1}^{M}\sum_{i=1}^{N} F(j,k)\hat{F}(j,k)$$

Discretized cross-correlation function

It may be normalized by the reference image energy to give unity as the peak correlation

$$\text{NK} = \frac{\sum_{j=1}^{M}\sum_{i=1}^{N} F(j,k)\hat{F}(j,k)}{\sum_{j=1}^{M}\sum_{i=1}^{N} [F(j,k)]^2}$$

(iv) Correlation quality

$$\text{CQ} = \frac{\sum_{j=1}^{M}\sum_{i=1}^{N} F(j,k)\hat{F}(j,k)}{\sum_{j=1}^{M}\sum_{i=1}^{N} F(j,k)}$$

(v) Structural content:

$$SC = \frac{\sum_{j=1}^{M}\sum_{i=1}^{N}[F(j,k)]^2}{\sum_{j=1}^{M}\sum_{i=1}^{N}[\hat{F}(j,k)]^2}$$

(vi) Normalized absolute error between the reference and degraded image fields:

$$NAE = \frac{\sum_{j=1}^{M}\sum_{i=1}^{N}|O\{F(j,k)\}-O\{\hat{F}(j,k)\}|}{\sum_{j=1}^{M}\sum_{i=1}^{N}|O\{F(i,k)\}|}$$

(vii) Normalized Mean Square Error:

$$\text{NMSE} = \frac{\sum_{j=1}^{M}\sum_{i=1}^{N}[O\{F(j,k)\} - O\{\hat{F}(i,k)\}]^2}{\sum_{j=1}^{M}\sum_{i=1}^{N}[O\{F(j,k)\}]^2}$$

(viii) Peak Mean Square Error:

$$\text{PMSE} = \frac{(1/MN)\sum_{j=1}^{M}\sum_{i=1}^{N}[O\{F(j,k)\} - O\{\hat{F}(j,k)\}]^2}{A^2}$$

Where A represents the maximum value of O{F(j,k)}.

The definitions used for the operator O{ · } in (vii) and (viii) are

(a) $F(j,k)$

(b) $[F(j,k)]^v$: power law

(c) $K_1 \log_b[K_2+K_3F(j,k)]$: Logarithmic

(d) $[F(x,y) \otimes H(x,y)]\delta(x-j\Delta x, y-k\Delta y)$: Convolution

(ix) Laplacian MSE

$$\text{LMSE} = \frac{\sum_{j=1}^{M-1}\sum_{k=2}^{N-1}[O\{F(j,k)\} - O\{\hat{F}(j,k)\}]^2}{\sum_{j=1}^{M-1}\sum_{k=2}^{N-1}[O\{F(j,k)\}]^2}$$

Where

$$O\{F(j,k)\} = F(j+1,k) + F(j-1,k) + F(j,k+1) + F(j,k-1) - 4F(j,k)$$

In many applications, the MSE(no matter how it is defined) is often expressed in terms of a SNR defined in dB.

(x) Image fidelity:

$$IF = 1 - \frac{\sum_{j=1}^{M}\sum_{i=1}^{N}[F(i,k) - \hat{F}(i,k)]^2}{\sum_{j=1}^{M}\sum_{i=1}^{N}[F(j,k)\}]^2}$$

(xi) Difference[j,k] = $F(j,k) - \hat{F}(i,k)$

(xii) $\sum_{j=1}^{M}\sum_{i=1}^{N} Difference[i,k] / MN$

(xiii) $Max\{| Difference[i,k] |\}$

(xiv) Histogram of the compression error

(xv) Hosaka plots

(xvi) Sensitivity and predictive value positive curve

(xvii) Rate-distortion curves

It is reported that image quality assessment can be improved by incorporating into the evaluation process some models of the HVS.

HVS + Quality Measure

(i) Attaching a weight to the image samples either in the spatial or frequency domain

(ii) Weighting the digital image power spectrum

(a) $O\{\cdot\} = H_L(x, y) \otimes O_N\{\cdot\}$

$H_L(x, y)$: impulse response of the lateral inhibition process

IEEE Trans. IT-20, pp.525-536, July 1974

$O_N\{\cdot\}$: point nonlinearity

models the eye's photoreceptors

In the Fourier domain H_L is defined as

$$a[c+(\frac{w}{w_o})^{K_1}]\exp\{-(\frac{w}{w_o})^{K_2}\}$$

Where $w=(w_1+w_2)^{1/2}$, and $O_N\{\cdot\}=\{\cdot\}^{1/3}$ is chosen.

Experiments show that

a = 2.6

c = 0.0192

w_o = 1/0.114

K_1 = 1

K_2 = 1.1

Are suitable parameter set.

(b) $$E_p = \{\frac{1}{m}\sum_{i=1}^{m} |e_i|^p\}^{1/p}$$, p=1,2,3,4,6

m = number of picture elements in a picture

$e_i = x_i - \hat{x}_i$

x_i = the value of the pel in the original picture

$\hat{x}_i$ = the value of the pel in the distorted picture

E_p is a very good estimate of impairment rating where the type of distortion is additive white noise.

$$EM_P = \{\frac{1}{m}\sum_{i=1}^{m} |e_i|^p / w_i\}^{1/p}$$

To reflect the masking effect (of w_i)

W_i : the value of the weighting function at pel i and is derived from an activity function that is a measure of variability of the signal in the neighborhood of pel i.

Three different forms of activity functions are studied:

Amax: measures the maximum signal change between any pair of pels in a neighborhood consisting of the pel being evaluated plus eight surrounding pels.

Aov: sums the deviation of the same neighborhood of points from the neighborhood average $\overline{X}$

Adf: provides the weighted sum of the magnitude of the surrounding element difference (slope) in both the horizontal and vertical directions.

Remarks:

In most distorted images, quality is determined mainly by the visibility of distortion in flat areas where it is more visible and consequently the effects of masking have little effect.

Consider the following quality measures:

$$NMSE = \frac{\sum_{m=1}^{N}\sum_{n=1}^{N}[f(m,n) - \hat{f}(m,n)]^2}{\sum_{m=1}^{N}\sum_{n=1}^{N}[f(m,n)]^2}$$

$$LMSE = \frac{\sum_{m=2}^{N-1}\sum_{n=2}^{N-1}[G(m,n) - \hat{G}(m,n)]^2}{\sum_{m=2}^{N-1}\sum_{n=2}^{N-1}[G(m,n)]^2}$$

where

$$G(m,n) = f(m+1,n) + f(n-1,n) + f(m,n+1) + f(m,n-1) - 4f(m,n)$$

$$PMSE = \frac{\sum_{m=1}^{N}\sum_{n=1}^{N}[z(m,n)-\hat{z}(m,n)]^2}{\sum_{m=1}^{N}\sum_{n=1}^{N}[z(m,n)]^2}$$

Where $z(m,n)$ and $\hat{z}(m,n)$ are given by

$$z(m,n) = \ln(f[m,n] \otimes hbp(m,n))$$

$$\hat{z}(m,n) = \ln(\hat{f}[m,n] \otimes hbp(m,n))$$

Hbp(m,n) is the rectangular coordinate form of the point spread function of the HVS.

⇨ the correlation between PMSE and the subjective ranking (obtained by bubble sort) of the data set is higher than that of { NMSE, LMSE }

IEEE Trans. on Communications, pp. 551-557 June 1985

Fourier Domain Quality Measure:

$$K^{-1}\sum_{i=1}^{B} W_i \sum_{u=0}^{M-1}\sum_{j=0}^{N-1} H^2(r)[F_i(u,v)-\hat{F}_i(u,v)]^2$$

Where

B = number of subimage blocks in scene

K = normalization factor such as total energy

H(r) = rotationally symmetric spatial frequency response of HVS, $r=\sqrt{u^2+v^2}$

$F_i, \hat{F}_i$ = Fourier Transform of unprocessed and processed subimage i, respectively.

M,N = number of Fourier coefficients+1, in orthogonal u, v directions.

W_i = subimage i structure weighting factor, proportional to subimages intensity level variance.

$$H(r) = (0.2 + 0.45r)e^{-0.18r}$$ Fourier Transform

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

(i) Combining the HVS model with the image cosine transform will result in better performance in image compression and image quality assessment applications.

(ii) Performance in quality assessment also be enhanced by inclusion of the subimage structure weighting.

Optical Engineering, vol. 28, No. 7, pp. 813-818, July 1989

Once an image V(x,y) and its reproduction have been subjected to the HVS model, then the mean square error

$$d(U,U') = \frac{1}{N}\iint [U(x,y) - U'(x,y)]dxdy$$

where N is the image area or the number of pixels, may be considered as a meaningful measure of image quality.

HVS model f(u) = $u^{0.33}$, u= the pixel intensity

$$A(fr) = [0.2 + 0.81(\frac{fr}{5.55})]\exp[-(\frac{fr}{5.55})]$$

where $fr = (f_x^{\ 2} + f_y^{\ 2})^{1/2}$

$$C(fr) = \left\{ \frac{1}{4} + \frac{1}{\pi^2}\left\{ \log_e\left[\frac{2\pi}{\alpha} fr + \left(\frac{4\pi^2}{\alpha} fr^2 + 1 \right)^{1/2} \right] \right\}^2 \right\}$$

$$ADCT = A(fr)C(fr)$$

Information content(IC): the sum of the magnitudes of its DCT spectral components after they have been appropriately normalized based on HVS sensitivity models for that particular solution.

The plot of IC v.s. the resolution provides some insight into the quality of a given image.