Objective measures of image quality

Institute for Perception Research
PO Box 513, 5600 MB Eindhoven

VK/vk 92/01
06.03.1992

Rapport no. 847

Objective Measures of Image Quality: A Review

V. Kayargadde

Objective Measures of Image Quality: A Review

Vishwakumara Kayargadde

March 6, 1992

1 Introduction

The need to have objective measures that correlate well with subjective image quality has led to many measures with different degrees of success. Early attempts were based purely on physical measures like signal-to-noise ratio or root-mean-square error, whereas later measures incorporate properties of the human visual system. Most of the early work attempted to arrive at a useful measure for image display sharpness. Sharpness measures based on the Modulation Transfer Function (MTF) are most popular. A survey of objective measures of sharpness and their performance is given by Westerink (1988). In general, image quality is a complex percept which not only depends on sharpness, but also on several other underlying subjective attributes (such as brightness, contrast, etc.). The study of how image quality can be predicted when multiple degradations occur is fairly new and very important for practical applications, such as in image processing and image coding.

In this report, a brief survey of literature on objective measures of subjective image quality is given. First the MTF-based sharpness measures are briefly reviewed. Most of the other methods reviewed here were developed mainly to predict the quality of images that underwent some kind of degradation. These measures attempt to objectively grade quality of decoded or restored images. Presently, these quality evaluations require expensive subjective testing. Multidimensional scaling techniques have been used by a few authors to identify the important image quality dimensions. A review of one such analysis is given. An image quality prediction method based on multidimensional scaling analysis is also reviewed. Finally, some positive aspects and drawbacks of the approaches so far are discussed.

2 Sharpness measures based on MTF

Many different objective measures have been proposed in the literature to predict the sharpness of images and imaging devices. The most successful among them are based on the Modulation Transfer Function (MTF). The majority of these measures are based on the assumptions that 1) the display system can be characterized by an MTF (at least for the quality evaluation purposes), 2) the overall system transfer function is given by the product of the imaging system MTF and the MTF of the eye, and 3) the perceived sharpness is given by the (weighted) sum of the overall transfer function over the spatial frequency range. The validity of these assumptions is questionable (Westerink (1988)).

For a survey of objective sharpness measures, see Westerink (1988). A recent objective measure is the square root integral (SQRI) measure (Barten 1989, 1990). Barten correlates the predictions of subjective quality obtained by the SQRI method for different display parameters and viewing distances, with experimentally determined subjective quality. A correlation coefficient between 0.95 and 0.99 is reported. Hultgren (1990) shows how the objective sharpness measures proposed by Higgins (1977), Granger and Cupery (1972), Carlson and Cohen (1980) and Barten (1989) can be unified in one general framework of log frequency sharpness measure. A brief outline of Hultgren's generalized sharpness measure is given in the appendix.

3 Quality measures based on pixel value differences

In this section we review some of the objective measures proposed to estimate the picture quality. All the measures reviewed here have a common feature: they all assume the existence of an 'original image' and processed or distorted image of the same scene. Such situations occur typically in image coding or transmission, where one would like to estimate the quality of the decoded picture with the original as a reference. Such an estimate may for example be used to pick a best coding scheme out of many competing schemes or to compare them in terms of output image quality. Hence, unlike the measures based on MTF (where the main goal was to evaluate the perceived sharpness of the display), here the interest is in studying how the picture quality is degraded in the presence of degradations such as blurring, noise, coding artifacts etc.

3.1 MSE Measures

Mean square error (MSE) between the original and the distorted image is often used as a quality criterion of processed images in image processing, mainly because of its simplicity. It has been shown (Limb (1979), Lukas and Budrikis (1982)) that in general MSE correlates poorly with subjective image quality. Marmolin (1986) proposed and tested 4 different modifications of the mean square error measure, based on a simple model. According to the model

1. the observer compares each pixel in the processed picture with the corresponding pixel in the original; subsequently

2. evaluates the objectionability of each perceived difference; and finally
3. sums up the perceived differences to an overall impression of quality

The proposed overall measure of quality is given by

$$E = \left[\frac{1}{n}\sum_{i=1}^{n}|D_i|^p\right]^{\frac{1}{p}}$$

where

$$D_i = a_i\, g(x_i - y_i)$$

where $g()$ is a processing function that determines the visibility of the error, a_i weighs the visibility of the error at pixel i, p is a factor that determines the relative importance of large and small errors and x_i and y_i represent the grey value of pixel i in the original and processed picture respectively.

Marmolin studied 4 different ways in which D_i could be chosen:

1. The unweighted error measure E is obtained when

   $$D_i = x_i - y_i$$

   We get the familiar root mean square (RMS) error measure if $p = 2$.

2. The error measure weighted by the background, E-ratio is obtained when

   $$D_i = \frac{x_i - y_i}{M_i}$$

   $$D_i = 0 \;\; for \;\; D_i < 0.1$$

   where M_i is the mean value in a 7x7window, centered on pixel i.

3. The E-mean measure corresponds to

   $$D_i = Mx_i - My_i$$

   where Mx_i, My_i are the mean levels in a 2x2 window 'surrounding' the pixel in the original and processed picture.

4. This error measure is motivated by the observation that a) contours are the most informative parts of an image and, b) noise is less visible when the background is highly textured. Hence, a_i is chosen to be proportional to contour strength and inversely proportional to the variation in the surrounding background, to obtain E-gradient measure, with

   $$D_i = \frac{(Mx_i - My_i)(1 + Gx_i)}{(1 + 2Sx_i)}$$

   where Gx_i is the gradient level of pixel i in the original picture, Sx_i is the standard deviation in a 7x7 window surrounding pixel i in the original picture.

Subjective tests were conducted by Marmolin using 6 original images of size 128 x 128 pixels. The 6 scenes contained a square, chess pattern, terrain, building and two portraits. These images were distorted with gaussian blur, additive noise, combinations of blur and noise and high-pass amplified noise, resulting in a total of 58 distorted pictures. Five subjects took part in the experiments. They graded the similarity between distorted images and the original image of the same scene.

The proposed error measures were computed for each distorted picture. In the data analysis polynomials of different degree were fitted to the relation between perceived similarity and computed error. The value of the RMS residual in polynomial fit was used as a criterion for the goodness of the quality measure.

Unweighted MSE gives a good prediction only in the case of the distorted chess pattern and the square image and performs poorly for other scenes. The E-ratio measure performs well only in case of the chess pattern, whereas the E-mean weighted measure performs better for the chess pattern and the portrait scenes. The gradient weighted error measure performs quite well in case of the square, the terrain and the building scenes but behaves very poorly for the chess pattern scene.

The main finding of Marmolin's study is that no single measure he tried gave valid prediction for all scenes. He concludes that these results suggest that picture properties have to be incorporated in the subjective error measures.

In fact, all measures of Marmolin except the unweighted MSE measure do incorporate local properties of the picture in them. The real drawback may be the global summing of the error over all pixels, which averages the errors. In the analysis, pictures with different kinds of distortions were grouped together. Hence it is not surprising that overall correlation with similarity ratings is poor.

3.2 Limb's study

Limb (1979) aimed at incorporating the threshold and masking properties of the visual system in picture quality prediction metrics. He performed subjective experiments with still pictures with different types of distortions. The predictions of a number of different distortion criteria are compared against the data obtained using the subjective experiments. In these experiments, 5 original still pictures (252 lines with 256 active elements per line) with 16 different types of distortions were used (see Table I). The picture quality was evaluated by subjects using a 5 point CCIR scale. In the analysis, Limb uses data obtained from the 'expert' subjects only (12 experts and 21 non-expert subjects took part in the experiment).

The global measure

$$E_p = \left[\frac{1}{n} \sum_{i=1}^{n} |e_i|^p \right]^{\frac{1}{p}}$$

is determined for all degraded images, where n is the number of pixels in a picture and $e_i = x_i - y_i$ with x_i and y_i the i th pixel values in the original and the distorted picture respectively. The root mean square (RMS) ($p = 2$) error is used as a reference.

Other measures were aimed at discovering more about filtering and masking. A measure based on masking

$$EM_p = \left[\frac{1}{n} \sum_{i=1}^{n} \frac{|e_i|^p}{w_i} \right]^{\frac{1}{p}}$$

where w_i is a weighting function derived from an activity function at pixel position i, is proposed. Three different forms of activity functions were investigated, one of them being similar to the one used by Netravali and Prasada (1977). Limb also studied linear filtered error summed over entire picture ('filtered error') as a measure of picture quality.

The distorted pictures were classified into 3 groups for analysis purposes. Set 1, pictures 7 8 9 10 and 11, which mainly contained pictures with additive noise, was used in the nonlinear processing study. Set 2, pictures 3 4 5 6 8 14 15 and 16 which mainly contained picture dependent distortion was used in the masking study. Set 3, pictures 1 2 3 8 10 11 and 12, which contained different spectral distributions of error was used in the filtering study. Throughout the analysis, mean square deviation from the quadratic regression was used to assess the prediction accuracy.

For pictures with additive white noise (set 1) the E_p error measure resulted in very good predictions. Incorporating an activity function did not improve the prediction of quality in case of set 2. Limb argues that this is because of the global summation of error, whereas masking is operative only at edges. For set 3, two dimensional separable low pass filters result in better predictions of picture quality when compared to unfiltered error.

Local measures: The local measures are based on the assumption that the viewer concentrates on those regions in the picture where the degradation is most visible, and rates the quality by weighting the distortions at 2 or 3 worst regions. The pictures were divided into a rectangular array of square regions and the local measure was calculated in each square. The squares had a size of 1 degree of subtended arc, such that it is approximately equal to the size of the human fovea. The final measure of error was taken to be the largest value of E_p over all squares or the average of the 2 largest values.

Local measures incorporating masking with a slightly different activity function were used to estimate the picture quality (of set 2 only). In all the cases the local error measures performed only marginally better than the global measures.

Proposed model: Based on these observations, Limb proposed the model of Fig.1. He points out that although the filtering stage precedes masking in the model, it is likely that in reality filtering is distributed throughout the processing.

Limb's study shows that RMSE performs very well in most cases except when the distortions are greater at edges. According to Limb, RMSE performs so well because, in most distorted pictures, quality is determined mainly by the visibility of distortions in flat areas where it is more visible, consequently masking has little effect. This is true only when the noise is picture independent and additive, and does not hold in other situations like blurring. It's worth noting that Limb's study mainly contained images distorted by noise and only two blurred images, of which one (image number 13, Table II) has been left out in the analysis. Applicability of Limb's conclusions in case of other important distortions, for example blurring, is not known. Limb concludes by saying that while his study strongly indicates a local measure for masking effects, perhaps it should be combined with a global measure incorporating spatial filtering when additive noise distortions are present.

3.3 Lukas and Budrikis's model

The study by Lukas and Budrikis (1982) is along similar lines as the one by Limb (1979). The distortion measures proposed by them are based on a spatiotemporal model of threshold vision that incorporates filtering and masking. The model used to compute the error measure is shown in Fig.2. The filtering is carried out by parallel excitation and inhibition paths, each of which is separately linear. They combine in a nonlinear fashion to account for the adaptation with background luminance. The spatial point spread function and temporal impulse response of filters $U(x, y)$ and $V(t)$ are Gaussian and second order exponential functions respectively. The pictures are transformed from the electrical domain to the luminance domain using monitor characteristics. The parameters of the model are adjusted such that the model fits the data on sinusoidal grating thresholds.

Picture quality is estimated using the filtered, picture weighted noise distribution $N(x, y, t)$ (see Fig. 2) as

$$N = \left[\int_T \int_Y \int_X |N(x, y, t)|^p \, dxdydt \right]^{\frac{1}{p}}$$

For $p = 2$ we get the familiar RMS measure and for $p = 10$ the distortion measure is virtually dependent on peak error alone.

Subjective tests: The stimuli consisted of 6 frames of a video sequence, a talking head-and-shoulder view of a girl. Only a central window of 120 pixels width, 150 pixel height was used in the computations. 24 video sequences with different distortions (see table II), derived from the original video sequence, were used in quality rating tests. A modified CCIR procedure with a 0-10 scale was used to rate picture quality. A total of 40 subjects took part in the experiment (20 experts and 20 non-experts).

Picture quality prediction: Three classes of error measures were considered 1) a class of physical error measures: mean absolute error, root mean square error etc., 2) a class of filtered error measures where the subjective weighting is according to the filtering properties of the visual system, 3) error measures including filtering and masking.

The picture quality data was fitted using a quadratic regression procedure. The performance of error measures was evaluated using an index called coefficient of determination (see eq.10 of Lukas and Budrikis (1982)).

It is observed that for all values of the error power p the filtered error measure is a better predictor of quality than the physical error measure. Masking (either temporal or spatial) leads to worse predictions of picture quality. This is identical to what was also observed by Limb (1979). On arguments similar to those given by Limb, locally averaged error measures were proposed and evaluated.

Local measures: In evaluating the local error measures, the pictures are divided into an array of pixel blocks (30 pixel height, 20 pixel width). The 3 classes of errors mentioned above are now summed over each block. The measures derived were a) the maximum local error over all segments of all frames b) the maximum local error over all segments in each frame, averaged over all frames.

The local measures result in much better quality predictions. Separate analysis of the different classes of coding schemes shows that for errors that occur in different coding schemes, different quality measures provide best fit.

The main conclusion of Lukas and Budrikis's study is: viewers draw their attention to the worst areas of the picture and base their quality ratings on these. The conclusions arrived at have to be viewed in light of the fact that the subjective test involved a video sequence of only one scene.

3.4 Multiple channel model of Zetzsche and Hauske

Zetzsche and Hauske (1989) propose a model for quality prediction based on image pyramids and orientation selective filtering. The two main parts of the model are:

1. An adaptive input stage realized as a (nonlinear) ratio of Gaussian (ROG) pyramid

2. Decomposition by orientation selective filters, including a saturating nonlinearity acting at each point of the filter output

The output value at each point of each filter are regarded as a feature vector or internal representation of the image. The distance between the internal representation of the original and distorted picture is used as an error estimate. The model sums the error over the entire picture.

Zetzsche and Hauske report results of a pilot study in which subjective tests were performed on pictures coded with an integrated cosine transform CODEC. Two pictures (of 256 x 256 pixels) were degraded by 9 kinds of distortions. The model predictions had a correlation of 0.95 with the subjective ratings, as opposed to a correlation of 0.55 when S/N is used as the quality estimate. Testing the model predictions on an extended data set with 8 pictures and 13 kinds of distortions resulting in a total of 200 images, lead to a correlation coefficient of only 0.61 as opposed to 0.46 in case of the S/N measure. Five subjects took part in this experiment. They also observed that 2 kinds of distortions, low pass and logarithmic amplitude quantization, are systematically underestimated by the model.

Other recent studies on quality measures include Ohtsuka et al. (1989), who propose a method to evaluate quality of multiple impaired images. The original and distorted pictures are amplitude and spatial frequency weighted prior to differencing. They also propose an additive law of multiple impairments, which often results in an overestimate.

3.5 Models motivated by coding applications

Many authors have attempted to arrive at models of image quality that can be used for (perceptually) optimal coding of images. Only a few models are reviewed here. Netravali and Prasada (1977) use the spatial masking property of the visual system in amplitude quantization of digitized pictures. Using subjective tests, a visibility function, which indicates the visibility of noise, as a function of the masking function, is obtained. The masking function is a measure of local spatial detail. The visibility function is used in deciding the quantization step adaptively for each region or location. Simulations show that using this adaptive quantization technique, one can increase the coding efficiency by about 30 to 50 percent over nonadaptive techniques. Anderson and Netravali (1976) use the same criterion in the restoration of images.

Safranek and Johnston (1989) and Safranek et al. (1990) use an empirically derived masking model to optimize the sub-band image coder. The model incorporates spatial frequency sensitivity, contrast sensitivity and spatial masking. Using subjective tests, RMS noise sensitivity in each sub-band is obtained. Using the model, noise level targets are set for each pixel in each sub-band, thus fixing the allowed quantization noise at that point in that sub-band. Using this technique, good coding gains are obtained (transparent output quality requires 0.5 to 2.5 bits/pixel).

Girod (1989) proposed a non-linear spatio-temporal model of human threshold vision. The model is illustrated in Fig.3 and Fig.4. The signal processing that is performed in the fovea of this model is identical to that used by Lukas and Budrikis (1982). Using a linearized version of the model, Girod predicts the spatial and temporal masking data. Girod uses this model to predict the coding gain that can be achieved by incorporating masking effects in video coding. The model predicts very small coding gains due to masking. It is pointed out by Girod that the eye might possess other masking phenomena not included in his linearized model.

4 Multidimensional scaling analysis

Marmolin and Nyberg (1975) used multidimensional scaling techniques to identify the independent dimensions of image quality. They generated a number of images with a variety of different point spread functions, noise levels and contrast characteristics. Marmolin used pictures of size 5x5 cm, printed on hard copy. The images were used in viewing experiments where subjects were asked to judge the dissimilarity. The data of the subjective experiments was analyzed using multidimensional scaling programs. Marmolin identified the following 4 dimensions to be most important:

1. the first and most important quality dimension is sharpness

2. the second important dimension is noise

3. the third important dimension is image contrast

4. and the fourth important quality dimension could not be interpreted

The results of Marmolin's study have to be interpreted in light of the fact that they were based on only one scene. Hence it is not known whether the dimensions arrived at by this study are as important for other scenes. The 4th dimension, when also found in other scenes, might be easily interpreted. An experiment on a video display with many different scenes, might answer some of these questions.

Knowledge about the nature of interaction between the distorting stimulus variables is essential for understanding how picture quality varies in presence of multiple distortions. Linde (1981) investigated the interaction between physical edge blur and random noise using similarity data. Similarity between distorted still pictures was rated by subjects in a factorial design experiment involving stimuli distorted with varying degree of blur and noise. Analysis of the similarity data revealed interaction between the physical variables representing blur and noise. Physically constant noise intervals increased subjectively with increasing blur and physically constant blur intervals tended to decrease subjectively with increasing noise. The nature of the interaction was scene dependent. MDS analysis of the data using INDSCAL also resulted in configurations that were scene dependent. Linde identified the spatial frequency content of the scene as the main factor in deciding the nature of the interaction. Further experiments involving a variety of scenes will be required to see how systematic is such an interaction and to characterize its nature.

5 Nakayama et al.'s model

Nakayama et al. (1980) propose a model of image quality evaluation based on multidimensional scaling. According to the model (see Fig.5), the evaluation process in a human observer is subdivided into 2 sub-processes: the sensory process and the synthetic process. Response to an image stimulus in the sensory process can be described by a multidimensional scale D_k as

$$D_k = \phi_k(T_i, T_j)$$

where T_i and T_j are the physical parameters of the image and ϕ_k is the function used to compute D_k (see Fig. 5). The overall image quality Q is given as a response in the synthetic process

$$Q = \sum W_k D_k$$

where W_k are the weights of the underlying quality dimensions (attributes).

By applying the multidimensional scaling method to similarity data (collected in a subjective experiment with stimuli varying in contrast and luminance), the following dimensions were interpreted: subjective sharpness, color saturation and brightness. Subjective sharpness and brightness can be expressed as a function of the contrast and the luminance respectively. Quality is expressed as a linear combination of the 3 dimensions mentioned above. Optimal weights are derived using the data of the subjective experiment.

The most important dimension, subjective sharpness is studied further using another experiment with abstract stimuli. The parameters that were varied in the stimuli were: contrast, luminance and the spatial frequency characteristics. Using multiple regression analysis, the subjective sharpness is expressed as a function of a sharpness function (defined as optical transfer function of the system and the eye), contrast and effective luminance ($L_{max} - L_{min}$).

This study takes a new, promising approach. The study as it is, is largely incomplete and there are no other studies along these lines. Further study along these lines is worth pursuing.

6 Discussion

The objective measures found in the literature can be broadly classified into two categories, a class based mainly on MTF like properties and a class based on pixel value differences between the original and the distorted image of the same scene.

MTF based measures are a class of measures that were developed mainly to evaluate image and display sharpness. These are a class of measures that performed better than many other measures developed for the same purpose. These studies often do not distinguish between quality and sharpness.

The measures based on pixel value differences mainly aim at predicting quality in the presence of different artifacts in the image. Almost all quality measures here try to incorporate properties of the human visual system into the model. Among these, the model of Lukas and Budrikis is most exhaustive and is based on sound arguments.

The multidimensional scaling analysis of Marmolin and Nyberg and Nakayama et al. helps identify underlying dimensions of image quality. The dimensions identified by them seem to be in general agreement, although a good comparison cannot be made mainly because details of the subjective tests by Nakayama et al. are not available. The model proposed by Nakayama et al. represents a new and promising approach.

There is no single objective measure that performs best for all scenes and distortions.

It is difficult to compare the performance of objective measures proposed by different authors, mainly because of two reasons:

1. different authors use different performance indices to assess the performance of their objective quality measures, and

2. different studies are based on different stimulus sets.

A conclusion that can be drawn from the studies of Marmolin, Limb, Lukas and Budrikis is that local measures that give due weight to local image properties, in general perform better than global measures. In some cases the gain obtained using local measures is high, while in other cases it is marginal (for example in the case of additive white noise distortions)

These observations support the postulate (Limb, Lukas and Budrikis) that viewers tend to evaluate the property or base their ratings mainly on critical areas rather than on the whole picture.

7 Concluding remarks

Most of the studies on quality prediction aim at predicting the quality directly from the image, implicitly assuming it to be a one dimensional quantity, contrary to what Marmolin and Nyberg and Nakayama et al. found in their multidimensional scaling analysis. The intermediate step to compute the values along these dimensions is missing in these methods, and that is probably the main reason for their poor performance. Hunt and Sera (1978) do use Marmolin's findings in selecting their test stimuli, but their data analysis does not reflect these ideas.

In almost all studies so far, the stimuli had different kinds of distortions, mainly additive noise, blurring, image coding artifacts etc. Most authors analyze the data from experiments collectively as one group, gaining little insight into the problem. Better insight into the problem would be gained if distortions are studied systematically, starting with simple cases like blurring, noise etc. Once the quality change under simple cases is understood, it will be simpler to predict the quality in complex situations, like coding artifacts.

In arriving at local measures, the images were divided into rectangular regions of about 1 degree extent. These divisions are arbitrary in the sense that they are not guided by the structure in the image, like edges or flat regions, which will actually be the regions of viewers interest, or regions where distortions are most easily detected (the 'critical areas' in Lukas's terms). The fact that even such crude divisions perform better than global measures demonstrates the power of local measures.

Appendix

The generalized sharpness measure (Hultgren 1990) is formulated as follows. If in the spatial frequency domain of the image, the amplitude at spatial frequency u is $A(u)$, then the amplitude detected by the visual system will be

$$V(u) = E(u)M(u)A(u)$$

where $M(u)$ is the MTF of the system and $E(u)$ is that of the visual system. The detected signal is transformed into

$$f(V(u)) = f(E(u)M(u)A(u))$$

where $f(V)$ is either a linear or a non-linear function of the detected amplitude V.

The Generalized Sharpness Measure (GSM) is given by the ensemble average of the perceived information, computed with non-uniform weighting $1/u$. The weighting function $(1/u)$ represents the frequency of occurrence of amplitude information, i.e,

$$GSM = \frac{1}{s_0} < f(V) >= \frac{1}{s_0} < f(E(u)M(u)A(u)) >$$

where the expectation value $<>$ is computed over the ensemble of information in a scene.

If we make use of the fact that

$$\frac{1}{u} du = d(logu)$$

and assume that the function $f(V)$ is linear, GSM reduces to Higgins measure

$$S_{Higgins} = \int_0^\infty E(u)M(u)d(log(u))$$

Additionally, when s_0 is taken equal to the expectation value of the perceived information in the original scene,

$$S_0 =< f(E(u)A(u)) >$$

and the MTF of the eye is assumed to be an ideal band pass filter, GSM reduces to Granger's measure

$$S_{Granger} = \frac{\int M(u)d(log(u))}{\int d(log(u))}$$

The GSM reduces to Cohen and Carlson's and Barten's measure when the scale value of S is chosen to be just noticeable differences (JNDs).

References

1. Anderson G. L. and A. Netravali (1976), Image restoration based on a subjective criterion, IEEE Syst. Man Cybern., vol SMC-6, pp 845-853.

2. Barten P. (1989), The effects of picture size and definition on perceived image quality, IEEE Trans. Electr. Dev., vol 36, pp 1865-1869.

3. Barten P. (1990), Evaluation of subjective image quality with the square-root integral method, J. Opt. Soc. Amn., vol 7, pp 2024-2031.

4. Carlson C. R. and R. W. Cohen (1980), A simple psychophysical model for predicting the visibility of displayed information, Proc. SID, vol 21, pp 229-246.

5. Girod B. (1989), The information theoretical significance of spatial and temporal masking in video signals, Human Vision, Visual Processing and Digital Display, Proc. SPIE vol 1077, pp 178-187.

6. Granger E. M. and K. N. Cupery (1972), An optical merit function (SQF) which correlates with subjective image judgements, Phot. Sci. Eng., vol 15, pp 221-230

7. Higgins G. C. (1977), Image quality criteria, J. App. Photo. Eng., vol 3, pp 53-60.

8. Hultgren B. (1990), Subjective quality factor revisited (A unification of objective sharpness measures), Human Vision and Electronic Imaging: Models, Methods and Applications, Proc. SPIE vol 1249, pp 12-22.

9. Hunt B. R. and G. F. Sera (1978), Power-law stimulus-response models for measures of image quality in nonperformance environments, IEEE Trans. Syst. Man Cybern., vol SMC-8, pp 781-791.

10. Limb J. (1979), Distortion criteria of the human viewer, IEEE Trans. Syst. Man Cybern., vol SMC-9, pp 779-793

11. Linde L. (1981), Similarity of distorted pictures: on the interaction between edge blur and random noise, FAO report C 53004-H9, Swedish National Defense Research Institute, Stockholm.

12. Lukas F. and Z. Budrikis (1982), Picture quality prediction based on a visual model, IEEE Trans. Commun., vol COM-30, pp 1679-1692

13. Marmolin H. and S. Nyberg (1975), Multidimensional scaling of subjective image quality, FAO report C 30039-H9, Swedish National Defense Research Institute, Stockholm.

14. Marmolin H. (1986), Subjective MSE measures, IEEE Trans. Syst. Man Cybern., vol SMC-16, pp 486-489.

15. Nakayama T., M. Kurosu, K. Honjyo and T. Nishimoto (1980), Evaluation and prediction of displayed image quality, Proc. SID. 80 Digest, pp 180-181.

16. Netravali A. and B. Prasada (1977), Adaptive Quantization of picture signals using spatial masking, Proc. IEEE vol 65, pp 536-548.

17. Ohtsuka S., M. Inoue and K. Watanabe (1989), Quality evaluation of pictures with multiple impairments based on visually weighted errors, Proc. SID, vol 30, pp 3-8.

18. Safranek R. and J. Johnston (1989), A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression, Proc. ICASSP 89, vol 3, pp 1945-1948.

19. Safranek R., J. Johnston and R. Rosenholtz (1990), A perceptually tuned sub-band coder, Human Vision and Electronic Imaging: Models, Methods and Applications, Proc. SPIE vol 1249, pp 284-293.

20. Westerink J. H. D. M. (1988), Perceptually weighted spatial resolution, IPO report no. 587.

21. Zetzsche C. and G. Hauske (1989), Multiple channel model for the prediction of subjective image quality, Human Vision, Visual Processing and Digital Display, Proc. SPIE vol 1077, pp 209-216.

TABLE I

DISTORTIONS ADDED TO FIVE ORIGINAL PICTURES*

1	Quantization to 64 gray scale levels (6 bits)
2	Quantization to 32 gray scale levels (5 bits)
3	Quantization to 16 gray scale levels with 4-level dither added (produces rapid switching between adjacent levels)
4	Eleven-level DPCM coding with 2-dimensional predictor (error larger at edges)
5	Five-level DPCM coding with 2-dimensional predictor
6	Five-level DPCM coding with one-dimensional predictor
7	Added Gaussian noise, 43db p-p signal/RMS noise
8	Added Gaussian noise, 40db p-p signal/RMS noise
9	Added Rectangular noise, 40db p-p signal/RMS noise
10	Added Gaussian noise, 35db p-p signal/RMS noise
11	Added Gaussian noise, 30db p-p signal/RMS noise
12	Low-pass filtering, one-dimensional
13	Low-pass filtering, 2-dimensional
14	Masking Noise - Rectangular noise added in proportion to activity function, A_{df}, at each pel
15	Same as 14, but with rectangular noise 7db greater in amplitude
16	Masking Noise - Rectangular noise added when activity function, A_{df}, is greater than a threshold

TABLE II

THE 24 TEST PICTURES

1.	Original	
2.	5-bit uniform quantization	
3.	4-bit uniform quantization	
4.	4-level dither/4-bit uniform quantization	
5.	DPCM pred. 1, quantized to 5 levels	
6.	DPCM pred. 1, quantized to 9 levels	
7.	DPCM pred. 2, quantized to 9 levels	
8.	DPCM pred. 3, quantized to 9 levels	
9.	Additive Gaussian noise (34db SNR)	(i) static
10.	Additive Gaussian noise (34db SNR)	(ii) dynamic
11.	Additive Gaussian noise (40db SNR)	(i) static
12.	Additive Gaussian noise (40db SNR)	(ii) dynamic
13.	Spatially filtered Gaussian noise (22db SNR)	(i) LPF BW = 1cpd
14.	Spatially filtered Gaussian noise (22db SNR)	(ii) BPF fo = 4cpd BW = 1cpd
15.	Spatially filtered Gaussian noise (22db SNR)	(iii) BPF fo = 8cpd BW = 1cpd
16.	Spatial low-pass filtering	(i) BW = 2cpd
17.	Spatial low-pass filtering	(ii) BW = 4cpd
18.	Temporal low-pass filtering (recursive 1st order exponential)	(i) t = 30 ms
19.	Temporal low-pass filtering (recursive 1st order exponential)	(ii) t = 60 ms
20.	Temporally filtered Gaussian noise (34db SNR)	(i) t = 30 ms
21.	Temporally filtered Gaussian noise (34db SNR)	(ii) t = 60 ms
22.	Temporal interpolation - two fields out of four	
23.	Edge enhancement scheme	
24.	Adaptive interframe coding based on 2-D masking function [31].	

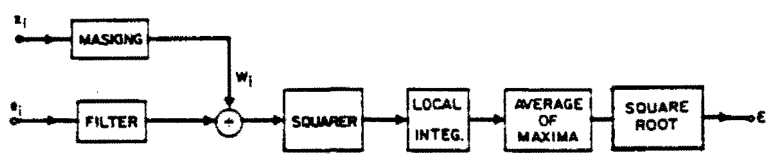
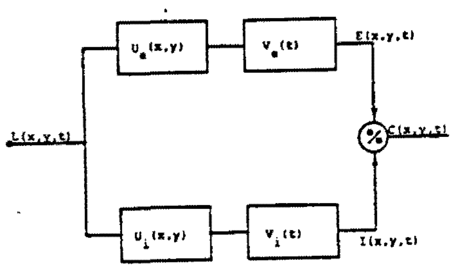


Fig. 1. Model assuming sequential processing of the error signal by a filtering stage and a masking stage.



The model of the visual filter.

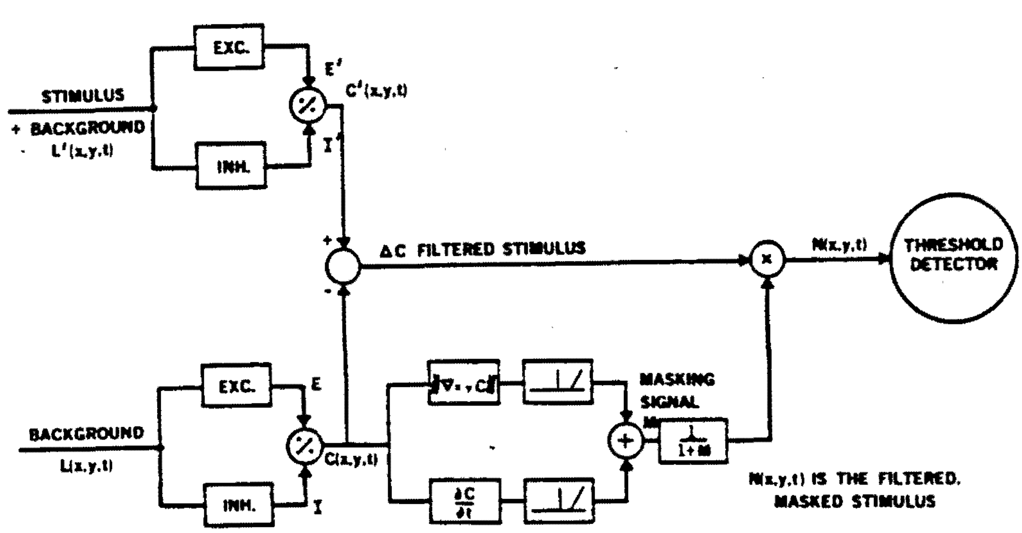


Fig. 2. Complete model of threshold vision.

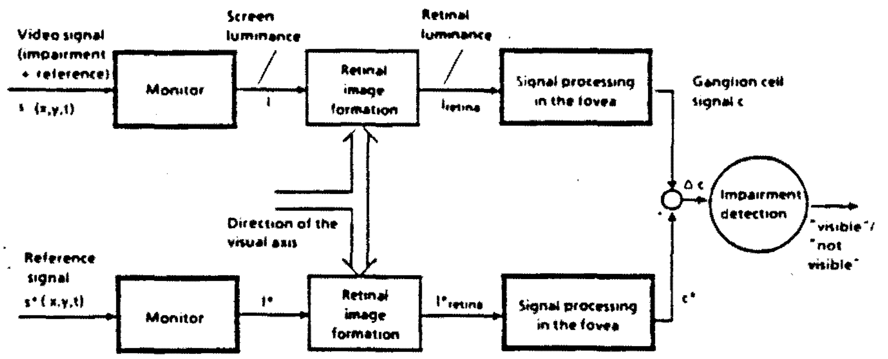


Fig. 3 Nonlinear 3-D threshold model of human brightness perception. Bold blocks indicate nonlinear system components.

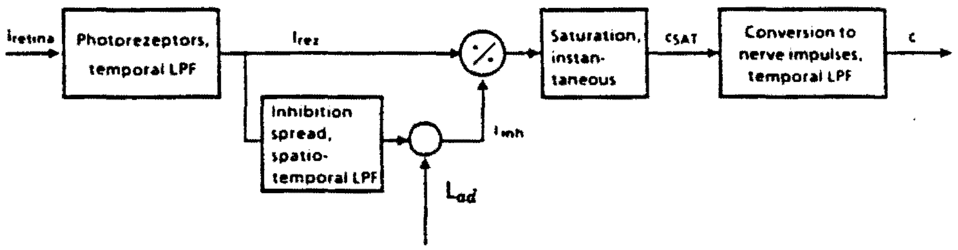


Fig. 4. Nonlinear model of the signal processing in the human fovea. LPF - lowpass filter.

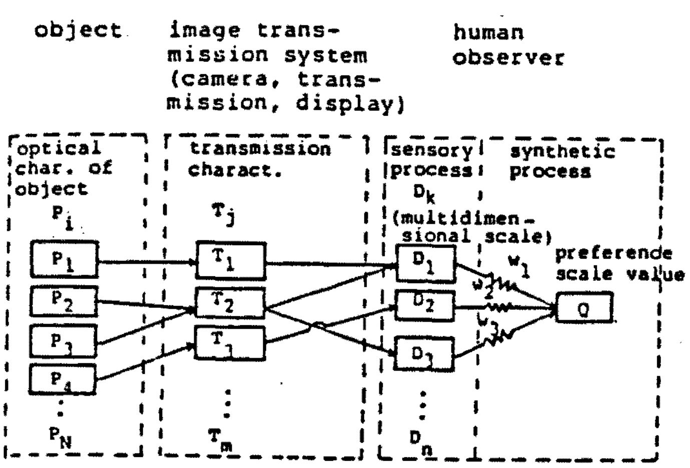


Fig. 5. Model of subjective image quality evaluation