# A CONSTANT SUBJECTIVE QUALITY MPEG ENCODER

*Fu-Huei Lin and Russell M. Mersereau*

School of Electrical & Computer Engineering,
Georgia Institute of Technology, Atlanta GA 30332
fuhuei@eedsp.gatech.edu, rmm@eedsp.gatech.edu

## ABSTRACT

In this paper, we present an video objective quality measure which has good correlation with subjective tests. We then introduce the objective measure into the design of an MPEG encoder. The new MPEG encoder extracts four features (bit rate, a feature that measures blockiness, one that measures false edges, and one that measures blurred edges) from the input and output video sequences and feeds those features into a four-layered back-propagation neural network which has been trained by subjective testing. Then the system uses a simple feedback technique to adjust the GOP (group of pictures) bit-rate to achieve a constant subjective quality output video sequence.

## 1. INTRODUCTION

Video quality measures play an important role in many fields of video processing, especially in video coding. The most widely used quality measure is mean squared error, which does not correlate well with human visual perception. Since a human observer is the end user of most of the video information, a video quality measure that is based on human visual perception is more appropriate for video quality prediction.

Miyahara's [1] pioneering image quality assessment system extracts five features from images. The first two features refer to random errors with different weighting, based on properties of perception. The third feature refers to the end of block disturbances. The first two features measure the structured errors, such as ringing, induced by image structure. The feature dimension is then reduced from five to three by a principal component analysis. The correlation between subjective scores and mean objective scores is 0.88. Webster *et al.* [2] developed an objective video quality assessment system that emulates human perception. Three features are extracted from input and output video sequences. The first feature measures the spatial distortion. The second and the third features both measure the temporal distortion. These features are combined in a linear model by using least squares error criterion. The correlation coefficient between the subjective scores and the estimated scores was 0.92 for the training set. With the testing set, the correlation coefficient was 0.94. Davies *et al.* [3] proposed an automated image quality assessment system for assessing image impairment. A bank of spatial and temporal filters and a 3-layer neural network are used to produce quantitative CCIR gradings that match those made by an expert human assessor. The network has 40 hidden units, and five output units each representing one of the five CCIR grades. It was trained with 20,000 iterations.

In Miyahara's system, the principal component technique is useful when features in the feature space are *jointly normal distributed.* It is noted that using more features results in excellent training, but poor testing results. This is because redundant features will track random errors (details) in the data during training. The linear model proposed by Webster does not fit the nonlinear relationship commonly observed between features and mean opinion scores. Davies's neural network is complicated and requires a large training set.

In this paper, a four-layered, fully connected, back-propagation neural network is trained by a subjective test. This network has only four input, six hidden (four for the first hidden layer, two for the second hidden layer) and one output units. In Section 2, the subjective test experiments will be described. Objective quality measures are presented in Section 3. The constant subjective quality MPEG encoder is shown in Section 4, followed by the conclusion in Section 5.

## 2. SUBJECTIVE TEST

In the subjective test, twelve video sequences containing scene changes, motion, details, lighting changes, zooming and panning were coded using MPEG at different bit rates resulting in 46 video sequences. Thirty

This work was supported in part by the Joint Services Electronics Program under contract DAAH-04-93-G-0027.

 0-7803-2431-5/95 $4.00 

untrained subjects from Digital Signal Processing Laboratory of Georgia Tech volunteered for the subjective test. A procedure similar to CCIR Recommendation 500-3's [4] high and low indirect anchoring with the change from indirect to direct anchoring was chosen.

The video sequences (color) were 160 by 128 pixels in size. Frame rate was 30 frames/second. The reason for choosing small size is to save the disk space. They were zoomed by two (i.e. 320x256 pixels) and displayed on a Hitachi monitor. The Video sequences were presented in a random order. The Viewing distance was six times the pictures height. A low ambient lighting condition was used. The subjective test contained two sections and each section did not last for more than 15 minutes.

A discrete quality scale is used in [1, 2, 3], but we employed a continuous quality scale. There are several reasons for choosing continuous quality scales. First, the distances between quality levels on discrete scales are not equal [5]. Averaging measurements from discrete quality scales is suspect. Further, different countries or regions may impart different interpretations to those scales [5]. Finally, it has a smaller quantization effect.

## 3. DETERMINING OBJECTIVE QUALITY

The four features extracted from the input output video sequence pairs are:

**Bit Rate :** The bit rate as encoded in the MPEG bitstream, which is a crude overall indicator of the video quality.

**False Edges Effect [6] :** False edges in the output image are more dense than the corresponding edges in the input image. The negative pixel values in the difference of the input image processed by a Sobel operator minus the output image processed by the same operator indicate the presence of false edges. The standard deviation of the negative pixel values is used as the feature.

**Blocking Effect :** This is a common coding artifact with block coding. First, the sum of all Sobel gradients at the block boundary and one pixel next to the block boundary is computed for the input and output frames separately. Then, the feature is computed as the difference of these two sums.

**Blurred Edges Effect [6] :** The edges in the input image are more dense (higher value) than the corresponding edges in the output image. The positive pixel values in the difference of the Sobel input image minus Sobel output image again indicate the presence of blurred edges. The standard deviation of the positive pixel values is chosen as the feature.

Each feature was computed on the Y frame, U frame, and V frame. The feature was weighted as 1.0, 0.3, and 0.3 on the Y frame, the U frame, and the V frame. Those weightings were obtained by trying several weightings and choose the one which had the best correlation with the subjective data. The average value of the features over all frames in the video sequence is used. These features were used for linear regression or were fed into a back-propagation neural network.

Linear regression uses least squares fitting to get regression coefficients on training data. Back-propagation neural network [7] (multi-layer perceptron) is a feedforward neural network, i.e. there are no feedback connections between its layers and the neurons of the layers themselves. It consists of nonlinear neurons. The neurons perform a weighted sum of their inputs (features) and pass the sum through a sigmoid non-linearity.

A four-layered, fully connected, back-propagation network is used. It consists of input, output and two hidden layers. Each hidden and output unit has a bias. In this experiment, four inputs, four units for the first hidden, two units for the second hidden and, one output units were used. The weights and offsets were initialized using small random values. The features and mean opinion scores were normalized and the learning rate was set at 0.1. It is noted that the error criterion can not be very small. This is to prevent the network from over learning (train), which would cause it to perform well on the training set but hurt performance on the testing set. This neural network converged in 500 iterations during training.

Both linear regression and the back-propagation neural network used 50% of the video sequences (i.e. six sets of MPEG encoded video sequences resulting in 23 video sequences) for training which were used to test the other 50%. Then, interchanging the training and testing sets, the procedures were repeated. The correlation coefficient between mean opinion scores of 30 subjects on the 46 video sequences of linear regression are 0.92 on the testing sets and 0.96 on the training sets. The neural network has nonlinear mapping capability between features and mean opinion scores; its correlation on testing and training video sequences are 0.95 and 0.98. Fig. 1 shows the results on test sets of the linear regression and neural network.

## 4. A CONSTANT SUBJECTIVE QUALITY MPEG ENCODER

Our results on subjective quality were then used to implement a constant subjective quality MPEG encoder. Video can take advantage of variable bit rate (VBR) operation because the activity in a video sequence is variable. Since features are averaged over all frames, it is reasonable to assume that the same quality measure can be applied to each Group Of Pictures (GOP,12 frames in our implementation.). The bit-rate control is done on GOPs. For each desired subjective quality, an approximation to the constant bit-rate encoding for that quality is used as an initial guess. The bit-rate is adjusted by examining the difference between the current and desired objective qualities. This scale or step_size is decreased as the iteration number increases, i.e. $scale/1 + log(iteration\_number)$. The bit-rate change is embedded in the quantizer_scale_code of the slice or macroblock [8]. Fig. 2 shows the test results on a video sequence which contains a scene change at GOP 9. All GOPs converge to the desired quality in 3–6 iterations.

## 5. CONCLUSION

In this paper, we first present a video subjective quality measure with a nonlinear mapping between features and mean opinion scores, a higher correlation coefficient, lower complexity and shorter training time than previously reported. The good performance partially comes from choosing a continuous subjective quality scale. The continuous quality scale reduces quantization effects and also simplifies the neural network structure.

We then introduce the video subjective quality measure into the design of an MPEG encoder resulting in a constant subjective quality output video sequence. A waveform coder, such as MPEG, has blocking artifacts, false edges, and blurred edges. For a general coder, CCIR [9] recommends a list of picture (video) quality degradation factors as: image blur, edge busyness, false contouring, granular noise, "dirty window" effect, movement blur, and jerkiness. These picture (video) quality degradation factors can be used as parameters in video quality assessment techniques which are similar to the case of the Diagnostic Acceptability Measure in speech subjective quality assessment [10].

The proposed constant subjective quality MPEG scheme can be implemented where low decoding complexity is desired but encoding complexity is inconsequential. These asymmetric applications are those that require frequent use of the decompression process, but for which the compression process is performed once at the production of the program, such as electronic publishing (education and training, travel guidance, videotext), games and entertainment.

## 6. ACKNOWLEDGEMENTS

The authors would like to thank those who volunteered for the subjective test.

## 7. REFERENCES

[1] M. Miyahara, K. Kotani, and V. R. Algazi, "Objective Picture-Quality Scale (PQS) for Image Coding," pp. 859–862, *SID Digest*, 1992

[2] Arthur A. Webster, Coleen T. Jones, Margaret H. Pinson, Stephen D. Voran and Stephen Wolf, "An Objective Video Quality Assessment System Based on Human Perception," *Human Vision, Visual Processing and Digital Display IV*, SPIE 1913, pp. 15–26, 1993.

[3] Ian R.L. Davies and David Rose, "Automated Image Quality Assessment," *Human Vision, Visual Processing and Digital Display IV*, SPIE 1913, pp. 27–36, 1993.

[4] *CCIR Recommendation 500-3*, "Method for the Subjective Assessment of the Quality of Television Pictures," 1986.

[5] *CCIR Report 1082-1*, "Studies Toward the Unification of Picture Assessment Methodology," pp. 384–414, 1990.

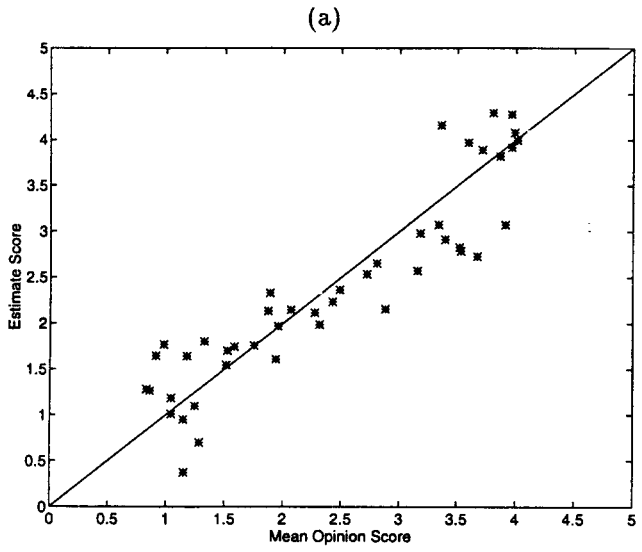
[6] S. Wolf, "Features for Automated Quality Assessment of Digitally Transmitted Video," U.S. Department of Commerce, National Telecommunications and Information Administration Report 90–264, June 1990.

[7] David E. Rumelhart, James L. McClelland, and the PDP Research Group, *Parallel Distribution Processing*, MIT Press, 1986.

[8] *ISO/IEC 13818-2.*

[9] *CCIR Rec. 813*, "Methods for Objective Quality Assessment in Relation to Impairments from Digital Coding of Television Signals," 1992.

[10] W. D. Voiers, "Diagnostic Acceptability Measure for Speech Communication Systems," *ICASSP*, pp. 204–207, 1977.



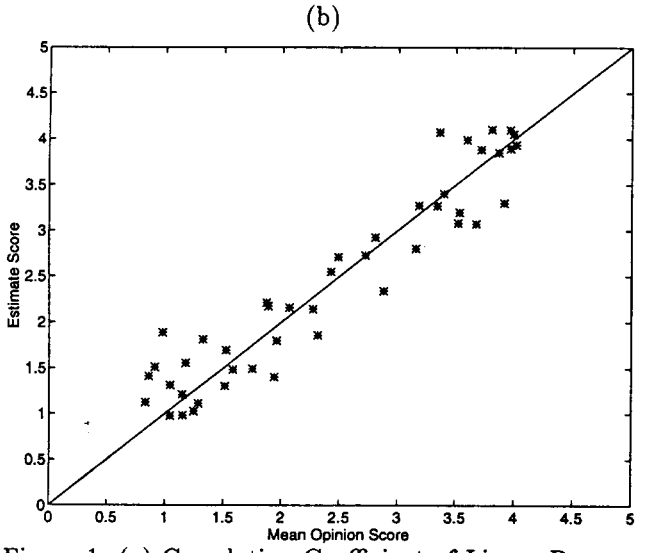
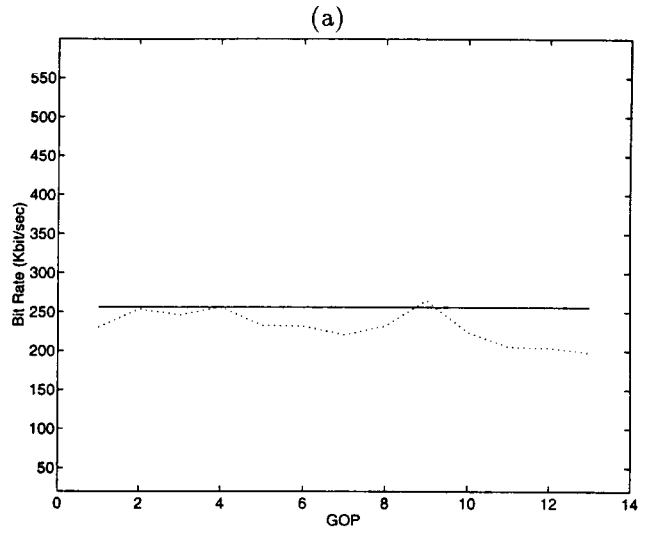




Figure 1: (a) Correlation Coefficient of Linear Regression on Test Sets = 0.920245. (b) Correlation Coefficient of BP Neural Network on Test Sets = 0.950364.



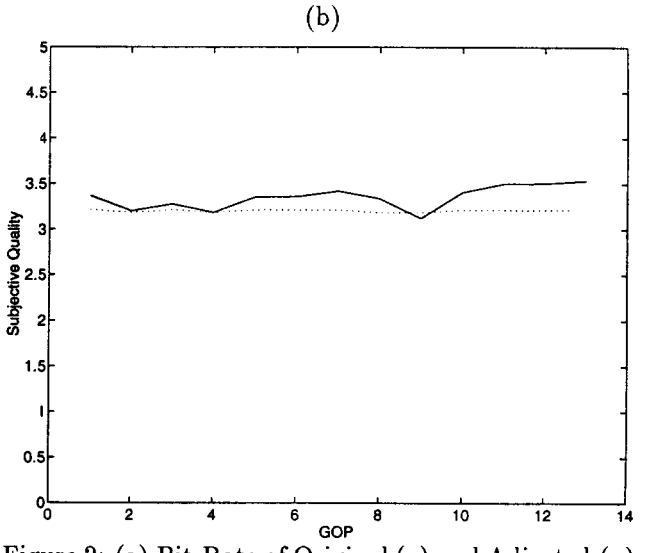




Figure 2: (a) Bit-Rate of Original (–) and Adjusted (..) video sequences. (b) Subjective Quality of Original (–) and Adjusted (..) video sequences. Values 5, 4, 3, 2, and 1 represent excellent, good, fair, poor, and bad qualities, respectively.