Intra-Mode Indexed Nonuniform Quantization Parameter Matrices in AVC/H.264

Jing Hu and Jerry D. Gibson
Department of Electrical and Computer Engineering
University of California, Santa Barbara, California 93106-9560
Email: jinghu@umail.ucsb.edu and gibson@ece.ucsb.edu

***Abstract*— We generate nonuniform quantization parameter(QP) matrices to improve the perceptual quality of reconstructed video in the AVC/H.264 standard. The resulting 9 QP matrices are indexed to the 9 intra-frame prediction modes of the 4×4 blocks and can be applied to the 16×16 macroblock(MB)s which engage only the first 4 out of the 9 intra-modes. We also studied how the nonuniform QP matrices are affected by the local luminance level and the video frame rate. Two subjective experiments exploiting the luminance, texture, and temporal masking properties of the human vision system are conducted to generate the results. A third subjective experiment is conducted to evaluate the derived scheme. The data we collect from this experiment show that using the QP matrices instead of a uniform QP considerably improves the perceptual quality of AVC/H.264 compressed video sequences. Some implementation details are discussed.**

I. Introduction

The recently finalized Advanced Video Coding(AVC) standard, designated ITU-T H.264 and MPEG-4 Part 10, offers a coding efficiency improvement by a factor of two over previous standards and its network abstraction layer (NAL) transports the coded video data over networks in a more "network-friendly" way [5]. Because of these two features, the AVC/H.264 standard is very likely to emerge as the method of choice for the next generation video networks.

In this paper we investigate the visibility of quantization errors in different frequency coeffocients after the integer transform, as a function of the intra-mode, local luminance and frame rate. Two subjective experiments are designed and conducted to explore the luminance, texture and temporal masking of the human vision system (HVS) [4] and the perceptually optimal QP matrices are calculated from the data collected in the experiments. A third subjective experiment is conducted to evaluate the derived scheme. The data we collect from this experiment and other simulation results show that using the QP matrices instead of a uniform QP considerably improves the perceptual quality of AVC/H.264 compressed video sequences.

The authors gratefully acknowledge discussions with Prof. Miguel Eckstein. This work was supported by the California Micro Program, Applied Signal Technology, Dolby Labs, Inc. and Qualcomm, Inc., by NSF Grant Nos. DGE-0221713, Nos. CCF-0429884 and CNS-0435527, and by the UC Discovery Grant Program and Nokia, Inc..

II. Intra-frame prediction

Intra-frame prediction is a new feature in AVC/H.264 which contributes to $6-9\%$ in bit saving by removing the spatial redundancy in neighboring 4×4 blocks or 16×16 MBs. If a block or MB is encoded in intra mode, a prediction block is formed based on previously encoded and reconstructed surrounding pixels. The prediction block P is subtracted from the current block prior to encoding. For the luminance samples, P may be formed for each 4×4 sub-block or for a 16×16 MB. There are a total of 9 optional prediction modes for each 4×4 luminance block as shown in Fig. 1 and 4 optional prediction modes (mode 0 to 3 in Fig. 1) for a 16×16 luminance MB.

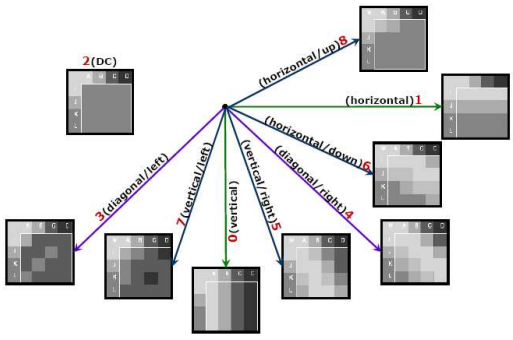


Fig. 1. The intra prediction modes for 4×4 blocks in AVC/H.264.

The exact integer transform (1) instead of the discrete cosine transform(DCT) is used to avoid the mismatch between encoder and decoder in the discrete cosine transform (DCT)-based codecs, as

$$\mathbf{Y} = \left\{ \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \mathbf{X} \begin{bmatrix} 1 & 2 & 1 & 1 \\ 1 & 1 & -1 & -2 \\ 1 & -1 & -1 & 2 \\ 1 & -2 & 1 & -1 \end{bmatrix} \right\} \otimes \mathbf{E_f}, \quad (1)$$

where the post-scaling factors contained in $\mathbf{E_f}$ are combined with the uniform quantization following the forward transform. A total of 52 values of quantization stepsizes are supported by AVC/H.264 and they are indexed by QPs as shown in Table I. Note that these values are arranged so that an increase of 1 in QP means an increase of quantization stepsizes by approximately 12% and a reduction of bit rate by approximately 12%.

When a MB is predicted using one of the 4 intra-modes, the residual 16×16 block is broken into 16 4×4 blocks and the transform and quatization process is the same as the residual 4×4 block after intra-frame prediction using one of the 9 intra-modes.

The quantized 4×4 block $\hat{\mathbf{Y}}$ is post-scaled by the components in $\mathbf{E_i}$ and then inverse transformed to get the reconstructed residual box $\hat{\mathbf{X}}$:

$$\hat{\mathbf{X}} = \begin{bmatrix} 1 & 1 & 1 & \frac{1}{2} \\ 1 & \frac{1}{2} & -1 & -1 \\ 1 & -\frac{1}{2} & -1 & 1 \\ 1 & -1 & 1 & -\frac{1}{2} \end{bmatrix} \hat{\mathbf{Y}} \otimes \mathbf{E_i} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & \frac{1}{2} & -\frac{1}{2} & -1 \\ 1 & -1 & -1 & 1 \\ \frac{1}{2} & -1 & 1 & -\frac{1}{2} \end{bmatrix}. \quad (2)$$

III. Nonuniform QP matrices indexed to intra-modes

We are interested in the visibility of the error in the reconstructed residue $\hat{\mathbf{X}}$ caused by quantizing the components of $\mathbf{Y}$. Therefore we first set only one entry in $\hat{\mathbf{Y}}$ to a nonzero value and get an 4×4 individual error block using Eq. (2). We repeat this process for all the entries in $\hat{\mathbf{Y}}$ and normalize all the 16 error patterns so that they have the same mean squared values. The 16 normalized quantization error patterns are shown in Fig. 2, where $I = i, J = j$ means this error pattern is obtained when only the entry (i, j) of $\hat{\mathbf{Y}}$ is nonzero.

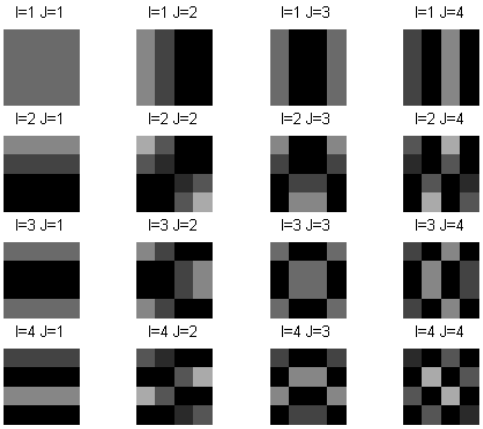
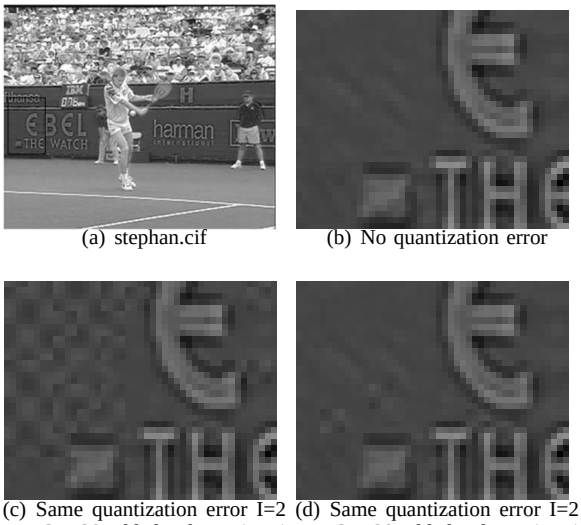


Fig. 2. Normalized quantization error patterns in AVC/H.264

Texture masking refers to the reduction in visibility of one image component caused by the presence of another image component with similar spatial location and frequency content [4]. By comparing the 16 individual error patterns to the 9 intra-prediction blocks in Fig. 1, we conjectured that the quantization error is texture-masked by the image content to different levels depending on the similarities between the error pattern and intra-prediction mode co-existing in one 4×4 block. For example, as shown in Fig. 3, when the same quantization error $I = 2$ $J = 2$ is added only to the 4×4 blocks coded with intra-mode 2 (Fig. 3(c)) or only to the 4×4 blocks coded with intra-mode 4 (Fig. 3(d)), the error with diagonal pattern is considerably masked by intra-mode 4 that also has a diagonal pattern while not masked by intra-mode 2 that has a monotone pattern.

We therefore designed and conducted a subjective experiment (details in Appendix I) to measure the QPs of each frequency coefficient, i.e., each entry in $\mathbf{Y}$, for each intra mode, at threshold when the quantization error is just noticeable. For each image, each subject and each intra-mode, we get a 4×4 matrix which contains the QPs at threshold for all the 16 error patterns, respectively. Correlations among all the 4x4 matrices are computed and the averages are calculated across all the subjects and all the tested images, as shown in Fig. 4. The 9 points on the left are the averages of correlations among the same intra-modes and the point on the right is the average of correlations across the intra-modes. We can see that good correlation is achieved between the same intra-modes, which shows the consistency of the QP matrices of each intra-mode in all the subjects and tested images. Only the 9 intra-modes for the 4×4 blocks are tested in this experiment and the QP matrices indexed to intra-modes 0 to 3 can be used for the 4×4 blocks in the residue block derived from 16×16 intra-predictions. The 9 QP matrices are presented in Table II.



(a) stephan.cif (b) No quantization error

(c) Same quantization error I=2 J=2 QP=38 added only to 4×4 blocks coded by intra-mode 2

(d) Same quantization error I=2 J=2 QP=38 added only to 4×4 blocks coded by intra-mode 4

Fig. 3. The same quantization error is texture-masked by different intra-modes to different levels.

For the inter-coded blocks/MBs, the residual blocks/MBs are generated after motion estimation. No intra-modes are calculated. Nevertheless by calculating the intra-modes of the to-be-inter-coded blocks/MBs the basic texture in the block/MB is detected and the 9 QP matrices can be applied. The quantization error has different perceptual effects when presented in a video sequence than when presented in still images, and it also depends on the local background luminance.

A second subjective experiment (details in Appendix II) exploited the three other masking types — luminance masking, baseline contrast masking and temporal masking, as well as texture masking as in the first subjective experiment. We found out that the correspondence between intra-modes and the QP matrices in experiment 2 is similar to that in experiment 1, except that QP values in experiment 2 are much smaller than

TABLE I

QUANTIZATION STEP SIZES IN AVC/H.264 CODEC.

QP	0	1	2	3	4	5	6	7	8	9	10	11	12	...
QStep	0.625	0.6875	0.8125	0.875	1	1.125	1.25	1.375	1.625	1.75	2	2.25	2.5	...
QP	...	18	...	24	...	30	...	36	...	42	...	48	...	51
QStep	...	5	...	10	...	20	...	40	...	80	...	160	...	224

the QP values obtained in experiment 1 because of temporal error propagation in experiment 2. So we only looked at the average QP values across all intra-modes and all subjects as a function of local luminance and frame rate, which are plotted in Fig. 5. From this experiment we can draw the conclusion that visibility of the error increases as the background luminance increases, and decreases, although less significantly, as the frame rate increases.

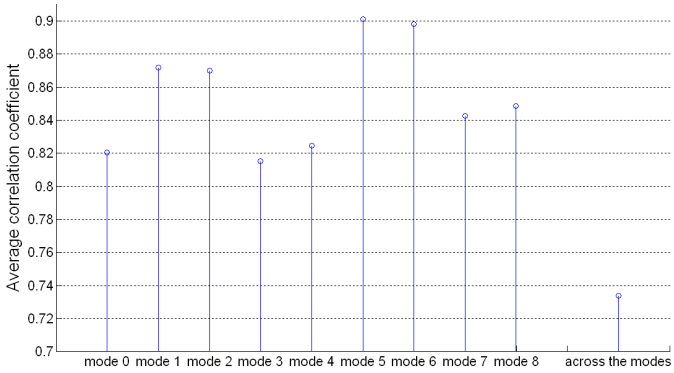


Fig. 4. Averages of correlations among all QP matrices in the first experiment.

IV. RESULTS AND IMPLEMENTATION

All the QP matrices are embedded in both the encoder and decoder. For the intra-coded frames, the intra-modes for each 4×4 block or each 16×16 MB are already calculated and transmitted from the encoder to the decoder, so the nonuniform QP matrices do not require additional rate. In Fig. 6 we show that using nonuniform QP matrices results in better perceptual quality of the images than using a uniform QP, which is set as the average of all the values in the nonuniform QP matrices, when either 4×4 block or 16×16 MB intra-prediction is involved. For the inter-coded frames/MBs, since the intra-modes are not available either at the encoder or at the decoder, we can calculate the intra-modes just as is done for the intra-coded frames/MBs and send the intra-modes from the encoder to the decoder. The computation consumed in calculating the intra-modes is negligible compared to the computation involved in motion estimation. Spending extra bits, averaging around 3 bits per 4×4 block to send the intra-modes, however, is unrealistic since the number of bits consumed by each 4×4 block is really small, usually under 15 in AVC/H.264. Therefore, it is advisable to calculate the intra-mode of each inter-coded MB even when smaller block sizes are engaged in motion compensation so that less than 2 bits need to be sent for each MB, which results in a rate increase of less than 2% for low motion videos and even less for medium and high motion videos. In Fig. 7 we show two inter-coded frames, one being coded using uniform QP and the other being coded using nonuniform QP matrices when 16×16 intra-modes are calculated.

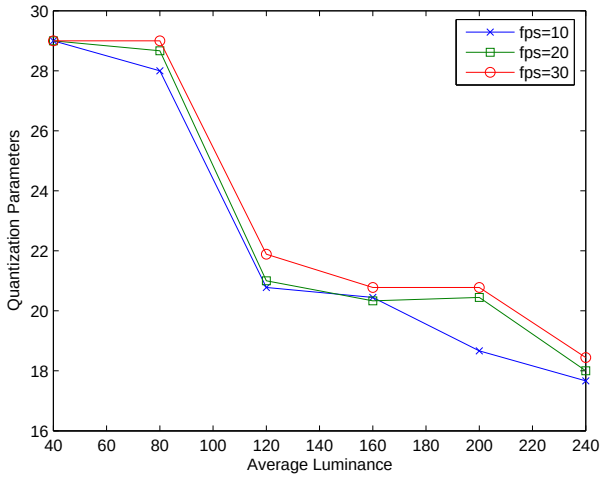


Fig. 5. The average QPs for error pattern I=1 J=1 over all the intra-coding modes.

TABLE II

GENERATED NONUNIFORM QP MATRICES INDEXED TO INTRA-PREDICTION MODES.

$$QP_0 = \begin{bmatrix} 33 & 33 & 34 & 38 \\ 32 & 34 & 35 & 40 \\ 32 & 33 & 36 & 40 \\ 34 & 36 & 38 & 43 \end{bmatrix}, \quad QP_1 = \begin{bmatrix} 31 & 28 & 30 & 33 \\ 28 & 30 & 31 & 36 \\ 31 & 32 & 34 & 39 \\ 33 & 36 & 38 & 42 \end{bmatrix},$$

$$QP_2 = \begin{bmatrix} 31 & 30 & 31 & 36 \\ 31 & 30 & 32 & 37 \\ 32 & 32 & 35 & 39 \\ 34 & 36 & 37 & 43 \end{bmatrix}, \quad QP_3 = \begin{bmatrix} 31 & 33 & 35 & 37 \\ 33 & 35 & 36 & 40 \\ 34 & 36 & 37 & 40 \\ 36 & 38 & 41 & 44 \end{bmatrix},$$

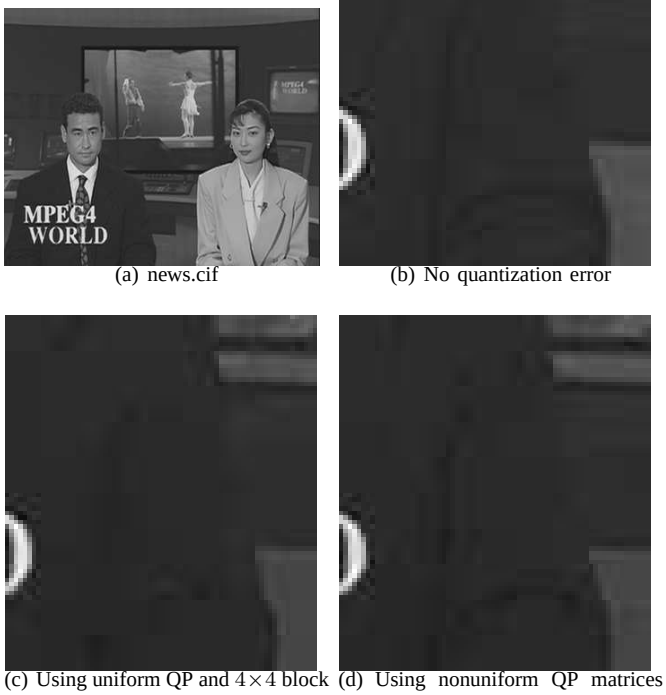
$$QP_4 = \begin{bmatrix} 35 & 34 & 35 & 38 \\ 36 & 36 & 38 & 42 \\ 36 & 37 & 39 & 44 \\ 38 & 40 & 42 & 46 \end{bmatrix}, \quad QP_5 = \begin{bmatrix} 34 & 34 & 35 & 38 \\ 35 & 36 & 37 & 41 \\ 35 & 37 & 38 & 43 \\ 36 & 39 & 41 & 46 \end{bmatrix},$$

$$QP_6 = \begin{bmatrix} 34 & 34 & 35 & 38 \\ 34 & 35 & 36 & 41 \\ 35 & 36 & 38 & 41 \\ 35 & 36 & 38 & 43 \end{bmatrix}, \quad QP_7 = \begin{bmatrix} 34 & 33 & 34 & 37 \\ 33 & 33 & 35 & 39 \\ 34 & 35 & 37 & 41 \\ 36 & 38 & 39 & 44 \end{bmatrix},$$

$$QP_8 = \begin{bmatrix} 33 & 31 & 33 & 35 \\ 32 & 34 & 34 & 38 \\ 32 & 34 & 34 & 39 \\ 34 & 37 & 39 & 44 \end{bmatrix}.$$

To evaluate the perceptual quality improvement using the derived QP matrices, a third subjective experiment is conducted (details in Appendix III). Figure 8 plots five human subjects' opinion scores of the quality of the compressed videos on a scale from 0 to 100 (Figure 10(a)), compared to the uncompressed videos. It shows that for all the tested 4 video sequences at 4 different bit rates, our scheme inceases the

average opinion scores by 10 to 20. Table III lists five human subjects' opinion scores of the quality of the video sequences compressed with 9 non-uniform QP matrices compared to the corresponding video sequences compressed with one uniform QP on a scale from -3 to 3 (Figure 10(b)). It shows that the new scheme considerably increases the perceptual quality of paris.cif, silent.cif, both of which are of low motion, at all tested bit rates and one of the two high motion videos stefan.cif at medium bit rate.



(a) news.cif (b) No quantization error

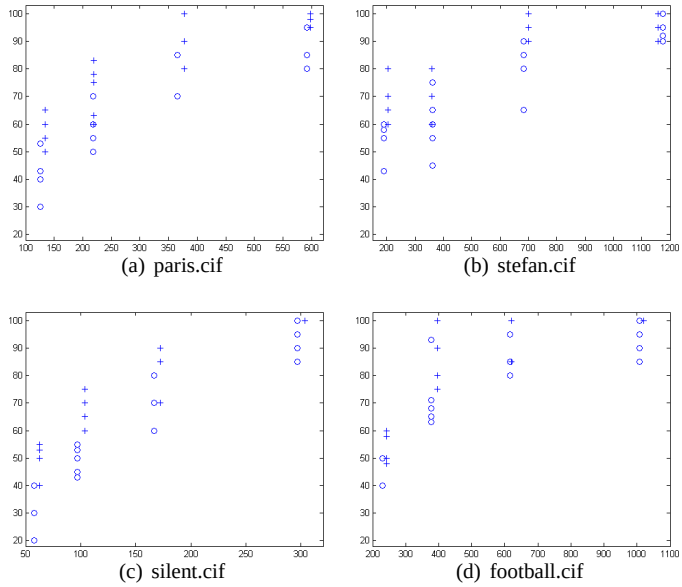
(c) Using uniform QP and 4×4 block intra-modes (d) Using nonuniform QP matrices and 4×4 block intra-modes

(e) Using uniform QP and 16×16 MB intra-modes (f) Using nonuniform QP matrices and 16×16 MB intra-modes

Fig. 6. Comparison of intra-coded frames with average QP = 32

(a) Using uniform QP (b) Using nonuniform QP matrices

Fig. 7. Comparison of inter-coded frames with average QP = 30



(a) paris.cif (b) stefan.cif

(c) silent.cif (d) football.cif

Fig. 8. Five human subjects' opinion scores of the quality of the compressed videos compared to the uncompressed videos. (X axis — bit rate in kbps, Y axis — opinion scores, '+' — using non-uniform QP matrices, 'o' — using one uniform QP)

REFERENCES

[1] ITU-R Recommendation BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," (2002).
[2] ITU-T Recommendation H.264, "Advanced video coding for generic audiovisual services," (2003).
[3] H.264/AVC Software Coordination, Version JM10.1, http://iphome.hhi.de/suehring/tml/, (2005).
[4] T. N. Pappas and R. J. Safranek, "Perceptual criteria for image quality evaluation," *in Handbook of Image & Video Processing (A. Bovik eds.)*, Academic Press, (2000).
[5] T. Wiegand, G. J. Sullivan, G. Bjøntegaard and Ajay Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol 13, pp. 560-576, (Jul. 2003).

APPENDIX I
SUBJECTIVE EXPERIMENT 1

The first subjective experiment measures the QPs of each frequency coefficient, i.e., each entry in $\mathbf{Y}$, for each intra-

TABLE III

FIVE HUMAN SUBJECTS' OPINION SCORES OF THE QUALITY OF THE VIDEO SEQUENCES COMPRESSED WITH 9 NON-UNIFORM QP MATRICES COMPARED TO THE CORRESPONDING VIDEO SEQUENCES COMPRESSED WITH ONE UNIFORM QP ON A SCALE FROM -3 TO 3 (SCALE SHOWN IN FIGURE 10(B)).

	paris				stefan				silent				football			
Bit rate (kbps)	595	370	219	129	1155	692	360	200	299	175	104	60	1020	618	385	236
Subject1	2	1	1	1	0	1	1	0	2	3	2	1	0	0	-1	2
Subject2	1	1	1	1	0	1	0	0	-1	1	1	1	0	0	1	0
Subject3	1	1	1	2	0	2	-1	-1	0	1	3	2	0	1	1	-1
Subject4	1	1	2	2	0	0	2	-1	0	3	1	2	0	0	0	-1
Subject5	1	1	1	2	0	1	2	0	1	2	2	1	0	0	-1	2

prediction mode, at threshold when the quantization error is just noticeable. This experiment involved three subjects and three images from the news, sports and sceneries categories. Only the 4×4 blocks which are intra-coded using one of the 9 intra-prediction modes in the image are chosen at one time to add only one out of the 16 distortion patterns. The magnitude of the distortion is decided by QPs which keep increasing until the distortion becomes perceivable to the subject and then the QPs at threshold are recorded. The three images are all of Common Intermediate Format(CIF) format. The experiment was conducted on a 17 inch monitor with resolution 1152×864 pixels and average background luminance $170cd/mm^2$. The viewing distance is about 6 times the height of the images.

APPENDIX II
SUBJECTIVE EXPERIMENT 2

In the second subjective experiment an artificial video sequence is generated for each error pattern and a certain QP value. Figure 9 shows the first frame in the artificial video sequence for the error pattern $I = 1$ $J = 1$ and $QP = 30$. All 9 intra-prediction blocks at 6 different average luminance levels are generated as the background image, which is the same for every frame in every video sequence. For each video sequence, only one error pattern at one QP value is added, but this error propagates spatially and temporally by inter-frame motion estimation and compensation. All the video sequences were played one by one at three different frame rates $10, 20, 30$ *frames per second(fps)*. The subjects were asked to record the QPs at threshold for each error pattern and each intra-mode, at each luminance level and each frame rate. Subjects and the viewing conditions were kept the same as in subjective experiment 1.

APPENDIX III
SUBJECTIVE EXPERIMENT 3

A third subjective experiment is conducted to evaluate the derived scheme where 4 video sequences of different content are coded at 4 different bit rates, using both the non-uniform QP matrices generated and one uniform QP as used in AVC/H.264. All sequences are generated using modified H.264 reference software JM10.1 [3]. No rate control scheme is used in either case. To set up a fair comparison, the bit rates are tuned by adjusting the average QPs to achieve a fluctuation of less than ±10kbps in each bit rate range tested. This experiment consists of two phases. Stimulus-comparison methods [1] are used in both phases, where two video sequences of the same content were presented to the subjects side by side and played simultaneously. In phase I the compressed video was presented together with the uncompressed video and the subjects were asked to pick a number representing the perceptual quality of the compressed video compared to the uncompressed (perfect) video from the continuous quality scale (Figure 10(a)). In phase II the compressed videos using both schemes were presented together and subjects were asked to pick a number representing the perceptual quality of the compressed video on the right compared to the compressed video on the left from the Adjectival categorical Comparison Scale (Figure 10(b)). 25% of the video sequences appear twice in this experiment to test the consistency of the subjects' decisions. Five naive subjects participated in this experiment and only one of them participated in the first two experiments.

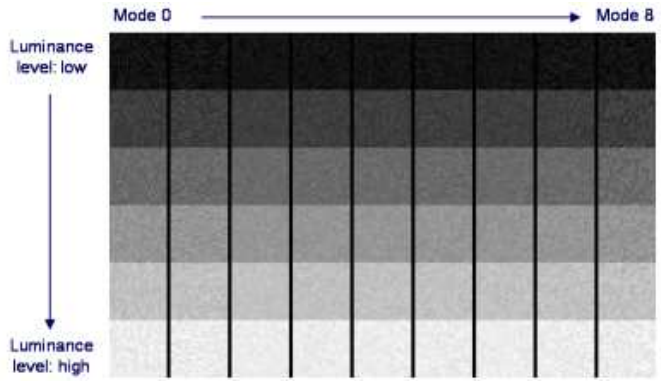
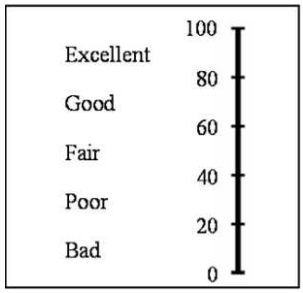


Fig. 9. One frame in subjective experiment 2— error pattern added: I=1 J=1, QP=30. Columns from left to right: intra-prediction modes 0 to 8. Rows from top to bottom: average local luminance from low to high.



(a) continuous quality scale

-3	Much worse
-2	Worse
-1	Slightly worse
0	The same
+1	Slightly better
+2	Better
+3	Much better

(b) adjectival categorical Comparison Scale

Fig. 10. Quality scales used in experiment 3