Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG
(ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6)
10th Meeting: Waikoloa, Hawaii, USA, 8-12 December, 2003

Document: JVT-J024
Filename: JVT-J024.doc

Title: **Method for Objective Video Quality Measurements**
Status: Input Document to JVT
Purpose: Proposal, Information

Author(s) or Contact(s): EePing Ong, Xiaokang Yang, Weisi Lin, Zhongkang Lu, and Susu Yao
Institute For Infocomm Research
21 Heng Mui Keng Terrace, Singapore 119613
Tel: (65)-6874-6856
Email: {epong, xkyang, wslin, zklu, ssyao}@i2r.a-star.edu.sg
Source: Institute For Infocomm Research, Singapore

---

1. ABSTRACT

This proposal encompasses an objective video quality measurement method to automatically measure the perceived quality of a stream of video images. The method is based on a combined measure of distortion-invisibility, block-fidelity, and content richness fidelity.

There is a need for automatic and objective video quality measurement method that is able to emulate the human vision to detect the perceived quality of a video stream. Traditionally, video quality is performed via a subjective test where a large number of human subjects are used to gauge the quality of a video but this process is not only time-consuming but and tedious and expensive to perform.

This proposal describes an objective video quality measurement method for automatically gauging the perceived quality of a stream of video images using a combined feature of distortion-invisibility, block-fidelity, and content richness fidelity. It basically replaces the need of a subjective test in order to be able to gauge the perceived quality of a video stream.

2. INTRODUCTION

Besides on-line and off-line visual quality evaluation, how distortion is gauged also plays a determinative role in shaping most algorithms for image and video manipulations, such as enhancement, reconstruction, data hiding, compression, and joint source/channel coding. Visual quality control within an encoder and distortion assessment for decoded signal is particularly of interests due to the widespread applications of H.26x/MPEG-x compression and coding. Since human eyes are the end receiver of most decoded images and video, it is desirable to develop visual quality metrics that correlate better with human eyes' perception than the conventional pixel-wise error (e.g., MSE, PSNR) measures.

Perceptual models based upon human vision characteristics can be constructed [2, 3]. In the metric proposed in [2], the colour-transformed original and decoded sequences are subjected to blocking and Discrete Cosine Transform (DCT), and the resultant DCT coefficients are then converted to the local contrast, which is defined as the ratio of the AC amplitude to the temporally low-pass filtered DC amplitude. A temporal recursive discrete second-order IIR filtering operation follows to implement the temporal part of the contrast sensitivity function (CSF). The results are then converted to measures of visibility by dividing each coefficient by its respective visual spatial threshold. The difference of two sequences is subjected to a contrast masking operation, and finally the masked difference will be pooled over various dimensions to illustrate perceptual error.

With the same paradigm, Winkler's metric [3] consists of colour conversion, temporal filters, spatial subband filters, contrast control, and pooling for various channels, which are based on the spatio-temporal mechanisms in the human visual system. The difference of original and decoded video is evaluated to give an estimate of visual quality of the decoded signal. The metric's parameters were determined by fitting the metric's output to the experimental data on human eyes.

Prevalent visual coding (e.g., DCT- or wavelet-based) schemes introduce specific types of artefacts such as blockiness, ringing and blurring. The metrics presented in [9, 10] evaluate blocking artefacts as the distortion measure. The metric in [7] measured five types of error (i.e., low-pass filtered error, Weber's law and CSF corrected error, blocking error, correlated error, and high contrast transitional error), and used Principal Component Analysis to decide the compound effect on visual quality.

In [6], switching is suggested between a perceptual model and a blockiness detector depending on the video under test.

In [1], a perceptual distortion metric architecture which is distinctively similar to the one used by Winkler [3] has been proposed. The method consists of opponent colour conversion, perceptual decomposition, masking, followed by pooling. In his method, both the spatial frequency and orientation-selective filtering and temporal filtering are performed in the frequency (spectral) domain. Here, the behaviour of human vision system is modelled by cascading a 3-D filter bank and the non-linear transducer that models masking. The filter bank used in this model is separable in spatial and temporal directions. The model features 17 Gabor spatial filters and 2 temporal filters. A non-linear transducer modelling of masking has been utilized.

In [4], a simplified version of the perceptual model proposed in [1] is applied to blockiness dominant regions.

In [14], a full reference video quality metric has been proposed. Here, for each frame, corresponding local areas are extracted from both the original and test video sequences respectively. For each selected local area, statistical features such as mean and variance are calculated and used to classify the local area into smooth, edge, or texture region. Next a local correlation quality index value is calculated and these local measures are averaged to give a quality value of the entire frame. The frame quality value is adjusted by 2 factors: the blockiness factor and motion factor. The blockiness measurement is evaluated in the power spectrum of the image signal. This blockiness measure is used to adjust the overall quality value only if the frame has relatively high quality index value but severe blockiness. The motion measurement is obtained by a simple block-based motion estimation algorithm. This motion adjustment is applied only if a frame simultaneously satisfies the conditions of low quality index value, high blurriness and low blockiness. Finally, all frame quality index values are averaged to a single overall quality value of the test sequence.

3. DESCRIPTION OF PROPOSAL

3.1. SUMMARY OF PROPOSAL

In this proposal, a method for measuring the objective video quality of video streams has been proposed. The proposed method consists of computation of a video quality rating using a video quality model made up of the following components: (1) content richness fidelity (F_{RF}), (2) block-fidelity (F_{BF}), and (3) distortion-invisibility (D).

The content richness fidelity feature measures the fidelity of the richness of test video's content with respect to the original reference (undistorted) video. This content richness fidelity feature gives higher values for test video which has better fidelity in content richness with respect to the original reference (undistorted) video.

The block-fidelity feature measures the amount of distortion at block-boundaries in the test video when compared with respect to the original reference (undistorted) video. The block-fidelity feature should give lower values when distortion at block-boundaries in the test video is more severe and higher values when distortion is very low or does not exist in the test video (when compared to the original reference (undistorted) video).

The distortion-invisibility feature measures the average amount of distortion that may be visible at each pixel with respect to a visibility threshold and gives higher values for lower visible distortions and lower values for higher visible distortions.

A combined measure is proposed and then demonstrated to measure visual quality for video with the video-quality features.

3.2. DETAILED DESCRIPTION OF THE PROPOSAL

3.2.1. Content Richness Fidelity

The content richness fidelity feature measures the fidelity of the richness of test video's content with respect to the original reference (undistorted) video. This content richness fidelity feature gives higher values for test video which has better fidelity in content richness with respect to the original reference (undistorted) video. This feature closely correlates with human perceptual response which tends to assign better subjective ratings to more lively and more colourful images and lower subjective ratings to dull and unlively images.

The image content richness fidelity feature for each individual frame of time interval *t* of the video can be defined as:

$$F_{RF}(t) = e^{(0.25)R_d(t)/R_o(t)}$$

$$R(t) = -\sum_{p(i)\neq 0} p(i)\log_e\left(p(i)\right)$$

where the subscript *o* refers to the original video sequence, *d* refers to the test video sequence, $t \in [1, N]$, *N* is the total number of image-frames in the video sequence, and:

$$p(i) = \frac{N(i)}{\sum_{\forall i} N(i)}$$

Here, *i* is a particular colour (either the luminance or the chrominance) value, $i \in [0,255]$, *N*(*i*) is the number of occurrence of *i* in the image frame, and *p*(*i*) is the probability or relative frequency of *i* appearing in the image frame.

3.2.2. Block-Fidelity

The block-fidelity feature measures the amount of distortion at block-boundaries in the test video when compared with respect to the original reference (undistorted) video. The block-fidelity feature should give lower values when distortion at block-boundaries in the test video is more severe and higher values when distortion is very low or does not exist in the test video (when compared to the original reference (undistorted) video).

The blocking effects, and its propagation through reconstructed video sequences, is one of the significant coding artefacts that often occur in video compression. The blocking effect is also a source of a number of other types of reconstruction artifacts, such as stationary area granular noise.

The block-fidelity measure for each individual frame of the video is defined as follows:

$$F_{BF}(t) = e^{(0.25)\left\{\left(B_d^h(t)+B_d^v(t)\right)-\left(B_o^h(t)+B_o^v(t)\right)\right\}/\left(B_o^h(t)+B_o^v(t)\right)}$$

where the subscript *o* refers to the original video sequence, *d* refers to the test video sequence, and:

$$B^h(t) = \frac{1}{H\left(\lfloor W/4 \rfloor - 1\right)} \sum_{y=1}^{H} \sum_{x=1}^{\lfloor W/4 \rfloor - 1} \left| d^h(4x, y, t) \right|$$

$$d^h(x, y, t) = I(x+1, y, t) - I(x, y, t)$$

$I(x, y, t)$ denotes the colour value of the input image frame *I* at pixel location (*x*,*y*) and time interval *t*, *H* is the height of the image, *W* is the width of the image, $x \in [1, W]$ and $y \in [1, H]$.

Similarly,

$$B^v(t) = \frac{1}{W\left(\lfloor H/4 \rfloor - 1\right)} \sum_{y=1}^{\lfloor H/4 \rfloor - 1} \sum_{x=1}^{W} \left| d^v(x, 4y, t) \right|$$

$$d^v(x, y, t) = I(x, y+1, t) - I(x, y, t)$$

3.2.3. Distortion-Invisibility

The distortion-invisibility feature measures the average amount of distortion that may be visible at each pixel with respect to a visibility threshold and gives higher values for lower visible distortions and lower values for higher visible distortions. The distortion-invisibility measure, *D(t)*, for each frame of the video is given by (see Figure 1):

$$D(t) = \left\{ 1 \Big/ \frac{1}{WH} \sum_{x=1}^{W} \sum_{y=1}^{H} \left[\gamma_1 + \frac{\hat{d}(x,y,t)}{\gamma_2 + T(x,y,t)} \right] \right\}$$

$T(x,y,t)$ is the visibility threshold at a particular pixel location (*x*,*y*) and time interval *t*, *W* and *H* are width and height of the video frame respectively, $1 \le x \le W$, $1 \le y \le H$, γ_1 is included for introducing linearity into the equation, and γ_2 prevents division by zero in the equation.

Also:

$$\hat{d}(x,y,t) = \begin{cases} d(x,y,t) & \text{if } d(x,y,t) \ge T(x,y,t) \\ 0 & \text{otherwise} \end{cases}$$

where d(*x*,*y*,*t*) is the difference between a frame in the test video I_d and the reference video I_o at the same pixel location (*x*,*y*) and time *t* and is defined as:

$$d(x,y,t) = \left| I_o(x,y,t) - I_d(x,y,t) \right|$$

Here, $I_o(x,y,t)$ denotes a pixel at location (*x*,*y*) at frame *t* of the original video sequence while $I_d(x,y,t)$ denotes a pixel at location (*x*,*y*) at frame *t* of the test video sequence.

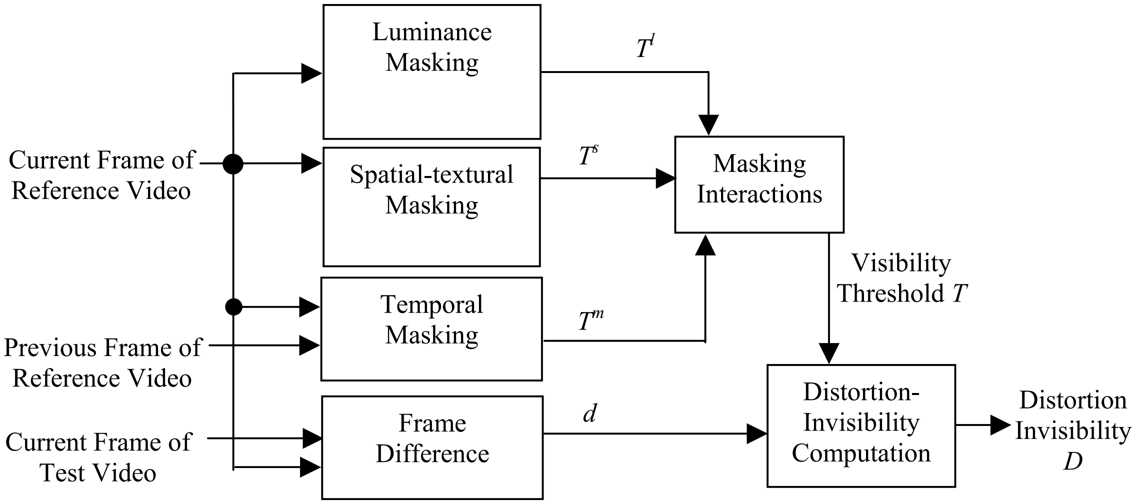

Figure 1: Process for Computing Distortion-Invisibility

The visibility threshold *T* is given by:

$$T(x,y,t) = \left(T^l(x,y,t) + T^s(x,y,t) - C^{ls}.\min\{T^l(x,y,t), T^s(x,y,t)\} \right) T^m(x,y,t)$$

The visibility threshold at a particular pixel located at position (*x*,*y*) and time *t*, denoted *T*(*x*,*y*,*t*), provides an indication of the maximum allowable distortions at a particular pixel in the image frame which will still not be visible to human eyes. Here, $T^l(x,y,t)$, $T^s(x,y,t)$ and $T^m(x,y,t)$ can be regarded as effects due to colour masking, spatial-textural masking, and temporal masking respectively at a particular pixel located at position (*x*,*y*) in the image frame at time interval *t* in the video sequence, while C^{ls} is a constant. The three masking effects interact in a manner as described by the above equation in order to provide a visibility threshold required for this objective video quality measurement method.

Masking is a very important visual phenomenon and explains why similar artifacts are disturbing in certain regions (such as flat regions) of an image frame while they are hardly noticeable in other regions (such as textured regions). In addition, similar artifacts in certain regions of different video sequences displaying different temporal characteristics will appear as disturbing in a particular video sequence but not in another. Here, these visual phenomenon have been modelled using colour masking, spatial-textural masking, and temporal masking which will be further described in the below section.

The temporal masking T^m attempts to emulates the effect of human vision's characteristic of being able to accept higher video-frame distortion due to larger temporal changes and can be derived as follow:

$$T^m(x,y,t) = e^{f_s \cdot f_r} \begin{cases} T_2^m & \text{if } \left|d_f(x,y,t)\right| \le T_3^m \\ T_1^m \left(z_2^{\left(1-\left(\left(L_m - d_f(x,y,t)\right)/\left(L_m - T_3^m\right)\right)\right)} - 1 \right) + T_2^m & \text{if } d_f(x,y,t) < -T_3^m \\ T_o^m \left(z_1^{\left(1-\left(\left(L_m - d_f(x,y,t)\right)/\left(L_m - T_3^m\right)\right)\right)} - 1 \right) + T_2^m & \text{Otherwise} \end{cases}$$

where $d_f(x,y,t)$ is the inter-frame difference at a particular pixel location (x,y) in time t between a current frame $I_o(x,y,t)$ and a previous coded frame $I_o(x,y,t\text{-}30/fr)$ (assuming that frames that has been coded at below full frame rate has been repeated in this video sequence) and is mathematically expressed as:

$$d_f(x,y,t) = I_o(x,y,t) - I_o(x,y,t - 30/f_r)$$

Here, f_r is the frame rate at which the video has been compressed, f_s is a scaling factor, while L_m, T_o^m, T_1^m, T_2^m, T_3^m, z_1, and z_2 are constants used to determine the exact profile of the temporal masking.

The luminance masking attempts to emulates the effect of human vision's characteristic of being able to accept higher video-frame distortion when the background luminance is above or below a certain mid-level threshold and can be derived as follow:

$$T^l(x,y,t) = \begin{cases} T_1^l \left(v_2^{\left(1-\left(\left(\lfloor L_l/2 \rfloor + b(x,y,t)\right)/\left(L_l - r\right)\right)\right)} - 1 \right) + T_2^l & \text{if } b(x,y,t) \le (L_l - r) \\ T_o^l \left(v_1^{\left(1-\left(\left(\lfloor 3L_l/2 \rfloor - b(x,y,t)\right)/\left(L_l - r\right)\right)\right)} - 1 \right) + T_2^l & \text{if } b(x,y,t) > (L_l + r) \\ T_2^l & \text{Otherwise} \end{cases}$$

Here, T_o^l, T_1^l and T_2^l, L_l, r, v_1, and v_2 are constants used to determine the exact profile of the luminance masking.

The spatial-textural masking attempts to emulates the effect of human vision's characteristic of being able to accept higher video-frame distortion when the particular point has richer texture or spatial profile and can be derived as follow:

$$T^s(x,y,t) = \left(m(x,y,t)b(x,y,t)\alpha_1 + m(x,y,t)\alpha_2 + b(x,y,t)\alpha_3 + \alpha_4 \right) W(x,y,t)$$

Here, α_1, α_2, α_3, and α_4 are constants used to determine the exact profile of the spatial-textural masking.

In the spatial-textural masking, $m(x,y,t)$ is the average of the average luminance $g_k(x,y)$ in four different orientations and it attempts to capture the textural characteristic of the small local region centred on pixel (x,y,t) and can be mathematically written as:

$$m(x,y,t) = \frac{1}{4} \sum_{k==1}^{4} \left| g_k(x,y,t) \right|$$

Also, $g_k(x,y,t)$ is the average luminance around a pixel located at position (x,y) of a frame in the original reference video sequence at time interval t and is computed by convolving a 7x7 mask, G_k, with this particular frame in the original reference video sequence. Mathematically, $g_k(x,y,t)$ can be expressed as:

$$g_k(x,y,t) = \frac{1}{19}\sum_{m=-3}^{3}\sum_{n=-3}^{3} f(x+m, y+n, t).G_k(m+4, n+4, t)$$

The four 7x7 masks, G_k, for $k=\{1,2,3,4\}$, are four differently oriented gradient masks used to capture the strength of the gradients around a pixel located at position (x,y,t).

Here, $b(x,y,t)$ is the average background luminance around a pixel located at position (x,y) of a frame in the original reference video sequence at time interval t and is computed by convolving a 7x7 mask, B, with this particular frame in the original reference video sequence. Mathematically, $b(x,y,t)$ can be expressed as:

$$b(x,y,t) = \frac{1}{40}\sum_{m=-3}^{3}\sum_{n=-3}^{3} f(x+m, y+n, t).B(m+4, n+4, t)$$

The 7x7 mask, B, acts like a low-pass filter when operated on a pixel located at position (x,y,t).

In addition, $W(x,y,t)$ is an edge-adaptive weight of the pixel at location (x,y) of a frame in the original reference video sequence at time interval t, and it attempts to reduce the spatial-textural masking at edge locations because artifacts that are found on essential edge locations tend to reduce the visual quality of the image frame. The corresponding edge-adaptive weight matrix W, obtained by convolving $\hat{E}$ with a 7x7 low-pass filter g, is given by:

$$W = \hat{E} * g$$

$$\hat{E} = 1 - (0.9E)$$

where $*$ is a convolution operator, E is the edge matrix of the original image frame computed with any edge detection technique and contains values of 1 and 0 for edge and non-edge pixels respectively.

3.2.4. Video Quality Measurement Method

The overall objective video quality rating for a video sequence, Q, is given by averaging the objective video quality rating for each frame $q(t)$ and can be expressed as (see Figure 2):

$$Q = \sum_{t=i.(30/f_r)}^{N} \left[q(t)\right] \Big/ N_t, \quad i = 1,2,...$$

where N is the total number of frames in the original video sequence, N_t is the total number of coded video sequences (which will be different if the video is coded at below the full frame rate) and is given by:

$$N_t = N/(30/f_r),$$

and $q(t)$ is the objective video quality rating for each frame, defined as follows:

$$q(t) = D(t).F_{BI}(t).F_{RF}(t)$$

Figure 2 shows a block diagram of the proposed video quality measurement system.

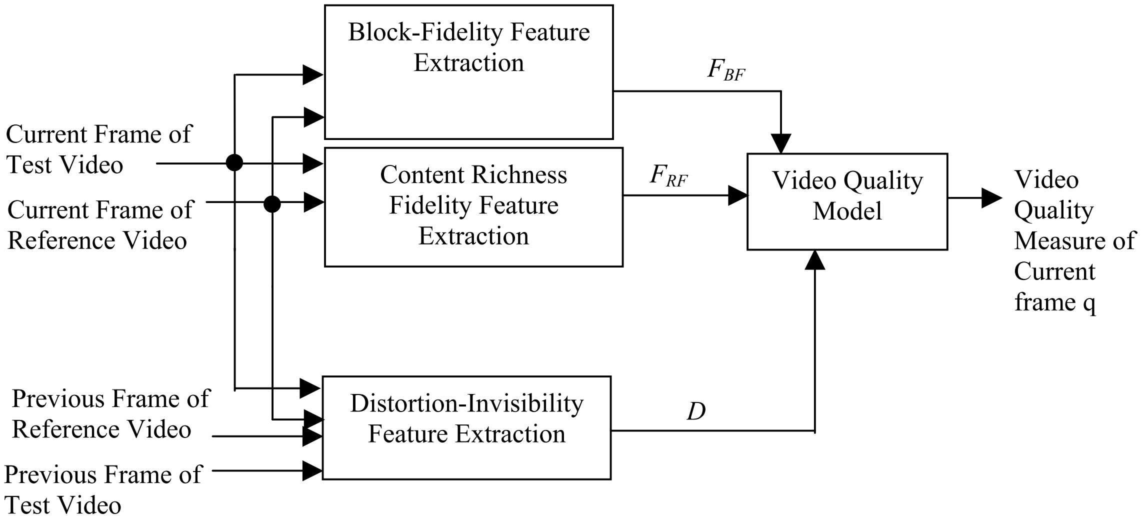

Figure 2: Block diagram of video quality measurement system

RESULTS

The proposed video quality measurement method has been tested on a test data set consisting of 90 video sequences that have been obtained by subjecting 12 original video sequences to various compression bitrates and frame rates. The performance of the proposed metric is measured with respect to the subjective ratings of test video sequences that have been obtained by subjective video quality experiment.

The test video sequences are generated by subjecting 12 different original undistorted CIF and QCIF video sequences (“Container”, “Coast Guard”, “Japan League”, “Foreman”, “News”, and “Tempete”) to H.26L video compression with different bit rates (from 24 kbps to 384 kbps) and frame rates (from 7.5Hz to 30Hz) [12]. The bit rates under test are much lower than those used in [13] after the image size factor is offset. Each of the video sequence consists of 250 frames.

The subjective video quality tests of the test video sequences have been carried as the tests conducted for the evaluation of video sequences [12]. The laboratory has been set up according to the indication contained in the ITU-R Recommendation 500-11. The decoded sequences with frame rate lower than 30 fps are displayed with repeated frames on the 30 Hz display device.

The Double-Stimulus Impairment Scale variant II (DSIS-II) subjective test method has been used in the subjective test experiment. The DSIS-II method presents two sequences to the assessors, first the original, then the processed, and then the sequences are repeated in sequel. The assessors are required to vote using a five-grade impairment scale: Excellent (5), Good (4), Fair (3), Poor (2), and Bad (1). Subjects vote by ticking the appropriate boxes on pre-designed voting forms. This subjective test has been performed by 20 subjects of which half of them are male and the other half are female.

More specifically, in the DSIS-II subjective test procedure, the presentation of the test material is announced by a message on the screen standing for one second; the letter A is displayed immediately before the reference is presented for the first time; the letter B is displayed immediately before the processed sequence is presented for the first time; the letter A* is displayed immediately before the reference is presented for the second time; the letter B* is displayed immediately before the processed sequence is presented for the second time; immediately after the processed sequence is presented for the second time the message “Vote N” is presented on the screen for 4 seconds. The votes have been collected by means of scoring sheets.

Performance is measured by comparing the metric output q_o (= Q as used above) with the subjective rating of subjective tests between the original and the distorted sequences. To facilitate monotonicity of prediction and a common analysis space of comparison, q_o is fitted via a 4-parameter cubic polynomial [11] to the corresponding subjective rating as:

$$q = a_o + a_1(q_o) + a_2(q_o{}^2) + a_3(q_o{}^3)$$

Two performance measures have been used for comparison here (as in [11]): (1) Pearson correlation coefficient (where a value of 1 indicates perfect correlation between tested data and subjective ratings and a value of 0 indicates no correlation), and (2) Spearman rank-order correlation coefficient (where a value of 1 indicates perfect match between the tested data and the subjective ratings and a value of 0 indicates no correlation).

Pearson correlation, r_p, which measures the prediction accuracy, i.e. the ability of the outputs of the measurement system to predict subjective ratings, is defined as:

$$r_p = \frac{\sum_k (q_k - \overline{q})(S_k - \overline{S})}{\sqrt{\sum_k (q_k - \overline{q})^2} \sqrt{\sum_k (S_k - \overline{S})^2}}$$

where q is the output of the no-reference image-quality measurement system, S is the value of the subjective rating, $\overline{q}$ and $\overline{S}$ are the means of q and S, and k is the index for the image under test. A value of 1 for Pearson correlation indicates perfect correlation between tested data and subjective ratings while a value of 0 indicates no correlation.

Spearman rank-order correlation, r_s, which measures the prediction monotonicity, i.e. whether the increases/decreases in one variable are associated with increases/decreases in the other variable independent of the magnitude of the increase/decrease, is defined as:

$$r_s = \frac{\sum_k (\chi_k - \overline{\chi})(\gamma_k - \overline{\gamma})}{\sqrt{\sum_k (\chi_k - \overline{\chi})^2} \sqrt{\sum_k (\gamma_k - \overline{\gamma})^2}}$$

where χ_k is the rank of q_k and γ_k is the rank of S_k in the ordered data series, and $\overline{\chi}$ and $\overline{\gamma}$ are the respective midranks. In the ideal match between a metric's outputs and subjective ratings, $r_p = 1$ and $r_s = 1$.

Table 1 shows the results of the proposed method with respect to PSNR. The upper bound and lower bound of Pearson correlation were obtained with a confidence interval of 95%. It can be seen that the proposed video quality measurement method (with a Pearson correlation of 0.897 and Spearman correlation of 0.902) performs much better than PSNR (which has a Pearson correlation of 0.701 and Spearman correlation of 0.676).

Figure 3 shows the scatterplot of subjective ratings versus the PSNR values, while Figure 4 shows the scatterplot of subjective ratings versus the video quality ratings estimated using our proposed method. In these two figures, the middle solid line portrays the logistic fit using the above-mentioned 4-parameter cubic polynomial, while the upper dotted curve and the lower dotted curve portray the upper bound and lower bound respectively obtained with a confidence interval of 95%.

	Pearson-Correlation			Spearman-Correlation
		Upper Bound	Lower Bound	
PSNR	0.701	0.793	0.578	0.676
Proposed	0.897	0.931	0.848	0.902

Table 1: Results of proposed video quality measurement method and that given by PSNR

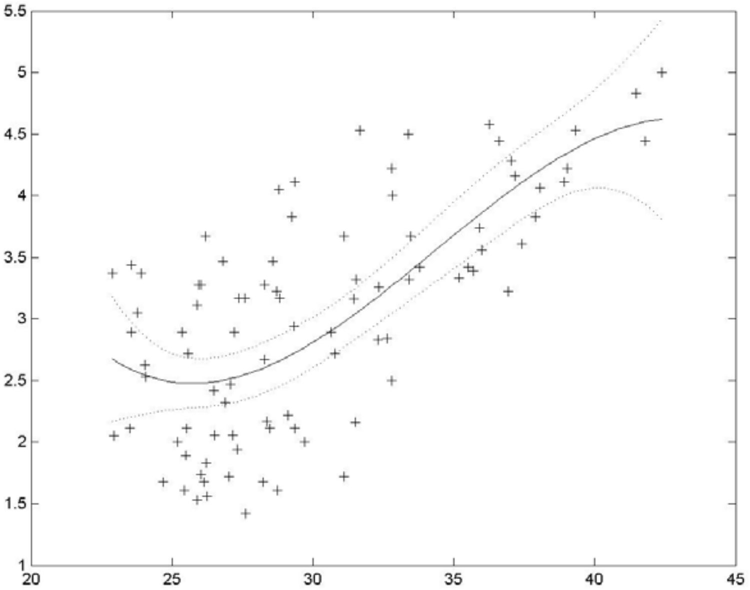

Figure 3: Scatterplot of subjective ratings vs PSNRs

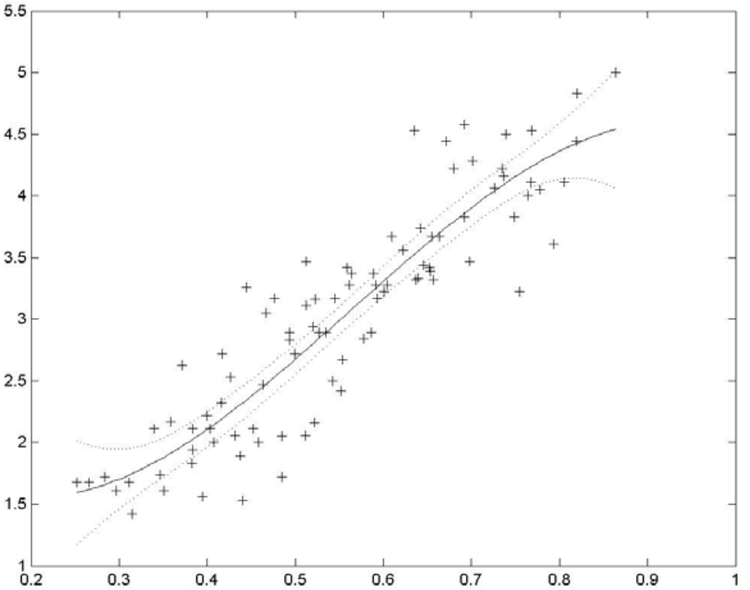

Figure 4: Scatterplot of subjective ratings vs metric outputs of proposed method

4. REFERENCES

[1] Lindh, P., and Lambrecht, C., "Efficient spatio-temporal decomposition for perceptual processing of video sequences", *International Conference on Image Processing*, 1996.

[2] Watson, A.B., Hu, J., McGowan III, J.F., "DVQ: A digital video quality metric based on human vision", *Journal of Electronic Imaging*, Vol. 10, No. 1, 2001, pp. 20-29.

[3] Winkler, S., "A perceptual distortion metric for digital color video", *SPIE Proc. Human Vision and Elect. Imaging IV,* Vol. 3644, pp. 175 – 184, 1999.

[4] Yu, Z, Wu, H.R., Winkler, S. and Chen, T., "Vision-model-based impairment metric to evaluate blocking artifacts in digital video", Proc. IEEE, Vol. 90(1), pp. 154-169, 2002.

[5] Teo, P.T. and Heeger, D.J., "Perceptual image distortion", *Proc. 1^{st} IEEE Int'l Conf. Image Processing*, vol.2, pp.982-986, 1994.

[6] Tan, K.T. and Ghanbari, M, "A multi-metric objective picture-quality measurement model for MPEG video", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 10, No. 7, Oct. 2000, pp. 1208-1213.

[7] Miyahara, M., Kotani, K. and Algazi, V.R., "Objective picture quality scale (PQS) for image coding", *IEEE Transactions on Communications*, Vol. 46, No. 9, Sep. 1998, pp. 1215-125.

[8] Lubin, J., Peterson, H.A., Spence, C.D., and Aries, A.d., "Method and apparatus for training a neural network to learn and use fidelity metric as a control mechanism", US patent 6,075,884, 2000.

[9] Wu, H.R. and Yuen, M., "A generalize block-edge impairment metric for video coding", IEEE Signal Processing Letters, Vol. 4, No. 11, November 1997, pp.317-320.

[10] Karunasekera, S. A. and Kingsbury, N. G. "A distortion measure for blocking artifacts in image based on human visual sensitivity", *IEEE Trans on Image Processing*, Vol. 4, No. 6, June 1995, pp.713-724.

[11] Evaluation of new methods for objective testing of video quality: objective test plan, ITU-T/COM-T/COM12/C, September 1998.

[12] Baroncini, V., ISO/IEC JTC 1/SC29/WG11, 4240, Sidney, July 2001.

[13] VQEG (Video Quality Expert Group), "Final Report from the Video Quality Expert Group on the validation of Objective Models of Video Quality Assessment", *www.vqeg.org*, March 2000.

[14] Lu, L., Wang, Z., and Bovik, A.C., "Full-reference video quality assessment considering structural distortion and no-reference quality evaluation of MPEG video", *IEEE International Conference on Multimedia and Expo*, August 2002, pp. 61-64.

[15] Hooke R., Jeeves T. A., "Direct Search" solution of numerical and statistical problems, *Journal of the associate computing machinery*, Vol. 8, 1961, pp. 212-229.

[16] Ong, E.P., Lin, W., Lu, Z., Yang, X, Yao, S., Lin, X., and Moschetti, F., "Quality evaluation of MPEG-4 and H.26L coded video for mobile multimedia communications", *International Symposium on Signal Processing and its Applications*, Paris, France, 1-4 July 2003.

(Append for Proposal Documents)

JVT Patent Disclosure Form

International Telecommunication Union
Telecommunication Standardization Sector

International Organization
for Standardization

International Electrotechnical Commission

Joint Video Coding Experts Group - *Patent Disclosure Form*

(Typically one per contribution and one per Standard | Recommendation)

Please send to:
JVT Rapporteur Gary Sullivan, Microsoft Corp., One Microsoft Way, Bldg. 9, Redmond WA 98052-6399, USA
Email (preferred): Gary.Sullivan@itu.int Fax: +1 425 706 7329 (+1 425 70MSFAX)

This form provides the ITU-T | ISO/IEC Joint Video Coding Experts Group (JVT) with information about the patent status of techniques used in or proposed for incorporation in a Recommendation | Standard. JVT requires that all technical contributions be accompanied with this form. *Anyone* with knowledge of any patent affecting the use of JVT work, of their own or of any other entity ("third parties"), is strongly encouraged to submit this form as well.

This information will be maintained in a "living list" by JVT during the progress of their work, on a best effort basis. If a given technical proposal is not incorporated in a Recommendation | Standard, the relevant patent information will be removed from the "living list". The intent is that the JVT experts should know in advance of any patent issues with particular proposals or techniques, so that these may be addressed well before final approval.

This is not a binding legal document; it is provided to JVT for information only, on a best effort, good faith basis. Please submit corrected or updated forms if your knowledge or situation changes.

This form is *not* a substitute for the *ITU ISO IEC Patent Statement and Licensing Declaration*, which should be submitted by Patent Holders to the ITU TSB Director and ISO Secretary General before final approval.

Submitting Organization or Person:

Organization name ____________________

Mailing address ____________________
Country ____________________
Contact person ____________________
Telephone ____________________
Fax ____________________
Email ____________________
Place and date of submission ____________________

Relevant Recommendation | Standard and, if applicable, Contribution:

Name (ex: "JVT") ____________________
Title ____________________
Contribution number ____________________

(Form continues on next page)

Disclosure information – Submitting Organization/Person (choose one box)

☐ 2.0 The submitter is not aware of having any granted, pending, or planned patents associated with the technical content of the Recommendation | Standard or Contribution.

or,

The submitter (Patent Holder) has granted, pending, or planned patents associated with the technical content of the Recommendation | Standard or Contribution. In which case,

☐ 2.1 The Patent Holder is prepared to grant – on the basis of reciprocity for the above Recommendation | Standard – a free license to an unrestricted number of applicants on a worldwide, non-discriminatory basis to manufacture, use and/or sell implementations of the above Recommendation | Standard.

☐ 2.2 The Patent Holder is prepared to grant – on the basis of reciprocity for the above Recommendation | Standard – a license to an unrestricted number of applicants on a worldwide, non-discriminatory basis and on reasonable terms and conditions to manufacture, use and/ or sell implementations of the above Recommendation | Standard.

Such negotiations are left to the parties concerned and are performed outside the ITU | ISO/IEC.

☐ 2.2.1 The same as box 2.2 above, but in addition the Patent Holder is prepared to grant a "royalty-free" license to anyone on condition that all other patent holders do the same.

☐ 2.3 The Patent Holder is unwilling to grant licenses according to the provisions of either 2.1, 2.2, or 2.2.1 above. In this case, the following information must be provided as part of this declaration:

- patent registration/application number;
- an indication of which portions of the Recommendation | Standard are affected.
- a description of the patent claims covering the Recommendation | Standard;

*In the case of any box **other than 2.0** above, please provide the following:*

Patent number(s)/status ____

Inventor(s)/Assignee(s) ____

Relevance to JVT ____

Any other remarks: ____

(please provide attachments if more space is needed)

(form continues on next page)

Third party patent information – fill in based on your best knowledge of relevant patents granted, pending, or planned by other people or by organizations other than your own.

Disclosure information – Third Party Patents (choose one box)

☐ 3.1 The submitter is not aware of any granted, pending, or planned patents *held by third parties* associated with the technical content of the Recommendation | Standard or Contribution.

☐ 3.2 The submitter believes third parties may have granted, pending, or planned patents associated with the technical content of the Recommendation | Standard or Contribution.

For box 3.2, please provide as much information as is known (provide attachments if more space needed) - JVT will attempt to contact third parties to obtain more information:

3rd party name(s) ____________________

Mailing address ____________________
Country ____________________
Contact person ____________________
Telephone ____________________
Fax ____________________
Email ____________________
Patent number/status ____________________
Inventor/Assignee ____________________
Relevance to JVT ____________________

Any other comments or remarks: