KDDI R&D LABS

INTRODUCTION TO KDDI R&D LABS' OBJECTIVE VIDEO QUALITY ASSESSMENT TECHNOLOGY

May 2009

Copyright © 2009 by KDDI R&D Laboratories Inc.

ABSTRACT

This paper shows a brief introduction to the objective perceived video quality measurement technology based on full reference framework, which is developed by KDDI R&D Laboratories Inc.

OUTLINE OF THE OBJECTIVE MODEL

Figure 1 shows a schematic diagram of the proposed objective model. The objective model consists of three stages: 1) registration (spatio-temporal pixel shift compensation between the reference and coded pictures, 2) image feature extraction, and 3) integration of image features and estimation of overall quality.
In the first stage, the model compares the reference and coded pictures to detect spatio-temporal shifts, gain and offset characteristics, and the existence of cropped regions. After the registration, the reference and coded pictures are correctly aligned in both spatial and temporal coordinates and the pixel values are calibrated to minimize the squared difference between the reference and coded pictures. In this paper, we assume that the reference and coded sequence are well aligned and thus calibration is not required. This is because we should focus on the performance of the objective quality metrics itself, independent of the performance of the calibration method. The second stage calculates seven image features considering the human visual system. The image features represent the degree of perceptual impairment in the coded pictures. The overall quality is defined as the weighted sum of these image features in the integration stage.

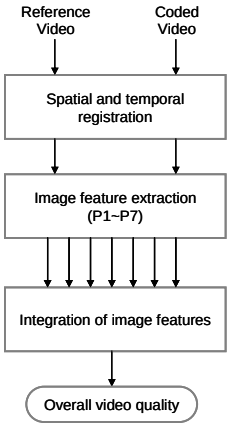

Figure 1 Schematic diagram of the objective model

Definitions of Image features

P_1: Blockiness

One of the major visible artifacts caused by compression coding is block distortion. Many objective quality metrics utilize blockiness as one of the major indices that approximates subjective quality since Karunasekera[1] et al. proved a strong correlation between blockiness and subjective quality. The proposed method first calculates $dDC(f)$, which is an average of the DC difference between the current 8x8 pixel block and four adjacent blocks (left, top, top left, and top right) in a frame as shown in Figure 2. Then, the

Copyright © 2009 by KDDI R&D Laboratories Inc.

maximum and minimum difference of $dDC(f)$ between the reference and coded the picture during the entire sequence is calculated as follows:

$$d_{\max} = \max_{f \in sequence} \{dDC_{\mathrm{Ref}}(f) - dDC_{Cod}(f)\}$$

$$d_{\min} = \min_{f \in sequence} \{dDC_{\mathrm{Ref}}(f) - dDC_{Cod}(f)\}$$

where $dDC_{\mathrm{Ref}}(f)$ and $dDC_{\mathrm{Cod}}(f)$ denote the average DC differences in frame number f in the reference and coded pictures, respectively. Image feature P1 is defined as the difference between $d_{\max}$ and $d_{\min}$.

$$P_1 = 20\log_{10}\sqrt{\frac{255^2}{d_{\max} - d_{\min}}}$$

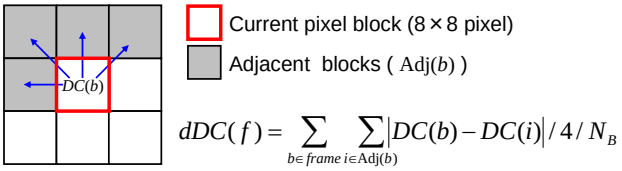

Figure 2 Derivation of image feature P_1 (Blockiness)

P_2: Deviation of frame-by-frame MSE in the sequence

The average PSNR and MSE (Mean Square Error) in the sequence are used as one of the general indices of coding quality, and it correlates with subjective quality to a certain extent. However, the correlation between PSNR/MSE and subjective quality may be weak when the deviation of PSNR/MSE in the sequence is large. The proposed method therefore defines image feature P2 as the deviation of maximum and minimum MSE from the average MSE in the sequence instead of applying PSNR to the quality index as it is. When $x_S(i, j, f)$ and $x_P(i, j, f)$ denote the luminance value at coordinate (i, j) in f-th frame of the reference and coded picture, P2 is expressed as follows:

$$e_2 = \sum_{i,j} \frac{\{x_S(f,i,j) - x_P(f,i,j)\}^2}{N_P}$$

$$P_2 = \log_{10}\left(\frac{e_{\max} - e_{\mathrm{ave}}}{e_{\mathrm{ave}} - e_{\min}}\right)$$

where $e_{\max}$, $e_{\min}$ and e_{ave} denotes maximum, minimum and average of $e_2(f)$ in the sequence, respectively.

***P_3*: Temporal local degradation of PSNR**

When applying low bitrate video coding, temporal local degradation of PSNR may occur according to the

Copyright © 2009 by KDDI R&D Laboratories Inc.

situation such as key-frame insertion, scene change and occurrence of rapid movement and this is expected to cause serious degradation of the subjective quality. Image feature P3 is utilized to detect such types of impairment. First, we define dPSNR(f) as the degree of the wedge-shaped temporal degradation of PSNR as shown in Figure 3, and the maximum value of dPSNR(f) in the sequence is defined as image feature P3.

$$P_3 = \max\{ f \in sequence \mid dPSNR(f)\}$$

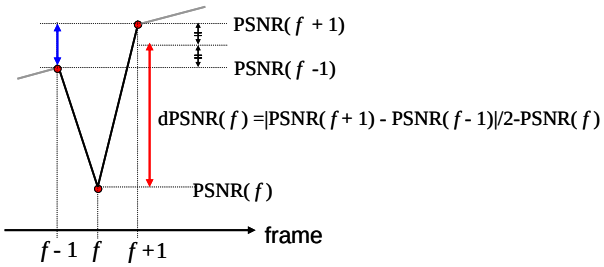

Figure 3 Definition of dPSNR(*f*)

P_4: Average MSE of the blocks having high variance

When we encode video sequences at lower bitrates, coarse quantization tends to be applied to realize higher compression ratio and thus it is difficult to sustain texture details that was originally watched in the reference pictures. To reduce artifacts caused by such coarse quantization, post filtering is often applied to the decoded pictures. However, Brotherton et al pointed out that such post processing sometimes makes the decoded pictures excessively blurred and causes serious degradation of subjective quality [2]. The proposed method therefore defines the average MSE of the blocks in which variance is higher than the given threshold as an index of reproduction of the textures. when B_{hivar} denote a set of 8x8 pixel blocks that have higher variance than given threshold, image feature P4 is expressed as follows:

$$e_4(f) = \sum_{i,j \in B_{hi\,var}} \frac{\{x_S(f,i,j) - x_P(f,i,j)\}^2}{N_B}$$

$$P_4 = \left\{ \sum_f 10\log_{10} \frac{255^2}{e_4(f)} \right\} \Big/ N_H$$

P_5: Average power of intra-frame differences

As it is defined in the ITU-T Recommendation P.910, Temporal Information (TI) is one of the major features which represent the characteristics of the motion in the picture. Although TI doesn't represent spatio-temporal distortions directly, this is a good measure to show the characteristics of the distortions caused by the motion compensation since the significance of motion well correlates to the coding efficiency. Image feature P5 is therefore defined as the average power of the inter-frame differences in the sequence.

$$e_5(f) = \sum_{i,j} \frac{\{x_P(f,i,j) - x_P(f-1,i,j)\}^2}{N_P}$$

Copyright © 2009 by KDDI R&D Laboratories Inc.

$$P_5 = \left\{ \sum_f 10 \log_{10} \frac{255^2}{e_5(f)} \right\} \Big/ N_F$$

P_6: Degradation of lower frequency components

To examine the quality of reproduction in lower frequency components in the decoded pictures, average of squared difference between the reference and coded pictures after applying low-pass filter is employed as the sixth image feature.

When $x_{SL}(f, i, j)$ and $x_{PL}(f, i, j)$ denote the luminance value of the reference and coded picture after applying the low-pass filter, the image feature P6 is expressed as follows:

$$e_6(f) = \sum_{i,j} \frac{\{x_{SL}(f,i,j) - x_{PL}(f,i,j)\}^2}{N_P}$$

$$P_6 = \sum_f \left\{ 10 \log_{10} \frac{255^2}{e_6(f)} \right\} \Big/ N_F$$

Table 1 shows the definition of the filter coefficients.

P_7: Degradation of higher frequency components

To examine the quality of reproduction in edge components in the decoded pictures, image feature P7 is defined as average of squared difference between the original and decoded pictures after applying Laplacian filter that extract edges of the objects. Table 2 shows the filter coefficients of the Laplacian filter.

When $x_{SE}(f, i, j)$ and $x_{PE}(f, i, j)$ denote the luminance value of the reference and coded picture after applying the Laplacian filter whose filter coefficients are shown in Figure 5, image feature P7 is expressed as follows:

$$e_7(f) = \sum_{i,j} \frac{\{x_{SE}(f,i,j) - x_{PE}(f,i,j)\}^2}{N_P}$$

$$P_7 = \sum_f \left\{ 10 \log_{10} \frac{255^2}{e_7(f)} \right\} \Big/ N_F$$

Integration of image features

The overall quality is defined as weighted sum of the image features. When w_{ki} and P_{ki} denote k-th weighting coefficients and k-th image feature respectively, the overall quality Q_{obj} is expressed as follows.

$$Q_{obj} = \sum_{k=1}^{7} w_k \cdot P_k$$

Then nonlinear mapping is applied to this objective score. This is because the subjective scores are often compressed at the end of the rating scales and thus considered to have nonlinear characteristics. A logistic function is exploited as the mapping function and thus the estimated subjective score $DMOS_p$ is expressed as follows:

$$DMOS_p = \frac{c_0}{c_1 + \exp(-c_2 \times Q_{obj})}$$

where c_k (k=0,1,2) denotes coefficients obtained by regression analysis between the objective and subjective scores.

Copyright © 2009 by KDDI R&D Laboratories Inc.

Table 1 Coefficients of low-pass filter for image feature P6
(center of the matrix is the current pixel)

	1/5	
1/5	1/5	1/5
	1/5	

Table 2 Coefficients of Laplacian filter for image feature P7
(center of the matrix is the current pixel)

	-1	
-1	4	-1
	-1	

PERFORMANCE OF THE PROPOSED MODEL

Simulation Conditions

To examine performance of the model, computer simulation experiment was conducted under the conditions shown in table 3. Total 22 title sequences were selected from the standard HDTV test materials recommended in ITU-R BT.1210 and they were coded by two major HD encoders (x264 software encoder for H.264 and SONY BDKP-E2001 hardware encoder for MPEG-2) at bitrates from 2.0 to 20Mbps. Total number of the processed sequences was 242 and subjective scores of those sequences were collected by ACR-HR test, which is recommended in ITU-T Recommendation P.910. The test sequences were categorized into training and test set. The training set is used for the optimization of the model and the test set is used to examine estimation accuracy of the model. Table 4 shows the selection and categorization of the test sequences.

Table 3 Experiment Conditions

Coding Conditions	H.264 MP@HL
	MPEG-2 MP@HL, MP@1440HL
Bitrates	2,4,6,8,10,14Mbps (H.264)
	12,14,18Mbps(HL1440), 16,20Mbps(HL)
Number of Subjects	20 (after screening)
Subjective Test	ACR-HR (ITU-T P.910Rev)

Copyright © 2009 by KDDI R&D Laboratories Inc.

Table 4 Test sequences

Sequence #	Set Category	Title
1	Training	Cognac and Fruits
7	Training	European Market
10	Test	Streetcar
13	Test	Church
14	Training	Yacht Harbor
16	Test	Whale Show
20	Test	Soccer Action
21	Training	Baseball
23	Test	Green Leaves
24	Training	Swinging
28	Training	Summertime Tanning
30	Training	Crowded Crosswalk
31	Training	Flamingoes
36	Test	Airplane Landing
38	Test	Skyscrapers
39	Test	Weather Report
43	Training	Bronze with Credits
44	Training	Chromakey(fishbowl)
46	Training	Chromakey(sprinkling)
90	Test	Flower and Lady*
91	Test	Ferris Wheel*
92	Test	Banshee Jump*

* Scene from KDDI's own footage

Simulation Results

Figure 4 shows the relation between the objective and subjective scores of the training set. The mapping function is obtained by regression analysis between objective and subjective data (i.e., Q_{obj} vs. DMOS). The regression curve in figure 4 has a correlation coefficient of 0.926.

Subjective scores of the test set are estimated using this mapping function. Figure 5 shows relations between estimated and actual DMOS for training and test set. The proposed model achieves a correlation coefficient of 0.912. This is equivalent to the estimation accuracy achieved by the metrics recommended in ITU-T J.144, which is a standard of perceptual quality measurement for standard definition sequences.

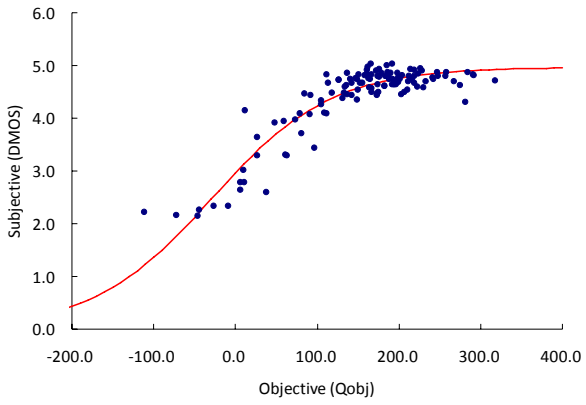

Figure 4 Nonlinear mapping function (R^2=0.857)

Copyright © 2009 by KDDI R&D Laboratories Inc.

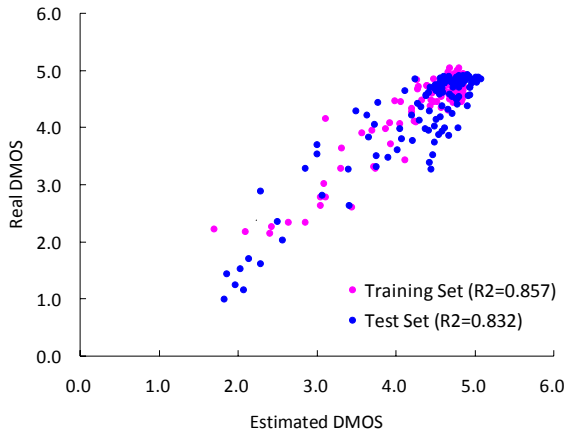

Figure 5 Correlation between subjective and objective score

CONCLUSIONS

In this paper we introduced KDDI's proprietary technology for objective perceived picture quality measurement applicable to evaluation of high-definition video. The objective model offers an easy way to measure perceived video quality and helps many video applications to manage their quality issues.

REFERENCES

[1] S.A. Karunasekera, N.G. Kngsbury, "A distortion measure for blocking artifacts in images based on human visual sensitivity", IEEE Trans. Image Processing 4(6): pp.713-724, 1995

[2] M.D. Brotherton, D.Bayart, D.S. Hands, "Subjective Quality Assessment of the H.264/AVC In-Loop De-Blocking Filter", IEICE Trans. Comm. VOL.E89-B NO.2

Copyright © 2008 by KDDI R&D Laboratories Inc.

KDDI R&D LABS

KDDI R&D Laboratories Inc.
Ohara 2-1-15, Fujimino City, Saitama, 356-8502 JAPAN
E-mail: inquiry@kddilabs.jp
Web: http://www.kddilabs.jp