NO-REFERENCE QUALITY ESTIMATION FOR COMPRESSED VIDEOS BASED ON INTER-FRAME ACTIVITY DIFFERENCE

Toru Yamada[†,††] *and Takao Nishitani*[††]
NEC Corporation[†] and Tokyo Metropolitan University[††]

ABSTRACT

This paper presents a no-reference (NR) based video-quality estimation method for compressed videos. The proposed method does not need bitstream information. Only pixel information is used for the quality estimation. An activity value which indicates a variance of luminance values is calculated for every given-size pixel block. The activity difference between an intra-coded frame and its adjacent frame is employed. In addition, blockiness and blur levels are also estimated at every frame and are taken into account. Experimental results show that the proposed method achieves accurate video-quality estimation. The correlation coefficient between subjective quality and estimated quality is 0.925. The proposed method is suitable for automatic quality check when the original videos cannot be used.

Index Terms—Video Quality Estimation, No Reference

1. INTRODUCTION

Broadcasting services which use digital videos are increasing today. Generally, these digital videos are compressed because of limitations of a storage size and a network bandwidth. Since lossy video-compression methods such as MPEG-2 and ITU-T H.264 are used, the video quality is generally degraded by the video compression.

For large-scale broadcasting services using the digital videos such as terrestrial TV broadcasting or IPTV, quality assurance is an important issue. It is necessary for the service providers to check the video quality before the video contents are delivered to end users. Currently, it is common that the quality check is conducted by human observers. This subjective approach is not only high cost, but also may result in quality-check leakages. Therefore, an objective quality check is preferable. The international standards for the objective video-quality metrics have been considered by the Video Quality Experts Group (VQEG) [1]. The published international standards [2][3] require a reference of original videos which do not include quality degradation. However, there exist some cases that the original videos cannot be used for the quality estimation. For example, content providers often deliver the video contents to the service providers after the video compression is applied. In this case, the service providers must conduct quality checks without the original videos.

For an approach without the original-video reference (no-reference (NR) model), it is hard to distinguish video quality degradation from features of the video itself. From this reason, it is difficult for the NR model to achieve accurate quality estimation. International standards based on the NR model have not been established yet. NR methods previously proposed in [4][5] do not estimate overall subjective quality but estimate the degree of blockiness or blur. To achieve accurate quality estimation with the NR models, methods which use bitstream information have been studied [6][7]. Such methods, however, depend on the compression algorithms and can only be used for video sequences using the specific video-compression algorithm.

This paper proposes an NR-based video-quality estimation method without compressed bitstream information. To exclude dependency of the compression algorithms, the proposed method only uses decoded-pixel information. First, the spatial-frequency information is analyzed to detect intra-coded frames. Then, signal difference between the intra-coded frame and its adjacent frame is calculated to estimate the amount of the quality degradation. For the signal, a value called "activity" which indicates a variance of luminance values is adopted. In addition, blockiness and blur levels are also estimated at every frame and are taken into account.

The subsequent sections of this paper are organized as follows: Section 2 describes the proposed method; Section 3 discusses a performance evaluation of the proposed method; and Section 4 summarizes our work.

2. THE PROPOSED METHOD

The proposed method estimates subjective video quality by analyzing pixel information of the degraded videos which were compressed by algorithms adopting an inter-frame prediction. First, intra-coded frames are detected and a signal difference between the intra-coded frame and the adjacent non-intra-coded frame is calculated by using an activity value of every given-size pixel block. The quality-estimation method based on the activity difference was originally introduced to a reduced-reference (RR) model [8]. In this method, the activity difference between an original video and the degraded video is calculated for the quality estimation. The proposed method applies this approach to successive frames of the degraded video. Generally, since intra-coded frames are used for inter-frame prediction in the successive-frame decoding, they tend to be compressed with small quality degradation. On the flip side, non intra-coded frames which are not used for inter-frame prediction tend to

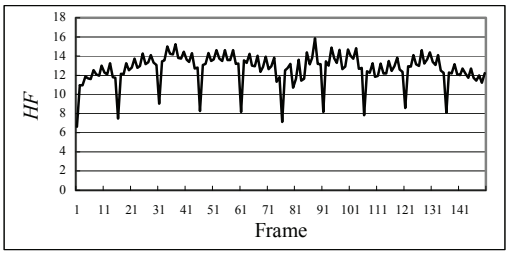


Fig. 1 *HF* value of each frame ("bus", MPEG-2 4 Mbps).

be compressed with relatively large quality degradation. In the proposed method, the intra-coded frames with small quality degradation are considered as the original video frames in the RR model. Then the quality of the adjacent non intra-coded frame is estimated with the information of the intra-coded frame. When the inter-frame activity difference is large, it can be considered that the video quality is largely degraded. In addition, blockiness and blur levels are also estimated at every frame and are taken into account. The detailed algorithm is described below.

2.1. Intra-coded Frame Detection

For the activity-difference calculation between an intra-coded frame and the adjacent frame, it is necessary to detect the intra-coded frames. In the proposed method, only the decoded pixels are used to detect them. The intra-coded frame detection employs the difference of spatial-frequency information between intra-coded and non-intra-coded frames. Generally, the intra-coded frames are compressed by applying quantization to the coefficients of an orthogonal transform such as Discrete Cosine Transform (DCT). In the quantization process, data compression is achieved by reducing high-frequency signals which the human vision hardly detects. As a result, the intra-coded frames include less amount of high-frequency information. On the flip side, non intra-coded frames are compressed by the quantization after applying the orthogonal transform to residuals from corresponding pixel information of reference frames. In general, the quantization process for the residual signals is uniformly applied to every frequency band. Therefore, amount of the spatial high-frequency signals for the non intra-coded frames is different from that for the intra-coded frames. The amount of the high frequency signals *HF* for a frame is defined as:

$$HF = \frac{1}{M_b}\sum_{k=0}^{M_b-1}\left[\frac{1}{16}\sum_{i=4}^{7}\sum_{j=4}^{7}\left|DCT_k(i,j)\right|\right], \quad (1)$$

where $DCT_k(i, j)$ is a DCT coefficient of kth 8x8 pixel block, M_b is the number of blocks in a frame. Figure 1 shows *HF* values for each frame when a video sequence "bus" is encoded at 4 Mbps of MPEG-2. The *HF* values periodically become small. These are intra-coded frames. In other words, intra-coded frames can be detected by analyzing *HF* values. In the proposed method, when the following equation is satisfied, the frame is considered as an intra-coded frame.

$$HF / HF_{Ave} < Th_1, \quad (2)$$

where HF_{Ave} is an average of *HF* values in last two seconds. From preliminary experiments, Th_1 has been set to 0.7. With this parameter, all intra-coded frames have been successfully detected for training video sequences which are used in experiments in the following section.

2.2. Inter-frame Activity Difference Calculation

The activity of the luminance values is defined as:

$$Act = \frac{1}{K}\sum_{i=0}^{K-1}\left|Y_i - Y_{Ave}\right|, \quad (3)$$

where K is the number of pixels in a block, Y_i is a luminance value in a block, and Y_{Ave} is an average of the luminance values in the block. As the RR model in [8] does, the proposed NR model also adopts 16x16 pixel block size. Therefore, K is equal to 256.

Before the inter-frame activity-difference calculation, motion compensation is applied in order to find a corresponding block position in the adjacent frame. As the motion compensation for video compressions, the mean of absolute difference (MAD) of luminance for each block is calculated in a predetermined search area and the block position where the MAD becomes the minimum is decided as the corresponding block position.

Video quality is estimated on the basis of the mean squared error (MSE) between activity values of blocks in the intra-coded frame ($ActIntra_{i,j}$) and those of corresponding blocks in the adjacent frame ($ActAdjacent_{i,j}$). The MSE is calculated as:

$$MSE = \frac{1}{N_{Intra}}\sum_{i=0}^{N_{Intra}-1}\left[\frac{1}{M_i}\sum_{j=0}^{M_i-1}\left(ActIntra_{i,j} - ActAdjacent_{i,j}\right)^2\right], \quad (4)$$

where N_{Intra} is the number of intra-coded frames and M_i is the number of blocks in which the activity difference is calculated. A peak signal to noise ratio (PSNR) value on the basis of the activity-difference is then calculated as:

$$VQ = 10 \times \log_{10}\frac{255 \times 255}{MSE}. \quad (5)$$

This value represents the tentative video-quality score.

When the minimum *MAD* is larger than a predetermined threshold Th_2, it indicates that there are no corresponding blocks in the adjacent frame. This may happen when there is a scene change, an object going out of the frame, and so on. In this case, the activity-difference calculation is skipped. From preliminary experiments, Th_2 has been set to 12.

2.3. Blockiness Estimation

Since blockiness, which is generated by video compressions with high compression ratios, is an annoyable artifact, subjective quality tends to be low for videos with a high blockiness level. In the proposed method, when a video sequence includes significant blockiness, the tentative video-quality score is modified to be lower.

To estimate the blockiness level, the proposed method adopts a method in [8], using activity values for 8x8 pixel

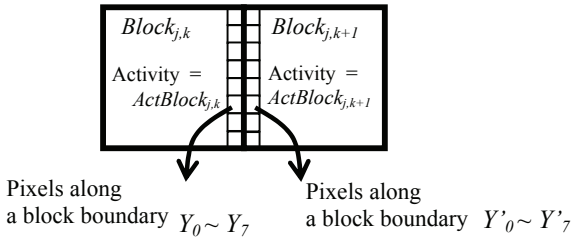


Fig. 2 Information for blockiness level estimation

blocks ($ActBlock_{i,j}$). In this method, every pair of horizontally adjacent blocks shown in Fig. 2 is processed in the following way. The average (Act_{Ave}) of the two activity values ($ActBlock_{j,k}$, $ActBlock_{j,k+1}$) is calculated by

$$Act_{Ave} = \frac{1}{2}\left(ActBlock_{j,k} + ActBlock_{j,k+1}\right). \quad (6)$$

Then, the absolute difference of the luminance values along the block boundary is calculated. As illustrated in Fig. 2, let Y_k and Y'_k represent luminance values along the boundary in both blocks. An average of the absolute difference (*DiffBound*) is expressed as:

$$DiffBound = \frac{1}{8}\sum_{k=0}^{7}\left|Y_k - Y'_k\right|. \quad (7)$$

Blockiness level ($BL_{j,k}$) of $Block_{j,k}$ is defined by the ratio of *DiffBound* to Act_{Ave}, i.e.,

$$BL_{j,k} = \frac{DiffBound}{Act_{Ave} + 1}. \quad (8)$$

Finally, as the blockiness level of the video sequence, the average of all the *BL*s is calculated by

$$BL_{Ave} = \frac{1}{N \times M_b}\sum_{j=0}^{N-1}\sum_{k=0}^{M_b-1} BL_{j,k}, \quad (9)$$

where N is the number of frames. This value is used for the tentative-score adjustment.

2.4. Blur Estimation

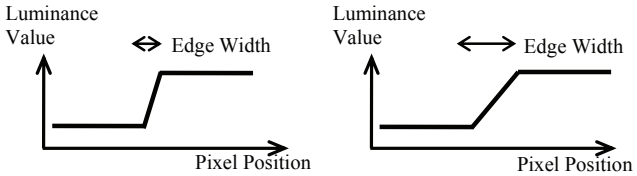
Blur, which is also generated by video compressions, is another annoyable artifact. An estimated blur level is also employed for the score adjustment. To calculate blur level, the proposed method analyzes an edge width. Figure 3 shows samples of the edge width. In a video with low blur level, edges tend to be steep as shown in Fig. 3 (a). On the flip side, edges tend to be gradual as shown in Fig. 3 (b) in a video with high blur level. The edge width $EW_{i,j}$ is defined as the number of pixels whose luminance values are monotonically increasing or decreasing in an edge region. The edge regions are detected by a high pass filter such as Sobel filter. The blur level (*BlurLv*) is calculated as an average of the edge width in a whole video and defined as:

$$BlurLv = \frac{1}{N}\sum_{i=0}^{N-1}\left[\frac{1}{L_i}\sum_{j=0}^{L_i-1} EW_{i,j}\right], \quad (10)$$

where L_i is the number of the edge pixels in a frame.

2.5. Subjective Video Quality Estimation

Subjective video quality is estimated by the activity difference with adjustments by both blockiness and blur



(a) Low blur level (b) High blur level

Fig. 3 Blur level estimation with edge widths.

Table 1 Subjective video quality test conditions

Test Methodology	ITU-T P.910 ACR-HR
The Number of Subjects	35
Video Codec	MPEG-2 and H.264
Video Bit Rate	1~6 Mbps (CBR)
I-Picture Period	15 Frames
Video Duration	5 Seconds
Video Resolution	720 x 480 pixels
Frame Rate	29.97 fps
Video Sequences	Training Set: ballet, bus, mobile and calendar, table tennis
	Test Set: cheer leaders, flower garden, foot ball, hockey

levels. In order to adjust the tentative video-quality score calculated by Eq. (5), the following functions are adopted:

$$MVQ = \frac{VQ}{(1 + BL_{Ave}^2)/W_{Block}} \times \frac{1}{W_{Blur}}, \quad (11)$$

$$W_{Block} = f(B_{Ave}), \quad (12)$$

$$W_{Blur} = g(BlurLv), \quad (13)$$

where f and g are functions to decide adjustment parameters W_{Block} and W_{Blur}, described in the next section. The *MVQ* value represents the estimated video-quality score by the proposed method.

3. EXPERIMENTAL RESULTS

The performance of the proposed method is evaluated by examining correlation to subjective quality. Subjective-quality testing was conducted under the conditions shown in Table 1. First, a training set is used to determine the functions in Eq.(12) and Eq.(13). Then, a test set is used for performance verification. The test set does not include the video sequences in the training set which are used to determine the functions.

Figure 4 shows a scatter plot for the estimated quality calculated by Eq. (5), without the adjustments by blockiness and blur levels. A certain extent of correlation is observed. However, there exist some cases which significantly lose the correlation. Subjective quality in these cases is very low and it can be considered that both the intra-coded frames and their adjacent frames are largely degraded and equally lose high-frequency signals. As a result, since the activity difference becomes relatively small and the estimated quality becomes high. In these video sequences, since blockiness and blur levels are significant, quality-estimation accuracy can be improved by adopting the adjustment in Eq. (11). First, the functions in Eq.(12) and Eq.(13) are

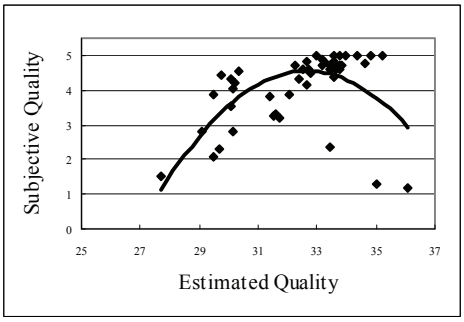


Fig. 4 Scatter plot of the activity difference. (Training set)

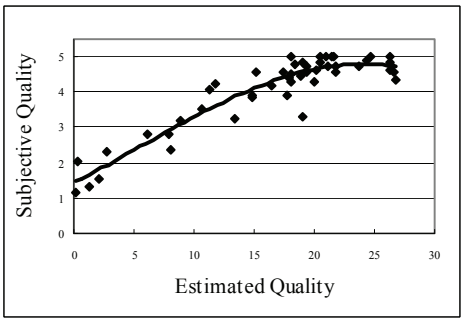


Fig. 5 Scatter plot of the proposed method (Training set)

Table 2 Functions to use score adjustments

Adjustment Operation Type	Parameter Values	
Function for Blockiness ($f(BL_{Ave})$)	BL_{Ave}> 0.9	W_{Block}=1.5
	Otherwise	W_{Block}=1.0
Function for Blur ($g(BlurLv)$)	$BlurLv$> 5.7	W_{Blur}=12.0
	$BlurLv$> 5.4	W_{Blur}= 3.5
	$BlurLv$> 3.5	W_{Blur}=1.7
	$BlurLv$> 3.3	W_{Blur}=1.6
	Otherwise	W_{Blur}=1.25

Table 3 Experimental results for the test set

	Model	CC	RMSE
Blockiness Level [4]	NR	0.812	0.602
Blur Level [5]	NR	0.864	0.518
Proposed Method	NR	0.925	0.390
ITU-T J.249 Annex B [3]	RR	0.931	0.368
PSNR	FR	0.690	0.576

determined by experiments with the training set in Table 1. Pearson correlation coefficients (CC) between the actual subjective quality and the estimated quality were calculated over changing the functions and a function set for the best CC was obtained. Table 2 shows the details of the functions in Eq.(12) and Eq.(13). Figure 5 shows a scatter plot when applying Eq. (11) with the functions shown in Table 2. As shown in Fig. 5, the correlation has been greatly improved. The correlation coefficient of 0.942 is achieved.

For a verification of this function set, the quality estimation accuracy is examined using the test set in Table 1. Table 3 shows the CCs and the root mean square errors (RMSE) between the subjective quality and the estimated quality. In Table 3, the results of the proposed method with the functions in Table 2, conventional approaches (NR and RR models), and PSNR which is a typical full-reference model are shown. The conventional NR methods simply estimate blockiness and blur levels. The CC of the proposed method is slightly lower than that of ITU-T J.249 Annex B which is an RR model. However, that is greatly higher than those provided by the conventional NR approaches. The RMSE of the proposed method is smaller than those of the conventional NR models. Besides, the proposed method achieves higher correlation than PSNR. In the discussions in the VQEG, PSNR was used as a reference measure for a performance verification of RR and NR models [9]. The models whose quality-estimation accuracy was statistically equivalent or higher than PSNR were considered as having good performance. These results show that the proposed method has high enough performance for an NR model.

4. CONCLUSION

The proposed NR-based quality-estimation method for compressed videos has been shown to achieve a high correlation with the subjective video quality. This is because the introduction of the activity difference between the decoded frames has resulted in similar scores of the subjective test, although the video quality is estimated without any bitstream information. Blockiness and blur effects are also counted in the final-score calculation by using only information of decoded pixels. These adjustments significantly reduce outliers in the difference of the subjective quality and the estimated quality. As the method does not use any compressed bitstream information, the proposed algorithm can be applied to a quality check in various systems which handle video compressions with inter-frame prediction.

5. REFERENCES

[1] The Video Quality Experts Group Web Site, http://www.its.bldrdoc.gov/vqeg/ .
[2] ITU-T Recommendation J.144, 2004.
[3] ITU-T Recommendation J.249, 2010.
[4] K.T. Tan, et al., "Frequency Domain Measurement of Blockiness in MPEG-2 Coded Video," Proc. of ICIP, Vol. 3, pp.977-980, Sept., 2000.
[5] P. Marziliano, et al., "A No-reference Perceptual Blur Metric," Proc. of ICIP, Vol.3 pp.III57-III60, June 2002.
[6] A. Ichigaya, et al., "A Method of Estimating Coding PSNR Using Quantized DCT Coefficients," IEEE Trans. on CSVT, Vol. 16, No. 12, pp.251-259, Feb., 2006.
[7] D. S. Turaga, et al., "No Reference PSNR Estimation for Compressed Pictures," Proc. of ICIP, Vol. 3, PP. III61-64, 2002.
[8] T. Yamada, et al., "Video-Quality Estimation Based on Reduced-Reference Model Employing Activity-Difference," IEICE Trans. on Fund., Vol. E92-A, No. 12, pp.3284-3290, Dec., 2009.
[9] "Final Report from the Video Quality Experts Group on the Validation of Reduced-Reference and No-Reference Objective Models for Standard Definition Television, Phase I," 2009.