Research on No-Reference Video Quality Evaluation Algorithm Based on H.264

Yangming Chen, Xiuhua Jiang
Digital Video Quality Assessment Lab
Communication University of China
Beijing, China
e-mail:1791036454@qq.com, jiangxiuhua@cuc.edu.cn

***Abstract*—In this paper, no reference video quality evaluation algorithm is researched and analyzed, and we implement a two-domain no-reference video quality assessment method: First, a compressed domain sub video quality assessment model is implemented based on the encode information are extracted from bit streams, then the video similarity between the distortion and original videos is calculated, so the preliminary evaluation of video quality is obtained. Second, two distortion artifacts, namely, block effect and blur effect are detected in the pixel domain. Finally, the video quality is given considering both video similarity and distortion. In this paper, the effect of the algorithm is evaluated by using the Consumer Digital Video Library. Numerous experiment results demonstrate that our implemented the objective evaluation model achieves near optimal tradeoff between consistency of subjective evaluation results and computational complexity.**

Keywords-no-reference; compressed domain; pixel domain; two-domain

I. INTRODUCTION

Recently, the research of digital video quality assessment method is a hot topic. In compressed domain, Hui Shi et al. [1], Rui Song et al. [2] and Cheng et al. [3], respectively, proposed a very similar no reference method based on H.264 code stream. These algorithms are both extract two parameters from code stream: quantization coefficient and number of encoding skipped macro blocks, then, establish a mathematical model between the extraction parameters and the video quality. Yogini et al. [4] proposed a no reference method based on compressed domain, the algorithm analyzes the relationship between the compression ratio, bit rate and video scene, calculate the quantization factor, motion factor and stream factor, and establishes a mathematical model between the factors and video quality. The computational complexity of these methods is low, but they are not accurate to evaluate some common distortion effects, such as block effect, blur effect and so on, the accuracy of evaluation results should be improved. In pixel domain, Wang et al. [5], proposed a method to detect the blur effect in video, which is based on the Canny operator for edge detection, and introduces the concept of gradient direction, edge extension and image blur. These methods have achieved good accuracy, but their computational complexity is high.

In order to improve the accuracy of no reference video quality assessment and adaptability, at the same time, reduce the computational complexity, we have made some changes to the algorithm that we refered to, we leaved the part of the neural network out and simplified the algorithm of image distortion in pixel domain. The improved algorithm reduces the running time and improves the algorithm efficiency.

The remainder of this paper is organized as follows. Section 2 presents the objective evaluation model and problem formulation. Numerical experimental results are provided in Section 3. Finally, conclusion will be stated in Section 4.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We first implement an objective evaluation model, then we make a detailed introduction to each part.

A. Objective Evaluation Model

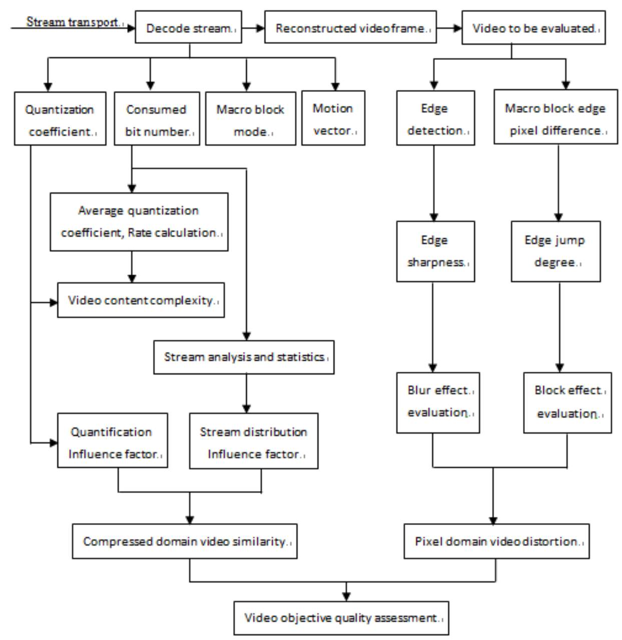


Figure 1. Framework of two-domain no reference video quality assessment method

The model is mainly divided into three parts: i) video similarity estimation in compressed domain, using encoded information obtained from the decoder, such as macroblock types, quantization coefficients, motion vectors and the number of bits per macroblock consumption, combined with video content complexity to measure the damage caused by compression coding, implemented the preliminary evaluation of video quality assessment with lower computational complexity, ii) video distortion estimation in pixel domain, according to the information contained in the adjacent pixels in the damaged video, the mathematical statistic characteristics of the video is analyzed, which is the variance, the local distortion of the video is measured, and then calculate the detail loss of the video after the Gauss filter to get the global distortion, and then combining these two measures the overall distortion of the video, at the same time, considering the influence of the video content on the sensitivity of the distortion effect, the intra prediction is used to calculate the complexity of the video content, combined with the overall distortion of video to get the video quality evaluation, and iii) objective quality evaluation. Fig. 1 shows its overall structure.

B. Video Similarity Estimation in Compressed Domain

In the process of video compression, the rate of video plays a decisive influence on video quality, and the quantization coefficient is directly related to the code rate. In order to make a more accurate analysis of the relationship between quantization coefficient and video quality, in this paper, the relationship between the quantization coefficient and the subjective quality Differential Mean Opinion Score (DMOS) obtained in [6] shows that the greater the DMOS value is, the worse the subjective quality of the video is, as shown in Fig. 2.

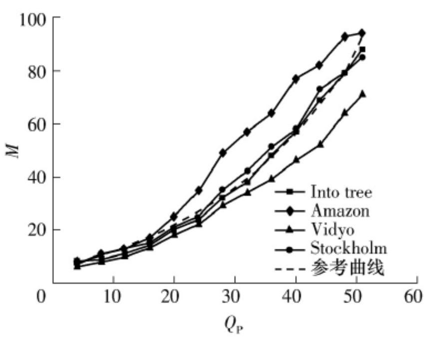


Figure 2. Relationship between quantization coefficient QP and DMOS value of subjective video quality

Analysis the four curves in Fig. 2 can be obtained, the bigger the QP, the bigger the value of DMOS, and the worse the video quality, the relationship between the subjective quality DMOS and the quantization coefficient QP is similar to the exponential relation, but these four curves are not the standard exponential curve, it shows that the DMOS value of video quality and the quantization coefficient QP are not exactly exponential, in addition to the quantization factor, there are other factors that affect the quality of the video. According to the above analysis, we can use (1) to describe the composition of the subjective quality of DMOS:

$$\mathrm{DMOS} = C^q + f \tag{1}$$

where, C stands for the video content complexity, q represents the factor associated with quantization coefficient, and f represents other factors that affect the quality of video. In order to better understand (1), a set of values of C，q and f were obtained by curve fitting in [6], then to produce a reference curve based on the values, as shown in Fig. 2 black curve, its shape is very similar to the other four, this also shows that the (1) can be more reasonable to reflect the composition of the subjective quality of DMOS. At the same rate, the rate control algorithm has a great influence on video quality. In addition, there are a number of other factors will have a certain impact on the quality of video, but its influence is far less than both of the above. Therefore, the method of this section uses the quantization coefficient influence factor $\mathrm{F_Q}$ and the code stream distribution influence factor $\mathrm{F_B}$ to reflect the compression domain video similarity $\mathrm{Q_S}$, which approximately reflects the video objective quality.

1) Quantization influence factor

In general, under the same compression ratio, a complex video after compression tend to get higher bit rate. Therefore, the complexity of video content can be reflected by the product of compression ratio and bit rate. The compression ratio is proportional to the quantization step in the quantization process, the quantization parameter is increased by 6, and the quantization step size is doubled. Based on the above information, content complexity C of a video (n frames) can be expressed as:

$$\mathrm{C} = \frac{B}{w*h} * 2^{\frac{qp}{6}} \tag{2}$$

where, B stands for the total number of bits consumed by n video frames, w and h represent the width and height of the video respectively, qp represents the average quantization parameter used in video compression of n frames. Finally, the calculation formula of the quantization influence factor $\mathrm{F_Q}$ is shown as follows:

$$F_Q = (a_1 * C + a_2)^{a_3 * qp} \tag{3}$$

where, $\mathrm{a_1}$, $\mathrm{a_2}$ and $\mathrm{a_3}$ are the weight coefficients of the value greater than zero, their values are obtained through training in the experiment. The bigger $\mathrm{F_Q}$, the lower the degree of similarity between the video to be evaluated and the original video, and the worse the quality.

2) Code stream distribution influence factor

The concentration degree of the code stream distribution is positively correlated with the video quality, and the video quality is relatively high when the code stream distribution is concentrated in the sensitive area of the HVS. Therefore, the method of this section uses the variance of the code stream distribution in each frame $\mathrm{D_B}$ to reflect the effect of the frame rate control, as shown in (4):

$$D_B = \frac{\sum (B_{MB} - B_{AVG})^2}{N} \tag{4}$$

where, $\mathrm{B_{MB}}$ represents the number of bits consumed per macro block, $\mathrm{B_{AVG}}$ stands for the average number of bits consumed per macro block, N represents total number of macro blocks. The formulas above (2), (3) and (4) are presented in [6]. At the same time, the accuracy of the code stream distribution $\mathrm{C_{RC}}$ is evaluated by the statistics of the probability that a macro block which belongs to the region of interest that consumes the number of bits more than the average value. First of all, from the whole frame of video to extract the Inter mode macro block which consumed the number of bits more than the average value, the total number of these macro is N_B, this part of the macro blocks determines the quality of the video, so they are called key macro blocks. And then determine whether each key macro block belongs to the visual attention region, as shown in (5):

$$I_M = \begin{cases} 1, if\left(\sqrt{\left(mv_x - mv_{x_{\_avg}}\right)^2 + \left(mv_y - mv_{y_{\_avg}}\right)^2} > T_{MV} * \sqrt{{mv_{x_{\_avg}}}^2 + {mv_{y_{\_avg}}}^2}\right) \\ 0, \qquad otherwise \end{cases} \tag{5}$$

Then, the method uses the total number of N_I with I_M value of 1 percentage of N_B as the accuracy of code stream distribution C_{RC}, as shown in (6):

$$C_{RC} = \frac{N_I}{N_B} \tag{6}$$

Finally, the effect of rate control on the video quality is reflected by the product of the rate control effect factor D_B and the accuracy of the bit stream distribution C_{RC}, that is code stream distribution influence factor F_B, the greater the F_B, the higher the degree of similarity between the video to be evaluated and the original video, the better the quality of the video.

$$F_B = D_B * C_{RC} \tag{7}$$

3) Video similarity

The total of 10 seconds in a total of 250 frames of video as a whole, and video similarity Q_S in compressed domain is used to evaluate preliminary objective quality of video, Q_S the greater the video quality is better, that is, the smaller the value of DMOS. The calculation formula of Q_S is as follows:

$$Q_S = b_1 * F_{B_{avg}} - b_2 * F_{Q_{avg}} \tag{8}$$

where, F_{Q_avg} stands for overall quantization influence factor, and it is obtained by the rate of the 250 frames of video and the quantization coefficient. F_{B_avg} represents overall code stream distribution influence factor, and it is the average value of the stream distribution impact factor F_B of each frame in n video frames. In (8), b_1 and b_2 are the weight coefficients of the value greater than zero, their values are obtained through training in the experiment.

C. Video Distortion Estimation in Pixel Domain

In pixel domain, the algorithm is used to evaluate the degree of video distortion from two aspects: block effects and blur effects, which is used to adjust the preliminary video quality evaluation results in the compressed domain.

1) Block effect distortion

The block effect appears in the compression algorithm of discrete cosine transform (DCT), and the rough quantization process also can increase the block effect. Different macro block by the DCT and quantify the impact is not the same, there are differences in the details of the loss, as a result, there is often a significant discontinuity in the macro block boundary, which forms block effect distortion, in addition, according to the characteristics of the HVS, the human eye is more sensitive to the block effect in the flat region with a relatively simple texture in a video. According to the above analysis, the distortion degree of the block effect is calculated from the horizontal and vertical directions respectively. The implementation process of the method is introduced in the horizontal direction as an example. As shown in Fig. 3, the white and gray squares represent the macro block A and B respectively. Firstly, the difference between the pixel points of the macro block boundary S_B and the difference between the pixels inside the macro block S_I are calculated:

$$\begin{aligned} S_{B(i)} &= \left|A_{(15,i)} - B_{(0,i)}\right| \\ S_{I(0,i)} &= \left|A_{(13,i)} - A_{(14,i)}\right| \\ S_{I(1,i)} &= \left|A_{(14,i)} - A_{(15,i)}\right| \\ S_{I(2,i)} &= \left|B_{(0,i)} - B_{(1,i)}\right| \\ S_{I(3,i)} &= \left|B_{(1,i)} - B_{(2,i)}\right| \end{aligned} \tag{9}$$

On the basis of this, the average difference between the pixels inside the macro block S_{I_AVG} is calculated :

$$S_{I\_AVG(i)} = \frac{S_{I(0,i)} + S_{I(1,i)} + S_{I(2,i)} + S_{I(3,i)}}{4} \tag{10}$$

13,0	13,1	13,2	13,3	13,4	13,5	13,6	13,7	13,8	13,9	13,10	13,11	13,12	13,13	13,14	13,15
14,0	14,1	14,2	14,3	14,4	14,5	14,6	14,7	14,8	14,9	14,10	14,11	14,12	14,13	14,14	14,15
15,0	15,1	15,2	15,3	15,4	15,5	15,6	15,7	15,8	15,9	15,10	15,11	15,12	15,13	15,14	15,15
0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,10	0,11	0,12	0,13	0,14	0,15
1,0	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	1,10	1,11	1,12	1,13	1,14	1,15
2,0	2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8	2,9	2,10	2,11	2,12	2,13	2,14	2,15

Figure 3. Macro block boundary pixel map

Then, according to the difference between the macro block edge and the inner pixel, the edge jump degree J is calculated, as shown in (11). Because the quantization process is based on the macro block, the adjacent macro blocks can be used with different quantization parameters. Therefore, the difference between the macro block edge pixels is often greater than the difference between the internal pixels, and with the increase of the difference, the discontinuity between the edges of the macro block is more obvious.

$$J_{(i)} = \begin{cases} S_{B(i)} - S_{I\_AVG(i)}, & if\left(S_{B(i)} > S_{I\_AVG(i)}\right) \\ 0, & otherwise \end{cases} \tag{11}$$

The research shows that the block effect distortion degree of different regions is different to the video quality, so it is needed to be different weights according to the texture features. The method is to use a relatively simple algorithm. Firstly, edge detection is performed by Sobel operator, and all pixels are divided into the edge region and non edge region. Based on this, the different weight w is given, which is used to modify the edge jump degree, and then take the average value of the edge jump degree of all pixel points, which is used as the macro block edge level block effect distortion D_{BLOCK_H}.

$$D_{BLOCK\_H} = \frac{\sum_{i=0}^{15} J_{(i)} * w_{(i)}}{16} \tag{12}$$

where, w is calculated as follows:

$$w_{(i)} = \begin{cases} w_1, & A_{(15,i)} \text{ or } B_{(0,i)} \text{ belongs to edge area} \\ w_2, & A_{(15,i)} \text{ and } B_{(0,i)} \text{ belong to non edge area} \end{cases} \tag{13}$$

where, w_1 and w_2 are the weight coefficients of the value greater than zero.

Finally, the average D_{BLOCK_H} value of each macro block $D_{BLOCK\_H\_AVG}$ in the frame video is counted. Using the same method to calculate the distortion degree of the block effect in the vertical direction $D_{BLOCK\_V\_AVG}$. The block effect distortion of the macro block D_{BLOCK} is the average of the horizontal and vertical distortion, as shown in (14). The greater the D_{BLOCK}, the more serious the block effect.

$$D_{BLOCK} = \frac{D_{BLOCK\_H\_AVG} + D_{BLOCK\_V\_AVG}}{2} \tag{14}$$

2) Blur effect distortion

Blur effect is a very common video distortion effect, from the subjective visual perspective, as shown in Fig. 4, blur effect is mainly reflected detail degradation near edge of object in video, and decrease in sharpness. The pixels on the edge of the object, if they are affected by the blur effect, generally will have the largest local gradient. Therefore, we reuse the results of Sobel edge detection mentioned in the last section. For each pixel at the edge of the object, the edge sharpness S of the gradient direction is detected, so as to estimate the distortion degree of the video affected by the blur effect.

(a) Image with sharp edge (b) Image affected by blur effect

Figure 4. Image with sharp edge and Image affected by blur effect

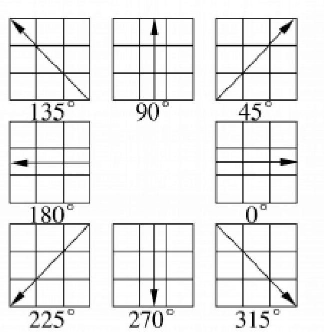


Figure 5. Eight gradient directions

Sobel operator to detect the edge of the object at the same time, we can also calculate the gradient direction of the pixel point. The direction of the gradient may be any value between 0° to 360°. In order to reduce the computational complexity, the gradient direction is divided into 8 directions, as shown in Fig. 5: 0ϵ{0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°}.

Any gradient direction is similar to the one of the closest of the above 8 directions, and calculated S in this direction. The smaller the S, the smaller the edge sharpness, the more serious the blur effect, then taking the direction of 315° as an example, the calculation method of S is introduced. As can be seen from Fig. 5, the edge of the blur image is a process of gradual change, the slower the gradient, the more serious the distortion caused by the blur effect. Wang et al. [6] proposed a fast algorithm for edge detection in image. First calculate the difference between the pixels around the edge of the image and the gray value of the pixels on the edge, and then use the ratio between the difference and the distance to estimate the gradient of the image edge gradient. On this basis, the method selects four points in the gradient direction, with their average gradient speed to calculate, as shown in (15):

$$S = \frac{1}{4} * \sum_{i=1}^{4} \frac{|I(i,i) - I(0,0)|}{i} \tag{15}$$

(i, i) position as shown in Fig. 6, where the gray point $I(0,0)$ is located on the edge of an object.

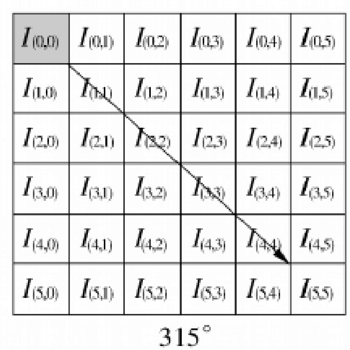


Figure 6. Gradient direction sharpness calculation

Finally, the S value corresponding to each edge pixel in the whole macro block is statistically, and the average value is taken as the blur effect distortion degree D_{BLUR}. The smaller the D_{BLUR}, the more serious the blur effect is.

3) Video distortion

In this section, the objective quality of the preliminary video in the compressed domain is modified by using the video distortion degree Q_D in the pixel domain. The greater the Q_D, represents the more serious distortion of the macro block, the worse the video quality, that is, the greater the value of DMOS. The calculation formula of Q_D is as follows:

$$Q_D = c_1 * D_{BLOCK} - c_2 * D_{BLUR} \tag{16}$$

where, c_1 and c_2 are the weight coefficients of the value greater than zero, their values are obtained through training in the experiment.

D. Objective Quality Evaluation

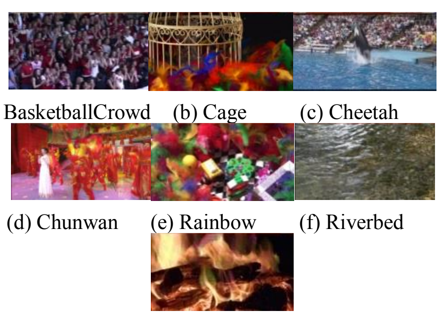
Human eye to video quality evaluation is a continuous process, the information contained between the frame and the frame should also be taken into account. Normally, each frame of video has similar content characteristics in a period of time. As a result, the whole sequence of a certain period of time is used as a whole, and the objective quality evaluation is given. Finally, the objective quality of the video Q_M is given by combining the video similarity of compressed domain and the distortion degree of the pixel domain. In order to have the same range of values and the numerical monotone of the subjective evaluation results, Q_M is revised as follows:

$$Q_M = (Q_D - Q_S + d_1) * d_2 \tag{17}$$

where, d_1 and d_2 are constant, the role are to limit Q_M to [0, 100], 0 represents the best quality, and the 100 represents the worst.

III. Numerical Analysis and Result

Subjective evaluation experiment on the basis of "double stimulus continuous quality standard degree method" in GY / T 134-1998 "the digital TV picture quality subjective evaluation method" [7]. In general, the method is validated by using a recognized standard public database in the academic field. We use the Consumer Digital Video Library (CVDL) database to evaluate the effectiveness of the algorithm. A total of 40 people participated in the experiment. The algorithm uses 7 sequences representing different features of the natural scene, as shown in Fig. 7, Each of the different sequences have 4M, 6M, 8M, 10M, 12M, 14M and 16M seven different bit rate of the code stream:



(a) BasketballCrowd (b) Cage (c) Cheetah

(d) Chunwan (e) Rainbow (f) Riverbed

(g) Tulip

Figure 7. Test video of CVDL database

According to the appendix GY/T 134-1998 a digital TV image quality subjective evaluation data statistical method [7], we confirm the validity of the evaluation data, and the original data were screened, in the test of the 40 people in the data to retain the data of 32 people.

The experimental process of this section is divided into the following four steps:

(1) Through the training to obtain the constant of the formula in the second chapter. In training, 5 video sequences: BasketballCrowd, Cage, Cheetah, Tulip and Chunwan are used as data sources, Rainbow and Riverbed will be two video sequences as test videos. Through the training, get the constant of the formula used in the second chapter, as shown in Table I.

TABLE I. The Constant of the Formula in the Second Chapter

Constant	Value	Constant	Value
a_1	0.0001	c_1	1
a_2	1.2	c_2	0.5
a_3	0.09	d_1	49
b_1	1.8	d_2	0.62
b_2	10.1		

As shown in Fig. 8 and Fig. 9, respectively expressed the DMOS value of 7 video sequences with the trend of the rate change chart and the DMOS average value of 7 video sequences with bit rate variation column chart:

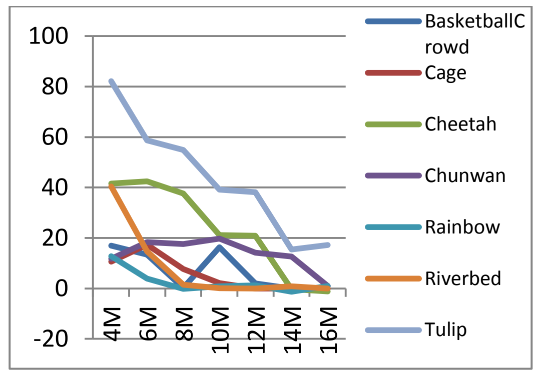


Figure 8. The DMOS value of 7 video sequences with the trend of the rate change chart

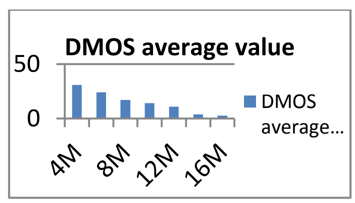


Figure 9. The DMOS average value of 7 video sequences with bit rate variation column chart

(2) Experiment with CVDL database as samples. According to the suggestion of [8], the Spearman Rank Order Correlation Coefficient (SROCC) and Pearson Linear Correlation Coefficient (LCC) were used to measure the monotone and accuracy of the method. As shown in Table II.

TABLE II. Two-Domain Experiential Results Based on CVDL Database

Prediction Model	SROCC	LCC
Two-domain	0.8649	0.8342

(3) By using two common video objective quality evaluation methods: PSNR and SSIM to evaluate the objective quality, and finally according to the above steps to calculate the SROCC and LCC, and compare the test results. As shown in Table III.

TABLE III. COMPARISON OF TWO-DOMAIN EXPERIENTIAL RESULTS BASED ON CVDL DATABASE

Prediction Model	SROCC	LCC
PSNR	0.7081	0.7857
SSIM	0.6748	0.8043
Two-domain	0.8649	0.8342

IV. CONCLUSION

In this paper, a two-domain no reference video quality assessment method based on [6] is implemented. In the compressed domain, the coding information is extracted from the code stream, and combined with the complexity of the video content to evaluate the video similarity. In the pixel domain, the two most common video distortion effects of the human eye vision system: loss caused by block effect and blur effect on video quality are calculated, and the video distortion is obtained, which is a supplement to objective video quality. Finally, the objective quality of the video is obtained by the compressed domain video similarity and the pixel domain video distortion [9]. The algorithm makes use of the coding information extracted in the process of video compression, and makes a preliminary evaluation of video quality with low computational complexity, then by calculating the effect of the block and blur to video quality as a supplement, gives a complete video objective quality [10]. The experimental results indicate that: the implementation of the algorithm is of high accuracy, has better generality, it also has low computational complexity.

In the future, we will extend the paper by considering looking for a better edge detector, it should have better noise suppression effect and higher edge detection accuracy, and we need to have a scene change detection, the video is divided into different scenes, when the video scene change occurs, the quality assessment of a period of the previous video immediately end, begin to evaluate the quality of a new video scenes, then get the overall video quality.

REFERENCES

[1] Hui S, Fang M, Xiuhua J. A high definition video quality evaluation algorithm based on H.264 code stream [J]. The film and television technology China institute of film and television technology, 2011

[2] Rui S, Xiuhua J, Hui S. Objective evaluation of video quality in the compressed domain of mobile multimedia broadcast [J]. TV Technology, 2010

[3] Yuqing Cheng, Xiuhua Jiang, Caihong Wang. HDTV Video Quality Objective Assessment Based on H.264 Compressed Domain [J]. 2011 Third International Conference on Communications and Mobile Computing, 2011.

[4] Yogini Kulkarni, Charudatta V. Kulkarni. No-Reference Video Quality Assessment [C]. Computer Systems Laboratory. HPL-98-193.1998

[5] Wang X, Tian B, Liang C, Shi D. Blind image quality assessment for measuring image blur [C]. Proceedings of International Congress an Image and Signal Processing, 2008: 467-470.

[6] Xiangyu Lin, Yaowu Chen. A Two-Domain No Reference Video Quality Assessment Algorithm [J]. Journal of South China University of Technology, 2012, 40(4); 64-70,78.

[7] GY/T 134-1998. Digital TV image quality subjective evaluation method.1998.

[8] Seshadrinathan K, Soundararajan R, Bovik A C et al.. A subjective study to evaluate video quality assessment algorithms [J]. IEEE Transactions on Image Processing, 2010, 19(6): 1427-1441.

[9] Xiangyu Lin, Yaowu Chen. Research on the method of quality assessment of no reference video [D]. Zhejiang: Zhejiang University, 2012

[10] written by Iain E. G. Richardson, translated by He Ouyang and Jun Han, H.264 and MPEG-4 video compression - a new generation of multimedia video coding technology [M]. Beijing: National University of Defense Technology press, 2004