No-reference video quality assessment model based on eye tracking datas

Lixiu Jia,Xuefei Zhong, Yan Tu*
Display R&D Center
School of Electronic Science and Engineering, Southeast University
Nanjing,China
email:jlxseu@gmail.com;zxf@seu.edu.cn;
corresponding author: email:tuyan@seu.edu.cn

***Abstract*—This paper describes a novel no-reference video quality assessment (VQA) model which is based on eye tracking datas for H.264 coding videos.This assessment model is based on the blur and the blockiness. The eye tracking datas were used to the no-reference video quality assesment.The experimental results show that the assessment model has better performance in terms of both the prediction accuracy and the computation complexity.**

Keywords—No reference Video quality assessment , blurring , blocking,H.264, Region of interest(ROI),eye tracking.

I. INTRODUCTION

With the advent of information era,the development of computer ,communication and network technology is advancing rapidly. Video compression techniques H.264 is widely used due to the limitation of bandwidth and storage capacity.The distortions were introduced in videos as the techniques decrease the bit rate in encoding.Generally speaking,the most accurate way to determine the quality of a video is by measuring it using psychophysical experiments with human subjects ,which is a time-consuming , expensive process and non-automatic.
Therefore,developing objective quality assessment metrics to measure the video quality without the presence of subjects is more urgent.[1]

Objective video quality assessment models play an important role in the video compression and communication field.Traditional objective quality metrics can be classified into the following three categories by its dependence on the original videos:Full-reference (FR) models need the original videos as reference[7]-[11],reduce-reference(RR)metrics depend on partial information associated with the original signal[12]-[15],and no-reference(NR)models that evalute the quality of videos only using the signals receive at the terminals[16]-[20].Actually,it is difficult to obtain the original video signals in most video received terminals,and the transmission of the signals will occupy certain limited bandwidth.Therefore,NR metrics have a better applicability in the real life. In recent years,a lot of no-reference algorithms were proposed for the estimation of compressed coded video artifacts in order to resolve the higher computation complexity of pixel domain[21].

In this paper the following video artifacts were considered to relate to compressed coded video signals:(i)blurring,(ii)blocking,and (iii)motion related.A way to design a single no-reference video quality assessment framework through using of eye tracking datas was presented.In section 2 ,the subjective experiment is described,while video quality assessment framework is represented in section 3. Saliency-based ROI metric is described in section 4. The experimental results is given in Section5 .Finally,the conclusion of this paper is presented in section 6.

II. DETAILS OF SUBJECTIVE EXPERIMENT

A. Sources Sequences

Four uncompressed , high-quality , source videos of natural scenes that are freely available for download from the Laboratory for Image and Video Engineering (LIVE) Video Quality Database[6] were used in this experiment.The videos which contain typical spatial features and time characteristics were chosen.All videos have a resolution of 1920×1080 pixels.The ITU-T organization suggests applying the Spatial Information (SI) and the Temporal Information (TI) as the measurement of complexity in spatial and temporal area[5] .The formula of the SI and the TI are shown below.

$$SI= \max_{time}\{std_{space}[Sobel(F_n)]\} \quad (1)$$

$$TI=\max_{time}\{std_{time}[M_n(i.j)]\} \quad (2)$$

$$M_n(i.j)= F_n(i.j)- F_{n-1}(i.j) \quad (3)$$

In the formular, F_n is the present frame of the sequence, $F_{n-1}$is the previous frame.

The SI as the Y-axis and TI as the X-axis are used to built spatial-temporal information coordinate system.The original videos seleceted should contain the four categories.

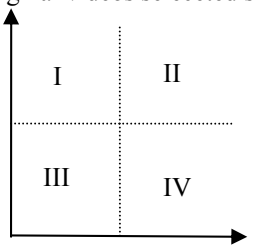


Fig.1. Spatial-temporal coordinate system

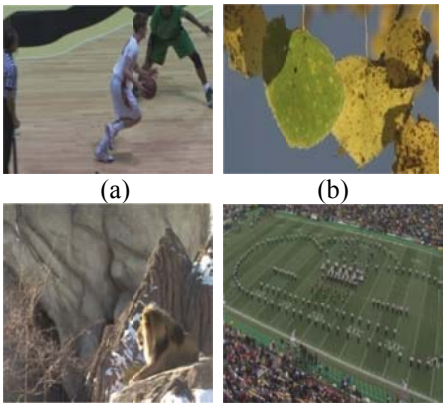


Fig.2. (a)Basketball (b)Leaf (c)Lion (d) Halftime music

Fig.2 shows the snapshot of each reference video in the LIVE Video Quality Databases.All videos are 10s long.The sequences have a frame rate of 30 frames per second.A short description of these videos is provided below.

- Basketball—The background and foreground are fast motion,camera lens move horizontally.(It belongs to the II category)
- Leaf—The background is still,the scenes change.(I)
- Lion—The background is complex,the foreground moves slowlly.(III)
- Halftime music—The background is still， the scenes change.(IV)

B. Test Sequences

24 test sequences were created from each the reference sequences using six different H.264 compressed bitrates including 9Mbps， 0.82Mbps,1Mbps,1.2Mbps,1.5Mbps,2 Mbps.Therefore,there were 24 videos in this experiment.

C. Subjective testing environment

In this paper,a single stimulus method was adopted to obtain subjective scores for different videos.Specifically, every video was played continuously twice,all of videos were viewed in a random order for the subjects everytime.At the end of the presentation of the video,a rating scale was continually displayed five seconds on the screen.The rating scale had five scales which are“Excellent” , “Good” , “Fair” , “Poor” , “Bad” and corresponded to 5 score,4 score,3 score,2 score,1 score and 0 score.The 20 scores of each video were averaged to a final Mean Opinion Score (MOS) of the video.All the videos were viewed by each subjects .The experimental time is twenty minutes for one subject.

Twenty colleague students,including ten males and ten females ,attended the experiment.They are nonprofessional with eye-tracking experiment.The average age is 23.7 years old. Each subject was told about the purpose of the three experiments which contains the blur assessment, blocking assessment and overallquality assessment,then viewed the referenced best and worst videos before starting the experiment.The six referenced videos were not recorded by eye tracking machine and were also different in the three experiments.

The SMI experiment center was used to present stimulus to the viewers.Then videos were viewed by the subjects on the Liquid crystal display (LCD) monitor.The LCD monitor is the 46 inch of HYUNDAI with a resolution of 1920×1080.In this experiment,the viewing distance was the four times screen height.The ambient light is 20lux on the front of the screen and 18Lux behind of the screen.In addition to this,the frequency of the eye tracking instrument is 250HZ.

III. VQA FRAMEWORK DESCRIPTION

A. Proposed blur metric

The blur metric was based on the analysis of the spread of the edges in an image.The novel metric had low computational complexity and was shown to perform well over a range of video content.[2]

The measurement of blur was always defined in the spatial domain.It was perceptually apparent along edges or textured areas.The blur metric in this paper was based on the smoothing effect of blur on edges,and the spread of the edges was served as the blur measurement.

Firstly, the edge detector was applied to find the horizontal and vertical edges in the image,then each row or column of the image was scaned.For pixels which corresponded to the edge location,the start and end positions of the edge were defined as the local extreme locations to the edge.The edge width was the difference between the end and the start and the end positions,and was defined as the local blur measure for this edge location.Finally,the global blur measure for the entire image was acquired by computing the average of the local blur values over all edge locations. More details refer to paper [2].

The following computational formula is the blur of one image

$$\begin{cases} B_0(i,j) = \mathrm{Diff}\big(\mathrm{pic}(x, i{:}j)\big) \\ B_p(i) = \frac{1}{\sum_{i,j=0}^{m,n} \mathrm{blur}(i,j)} \sum_{i,j=0}^{m,n} B_0(i,j) \times \mathrm{blur}(i,j) \end{cases} \quad (4)$$

$B_0(i,j)$ represents the original edge gradient,Diff() is the function to compute gradient,picc(x,i:j) represents the horizontal direction function in the some pixels of the image,blur(i,j) is the blur function, $B_p(i)$ expresses the final blur value of whole image .

For videos,the above description is straightforward served as the blur of each frame in every video.The blur value of the video is described in the following formula.

$$B_v = \frac{1}{N} \sum_{i=0}^{N} B_p(i) \quad (5)$$

N is the frame number in this video. B_v is the blur value of the video.

B. Proposed blockiness metric

In this paper, a blockiness assessment metric of image[3]were developed.Firstly, the effect of motion to the frame image of the video was weighted,and then different weights to blockiness value of each frame were

given,lastly the blockiness value of whole video was received.

Firstly,the accurate position of block boundaries using a detection phase metric were decided.Block's surrounding content with a restricted extent after confirming the all possilble blocking artifacts in each image was analysed.It is made up of two parallel procedures: (1)a local blockiness metric(LBM),computing the scale of distortion on the pixel level; and (2) human vision model(HVS), evaluating the degree of artifact and obtaining a visibility coefficient(VC),then integrated the LBM value and VC value into the local perceptual blockiness metric(LPBM). At last,the LPBM detected in the whole image were averaged to receive the blockiness value of the whole image(NPBM).More details refer to paper [3].

The human eyes are more interested in the mobile objects and neglecting the most background information. Therefore,we not only applied the blockiness value of the frame,but took into account the effect of motion information to visual perception.Based on the different degree of motion in each frame to give various weight , finally the blockiness value of the video was obtained.Specific equations are shown below.

$$\begin{cases} \mathrm{Block_f}(i) = \mathrm{M_{NPBM}}(i) \times \left[1 + \dfrac{\mathrm{M_f}(i)}{M \times N} + \mathrm{F_f}(i)\right] \\ \mathrm{Block_v} = \dfrac{1}{N_v}\sum_{i=0}^{N_v} \mathrm{Block_f}(i) \end{cases} \quad (6)$$

$\mathrm{Block_f}$ is the weighted improved blockiness of every frame,$\mathrm{M_{NPBM}}(i)$ is the original blockiness of this frame, $\mathrm{M_f}$ is the numbers of all mobile pixels.M and N are the numbers of horizontal and vertical pixels,respectively.$\mathrm{F_f}$ is the average of motion

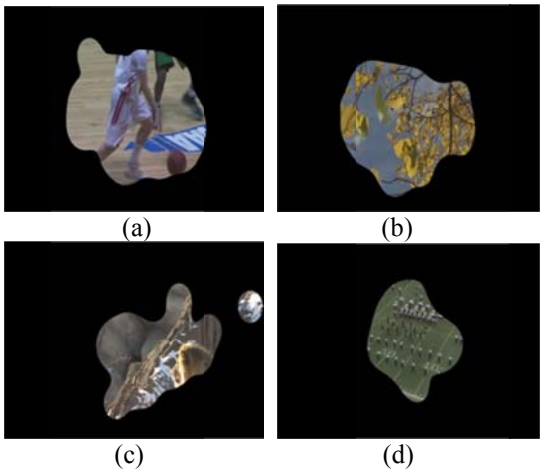
IV. SALIENCY-BASED MODEL OF VISUAL ATTENTION

In this paper we used a computational model to form fixation density map from eye movement experiment [4]. Firstly,the eye tracking datas which contained all fixations in the experiment were extracted from the SMI Begaze software.Then the eye tracking datas were dealed with using special characteristics,such as the position of x and y.Finally,the perfert datas were computed by the saliency-based model to form a human attention map.

The saliency-based algorithm was made up of two steps.Firstly,the synchronism question between positions and frames of the image using the start time and the end time was solved.The corresponding fixations in each frame were acquired based on the condition that $t_{start} \leq nT \leq t_{end}$.n is the frame number of the video.T is the duration time of one frame video.Secondly,it is the question that we visualized the eye tracking datas to receive weighted matrix of ROI area in each frame video.We established a zeros matrix of 1920×1080, which is similar to the original image.We conducted Gaussian flitering to the ROI map , then normalized the ROI weighted matrix to [0,1].The following formula is the Gaussian flitering function.

$$f(x,y,t)=(\alpha t+(1-\alpha))\exp(-\frac{x^2+y^2}{2\sigma^2}) \quad (7)$$

σ is the standard deviation of Gaussian function using simulating fovea centralis,it is 72 in this paper. α is the weight of duration time in the ROI weighted matrix,we took it 0.Fig.3 describes the ROI image of the original video.



(a) (b) (c) (d)

Fig.3. ROI image (a)Basketball ROI image (b)Leaf ROI image (c)Lion ROI image (d) Halftime music ROI image

V. EXPERIMENTAL RESULTS

We tested performance of the proposed objective model using the Pearson Linear Correlationm Coeficienr(PLCC) in the SPSS.

Table I shows the performance of proposed blur model in the terms of the PLCC for each original video and each ROI video.

TABLE I.

Videos	Original video(PLCC)	ROI video(PLCC)
Basketball	-0.967	-0.947
Leaf	-0.984	-0.982
Lion	-0.928	-0.825
Halftime	-0.983	-0.874
Overall	-0.685	-0.805

Table II shows the performance of proposed block model in the terms of the PLCC for each original video and each ROI video.

TABLE II.

Videos	Original video(PLCC)	ROI video(PLCC)
Basketball	-0.962	-0.938
Leaf	-0.976	-0.990
Lion	-0.952	-0.763
Halftime	-0.995	-0.921
Overall	-0.624	-0.727

Table III shows the average running time of whole original videos and whole ROI videos in the terms of the blur algorithm.

TABLE III.

	Original video	ROI video
Average time(min)	15.32	9.43
improved	38.5%	

Table IV shows the average running time of whole original videos and whole ROI videos in the terms of the block algorithm.

TABLE IV.

	Original video	ROI video
Average time(min)	8.13	7.94
improved	2.3%	

VI. CONCLUSIONS

A novel NR VQA algorithm based on eye tracking datas is proposed in this paper.The objective video quality is measured considering the impacts of two key artifacts during the compression process:blur and blocking.The experimental results demonstrate that the proposed algorithm shows not only higher correlation,but lower computation complexity.

As part of the experiment,we recorded eye tracking datas of every frame in each video as the subjects viewed the videos.On the one hand,we intended to make use of the datas to form ROI videos which were computed by the blur and blocking model later.Because the performance of the no-reference VQA is far less mature than full-reference VQA, the goal to improve subjective and objective correlation and reduce calculation complexity is mainly accomplished using eye tracking datas in this paper.On the other hand,we will intend to develop objective ROI model in the future and the subjective eye tracking experiment will be served as a verification.

ACKNOWLEDGMENT

We thank the National Program on Key Basic Research Project (973 Program) project (2010CB327705), the National High-tech R&D Program of China (863 Program) (2012AA03A302) and the Ph.D Program Foundation of Ministry of Education of China(20120092120024).

REFERENCE

[1]. Xiangyu Lin,Hanjie Ma,Lei Luo and Yaowu Chen, "No-reference Video Quality Assessment in the compressed Domain".IEEE pp505-511,Jun,2012

[2]. PinaMarziliano,FredericDufaux,Stefan Winkler and Touradj Ebrahimi, "A no-reference perceptual blur metric".IEEE ppIII-57-60,2002

[3]. Hantao Liu and Ingrid Heynderickx, "A no-reference perceptual blockiness metric".IEEE,2008.

[4]. Nabil Ouerhani,Roman von Wartburg,Heinz Hugli and Rene Muri, "Empirical Validation of the Saliency-based Model of Visual Attention."Electronic Letters on Computer Vision and Image Analysis 3(1):13-24,2004

[5]. ITU-T P. 910. "Subjective Video Quality Assessment Methods for Multimedia Applications"[S]. 2000.

[6]. LIVE Video Quality Database ,2009 [Online].Available:http://live.ece.utexas.edu/research/quality/live_video.html

[7]. Zhou Wang,Alan Conrad Bovik,Hamid Rahim Sheikh and Eero P.Simoncelli, "Image Quality Assessment:From Error Visibility to Structural Similarity."IEEE transactions on image processing,vol.13,NO.4,April 2004.

[8]. S.Winkler,"A perceptual distortion metric for digital color video",Proc.SPIE HumanVision Electron.Imag.1999,pp175-184

[9]. E.Ong,X.Yang,W.Lin,Z.Lu,and S.Yao, "Perceptual quality metric for compressed videos,"inProc,IEEE Int.Conf.Acoust.,Speech Signal Process.,vol.2.Mar.2005,pp581-584

[10]. Z.Wang,L.Lu,and A.C.Bovik, "Video quality assessment based on structural distortion measurement," Signal Processing :Image Communication,vol. 19,no.2,pp.121-132,Feb.2004

[11]. K.Seshadrinathan and A.C.Bovik, " Motion-based Perceptual Quality Assessment of Video,"Proc.SPIE – Human Vision and Electronic Imaging,2009

[12]. I.P.Gunawan and M.Ghanbari, "Reduced-Reference Video Quality Assessment Using Discriminative Local Harmonic Strength With Motion Consideration,"IEEE Trans.Circuits Syst.Video Technol,vol.18,no.1, pp.71-83,Jan.2008

[13]. I.P.GunawanandM.Ghanbari, "Efficient Reduced-Reference Video Qualitymeter,"IEEETrans.Broadcasting,vol.54,no.3,pp.669-679, Sep.2008

[14]. Z.Wang,G.Wu,H.Sheikh,E.simoncelli,E.-H.Yang,and A.Bovik, "Quality-awareimages".IEEETrans.Image Process.,vol.15,no.6, pp.1680-1689,Jun.2006

[15]. T.Oelbaum and K.Diepold, "A reduced reference video quality metric for AVC/H.264," in Proc.EUSIPCO,Sep.2007,pp.1265-1269

[16]. H.Liu,N.Klomp,and I.Heynderickx, "A No-Reference Metric for Perceived Ringing Artifacts in Images"IEEE Trans.Circuits Syst.Video Technol.vol.20,no.4,pp.529-539, Apr.2010

[17]. M.Naccari,M.Tagliasacchi,and S.Tubaro, "No-Reference Video QualityMontoringforH.264/AVC CodedVideo,"IEEETrans. Multimedia,vol.11,no.5,pp.932-946, Aug. 2009

[18]. A.Eden, "No-Reference Estimation of the coding PSNR for H.264-CodedSequences,"IEEETrans.Consumer Electron,vol.53,no.2, pp.667-674,Aug.2007

[19]. T.Brandao and M.P.Queluz, "No-Reference PSNR estimation algorithm for H.264 encoded video sequences," in Proc.EUSIPCO,Aug.2008

[20]. P.L.Correiaand F.Pereira, "Objective Evaluation of Video Segmentation Quality"IEEE Transactions on Image Processing,vol.12,no.2,pp.186-200,Feb.2003