Hybrid No-Reference Video Quality Assessment Focusing on Codec Effects

Xingang Liu[1, 2], Min Chen[3*], Tang Wan[4] and Chen Yu[5]
[1]School of Electronic Engineering, University of Electronic Science and Technology of China, China
[2]School of Electrical and Electronic Engineering, Yonsei University, South Korea
[e-mail: hanksliu.xg@gmail.com]
[3]School of Computer Science and Engineering, Seoul National University, South Korea
[email: minchen@ieee.org]
[4]College of Computer Science, South-Central University for Nationalities
[email: tangwan@scuec.edu.cn]
[5]School of Computer Science and Technology, Huazhong University of Science and Technology, China
[email: yuchen@hust.edu.cn]
[*]Corresponding author: Min Chen

Abstract

Currently, the development of multimedia communication has progressed so rapidly that the video program service has become a requirement for ordinary customers. The quality of experience (QoE) for the visual signal is of the fundamental importance for numerous image and video processing applications, where the goal of video quality assessment (VQA) is to automatically measure the quality of the visual signal in agreement with the human judgment of the video quality. Considering the codec effect to the video quality, in this paper an efficient non-reference (NR) VQA algorithm is proposed which estimates the video quality (VQ) only by utilizing the distorted video signal at the destination. The VQA feature vectors (FVs) which have high relationships with the subjective quality of the distorted video are investigated, and a hybrid NR VQA (HNRVQA) function is established by considering the multiple FVs. The simulation results, testing on the SDTV programming provided by VCEG Phase I, show that the proposed algorithm can represent the VQ accurately, and it can be used to replace the subjective VQA to measure the quality of the video signal automatically at the destinations.

1. Introduction

With the development of high-speed communication systems, video programs offered through network services like video chatting, IPTV, video-on-demand (VoD) etc, have become more popular [1]. However, the quality of programs must be of sufficient quality that customers can enjoy their communication or entertainment time. Therefore, VQA techniques have become a principle requirement for the testing of video programming services. The most direct way of VQA for digital video signals is the subjective VQA, meaning that VQ is measured by Human Visual System (HVS) and given a quality score as the mean opinion score (MOS). The MOS for each video sequence is generated by averaging the results of a set of standard, subjective tests whereby a number of appraisers, each with professional QoE for VQ judgment, rates the quality on a five point scale ranging from 1 (bad) to 5 (excellent). The alternate form of the MOS is the Difference MOS (DMOS), which has the same generation principle as MOS but measures the difference between the original and distorted video frames. MOS and DMOS can reflect the quality of the video signal correctly because they are obtained by directly using HVS. However, the subjective VQA is complex and not always available. People would prefer to use the objective QoE which abstracts useful information only from the video frame to establish the VQA functions.

The NR VQA is a relatively new topic in the related research fields. In fact, it is very difficult to design an objective NR VQA function because of our limited knowledge of the HVS and the cognitive aspects of the brain. However, the practicability is significant since it can be established only by using the distorted video frames at the destination as shown in **Fig. 1**. To realize NR VQA, the research is active, and numerous algorithms have been proposed in [2-7] which have shown their contributions to the research of QoE. However, their collective shortcoming is that any proposed algorithm only focuses on the unilateral impact and can measure the VQ corrected only under special conditions rather than pervasively. For instance, [2] measures VQ by detecting the impact of the network service; however, the codec effects are ignored. M. Tao et al considers the multiple impact of the quality loss in [3], but it is suitable only for low bit rate network transmission. Algorithms in [4][5][6][7] deal with the quality measurement, but only with partition of the impacts. To help realize a ubiquitous application, in this paper our research focuses on the objective NR VQA function establishment by mainly considering the possible codec effects. Firstly, the FVs which have high relationships with the subjective DMOS are analyzed, and then, the HNRVQA algorithm which can be used to measure the quality of video signals is investigated by using the selected FVs.

This paper is organized as follows. Related work is described in section 2 briefly to introduce the basic knowledge and current situation concerning VQA research. In section 3, the characteristics of selected FVs are analyzed and utilized to establish the proposed HNRVQA algorithm which can correctly reflect the perceptual VQ. To evaluate the proposed algorithm, section 4 presents the simulation results, and the conclusions are drawn in section 5.

2. Related Works

In past decades, the research on objective VQA was the subject of much attention. Many methods have been investigated and are divided into three categories, as follows.

- Full reference (FR): assess the VQ by making a comparison between the original and distorted video sequences;

- Reduced reference (RR): not all but some parameters of the original and distorted video sequences are compared to establish the VQA function;
- Non-Reference-Free (NR): Only the information of the distorted video sequence is utilized in the VQA function.
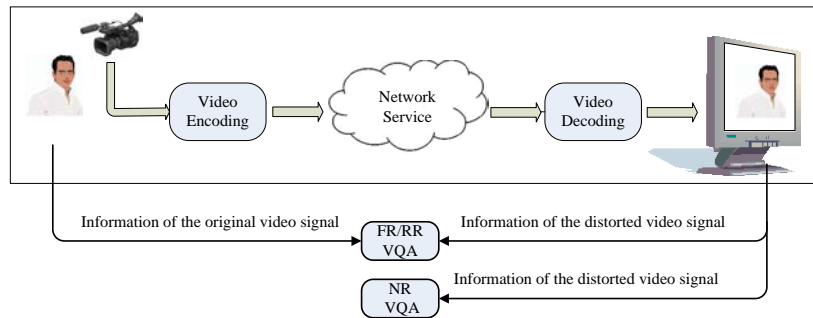

Fig. 1. Structures of the FR/RR and NR VQA

Among the three categories, FR VQA has been studied extensively, and numerous algorithms have been proposed. In general, the FR VQA method is based merely on the pixel difference between the original and distorted video frames, e.g., mean square error (MSE), peak signal-to-noise ratio (PSNR) [8] and their variations, such as Minkowsky metric (MM) [9], Czenakowski distance (CZD) [10] and so on. The VQA research community realizes the importance of validating the performance of algorithms by using extensive ground truth data. ITU-T Standard J.144 [11] provides four famous FR VQA algorithms which can serve for both PAL and NTSC. In [12], a novel image quality assessment algorithm named as MSSIM, calculates the structural similarity between the original and distorted video signals. However, as shown in **Fig. 1**, the holistic disadvantage of considering all the FR VQA algorithms is that both the original and the distorted video signals are needed to establish the FR VQA function. For RR VQA, part of the original video signal is still needed. It implies that additional bandwidth of the channel is required to load the transmission redundancy of the original signal from the server to the destination. But the fact is that neither the full nor part of the original video frame information is available at the destination in the multimedia communication systems.

The appearance of NR VQA solves the problem of the requirement to the original video information. The VQA can be established only by using the received video signal at the destination. However, human visual sense is controlled by the brain which is a very complex and mysterious system. In general, the eyes scan the environment items with 500o/sec and transmit the information to the brain [13]. From a physics perspective, the surface of the outside world should be fuzzy because rapid eye movement causes object images to pass the surface of the retina quickly. However, we do not see the world as flickering: objects are still clearly visible. As shown in **Fig. 2**, the human brain skillfully handles a pair of visual contradictions in physics and biology.

Since actual brain mechanisms are not completely understood currently, it is very hard to establish an accurate objective NR VQA function. Fortunately, the realized practice is significant and many related studies promoted the research on NR VQA in different areas. In [2], the authors described an NR VQA algorithm which can be used as a quality of service (QoS) monitoring strategy in order to control the end-user perceived quality. This algorithm considers several feature factors, such as frequency content features, power of frame

difference and blocking effect, and measures them independently to make up the final NR VQA function. The performance is quite well under the special simulation environments such as the bit rate limitation, video sequence selection and so on. However, it does not fully consider the feature factors caused by the video compression codec. Moreover, the computing load is a little high which will take additional burden to the terminals.
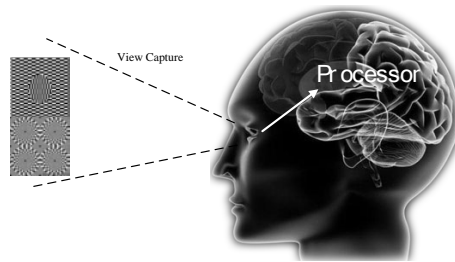

Fig. 2. The brain visual reorganization system

More recently, M. Tao et al proposed a home video visual quality assessment method in [3] where a set of spatiotemporal visual artifacts are mined from each sub-shot based on the particular characteristics of the home video signal. The factors, such as unstableness, infidelity, brightness, orientation and so on, which are frequently caused during the transmissions, are considered in detail. Because of the thoughtful proposal, this algorithm can measure the quality of the home monitor signal, video conference signal, very well. However, the obvious limitation blocks its applications because it is typically suitable for the VQA of the low bit rate video transmissions.

More NR VQA algorithms were proposed which attempted to evaluate the visual quality in terms of different aspects caused by the codec and transmissions frequently. For instance, [4][5][6][7] measure the VQ of the video signals which are distorted by blur, jitter and blocking effects, respectively. All of them did the significant contributions to the research of the NR VQA but anyone just focused on one or two kinds of the distortions. Therefore, not all factors affecting the perceived quality are considered.

It is the fact that any transmitted video sequence should be encoded in to bit stream at the server for compression [14]. Although various video coding standards exist, for example, H.264/AVC, MPEG-2, etc, the effect to the video quality by the lossy codec cannot be avoided due to codec's characteristic and the coding requirements. Therefore, the codec effect has become a major factor which can affect the video quality during the communications. According to the current situations of the NR VQA, our study in this paper focuses on the VQA establishment with the analysis and measurement of the FVs caused by the characteristics of the general video compression codec. The VQA function is created for each FV to establish the relationship with the subjective DMOS. The functions are integrated by using multiple regression technique [15] to generate the final NR VQA function which can achieve the best correlation with the DMOS.

3. Proposed HNRVQA Algorithm

To take place the research more efficiently, a special set of subjective VQA experiments is firstly implemented to find the FVs which can affect human judgment regarding the quality of the received video in theory. Fifty undergraduate students with experience in subjective VQA participated in our experiments. The SDTV video programs provided by VQEG [16] which

have multiple characteristics of codec effects were selected as the test video sequences. The students were ordered to view the SDTV sequences to select and mark the impacts which could affect their judgment to the VQ. According to the experimental results, several FVs, including blur, blocking, jitter/jerkiness and color were selected as the important FVs which can affect human judgment of the VQ, as shown in the highlighted "factor part" in **Fig. 3**.
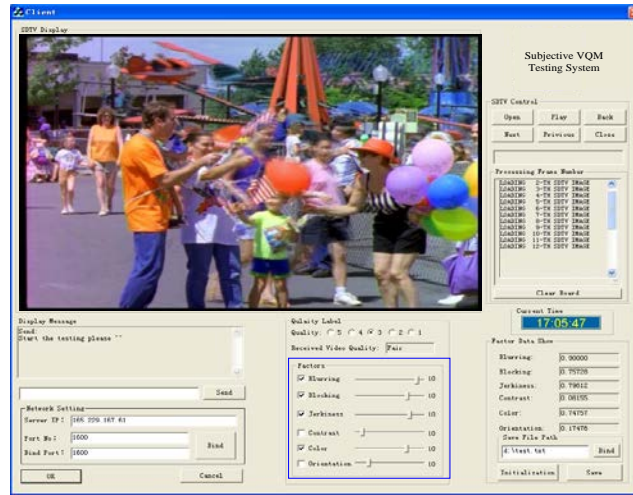

Fig. 3. The demo used in the subjective VQA experiments

3.1 Feature Vectors Detection and Analysis

3.1.1 Blur Effect Detection

According to the general video coding standard, the original video signal is transformed into the frequency domain to distinguish the information into different frequency regions [17]. It is the fact that high frequency region includes little useful information in general case. Therefore, it is more likely to be cut than the information in the low frequency domain to achieve higher compression ratio while maintaining the original video quality. However, the lossy compression generally causes many video artifacts, and blur is one of them. It is the global distortion of the entire video frame caused by the removal of high-frequency content from the original video signal. The primary characteristic of blur is the reduced sharpness of edges and limited spatial detail. There are different types of blur effects; for example, motion blur is due to the relative motion between the camera and the scene, and out-of-focus blur is caused by an unfocused camera and lens aberrations. In addition, use of an edge-attenuating filter, overlapped block motion compensation and packet loss can also result in significant blur. **Fig.4-(b)** shows the effect of dynamic blur which is caused by movement.
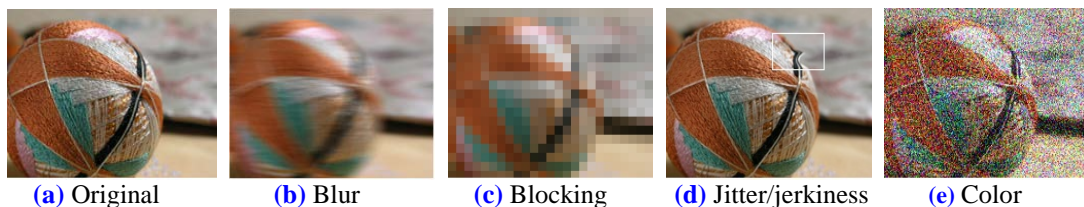

(a) Original   **(b)** Blur   **(c)** Blocking   **(d)** Jitter/jerkiness   **(e)** Color

Fig. 4. Impact examples of the distorted video frame

According to the QoE of the HVS, people are more sensitive to the edge of the video frame. The effect of blur can be scaled by detecting the edge information of the distorted video signal. In this paper, the famous image edge detector, the Sobel filter [18], is employed because of its strong ability on image edge detection. With the data statistical analysis, the blur effect occurs both in the horizontal and vertical directions along the frame edges. Filtering operations should be implemented separately (rather than the entire two-dimension implementation) to guarantee QA accuracy. **Fig. 5** gives this operation where the Sobel filter in horizontal direction is implemented to the luminance (luma) coefficients of the distorted video fame to calculate the gradient of the image intensity at each point and give the direction of the largest possible increase from light to dark and the rate of change in that direction. If the filtered pixel value is larger than a pre-determined threshold (T), this filtered pixel is set to be the horizontal edge point.
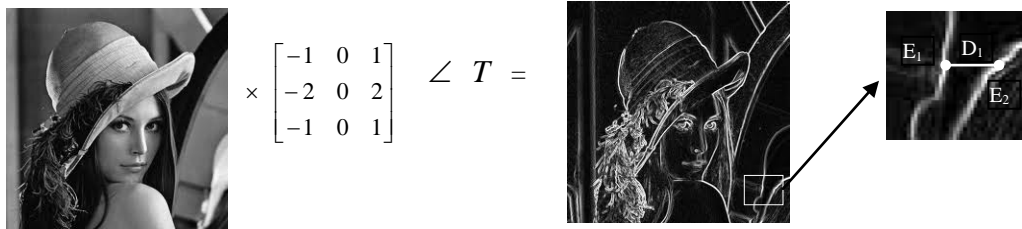

Fig. 5. Edged image generation by horizontal Sobel filter

Since blur will affect the image smoothness directly, the VQA function for blur is to measure the spread of the edges. As shown on the right of **Fig. 5**, the distance D_1 from one edge point E_1 to the next one E_2 along the horizontal direction is set to be the edge width of E_1. Deucedly, the widths of all the edge points are calculated and the average value of the widths A_{hor_spread} shown in eq. (1) is set as the image edge spread for the horizontal direction [5], where N in eq. (1) is the number of the total edge points.

$$A_{hor\_spread} = \sum_{n=0}^{N-1} D_n \tag{1}$$

By using the similar operations, the spread of edges for vertical direction (A_{ver_spread}) can be detected easily. And the entire blur detection VQA_{blur_luma}, where the subscript 'luma' means for 'luminance', could be established by using A_{hor_spread} and A_{ver_spread} as shown in eq. (2), where w_{hor_spread} and w_{ver_spread} are the weighted indices for the horizontal and vertical edge spreads, respectively. During the practical simulations, we found that the blur along the vertical direction is grander to the quality judgment than that along the horizontal direction. Therefore, the value of w_{ver_spread} is much larger than that of w_{hor_spread}.

$$VQA_{blur\_luma} = w_{hor\_spread} \times A_{hor\_spread} + w_{ver\_spread} \times A_{ver\_spread} \tag{2}$$

$$where \ w_{hor\_spread} + w_{ver\_spread} = 1$$

3.1.2 Blocking Effect Detection

To reduce the data load for the bandwidth-limited network services, the video signal should be compressed by video coding standards before the transmissions [19]. Although a variety of coding standards exists, the block-based transform operation is available in almost all codec systems. According to **Fig. 4-(c)**, the blocking artifact always happens because of the codec operation which can affect the judgment of HVS to the VQ. Although a variety of coding

systems exist, the block-based operation is available in almost all codec systems. The 2-D DCT codec segments the video frame into blocks, transforms them into coefficients and implements the quantization which is a lossy compression operation. At the destination, the output of certain decoded blocks makes the surrounding pixels appear averaged together and gives them the appearance of larger blocks. As TV screens become larger, the blocking artifact becomes more noticeable. Although the blocking effect always happens along the boundaries of the blocks, the artifact will extend to the inner block since the pixel values have been averaged by the lossy compression codec. As a result, the blocking impact can be scaled by using the following two steps.

3.1.2.1 Boundary Smoothness Detection between 8x8 Blocks

The conventional video signal encoding is based on the 8x8-block operation which is popularly used for multimedia signal compression. Therefore, blocking artifact detection focuses on the boundaries of 8x8 blocks and can significantly affect HVS VQA. Generally speaking, except for sudden scene changes, the difference between neighboring pixels is not distinct even if these two pixels belong to different 8x8 blocks. It implies that the boundary between blocks should be smooth enough in the original video frame to satisfy the video content scene. However, when blocking happens, the boundary of the blocks becomes so clear that the original video frame is seriously distorted due to the block average operation during the video signal compression. Therefore, detecting the boundary smoothness between each pair of 8x8 blocks is necessary.
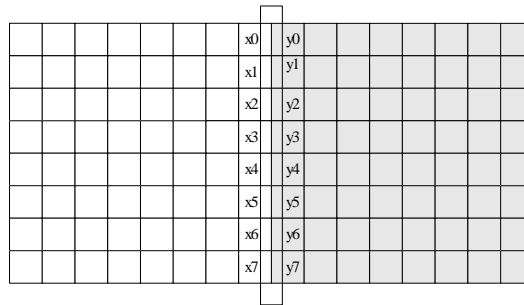

Fig. 6. Boundary smoothness detection between the blocks (horizontal direction)

Referring to **Fig. 6**, the absolute boundary pixel difference between two 8x8 blocks in the horizontal direction is detected at first, and the average value is set as the smoothness detection result ($S_{hor\_boundary}$) of the horizontal blocking artifact as shown in Eq. (3), where i is the boundary pixel index, $i = 0, 1, \ldots, 7$, and N is the number of boundary pixel, $N = 8$.

$$S_{hor\_boundary} = (\sum_{i=0}^{N} |x_i - y_i|)/N \tag{3}$$

In our study, boundary smoothness for the horizontal and vertical directions is detected separately, and the blocking artifact detection $S_{ver\_boundary}$ in the vertical direction can be found by using the similar operation. Following this, the two detected values can be utilized to establish the entire boundary smoothness detection resulting from the blocking ($S_{blocking}$) as shown by Eq. (4), where $w_{hor\_boundary}$ and $w_{ver\_boundary}$ are the weighted indices for horizontal and vertical directions, respectively.

$$S_{blocking} = w_{hor\_boundary} \times S_{hor\_boundary} + w_{ver\_boundary} \times S_{ver\_boundary} \tag{4}$$

$$where \ w_{hor\_boundary} + w_{ver\_boundary} = 1$$

3.1.2.2 Block Visibility Detection

The visibility of a block edge is determined by the contrast between the local gradient and the average gradient of the adjacent pixels. Let's consider a frame I with elements $I(i,j)$, where i and j denote the line and pixel position, respectively. To express the similarity between the local gradient and its spatial neighbors, a normalized horizontal gradient ($G_{hor\_visibility}$) is introduced as the ratio of the absolute gradient and the average gradient calculated over N adjacent pixels to the left and to the right [7]. Because block edges occur at regular intervals in the horizontal plane, they can be further highlighted by totaling the $G_{hor\_visibility}$ over all image lines and defining them as $S_{hor\_visibility}$. The visibility strength of the blocking artifacts for the horizontal direction ($V_{hor\_blocking}$) can be determined by averaging $S_{hor\_visibility}$ over the block edge and intermediate positions as follows:

$$V_{hor\_blocking} = \frac{S_{hor\_visibility}(block)}{S_{hor\_visibility}(no-block)} \tag{5}$$

Where $S_{hor\_visibility}(block)$ and $S_{hor\_visibility}(non\text{-}block)$ denote the average values of $S_{hor\_visibility}$ at the block edged and intermediate positions, respectively. The block size in the video signal corresponds to 8 pixels times the scaling factor in both horizontal and vertical directions, therefore, the identification of vertical blocking artifacts ($V_{ver\_blocking}$) is accomplished in a similar fashion. Subsequently, the entire blocking visibility detection can be established be using $V_{hor\_blocking}$ and $V_{ver\_blocking}$ as in Eq. (6), where $w_{hor\_visiblity}$ and $w_{ver\_visibility}$ are the weighted indices for horizontal and vertical directions, respectively.

$$V_{blocking} = w_{hor\_visibility} \times V_{hor\_blocking} + w_{ver\_visibility} \times V_{ver\_blocking} \tag{6}$$
$$where \ w_{hor\_visibility} + w_{ver\_visibility} = 1$$

3.1.2.3 Entire Blocking Effect Detection

Following the detection of boundary smoothness and block visibility, these can be utilized as the foundations of the proposed blocking detection method. Eq. (7) gives the entire blocking detection algorithm by combining the boundary smoothness and blocking visibility detection, where w_{smooth} and $w_{visibility}$ are the weighted indices for block smoothness and visibility detection, respectively.

$$VQA_{blocking\_luma} = w_{smooth} \times S_{blocking} + w_{visibility} \times V_{blocking} \tag{7}$$
$$where \ w_{smooth} \times w_{visibility} = 1$$

3.1.3 Jitter/Jerkiness Effect Detection

In the video display system, the perception by HVS is that continuous motion is comprised of a sequence of distinct "*snapshots*". However, the perception of continuous motion by human visual faculties is a manifestation of complex functions, i.e. characteristics of the eyes and brain. When presented with a sequence of fixed, still images of sufficient continuity at a sufficiently frequent update rate, the brain interpolates the intermediate images, and the observer subjectively appears to see the continuous motion that in reality does not exist. Actually, a very important source of video impairment comes from the transmission of a video stream over an error-prone channel, especially in a packet-switched network service where the data packets can be lost or delayed to the point where they are not received in time for

decoding. The visual impact of such losses varies between video decoders depending on their ability to deal with such corrupted streams.

However, some decoders will choose to entirely discard a frame that has corrupted or has information missing and repeat the previous video frame instead, until the next valid decoded frame is available. Not only is additional spatial degradation effectively introduced but frame repetition and frame drop also occur. This phenomenon is referred to as video jitter. Because of the error-prone network service, packet loss frequently happens; **Fig. 4-(d)** shows the effect of the packet loss after error concealment (highlighted region). The loss of video packets often results in the loss of slice-unit information, which in turn results in corruption of visual information along the macro block (MB) and slice edges. With the general error concealment method [20], the decoder replaces the lost packets by using the corresponding temporal MB from the previous frame. Therefore, visible discontinuity results when the MB data from regions with considerable motion between the consecutive frames is lost.

In this paper, we detect the jitter effect by using the codec operation in the decoder part. According to the conventional codec of multimedia communication, the MB with 16x16 pixel segment in the video frame is regarded as the basic coding unit. The MBs within a slice are ordered from left-to-right and top-to-bottom [4]. As shown in **Fig. 7**, each block denotes the 16x16 MB; the upper white blocks are the MBs belonging to the same slice and the bottom brown blocks are the MBs belonging to another slice. If data loss occurs during the transmission, the slice-based packet loss also occurs in general situation. For the VQA, energy strength among every slice boundary should be scaled. We define $P(i)$ as the ith row, where i is the last row of each MB, $i \in (16, 32 \ldots \text{m-16})$.
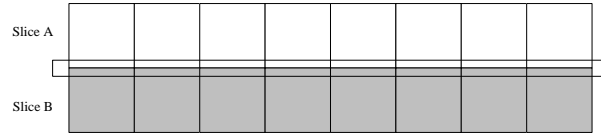

Slice A

Slice B

Fig. 7. Energy strength calculation between slices

$$S_{jitter}(j) = \left| (P(i-1) - P(i+1)) * F \right|$$ (8-1)

$$S'_{jitter}(j) = \left| (P(i-2) - P(i)) * F \right|$$ (8-2)

In Eq. (8), * denotes the convolution operation, and j is the boundary number which equals $i/16$. F is a low pass filter, $F = [1, 1, 1]/3$. The boundary strength between the MB j and $j+1$ is noted as $S_{jitter}(.)$. $S'_{jitter}(.)$ is also the boundary strength which is very close to $S_{jitter}(.)$. To further avoid boundary noise, a pre-threshold (T_{jitter}) is needed to pick up visible horizontal edges. $S_{jitter}(.)$ and $S'_{jitter}(.)$ are set to be 1 if they are larger than T_{jitter}. Otherwise, they are set to 0 which means the boundary strength is so slight that it can be ignored. With the statistical analysis, T_{jitter} is given a value of 72 for luminance and 36 for chrominance in our simulation. The reason for calculating the boundary strength $S_{jitter}(.)$ and also its neighboring boundary strength $S'_{jitter}(.)$ is to avoid eliminating the real boundary in the frame because they must have similar boundaries in a regular undistorted frame. Hence, the effect of the packet loss artifact along the MB row j is to compute the difference between $S_{jitter}(.)$ and $S'_{jitter}(.)$. The jitter effect for the NR VQA is established as follows.

$$VQA_{jitter\_luma} = \left| \sum_{j} S_{jitter}(j) - S'_{jitter}(j) \right|$$ (9)

3.1.4 Color Effect Detection

Color is also an important factor affecting people's subjective quality judgment. HVS is very sensitive to the video frame which has extra color. In **Fig. 4-(e)**, data loss and additional noise caused by the network affect the original video color so that the VQ is obviously reduced. Most existing VQA algorithms abstract some special information from the distorted video frames to build the color detection factors. In our proposed algorithm, the other thinking is followed to detect the color effect to the VQA.

In general video frames, the trends of luminance and chrominance are very similar [21]. Therefore, the visual effects on luminance also affect chrominance (color). Therefore, in this paper the operations of blurring, blocking and jerkiness can be repeated to the chrominance coefficients to detect their effects as the color factor.

Until now, the entire VQA of each FV could be established by using these effects to determine luminance and chrominance coefficients. For instance, Eq. (10) gives the entire VQA function for blur, where w_{blur_luma} and w_{blur_chroma} are the weighted factors for the luminance and chrominance coefficients, respectively.

$$VQA_{blur} = \frac{(w_{blur\_luma} \times VQA_{blur\_luma} + w_{blur\_chroma} \times VQA_{blur\_chroma})}{w_{blur\_luma} + w_{blur\_chroma}} \tag{10}$$

3.2 Proposed HNRVQA Algorithm

Fig. 8 shows a flowchart of the proposed HNRVQA algorithm and can be described as follows.
- Step 1: detect the VQA function of each FV (blur, blocking, and jitter) for luminance.
- Step 2: detect the VQA function of each FV (blur, blocking, and jitter) for chrominance.
- Step 3: Combine the detected VQA functions of luminance and chrominance with eq. (10) to establish the entire VQA function for each FV.
- Step 4: Combine the VQAs of the three FVs by using the multiple regression method [22] to establish the final HNRVQA function.

With the principles of the proposed algorithm, the HNRVQA algorithm is established using the selected FVs. The final function of the proposed method is shown by Eq. (11), where α, β and γ are the weighting coefficients for the different FVs, and λ is the offset constant.

$$HNRVQA = \alpha \times VQA_{blur} + \beta \times VQA_{blocking} + \gamma \times VQA_{jitter} + \lambda \tag{11}$$

It should be noted that the multiple regression is utilized here to generate the weights for the FVs. The general purpose of multiple regressions is to learn more about the relationship between several independent or predictor variables and a dependent or criterion variable [15]. In this paper, DMOS is regarded as the criterion variable, and the HNRVQA which is calculated by the FVs with their multiple weights is the independent variable. To find the best relationship between DMOS and HNRVQA, the linear regression is implemented to get the weights (α, β and γ) which achieve the maximum correlation between DMOS and HNRVQA.
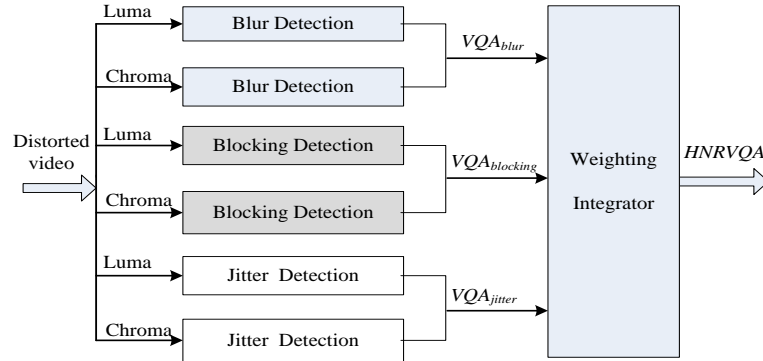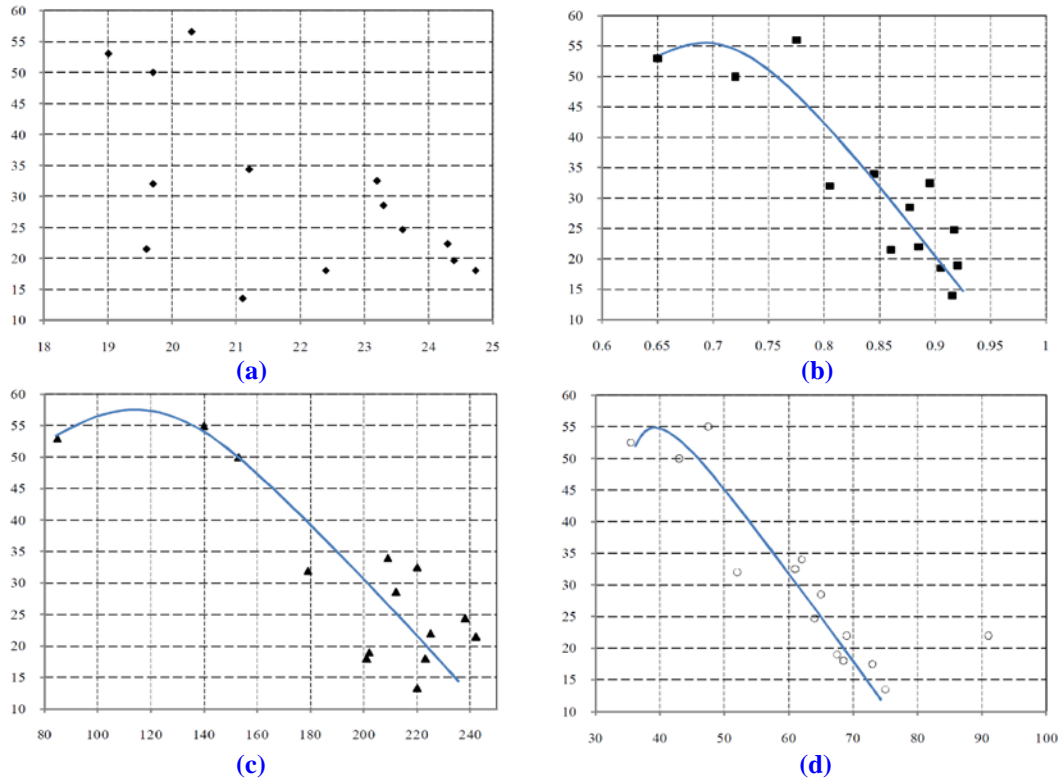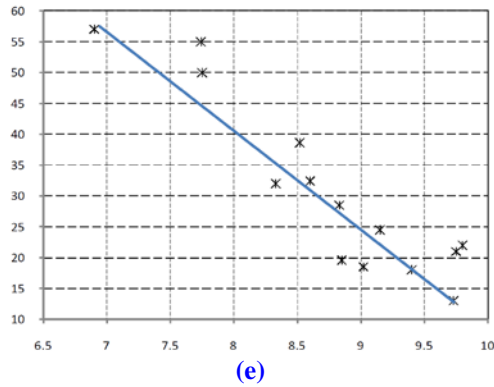
Fig. 8. Proposed HNRVQA algorithm

4. Performance Evaluation and Analysis

To confirm our analysis and the proposed HNRVQA algorithm, the existing VQA algorithms are studied and the performances are compared with that of the proposed HNRVQA algorithm. For universality, the public video sequences databases, the SDTV sequences (size: 720x486, format: 4:2:2) of VQEG Phase I which were given the subjective quality score (DMOS) are utilized in our simulations, and the results of the 14th and 15th SDTV sequences named as SRC14 and SRC15 are selected to be recorded. To evaluate the performance, several existing FR, NR and our proposed HNRVQA algorithms are implemented.

(a) PSNR vs. DMOS

(b) MSSIM vs. DMOS

(c) Ref [4] vs. DMOS

(d) Ref [5] vs. DMOS

(e) Proposed HNRVQA vs. DMOS

(e)

Fig. 9. Comparisons of scatter grams of VQA algorithms

Table 1. Correlation Summary (VQA Algorithm vs. DMOS)

Video sequence	Method	Correlation
VQEG SRC14	PSNR	0.5621
	SSIM	0.6648
	Ref[4]	0.8023
	Ref[5]	0.7658
	HNRVQA	0.8907
VQEG SRC15	PSNR	0.6081
	SSIM	0.8806
	Ref[4]	0.8569
	Ref[5]	0.8382
	HNRVQA	0.9002
Average	PSNR	0.5851
	SSIM	0.7727
	Ref[4]	0.8296
	Ref[5]	0.802
	HNRVQA	0.89545

According to **Fig. 9** and **Table 1**, as the well-known VQA algorithm, PSNR shows its strong ability for image/video quality measurement. However, its limitations determine that it cannot give perfect VQA results which can correctly mesh with the subjective perceptual quality scores because of the wide impacts from the compression codec. Although MSSIM has its own advantage, unfortunately it is still not a good QA method for evaluating the codec effect to the VQ since it was designed for still image QA and used for the various characteristics of moving video signal. In addition, the reference video signal should be available during the FR VQA processing which is generally impractical. References [5] and [6] serve NR VQA algorithms which only consider unilateral effects on VQ distortions. However, they can not assess the VQ accurately for the video signals impacted by multiple coefficients and therefore cannot deal with real-life situations. With the simulation results, we have determined that the proposed HNRVQA algorithm almost demonstrates a linear relationship with the DMOS and always gives the best correlation results because it is able to forecast the practical situations of the multimedia communication systems.

5. Conclusions

With the development of multimedia communications, video programs have become essential to personal entertainment. Therefore, the QoE of the received video signal is very important to people's quality of life. Actually, so many parameters can affect the judgment to the VQ because of the various environments and the un-well-known human perceptual system. Among them, the codec effect cannot be avoided in current multimedia communication systems since any video signal should be compressed by the video coding standards for transmission. In this paper, our study focuses on the VQA establishment by considering the video compression codec effects. We analyzed the FVs which have relationships with the judgment of video quality and established the VQA function for each of them. Based on the essentiality and generation of the selected FVs, the final objective NR VQA algorithm for quality detection was constructed. According to the simulation results, the proposed HNRVQA algorithm is able to measure SDTV quality correctly and has a higher correlation with the subjective VQ score than the existing FR and NR VQA algorithms.

Acknowledgement

References

[1] L. Zhou, X. Wang, W. Tu, G. Mutean and B. Geller, "Distributed Scheduling Scheme for Video Streaming over Multi-Channel Multi-Radio Multi-Hop Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 3, pp. 409-419, Apr. 2010. Article (CrossRef Link)

[2] P. Lecallet, C. Viard-Gaudin, S. Pechard and E. Caillault, "No reference and reduced reference video quality metrics for end to end monitoring," *IEICE Transactions on Communication*, vol. E89-B, no. 2, pp. 289-296, Feb. 2006. Article (CrossRef Link)

[3] M. Tao, X. Hua, C. Zhu, H. Zhou and S. Li, "Home video visual quality assessment with spatiotemporal factors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 6, pp. 699-706, June 2007. Article (CrossRef Link)

[4] Q. Huynh-Thu and M. Ghanbari, "Impact of jitter and jerkiness on perceived video quality," in *Proc. of International Workshop on Video Processing Quality Metrics*, Scottsdale, USA, Jan. 2006. Article (CrossRef Link)

[5] P. Marziliano, F. Dufaux, S. Winkler and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proc. of the International Conference on Image Processing*, vol. 3, pp. 57-60, 2002. Article (CrossRef Link)

[6] Z. Wang, H. Sheikh, and A. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. of IEEE International Conference on Image Processing*, pp. 477-480, September 2002. Article (CrossRef Link)

[7] R. Muijs and I. Kirenko, "A No-Reference Blocking Artifact Measure for Adaptive Video Processing," in *Proc. of the 13th European Signal Processing Conference*, Antalya, Turkey, September 2005. Article (CrossRef Link)

[8] D. Richard and X.D. Yang, "Existing and emerging image quality metrics," in *Proc. of Canadian Conference on Electrical and Computer Engineering*, pp. 1906-1913, May 2005. Article (CrossRef Link)

[9] J.S. Pan, F.R. Mclnnes and M.A. Jack, "Bound for Minkowski metric or quadratic metric applied to VQ codeword search," *IEE Proc. of Vision, Image and Signal Processing,* vol. 143, pp. 67-71, Feb. 1996. Article (CrossRef Link)

[10] I. Avcibas, B. Sankur and K. Sayood, "Statistical evaluation of image quality measures," *Journal of Electronic Imaging*, vol. 11, no. 2, pp. 206-23, April 2002. Article (CrossRef Link)

[11] "Objective perceptual video quality measurement techniques for digital broadcast television in the presence of a full reference, " *ITU-R Draft New Recommendation J.144*, 2003

[12] Z. Wang, A.C. Bovik, H.R. Sheikh and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, April 2004. Article (CrossRef Link)

[13] Y. Yang, P. Cao, Y. Yang and S. Wang, "Corollary discharge circuits for saccadic modulation of the pigeon visual system," *Nature Neuroscience*, vol. 11, pp. 595-602, 2008. Article (CrossRef Link)

[14] C. F. Lai, Y. M. Huang and H.C. Chao, "DLNA-based Multimedia Sharing System for OSGI Framework with Extension to P2P Network," *IEEE Systems Journal*, vol. 4, No. 2, pp 262-270, 2010. Article (CrossRef Link)

[15] L.S. Aiken and S.G. West, "Multiple Regression: Testing and Interpreting Interactions," *Sage*, Thousand Oaks, Ca, 1991.

[16] VQEG Phase I SDTV video sequences and DMOS. Article (CrossRef Link)

[17] L. Zhou, N. Xiong, L. Shu, A. Vasilakos and S.-S. Yeo, "Context-Aware Multimedia Service in Heterogeneous Networks," *IEEE Intelligent Systems*, vol. 25, no. 2, pp. 40-47, Mar. 2010. Article (CrossRef Link)

[18] K. Krajsek and R. Mester, "Signal and Noise adapted Filters for Differential Motion Estimation," *Lecture Note in Computer Science,* vol. 3663, pp. 476-484, Springer Berlin, Sep. 2005. Article (CrossRef Link)

[19] L. Zhou, B. Geller, B. Zheng, A. Wei and J. Cui, "System Scheduling for Multi-Description Video Streaming Over Wireless Multi-Hop Networks," *IEEE Transactions on Broadcasting*, vol. 55, no. 4, pp. 731-741, Dec. 2009. Article (CrossRef Link)

[20] J. Wu, X. Liu and K.Y. Yoo, "A temporal error concealment method for H.264/AVC using motion vector recovery," *IEEE Trans. on Consumer Electronics*, vol. 54, no. 4, pp. 1880-1885, Nov. 2008. Article (CrossRef Link)

[21] X. Liu, K. Yoo and S. Kim, "Low complexity Intra prediction algorithm for MPEG-2 to H.264/AVC transcoder," *IEEE Transaction on Consumer Electronics*, vol. 56, no. 2, pp. 987-994, May, 2010. Article (CrossRef Link)

[22] X. Liu, J. Wu, J. Liu and K.Y. Yoo, "Hybrid full reference video quality measurement for SDTV service", in *Proc. of the 2007 KSPC*, vol. 20, no. 1, Oct. 2007.

Xingang Liu is currently an associate professor in the School of Electronic Engineering (EE) at University of Electronic Science and Technology of China (UESTC). Meanwhile, he is a BK21 research fellow in the school of Electrical and Electronic Engineering in Yonsei University of South Korea. Dr. Liu has served as an organization committee, technical program committee and session chair of more than 20 IEEE International conferences, such as IPC'07, ICESS'08, HPCC'09, CSE'10 etc. And he is also a reviewer for many International journals, such as IET Communication, JVCI, IJCS and so on. He serves as a guest editor for several journals, for instance, EURASIP Journal on Wireless Communications and Networking, Springer Journal of Multimedia Tools and Applications. His is research interests are heterogeneous and homogenous video transcoding, quality of experience (QoE), error concealment, mode decision and 3D video codec. Dr. Liu is a member of IEEE, TIIS and KICS.

Min Chen is an assistant professor in the School of Computer Science and Engineering at Seoul National University (SNU). Before joining SNU, he was a post-doctoral fellow at the University of British Columbia (UBC) for three years. He received the Best Paper Runner-Up Award from QShine 2008. He serves as TPC Co-Chair and Web Chair for BodyNets 2010, and Workshop Co-Chair for CHINACOM 2010. He is Co-Chair of MMASN '09 and UBSN '10. He was TPC Chair of ASIT '09, TPC Co-Chair of PCSI '09 and PCSI '10, and Publicity Co-Chair of PICom '09. He has served as a guest editor for several journals, such as ACM/Springer Mobile Networks and Applications, International Journal of Communication Systems, International Journal of Sensor Networks, International Journal of Communication Networks and Distributed Systems, and International Journal of Autonomous and Adaptive Communications Systems. Dr. Chen is an IEEE Senior Member.
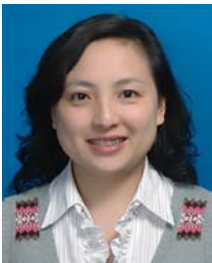
Tang Wan received the B.S. and M.S. degrees in Computer Application Technology from South-Central University for Nationalities (SCUN), Wuhan, China, in 1995 and 2001, respectively, and received the Ph.D. degree in Communication and Information System from Wuhan University, China in 2009. She is currently an associate professor in the School of Computer Science of SCUN. Also, from 2001 to 2002, she worked as a visiting researcher at the Advanced Communications and Networks Laboratory at Chonbuk National University, Jeonju, South Korea. Her research interests include protocols for optical/wireless communication networks, network security, and computational intelligence.

Chen Yu received his B.S. in Mathematics and M.S. in Computer Science from Wuhan University in 1998 and 2002 respectively. He received Ph.D. in Information Science from Tohoku University 2005. He had worked as a Japan Science and Technology Agency postdoctoral researcher in Japan Advanced Institute of Science and Technology from 2005 to 2006. From 2006, he had worked as a Japan Society for the Promotion of Science postdoctoral fellow in Japan Advanced Institute of Science and Technology to 2008. Since 2008, he has been with the School of Computer Science and Technology, Huazhong University of Science and Technology, China, where he is an associate professor and a special research fellow, working in the areas of Wireless Sensor Networks, Ubiquitous Computing and Green Communications. He has published more than 20 referred professional research papers and received the best paper award in ICC 2005 (the IEEE International Conference on Communication) and the nominated best paper award in DS-RT 2007 (the 11th IEEE International Symposium on Distributed Simulation and Real-Time Applications). He has served as the workshop co-chair for CIT 2009 (the 9th IEEE International Conference on Computer and Information Technology) and NPC 2010 (the 7th IFIP International Conference on Network and Parallel Computing) and PC co-chair for CUTE2010 (the 5th International Conference on Ubiquitous Technologies and Applications).