VIDEO QUALITY MONITORING OF STREAMED VIDEOS

Ee Ping Ong, Shiqian Wu, Mei Hwan Loke, Susanto Rahardja, Jason Tay, Cheng Kok Tan*, Lei Huang**

Institute For Infocomm Research, (A*STAR) Agency for Science Technology and Research, Singapore
*Rohde & Schwarz Systems & Communications Asia Pte Ltd

ABSTRACT

This paper describes a video quality analysis system for in-service monitoring of streamed videos, particularly over mobile/wireless networks. The algorithm adopts the no-reference method, and enables real-time measurement of video quality at any point in the content production and delivery chain using any given video. The technologies developed include no-reference methods for measuring picture freeze, picture loss, and blockiness. The developed system (where the software has not been optimized for speed) is able to process video of CIF size (352x288 pixels) at more than 30 fps on a Pentium-IV 3GHz computer. The experimental results show that the proposed video quality analysis system gives good accuracy for picture freeze, picture loss, and blocking detections.

Index terms – Image sequence analysis, image transmission, video codecs

1. INTRODUCTION

Although the literature contains large number of perceptual-based video quality metrics, these are actually full-reference approaches – i.e. where the original undistorted video sequences need to be present and used for the computation of the video quality measure [2, 11, 6, 13, 14, 9, 12] (including those in VQEG [7]).

With the migration of mobile services from 2G to 3G that is able to provide video streaming/transmission/reception and mobile video conferencing, there will be increasingly more applications and services of mobile visual signals, and therefore this creates the need for measuring/monitoring the quality of compressed videos, particularly by the content service providers during the creation of the digital content to be archived and streamed. Also, the use of no-reference technique for video quality measurement will become more and more important in the multimedia/broadcast industry because there is an eminent need and requirement for in-service analysis and quality measurement of streaming video where the original undistorted video are not available for comparison. Such in-service operation is necessary in order to ensure reliable maintenance of the perceived quality of video at various points in their creation, encoding, transmission / delivery, display, and storage. To maintain digital video quality requires measuring their quality quickly and reliably.

The literature on no-reference video quality measurements is much smaller and the emphasis is always on measurement of blockiness [13, 3, 4]. Blockiness is an important artifact which is usually created due to lossy compression. Due to limitation of computational resources and bandwidth, the video signal to be transmitted or stored is usually subjected to varying degrees of compression to decrease the bandwidth required for each video channel. The compression standards typically used are block transform-based techniques, such as JPEG, MPEG-1, MPEG-2, H.261 and H.263. In order to achieve low bit rates, quantization is normally used during encoding to compress the transform coefficients of the original image/video. The quantization process is lossy. As a result, the compressed image/video cannot be exactly reconstructed at the decoding side. The decompressed image/video exhibits various kinds of distortion artifacts such as blocking and blurring. In general, the human visual sensitivity to different types of artifacts is quite different. It is shown that the blocking effect and its propagation through reconstructed video sequences is one of the most significant of all coding artifacts.

To cater for situations where the original undistorted reference image may not be available, there were also some work on no-reference image quality measurements, which again is targeted at measurement of blockiness [5, 10, 1, 8]. All these techniques give a single measure to the blockiness, do not seem to perform well on images where blockiness is not the predominant distortion, and do not give information on area of video affected by presence of blockiness.

With the emergence of video services as a prominent component of mobile networks, video performance quality becomes increasingly important. Existing Quality of Service indicators for video quality, such as signal-to-noise or network statistics like packet loss rate and block error rate are not sufficient to measure the quality that the user perceives. There are basically 2 reasons:

1. Depending on which part of the bit stream is affected by errors or losses, the same amount of losses can thus have completely different effects on the decoded video content.
2. The human visual system is highly complex and processes information in a very adaptive and non-uniform fashion. This suggests that the visibility and annoyance of artifacts depend on the type of artifacts as well as the actual video content.

Currently, the testing requires a human operator to view a set of moving images and this leads to subjective measurement outcome, which tends to deteriorate in accuracy and consistency over time due to fatigue of the human operator.

Video quality measurement can be performed with or without reference to the original video image. With full reference, the video displayed is compared with the original. There are 2 important shortcomings for the reference method:

1. This approach is restricted to measurements of video quality at locations where both the reference video and the test video are readily available. This is difficult to realize in the case of live-content streaming where the content is encoded on the fly.
2. The need for aligning the test video with the reference video for a relative analysis will be very challenging. For this reason, the relative approach is limited to out-of-service testing.

2. PROPOSED VIDEO QUALITY ANALYZER

This paper proposed a Video Quality Analyzer (VQA) system which facilitates analysis of a video sequence on a mobile receiver device. The 3 artifacts that will be specifically measured are picture freeze, picture loss, and blocking.

2.1. Usage of Video Quality Analysis System

Figure 1 shows one of the possible scenario usage for our proposed video quality analysis system. The input video sequence used is initially passed to a source coder (such as a MPEG-2 video coder) to encode the video sequence. The output is then passed through the channel simulator to generate the wireless video signal for transmission and also inject channel transmission errors. The channel simulator acts as a channel transmitter in real life and also performs the task of injecting noise to simulate real life situation where noise will be added to the wireless video signal when they are being delivered through the channel. The wireless video signal is then received by a portable device and subsequently this video signal is analyzed by our proposed video quality analysis system to measure the video quality.

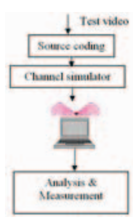

Figure 1: Usage of Video quality analysis system

2.2. Detecting Picture Freeze

Picture freeze is a temporal video artifact where there is a lack of motion between consecutive frames for a perceptually noticeable period of time. Often, the resultant video will appear to have a jerky effect. Picture freeze is caused by the process of video transmission when substantial parts of the video stream are lost. These missing data packets are replaced by the decoder by replacing them with the last good picture until they can resume playback.

The freeze artifact is defined to be a series of consecutive video frame sequences occurring with no visible change in content. The picture freeze detection algorithm makes use of the absolute differences between 2 consecutive frames. The algorithm computes a discriminate value that is representative of the amount of change between the current and previous frame. If the discriminate value is smaller than the *FreezeThreshold*, it indicates that there is not enough change in image content between the 2 frames to be visible to the eyes and the current video frame is marked as a possible picture freeze frame. The higher the *FreezeThreshold* value is, the lesser noise and motion is tolerated before the frame is detected to have a possible freeze artifact. When more than 5 consecutive frames are marked as possible picture freeze, it can be assumed that a freeze artifact has occurred in these series of frames.

2.3. Detecting Picture Loss

Loss artifacts affect both the temporal and spatial aspect of the video sequence. The output video sequence would appear to have a momentary flicker on the screen if the loss artifact is for a short duration. Otherwise, it appears to have a blank screen.

The picture loss artifact can be thought to be a sudden loss of data from the video screen. For example, Figure 2 shows a normal video frame and Figure 3 shows a partial picture loss video frame in one of the video sequences. Similar to the picture freeze artifact, the loss artifact makes use of the change in image content between the previous and the current frame.

In order for a loss artifact to be detected, two conditions must be present: (1) There must be a significantly large amount of content change between 2 consecutive frames. This is controlled by the change in mean gray value between consecutive frames and the parameter *LossThreshold*; (2) The pixels that experience change must become a low grey level values.

If the change in mean gray values is larger than the *LossThreshold*, it indicates that there is a large change in image content between the 2 frames. The pixels that have changed between the 2 frames are checked for low grey values and if they fulfill the condition, the current video frame is marked as a picture loss frame. The higher the *LossThreshold* value is, the more data that must be lost between the frames before it will check for possible picture loss.

Figure 2: Normal frame

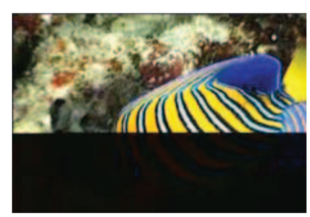

Figure 3: Half Loss Frame

2.4. Detecting Blockiness

Blockiness is a measure of block structure that is common to all DCT-based image and video compression techniques. Due to the regularity and extent of the resulting pattern, the blocking effect is easily noticeable. The blocking effect can be caused by lossy compression (affecting the whole video frame), as well as transmission errors (affecting a portion of the video frame).

Blocking artifacts often appear as sharp artificial edges marking out distinctively the outline of a small square block area in the video frame. Detection and measurement of blocking artifacts is a very challenge problem. Firstly, information about the encoder/decoder information for mobile video communication is not available, and accordingly, the quantization information that could be used to determine the blocking strength is unknown to us. Secondly, the amount of distortion is normally affected by both the video complexity and the bit rate, and hence must be determined using a combination of these two factors. Thirdly, in the case of video streaming and transmission through mobile/wireless networks, blocking effect can be caused by transmission errors and these blocking artifacts will appear in just a portion of the video frame (and not throughout the whole video frame as in the case of blocking artifacts generated due to lossy compression).

The blocking detection algorithm includes two parts: (1) blocking location detection and (2) blocking measurement. Details of these 2 stages will be described in more details in the following sub-sections.

2.4.1. Blocking Location Detection

Considering human visual system, the visibility of a block edge is determined by the contrast between the local gradient and the average gradient of the adjacent pixels. For example, Figure 4 shows one of the blocking frames in one of the video sequences. Note that parts of the video frame exhibits more significant blockiness but there is negligible blockiness on the body of the fish.

Figure 4: Blocking Frame

Therefore, we introduce the normalized gradient $D_{H,norm}$ (for horizontal direction) and $D_{V,norm}$ (for vertical direction):

$$D_{H,norm}(i,j) = \frac{|I(i+1,j) - I(i,j)|}{\sum_{k=-n}^{n} |I(i+k+1,j) - I(i+k,j)| / (2n+1)} \quad (1)$$

where n is the number of neighboring pixels considered, and $I(x,y)$ denotes the luminance of the video frame at pixel location x and y (in horizontal and vertical directions respectively). According to above equation, each frame is converted to horizontal / vertical gradient image.

If a frame has blockiness, the next step is to extract the blocking position. The blocking information is highlighted by summing $D_{H,norm}$ and $D_{V,norm}$.

$$S_H(i) = \sum_{j=1}^{N} D_{H,norm}(i,j) \qquad S_V(j) = \sum_{i=1}^{N} D_{V,norm}(i,j) \quad (2)$$

Figure 5(a) demonstrates the S_H which is larger than their average values. From Figure 5(a), it can be observed that it is still not easy to find the blocking locations or localization of blocking boundaries by setting a threshold because the threshold is frame-dependent. To alleviate the random choice of threshold, a new idea to determine the blocking positions or boundaries (BB_H and BB_V) was proposed based on the following assumptions:

1) The blocking boundaries correspond to positive peak values of S_H, S_V.

2) The distance of blocking boundaries should be around BS. (Note that BS is the blocking size in the spatial domain and BS = 256 / BF, where BF is equal to eight (assuming that the blocking occurs at an interval of 8 pixels, similar assumption to other no-reference blocking measurement techniques; We can extend this technique where F may not be 8 in the case of test videos captured externally from the screen of portable mobile display devices)).

3) For the positive peaks whose distance is small, say, less than BS/2, then, the largest one corresponds to potential location of blockiness.

After determining the most confident peaks to be regarded as blocking boundaries, it is straightforward to search for other locations of blocking boundaries. Figure 5(b) shows the locations of blockiness. It should be highlighted that the blocking boundaries extracted in Figure 5(b) is based on the S_H. It does not mean that all pixels on the potential blocking positions can be observed by human.

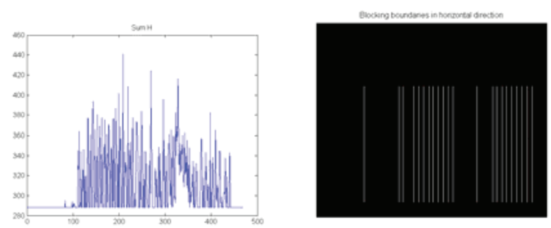

Figure 5: The information of blocking boundaries in horizontal direction: (a) The S_H above average; (b) Extracting the blocking boundaries in horizontal direction

2.4.2. Blocking Measurement

A blocking is generally caused by luminance discontinuities across the block boundaries as perceived by human eyes. By setting a threshold d, only the locations with gradients larger than d are considered as the visible blocking boundaries (VBB_H and VBB_V). By combining the blocking extracted in horizontal and vertical direction, we obtain the final result as shown in Figure 6.

We define the measurement of blocking amount BLK as a percentage of the extracted blocking pixels TBK over the total blocking pixels T as follows:

$$BLK = \frac{TBK}{T} \times 100\% \quad (3)$$

where:

$$TBK = TBK_H + TBK_V; \qquad T = T_H + T_V \quad (4)$$

$$TBK_H = \sum_{i=1}^{w}\sum_{j=1}^{h} VBB_H(i,j) \qquad TBK_V = \sum_{i=1}^{w}\sum_{j=1}^{h} VBB_V(i,j) \quad (5)$$

$$T_H = \sum_{i=1}^{w}\sum_{j=1}^{h} BB_H(i,j) \qquad T_V = \sum_{i=1}^{w}\sum_{j=1}^{h} BB_V(i,j) \quad (6)$$

where w and h are the width and height of the frame size.

In the whole system, there are 2 parameters that affect the efficiency of the blockiness algorithm: blocking threshold d and

SPicBlockiness. Low value of d indicates more blocking extracted. *SPicBlockiness* is the blockiness sensitivity and it is compared to *BLK* computed above. When *BLK* is above *SPicBlockiness*, then the frame is flagged to be a blocky frame.

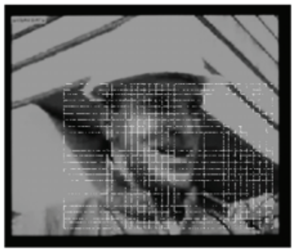

Figure 6: Final blocking extraction from an image

3. RESULTS

3.1. Test Sequences

The original undistorted video sequences have been selected to have differing contents and characteristics. Each of the test video sequences is restricted to 250 frames for easier management and conducting of subjective tests.

The test video sequences used are generated by compressing original undistorted video sequences using MPEG-2 video compression and then subjecting them to channel transmission errors such as additive white noise and packet loss etc before decompressing them for testing.

The video display used is a LCD monitor. The assessors have been seated at a distance of 4H during the subjective test sessions. The monitor used for the experiment was a Tobii 1750 LCD monitor.

These test video sequences are then scrutinized by several subjects to identify the number of picture freeze frames, picture loss frames, and blocking frames and their corresponding locations. The majority vote is then used to mark these frames as whether picture freeze, picture loss, or blocking existed and these results will be used as the ground truth in our experiments.

3.2. Objective Results

Numerous video sequences subjected to picture freeze, picture loss, and blocking artifacts have been tested using the proposed video quality analysis system. Without performing software optimization, the developed system is able to process video of CIF size (352x288 pixels) at more than 30 fps on a Pentium IV 3GHz machine.

Assuming that the test video has p freeze/loss/blocking frames (considered as the ground truth) but VQA measured q freeze/loss/blocking frames correctly while making r frames of respective false alarms, then, the accuracy of the respective algorithm $= (q - r)/p*100\%$. The video quality analysis system is found to give an average accuracy of 91.5% for picture freeze, picture loss, and blocking detections from the tests.

4. CONCLUSIONS

This paper described a video quality analysis system for in-service monitoring of streamed videos, particularly over mobile/wireless networks. Detailed descriptions of the no-reference methods for measuring picture freeze, picture loss, and blockiness have been provided. The developed system (where the software has not been optimized for speed) is able to process video of CIF size (352x288 pixels) at more than 30 fps on a Pentium-IV 3GHz machine.

Due to the unique design of the proposed video quality analysis system, it can be extended to analyze and measure the received video quality of handheld portable display devices by capturing the displayed video off the screen of these devices (this is necessary due to the absence of output signal port to tap received video signal directly from these devices).

5. REFERENCES

[1] Bovik, A.C., and Liu, S., "DCT-domain blind measurement of blocking artifacts in DCT-coded images", *IEEE International Conference on Acoustic, Speech, and Signal Processing*, Vol. 3, May 2001, pp. 1725-1728.

[2] Lindh, P., and Lambrecht, C., "Efficient spatio-temporal decomposition for perceptual processing of video sequences", *International Conference on Image Processing*, 1996.

[3] Lubin, J., Brill, M.H., and Pica, A.P., "Method and apparatus for estimating video quality without using a reference video", US Patent 6285797, Sep. 2001.

[4] Muijs, R., and Kirenko, I., "A no-reference blocking artifact measure for adaptive video processing", *Proc. 2005 European Signal Processing Conference*, Antalya, Turkey, 2005.

[5] Meesters, L., and Martens, J.B., "A single-ended blockiness measure for JPEG-coded images", *Signal Processing*, Vol. 82, 2002, pp. 369-387.

[6] Tan, K.T., Ghanbari, M., "A multi-metric objective picture-quality measurement model for MPEG video", *IEEE Trans. on Circuits & Syst. for Video Tech.*, Vol. 10(7), 2000, pp. 1208-1213.

[7] VQEG (Video Quality Expert Group), "Final Report from the Video Quality Expert Group on the validation of Objective Models of Video Quality Assessment", *www.vqeg.org*, 2000.

[8] Wang, Z., Bovik, A.C., and Evans, B.L., "Blind measurement of blocking artifacts in images", *IEEE International Conference on Image Processing*, Sep. 2000, pp. 981-984.

[9] Wang, Z., Lu, L., and Bovik, A.C., "Video quality assessment based on structural distortion measurement", *Signal Processing: Image Communication*, Vol. 19, No. 2, pp. 121-132, February 2004.

[10] Wang, Z., Sheikh, H.R., and Bovik, A.C., "No-reference perceptual quality assessment of JPEG compressed images", *International Conference on Image Processing*, September 2002.

[11] Watson, A.B., Hu, J., and McGowan III, J.F., "DVQ: A digital video quality metric based on human vision", *Journal of Electronic Imaging*, Vol. 10(1), 2001, pp. 20-29.

[12] Winkler, S., "A perceptual distortion metric for digital color video", *SPIE Proc. Human Vision and Elect. Imaging IV,* Vol. 3644, pp. 175 – 184, 1999.

[13] Wu, H.R. and Yuen, M., "A generalize block-edge impairment metric for video coding", *IEEE Signal Processing Letters*, Vol. 4(11), 1997, pp.317-320.

[14] Yu, Z, Wu, H.R., Winkler, S. and Chen, T., "Vision-model-based impairment metric to evaluate blocking artifacts in digital video", *Proc. IEEE*, Vol. 90(1), pp. 154-169, 2002.