Video Quality Assessment for Web Content Mirroring

Ye He, [1] Kevin Fei, [2] Gustavo A. Fernandez [2] and Edward J. Delp [1]

[1] *Video and Image Processing Lab (VIPER),*
School of Electrical and Computer Engineering, Purdue University
[2] Google Inc.

ABSTRACT

Due to the increasing user expectation on watching experience, moving web high quality video streaming content from the small screen in mobile devices to the larger TV screen has become popular. It is crucial to develop video quality metrics to measure the quality change for various devices or network conditions. In this paper, we propose an automated scoring system to quantify user satisfaction. We compare the quality of local videos with the videos transmitted to a TV. Four video quality metrics, namely Image Quality, Rendering Quality, Freeze Time Ratio and Rate of Freeze Events are used to measure video quality change during web content mirroring. To measure image quality and rendering quality, we compare the matched frames between the source video and the destination video using barcode tools. Freeze time ratio and rate of freeze events are measured after extracting video timestamps. Several user studies are conducted to evaluate the impact of each objective video quality metric on the subjective user watching experience.

Keywords: video quality, image quality, web content mirroring

1. INTRODUCTION

Over the last few years, video traffic has become a significant fraction of the Internet data traffic.[1–3] Studies show that 57% of all consumer Internet traffic was used for video traffic in 2012, and this number will increase to 69% by 2017.[1] These percentages are for streaming video traffic (including both live and video-on-demand services) and does not include video exchanged through peer-to-peer (P2P) file sharing. Projections show that the sum of all forms of video traffic will be in the range of 80% to 90% of global consumer traffic by 2017.[1] According to the 2013 global internet phenomena report, Netflix and YouTube account for over half of the downstream traffic during peak period in North America.[4]

With the dramatic growth in the online video content, user expectation for high quality streaming video and viewing experience is continuously increasing. Studies show that Internet video to TV doubled in 2012, and will continue to grow at a rapid pace, increasing fivefold by 2017.[1] Many video streaming devices provide the capability of projecting web video content from the screens in mobile devices to a larger TV screen. We call this feature "web content mirroring." However, fundamental questions, such as whether the mirrored web content quality is good enough in terms of user satisfaction, have not been formally addressed. One of the major challenges lies in the lack of an objective score to quantify the degree of user satisfaction. In this context, it is crucial to develop video quality metrics and understand how these metrics affect user viewing experience in order to best utilize internet resources to optimize user experience.

In this paper, we propose an automated scoring system, geared to Google's Chromecast product, but generalizable to other video streaming services, to quantify user satisfaction. We compare the quality of source videos with the videos transmitted to the TV. The general diagram of video transmission is shown in Figure 1. To be displayed on a TV, the video data is first captured and encoded in the Chrome web browser, transmitted over a Wi-Fi network, received and decoded in Chromecast, and finally displayed on the TV screen through HDMI connection.

To focus on the evaluation of web content mirroring quality and reduce the effect of a source video as much as possible, we choose to play the source video locally. While capturing the destination video, we want the video to be as similar as shown in TV. Due to High-Definition Copyright Projection (HDCP), we can not capture the destination video in the HDMI output port. In our system, we save the video raw data before it is rendered to

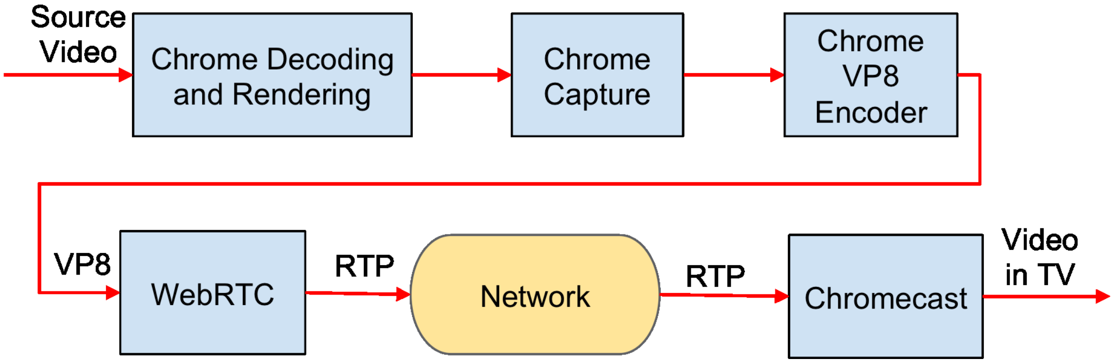


Figure 1. A diagram of transmitting source videos to the TV through Chromecast.

Chromecast's Marvell hardware and use it as the destination video. The source video and the destination video are compared to evaluate the quality of a web content mirroring session.

To estimate a user's viewing experience of a video session, we first measure the video quality for the video session. While there are several video quality metrics that can be used to characterize the performance of a video session, we focus on the following video quality metrics:[3, 5–7]

- Image quality: The similarity between the original video frames and the available matching video frames in the captured video.
- Rendering quality: The ratio of the rendered frames per second to the encoded frames per second.
- Freeze time ratio: The fraction of the total video session time spent in screen freeze.
- Rate of freeze events: the total number of freeze events divided by the whole video session duration.

We choose these video quality metrics to estimate subjective viewing experience because earlier work showed that they have a significant impact on user engagement[3, 5–7] . According to,[7] buffering ratio is the most important metric with respect to impact on user engagement. In this paper we estimate the buffering ratio with two metrics: freeze time ratio and rate of freeze events.

Several user studies are conducted to evaluate the impact of each individual video quality metric on subjective viewing experience. We scaled the user viewing experience from 1 to 5 corresponding to unwatchable, poor, acceptable, good and perfect. For each video quality metric, we generate corresponding artifacts to videos to such an extent that users score the videos from unwatchable to perfect. In this way, we find the user's tolerance threshold for each video quality metric. We use the results from user studies to measure the impact of each objective video quality metric on subjective viewing experience. The video quality metrics used in our system is unique in that 1) it is not derived from network-level metrics, such as the bit rate and packet drop ratio, 2) each video quality metric is controlled and measured individually to test user viewing experience, and 3) the objective score is matched to subjective score through user studies.

2. VIDEO QUALITY METRICS

According to the viewer experience report, viewers are less patient with poor video quality.[8] Studies show that viewers with a buffer-free experience watch 226% more and viewers receiving better picture quality watch 25% longer.[8] In practice, however, subjective evaluation is usually inconvenient, expensive and time-consuming. The goal of this paper is to develop objective video quality metrics to automatically predict viewers' subjective viewing experience. Given a source video, we measure the quality of the destination video in TV using four video quality metrics: image quality, rendering quality, freeze time ratio and the rate of freeze events. In this section, we will explain the video quality metrics in detail and the tools we used to measure them.

2.1 Image Quality

Image quality measures the similarity between the original video frames and the available matching video frames in the captured video. Given a reference video frame (F_r) in the original video and the corresponding frame (F_d) in the captured video, we analyze two well-known objective image quality metrics, the peak-signal-to-noise radio (PSNR) and the structural similarity index measure (SSIM).

PSNR measures the average squared differences between a distorted image and a reference image. PSNR is widely used as an image quality metric because it is simple to calculate and has clear physical meanings. Suppose two frames are both of size $M \times N$, the PSNR between F_r and F_d is defined by:

$$PSNR(F_r, F_d) = 10log_{10}(\frac{255^2}{MSE(F_r, F_d)}) \tag{1}$$

where

$$MSE(F_r, F_d) = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (F_r(i,j) - F_d(i,j))^2 \tag{2}$$

Wang et al.[9] describe a new paradigm for image quality assessment, SSIM, based on the hypothesis that the human visual system (HVS) is highly adapted for extracting structural information. The SSIM is designed by modeling any image distortion as a combination of three factors: luminance distortion, contrast distortion and structure distortion. The SSIM is defined as:

$$SSIM(F_r, F_d) = l(F_r, F_d) c(F_r, F_d) s(F_r, F_d) \tag{3}$$

where

$$l(F_r, F_d) = \frac{2\mu_{F_r}\mu_{F_d} + C_1}{\mu_{F_r}^2 + \mu_{F_d}^2 + C_1} \tag{4}$$

$$c(F_r, F_d) = \frac{2\sigma_{F_r}\sigma_{F_d} + C_2}{\sigma_{F_r}^2 + \sigma_{F_d}^2 + C_2} \tag{5}$$

$$s(F_r, F_d) = \frac{2\sigma_{F_r F_d} + C_3}{\sigma_{F_r} * \sigma_{F_d} + C_3} \tag{6}$$

$l(F_r, F_d)$ is the luminance comparison function which measures the closeness of the two images' mean luminance (μ_{F_r} and μ_{F_d}). $c(F_r, F_d)$ is the contrast comparison function which measures the closeness of the contrast of the two images through the standard deviation σ_{F_r} and σ_{F_d}. $s(F_r, F_d)$ is the structure comparison function which measures the correlation coefficient between the two images. $\sigma_{F_r F_d}$ is the covariance between F_r and F_d. The possible values of the SSIM index are in [0,1]. 0 means no correlation between images, and 1 means the two images are the same. The positive constants $C1$, $C2$ and $C3$ are used to avoid a null denominator.

Given the image quality metrics and a reference frame, we need to find the matching frame in the captured video. Video encoder/decoder and video transmission can produce pauses in the video presentation that result from dropped or repeated video frames.[10] In our system, we overcome the problem by overlaying the input video with barcodes. For the implementation of barcode encoding and decoding, we use an open-source, multi-format 1D/2D barcode image processing library, ZXing, in our system.[11] Many barcode formats are supported in the library, from which we select the most commonly used UPC-A barcode. This can be easily changed to other formats of barcodes. An example of overlaying an original video frame with an UPC-A barcode is shown in Figure 2.

The overall workflow of barcode encoder is shown in Figure 3. Given a source video, we first extract the video frames in the YUV format. Then we use the ZXing library to generate barcodes for each extracted video frame. The output of the ZXing barcode generator is an image barcode in the PNG format. We need to convert it to the YUV format and then write it on the base YUV frames. Finally, the YUV frames with barcode information are combined together and converted to the original video format for future transmission and analysis.

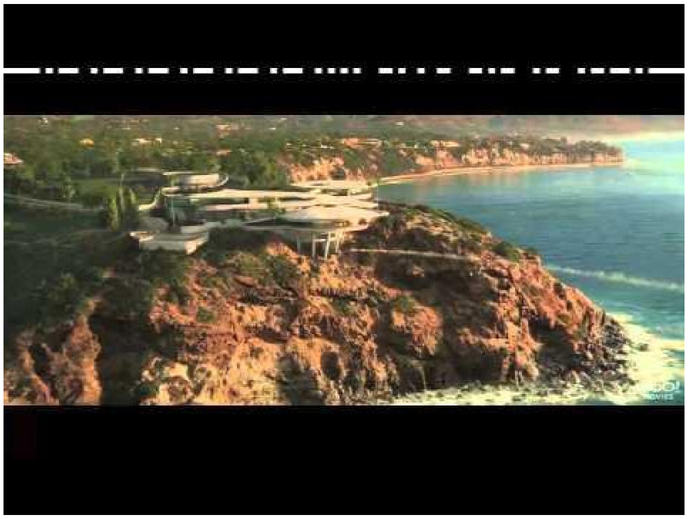


Figure 2. An example of overlaying an original video frame with an UPC-A barcode.

The barcode decoder is a reverse process of the barcode encoder as shown in Figure 4. Given a captured video, we first extracted the video frames in the YUV format. Then we extract a barcode from each video frame. The ZXing barcode decoder is used to decode each barcode. We match the decoded barcodes with the original barcodes to find the matching frames between the source video and the captured video. The average PSNR/SSIM of all the matching frames in the captured video is calculated as the image quality index for the whole video mirroring session.

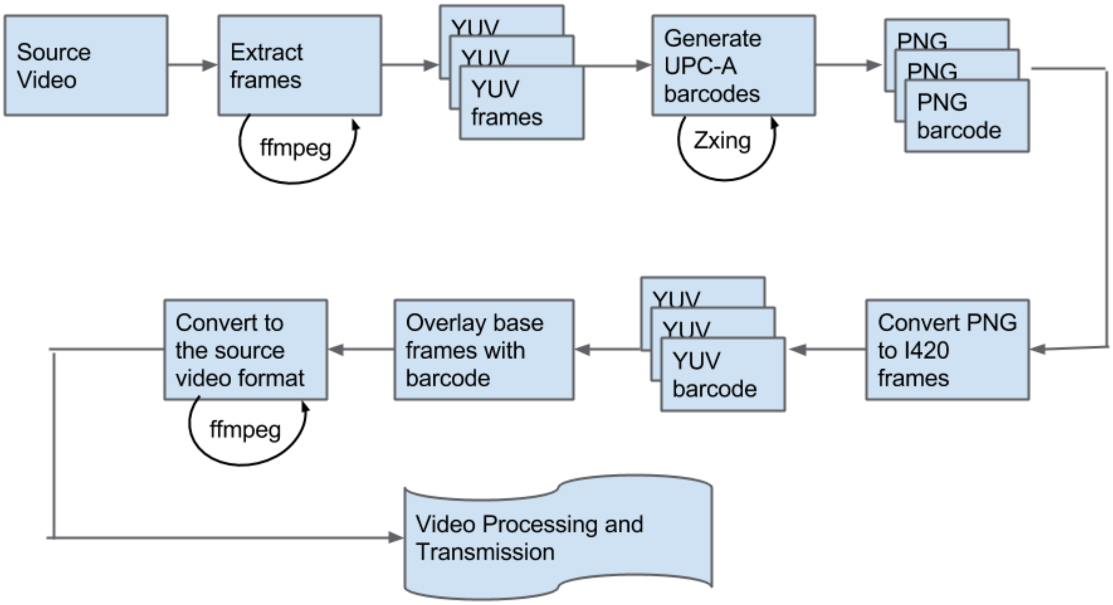

Source Video
Extract frames
ffmpeg
YUV
YUV
YUV frames
Generate UPC-A barcodes
Zxing
PNG
PNG
PNG barcode
Convert PNG to I420 frames
YUV
YUV
YUV barcode
Overlay base frames with barcode
Convert to the source video format
ffmpeg
Video Processing and Transmission

Figure 3. The diagram of the barcode encoder.

2.2 Rendering Quality

The rendering quality is measured as the ratio of the rendered frames per second to the encoded frames per second. Rendering quality may drop due to several reasons.[7] For example, the video player may drop frames to keep up with the stream if the CPU is overloaded. Also if the buffer becomes empty due to network congestion, the rendering quality will drop to 0. Note that most Internet video streaming uses TCP (e.g., RTMP, HTTP

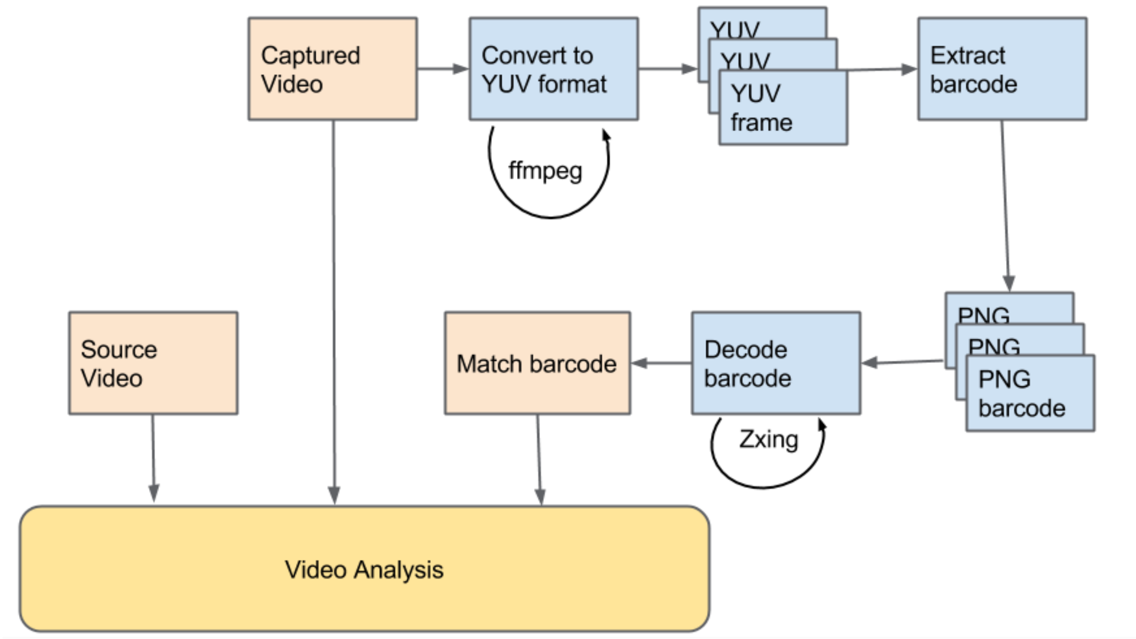


Figure 4. The diagram of the barcode decoder.

chunk streaming), thus network packet loss does not directly cause a frame drop, but it could deplete the client buffer due to reduced throughput.[7]

Using the barcode tools described in Section 2.1, we can find the matching frames between the captured video and the source video. Thus, we can calculate the dropped frames by counting the frames in the source video that can not find matches in the captured video.

2.3 Video Freeze

Since viewers are very sensitive to video freeze,[7] we measure video freeze for web content mirroring with two metrics: freeze time ratio and rate of freeze events. For a video mirroring session of T seconds, the freeze time ratio (FTR) is defined as the fraction of the total video session time spent in screen freeze:

$$FTR = \frac{\sum_{i=1}^{N} t_i}{T} \tag{7}$$

where N is the number of screen freeze happened through the whole video mirroring session. t_i is the freezing time of the i^{th} screen freeze. The rate of freeze events (FER) is defined as the total number of freeze events divided by the total video session duration:

$$FER = \frac{N}{T} \tag{8}$$

In our system, we measure the freeze time ratio and rate of freeze events by extracting the timestamp of every frame in the captured video. Given the frame rate of the source video, we can calculate the ideal time difference between two continuous frames. If the time difference between two continuous frames is above a max threshold, it is considered as a screen freeze event.

3. EXPERIMENTAL RESULTS

Several user studies are conducted to evaluate the impact of video quality metrics on subjective viewing experience. For each video quality metric, we generate corresponding artifacts to videos to such an extent that users score the videos from unwatchable to perfect. In this way, we find the user's tolerance threshold for each video quality metric. For each user study, we ask the participants to watch some videos in a random order and give a score describing the video quality after watching a video. The scores are designed from 1 to 5, corresponding to unwatchable, poor, acceptable, good and perfect. In this section we describe our experimental results of each video quality metric.

To measure the influence of image quality on the subjective viewing experience, 8 video clips are created with different image quality. Half of the video clips contain slow motion video content, while the other half of video clips contain fast motion video content. There is no frame drop or screen freeze issues with these video clips. In total 30 participants watched the video clips in a random order and gave a score to each video clip. We calculated the PSNR and SSIM of these video clips as described in Section 2.1. For each video clip, we calculated the PSNR and SSIM both for all YUV components and for only the Y component. The results of the viewers' subject score and the normalized PSNR/SSIM scores are shown in Figure 5.

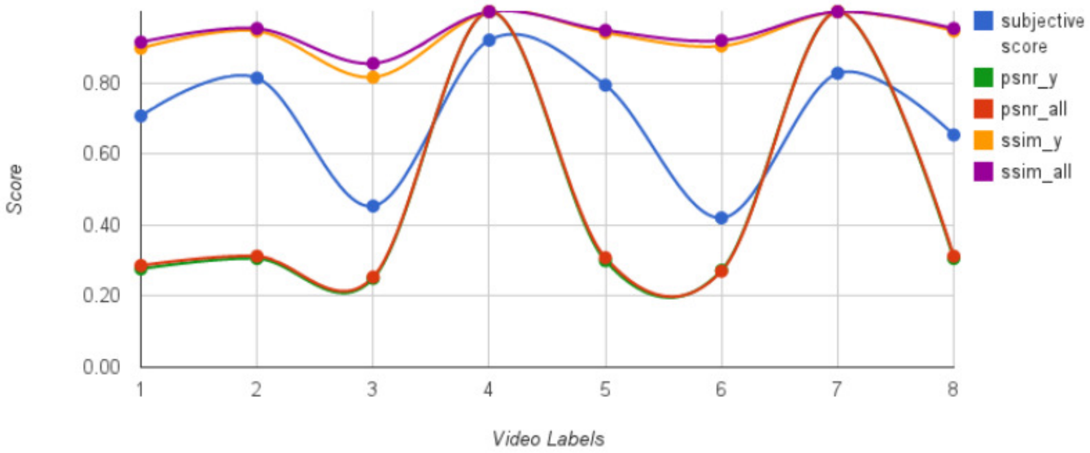
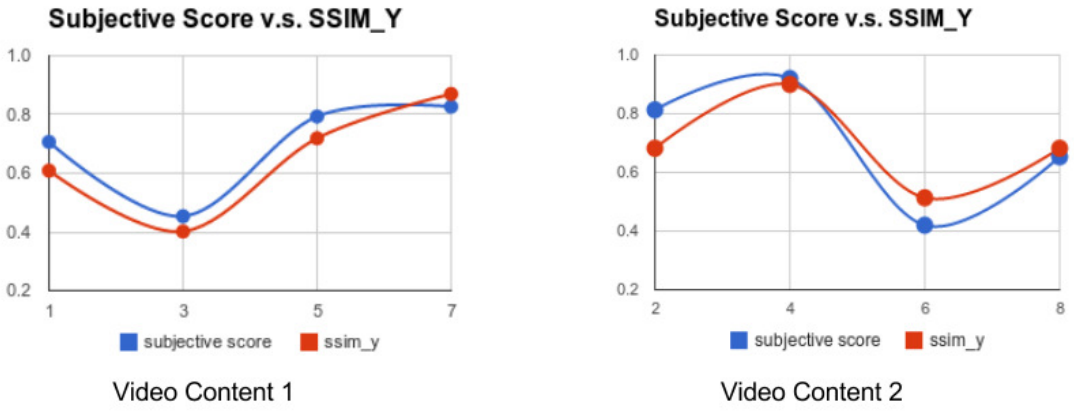


Figure 5. Experimental results of the image quality metric.

As we can see from Figure 5, PSNR for all YUV components and PSNR for only the Y component are similar. SSIM for all YUV components and SSIM for only the Y component are also similar. Thus to reduce the computation burden, we use PSNR and SSIM for the Y component of the video frames. In our studies, PSNR and SSIM scores are both consistent with the human observation, but SSIM score is more adaptive to human observation with regard to the change of video image quality. The PSNR and the SSIM have been discussed extensively in many studies.[12–14] There are no precise rules for selecting the SSIM or the PSNR for the evaluation of image quality. In our system, we use the average SSIM for the Y component (SSIM_Y) of all matched video frames as the image quality index. In our studies, using the SSIM_Y to approximate human subjective scores shows nearly linear relationship between the SSIM_Y and the subjective score, as shown in Figure 6.

To measure the influence of rendering quality on the subjective watching experience, we prepare 7 video clips with different frame loss rate. Other factors such as image quality are kept the same. The video clips are watched by 20 participants in a random order. The result of the average viewers' subjective score for videos with different frame loss rate is shown in Figure 7. As we can see from the experimental result, the average subjective score is between "good" and "perfect" for a video with frame loss rate around 4%. For videos with frame loss rate between 10% and 30%, the average subjective score is between "acceptable" and "good". Videos with frame loss rate above 30% are not acceptable for the majority of viewers.

To measure viewers' watching experience on videos with screen freeze, we prepare 9 videos with different number of freeze events and freeze time. The videos are watched by participants randomly. Each video was watched by about 20 participants. The experimental results is shown in Figure 8. As we can see the result, videos with more freeze events and videos with longer freeze time receive less subjective score. Comparing the result for videos with 1 freeze event lasting 5 seconds and videos with 5 freeze events with each one lasting 1 second, we can see that users prefer longer freeze time instead of more freeze events.



Video Content 1

Video Content 2

Figure 6. Experimental results of using linear transform from SSIM_Y to the subject score.

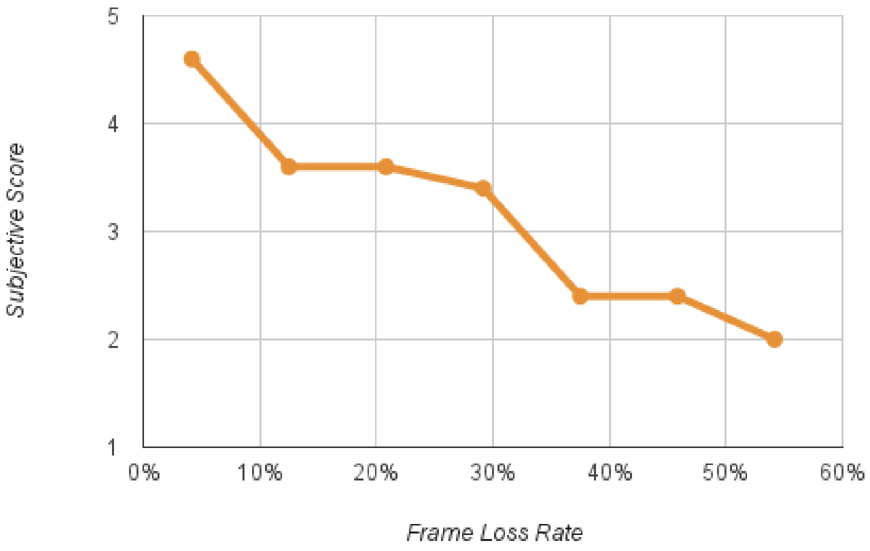
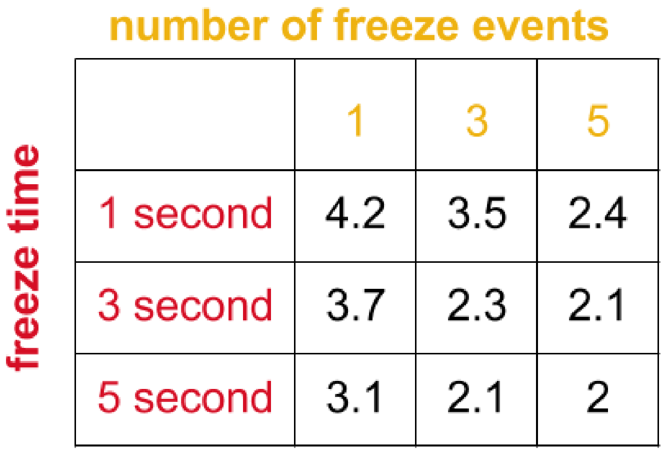


Figure 7. Experimental results of the rendering quality metric.

4. CONCLUSIONS

Some video streaming devices provide the capability of projecting web video content from the small screens in mobile devices to a larger TV screen. The goal of this work is to develop video quality metrics and understand how these metrics affect user viewing experience in order to best utilize internet resources to optimize user experience. We proposed to use four video quality metrics, namely Image Quality, Rendering Quality, Freeze Time Ratio and Rate of Freeze Events, to measure the video quality change for web content mirroring. To measure image quality and rendering quality, we compare the matched frames between the source video and the destination video using barcode tools. Freeze time ratio and rate of freeze events are measured after extracting video timestamps. Several user studies are conducted to evaluate the impact of each individual video quality metric on the subjective viewing experience. In particular, we find that users are less sensitive to rendering quality than other quality metrics and the rate of freeze events has larger impact on the user experience than



		number of freeze events		
		1	3	5
freeze time	1 second	4.2	3.5	2.4
	3 second	3.7	2.3	2.1
	5 second	3.1	2.1	2

Figure 8. Experimental results of the freeze time ratio metric and the rate of freeze events metric.

the freeze time ratio.

REFERENCES

[1] "Cisco visual networking index: Forecast and methodology, 2012-2017," tech. rep., Cisco Systems Inc. (May 2013). http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html.

[2] "The zettabyte era -trends and analysis," tech. rep., Cisco Systems Inc. (May 2013). http://www.cisco.com/en/US/netsol/ns827/networking_solutions_white_papers_list.html.

[3] Liu, X., Dobrian, F., Milner, H., Jiang, J., Sekar, V., Stoica, I., and Zhang, H., "A case for a coordinated internet video control plane," *Proceedings of the ACM SIGCOMM 2012 conference on Applications, technologies, architectures, and protocols for computer communication* , 359–370 (August 2012).

[4] "Global internet phenomena report," tech. rep., Sandvine (November 2013). https://www.sandvine.com/downloads/general/global-internet-phenomena/2013/2h-2013-global-internet-phenomena-report.pdf.

[5] Ganjam, A., Pappu, P., Stoica, I., Zhan, J., and Zhang, H., "Impact of delivery eco-system variability and diversity on internet video quality," *IET Journals* **4**, 36–42 (2012).

[6] Chen, K.-T., Huang, C.-Y., Huang, P., and Lei, C.-L., "Quantifying skype user satisfaction," *ACM SIGCOMM Computer Communication Review* **36**(4), 399–410 (2006).

[7] Dobrian, F., Sekar, V., Awan, A., Stoica, I., Joseph, D. A., Ganjam, A., Zhan, J., and Zhang, H., "Understanding the impact of video quality on user engagement," *ACM SIGCOMM Computer Communication Review* **41**(4), 362 – 373 (2011).

[8] "Viewer experience report," tech. rep., Conviva (February 2013). http://www.conviva.com/vxr/.

[9] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing* **13**(4), 600–612 (2004).

[10] Wolf, S. and Pinson, M., "A no reference (nr) and reduced reference (rr) metric for detecting dropped video frames," *Proceedings of the Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics* , 1–6 (January 2009).

[11] Google, "Zxing." https://code.google.com/p/zxing/.

[12] Eskicioglu, A. M. and Fisher, P. S., "Image quality measures and their performance," *IEEE Transactions on Communications* **43**(12), 2959–2965 (1995).

[13] der Weken, D. V., Nachtegael, M., and Kerre, E. E., "Image quality evaluation," *Proceedings of the 6th International Conference on Signal Processing* **1**, 711–714 (August 2002).

[14] Hore, A. and Ziou, D., "Image quality metrics: Psnr vs. ssim," 2366–2369 (August 2010).