GAZE PATTERN ANALYSIS FOR VIDEO CONTENTS WITH DIFFERENT FRAME RATES

Manri Cheon, Jong-Seok Lee

School of Integrated Technology
Yonsei University, Korea
{manri.cheon, jong-seok.lee}@yonsei.ac.kr

ABSTRACT

This paper presents a study investigating the viewing behavior of human subjects for video contents having different frame rates. Frame rate variability arises when temporal video scalability is considered for adaptive video transmission, and the gaze pattern variation due to the frame rate variability would eventually affect the visual perception, which needs to be considered during perceptual optimization of such a system. We design an eye-tracking experiment using several high definition contents having a wide range of content characteristics. By comparing the gaze points for a normal frame rate condition and a low frame rate condition, it is shown that, although the overall viewing pattern remains quite similar, statistically significant difference is also observed for some time intervals. The difference is analyzed in terms of two factors, namely, overall gaze paths and subject-wise variability.

Index Terms— Eye-tracking, frame rate, perception, quality of experience

1. INTRODUCTION

Video content delivery services over networks have become popular these days. We are familiar with many online video distribution services such as YouTube, Netflix, and Vimeo, which provide high quality videos over networks. In terms of the resolution, high definition (HD) contents are popularly consumed in currently available video applications. Not only TV, PC, or laptop but also mobile devices (such as smart phones and tablets) having HD resolutions are becoming more and more popular. As this trend continues, it can be prospected that the video traffic will continuously increase.

Therefore, there exist challenges in these video services, including the limit in the network capacity and user heterogeneity in terms of the network environment and terminal capability. Video scalability provides a flexible solution to these challenges, i.e., a scalable video can adapt its data rate and quality parameters without necessity of re-encoding of the video data [1]. There are several dimensions of video scalability, e.g., temporal scalability, spatial scalability, and quality scalability. When delivering video using scalable video coding, service providers can adapt several options from video scalability.

Among the video scalability dimensions, the temporal scalability refers to the possibility of changing frame rate adaptively. In order to exploit temporal scalability effectively, it is important to understand the effect of variations in frame rate on human viewers' perception [1]. In [2], effects of the video frame rate on human performance were investigated. It was concluded that around 15 Hz was preferred generally for a threshold of the low frame rate for many tasks including those that are psychomotor and perceptual in nature. In [3], the authors noted that the eye-path is not affected by frame rate variations, which was based on the observation that the median horizontal eye-path over viewers is consistent independently of the presentation frame rate varied among 5 fps, 15 fps, and 25 fps.

In general, the gaze pattern for a video sequence affects perception of the content. Especially, quality perception depends on where the viewers pay attention in the scene, which needs to be considered in applications using temporal scalability for optimizing users' quality of experience (QoE). It is known that the bottom-up visual attention mechanism is influenced by motion information in the given visual stimulus [4]. Changing the frame rate alters perceived motion information, which consequently affects the viewer's focus of attention.

In this paper, we investigate gaze pattern variations with respect to video frame rate variations through an eye-tracking experiment. In particular, we consider HD contents that are increasingly popular as aforementioned. Since HD contents have larger image sizes than SD, it is expected that the focus of attention effect is more prominent for the former, while the previous work [3] did not observe significant difference in the viewing behavior for different frame rates for non-HD. We conduct an eye-tracking experiment to examine the gaze patterns for the contents having normal frame rates and low frame rates. The analysis of the collected eye-tracking data aims at answering the following two research questions: (1) Does the frame rate influence the eye-path in an average

This work was supported by the Ministry of Science, ICT and Future Planning (MSIP), Korea, under the "IT Consilience Creative Program" supervised by the National IT Industry Promotion Agency (NIPA-2013-H0203-13-1002), and by the Basic Science Research Program through the National Research Foundation of Korea funded by the MSIP (2013R1A1A1007822).

sense over the subjects? (2) Does the frame rate influence the subject-wise variability of the eye-path?

The rest of the paper is organized as follows. The following section explains how the eye-tracking experiment was designed. Section 3 presents the experimental results and analysis. Finally, concluding remarks are given in Section 4.

2. EXPERIMENT DESIGN

2.1. Test sequences

We designed and conducted an eye-tracking experiment with eighteen HD sequences that have a frame size of 1920×1080 pixels. Their frame rate is 25 fps or 30 fps, which we refer to normal frame rate (NFR). We obtained most of the sequences from the contents available in online video-sharing websites, i.e., YouTube and Vimeo, and these have creative commons license [5]. The source clips are of high quality without visible artifacts. Various content types are included in the test sequences, e.g., talking person, moving cars, moving people, etc., to cover a wide range of visual content characteristics. The sequences have lengths of 5 to 10 seconds, which correspond to 175 to 300 frames for NFR videos. No scene cut was included in each sequence in order to investigate the effect of frame rate on perception properly. We chose low frame rate (LFR) as a quarter of NFRs, i.e., 6.25 fps or 7.5 fps. The test sequences were shown in random orders in the experiments and sorted in the alphabetical order in this paper for convenience. Fig 1 shows representative frames of the test sequences.

2.2. Participants

Twenty four subjects (six females and eighteen males) participated in the experiment. They are mostly graduate students and not familiar with experiments related to eye-tracking and visual perception. Their ages are between 24 and 34 with a mean 28.0. All of them reported normal or corrected-to-normal vision. The subjects were divided into two groups: twelve subjects watched NFR videos, while LFR videos were for another twelve subjects. We assigned separate groups for the two conditions to avoid the memory effect [6, 7] affecting the gaze pattern in the multiple viewing[1].

2.3. Procedure

The eye-tracking experiment was conducted using Smart Eye Pro 5.8 equipment and Samsung 24-inch LCD monitor having a resolution of 1920×1080 pixels. The eye-tracker recorded gaze information at a rate of 60 Hz.

[1]One many argue that it is fair to let each subject watch both NFR and LFR videos and to compare the results. However, we emphasize that the viewing pattern change due to the memory effect in the multiple viewing is strong enough to obscure the fairness, which led us to use separate groups for each frame rate condition.

Each subject sat in front of the monitor at a distance of 2-3 times the height of the monitor. Before tracking, gaze calibration was performed in order to enable accurate tracking of the each subject's gaze. Then, the subject was told about the test procedure with some example video sequences that were not used in the experiment. During the test, the subject watched eighteen test sequences continuously, where the viewing order was set random for each subject. Before display of each sequence, a red cross on the gray background was shown at the center of the screen for three seconds in order to set the starting gaze point to be at the center. No particular task was assigned to the subjects so that a free viewing scenario was considered. And the information about the frame rate of the video sequences, was not given to the subjects for unbiased results.

2.4. Data Processing

Since the sampling rate of the eye-tracker was 60 Hz, the recorded data were interpolated so as to match the video frame rates. First, each frame of a LFR video sequence was repeated four times so that the NFR and LFR videos have the same number of frames. Then, the eye-tracking data were resampled at a rate of 25 or 30 Hz for each of the NFR and LFR cases.

3. RESULTS AND ANALYSIS

In order to compare the gaze patterns for NFR and LFR videos, the coordinates of the recorded gaze points are examined. Figs. 2 and 3 show representative examples of median eye-paths (x-coordinates and y-coordinates) of all subjects for NFR and LFR and their variations (standard deviations). On average, it can be said that the overall gaze patterns for NFR and LFR are similar in terms of the overall shape of the median eye-paths and variations. At the same time, however, locally appearing discrepancy between the two cases is also observed, e.g., around frame #240 in Fig. 2(a) and around frame #160 in Fig. 2(b).

The local discrepancy between NFR data and LFR data can be analyzed from two aspects, namely, median and variation. The discrepancy between median eye-paths of NFR and LFR means that there exists difference in usual gaze patterns for the two cases. In terms of variations, the discrepancy between NFR and LFR means that the levels of agreement of the gaze patterns across the subjects are different for NFR or LFR.

In order to examine the discrepancy of median eye-paths further, we carried out statistical tests (non-parametric Wilcoxon-Mann-Whitney tests) under the null hypothesis that the gaze x-coordinates (or y-coordinates) for NFR and those for LFR are samples from distributions with equal medians. The tests were conducted for x-coordinates and y-coordinates separately. From the test results, it was found that the gaze

Fig. 1. Eighteen test sequences used in our experiment (shown in the alphabetical order).

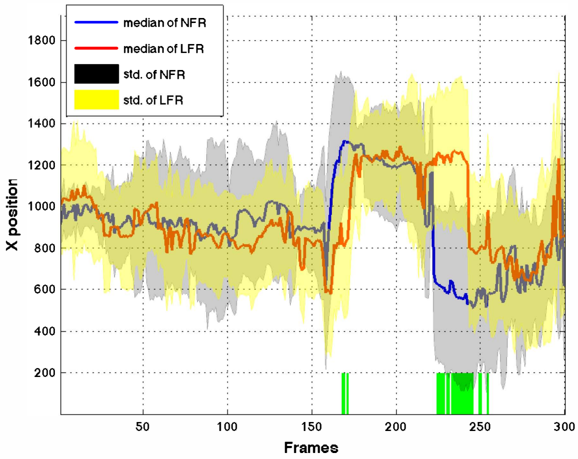
point locations for NFR and LFR are significantly different for some frames (marked with green bars in Fig. 2), e.g., around the 240th frame in x-coordinates (Fig. 2(a)) and around the 160th frame in y-coordinates (Fig. 2(b)). Fig. 4 shows the percentages of the number of those frames (either in x-coordinates or y-coordinates) for each sequence when the significance level is 5%. On average, 5.03% of all frames show statistically significant difference. The values are different for each sequence, and the maximum is nearly 17% (sequence #6). Certainly, this amount is not negligible and needs to be considered in perception-driven video processing when different frame rates are involved.

In order to examine the difference of variations of subject-wise eye-paths between NFR and LFR further, we conducted statistical test (non-parametric Ansari-Bradley tests) under the null hypothesis that the gaze x-coordinates (or y-coordinates) for NFR and those for LFR are from distributions with the same median and shape but different dispersions. As the statistical tests for the median, the tests for the dispersion were conducted for x-coordinates and y-coordinates separately. In this case, however, it is meaningful to observe which condition (NFR or LFR) has the larger dispersion value (marked with cyan or magenta bars in Fig. 3). Fig. 5 shows the percentages of the numbers of frames (either in x-coordinates or y-coordinates) that NFR or LFR have larger dispersion values than the other one for each sequence when the significance level is 5%. On average, 4.25% of all frames show statistically significant larger dispersions for NFR and 8.47% of all frames show statistically significant larger dispersions for LFR. Thus, it can be concluded that the level of inter-subject eye-path agreement is significantly influenced by the frame rate change for 12.72 % of all frames. Although content-dependence is observed, the dispersion of LFR is larger than that of NFR for a larger number of frames. It is probable that jerky motion in LFR videos, which is unnatural and unexpected, can make subjects' focus deviated variously from the usually focused region for NFR.

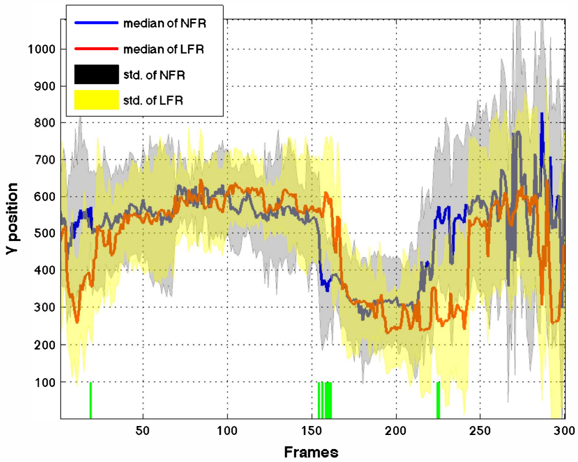
Finally, we examine whether the two types of discrepancy between NFR and LFR (i.e., median and dispersion) occurs at the same time or not. Fig 6 shows the proportions of the numbers of frames showing only median-discrepancy, only dispersion-discrepancy, and both types of discrepancy to the number of frames showing either type of discrepancy. The results show that two types of discrepancy are exclusive in most cases (about 97% on average). This implies that for some frames of LFR videos, either most of viewers focus on a particular region different from that focused for NFR, or viewers look at various regions around the region focused for NFR.

4. CONCLUSION

In this paper, we have presented an eye-tracking experiment with the aim of investigating the influence of the frame rate variations on viewing patterns and visual perception. Eye-tracking data were recorded from two groups of subjects for eighteen HD contents, one group for normal frame rates and the other group for low frame rates. It was shown that for about 5% among all frames on average, the gaze pattern is significantly affected by the frame rate. Then, it was shown that the dispersion of the gaze patterns of all subjects is affected by the frame rate for about 13% of all frames on average. Additionally, we observed that the two types of variations tend to occur exclusively. These findings would be useful for understanding users' perception in temporal scalability-based video communication systems and consequently perceptual optimization of such systems. In the future, we will analyze the obtained results further, e.g., dependence of the results on the content characteristics such as motion information.

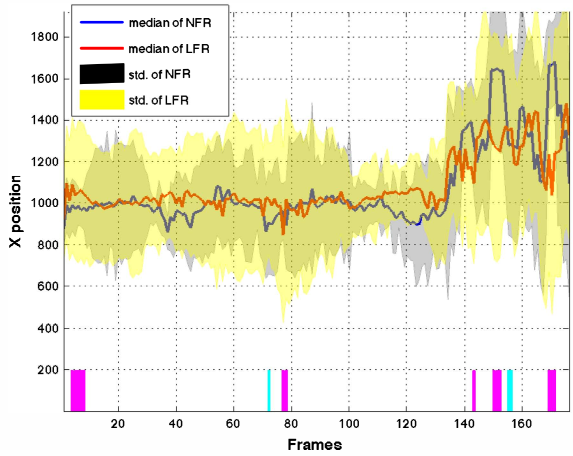


(a)

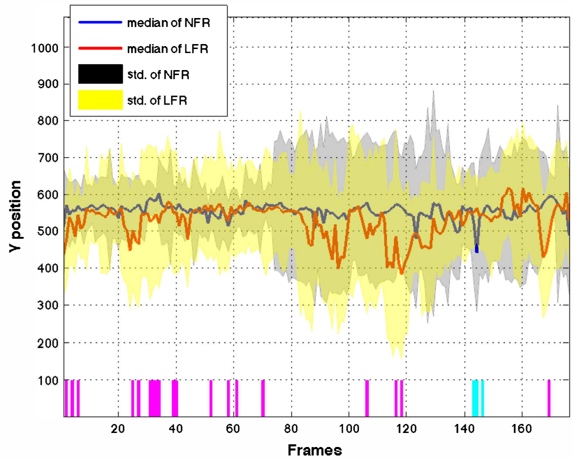


(b)

Fig. 2. Eye-path of (a) the x-coordinate and (b) the y-coordinate for sequence #17. Bold lines in the shaded area indicate median eye-paths for NFR (blue) and LFR (red) and shaded areas indicate their standard deviations over subjects for NFR (gray) and LFR (yellow). Green bars indicate frames for which gaze points of NFR and LFR are significantly different (at a significance level of 5%).



(a)



(b)

Fig. 3. Eye-path of (a) the x-coordinate and (b) the y-coordinate for sequence #15. Bold lines indicate median eye-path for NFR (blue) and LFR (red) and shaded areas indicate their standard deviations over subjects for NFR (gray) and LFR (yellow). Cyan (or magenta) bars indicate frames that the variance of gaze points of all subjects is larger (or smaller) for NFR than LFR .

5. REFERENCES

[1] J.-S. Lee, F. De Simone, N. Ramzan, E. Izquierdo, and T. Ebrahimi, "Quality assessment of multidimensional video scalability," *IEEE Communications Magazine*, vol. 50, no. 4, pp. 38–46, 2012.

[2] J. Y. C. Chen and J. E. Thropp, "Review of low frame rate effects on human performance," *IEEE Transactions on Systems, Man and Cybernetics - Part A : Systems and Humans*, vol. 37, no. 6, pp. 1063–1076, 2007.

[3] S. R. Gulliver and G. Ghinea, "Stars in their eyes: What eye-tracking reveals about multimedia perceptual quality," *IEEE Transactions on Systems, Man and Cybernetics - Part A : Systems and Humans*, vol. 34, no. 4, pp. 472–482, 2004.

[4] L. Itti, N. Dhavale, and F. Pighin, "Realistic avatar eye and head animation using a neurobiological model of visual attention," in *Proc. SPIE 48th Annual International Symposium on Optical Science and Technology*, B. Bosacchi, D. B. Fogel, and J. C. Bezdek, Eds., Bellingham, WA, August 2003, vol. 5200, pp. 64–78, SPIE Press.

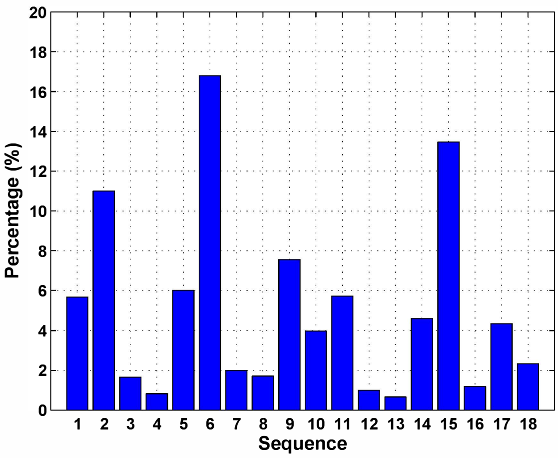


Fig. 4. Percentage of the number of frames where the gaze points for NFR and LFR are significantly different (at a significance level of 5%).

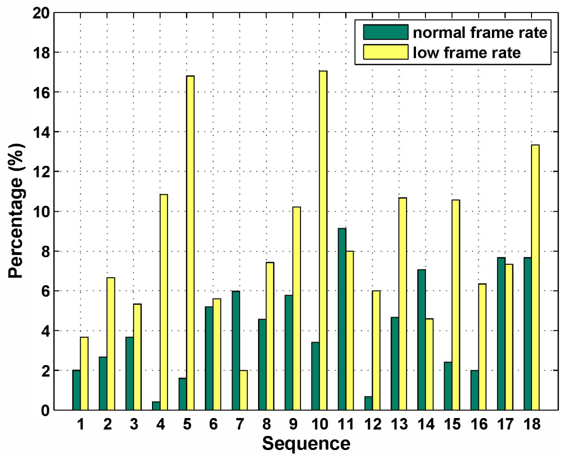


Fig. 5. Percentage of the number of frames where variations of gaze points across subjects for NFR or LFR are significantly larger than the other one (at a significance level of 5%).

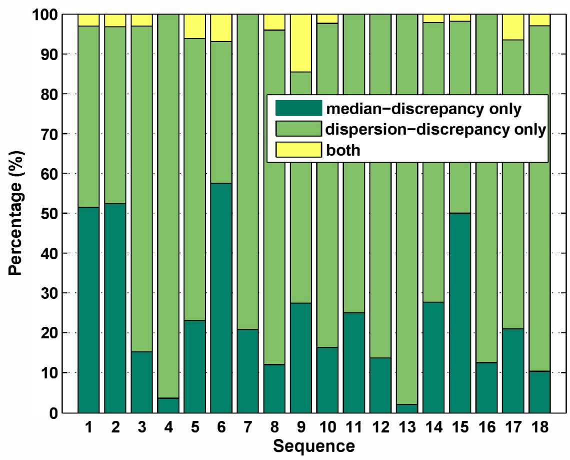


Fig. 6. Proportions of the numbers of frames corresponding to different types of significant discrepancy.

[5] "Creative commons license," `http://www.creativecommons.org/licenses/`.

[6] J.-S. Lee, F. De Simone, and T. Ebrahimi, "Subjective quality evaluation of foveated video coding using audio-visual focus of attention," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 7, pp. 1322–1331, November 2011.

[7] M. Nyström and K. Holmqvist, "Effect of compressed offline foveated video on viewing behavior and subjective quality," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 6, no. 1, pp. 1–14, February 2010.