Mixing Tile Resolutions in Tiled Video: A Perceptual Quality Assessment

Hui Wang, Vu-Thanh Nguyen, Wei Tsang Ooi and Mun Choon Chan
School of Computing, National University of Singapore
Email:{wanghui,nguyenvt,ooiwt,chanmc}@comp.nus.edu.sg

ABSTRACT

The mismatch between increasingly large video resolution and constrained screen size of mobile devices has led to the proposal of zoomable video systems based on tiled video. In the current system, a tiled video frame is constructed from multiple tiles in a single resolution stream. In this paper, we explore the perceptual effect of mixed-resolution tiles in tiled video, in which tiles within a video frame could come from streams with different resolutions, with the aim to tradeoff bandwidth and perceptual video quality. To understand how users perceive the video quality of mixed-resolution tiled video, we conducted a psychophysical study with 50 participants on tiled videos where the tile resolutions are randomly chosen from two resolution levels with equal probability. The experiment results show that in many cases, we can mix tiles from HD (1920×1080p) stream and tiles from 1600×900p stream without being noticed by the viewers. Even when participants notice quality degradation in videos combined with tiles from HD stream and tiles from 960×540p stream, the majority of participants still accept the degradation when viewing videos with low and medium motion; and greater than 40% of participants accept the quality degradation when viewing video with dense motion.

Categories and Subject Descriptors

H.4.3 [**INFORMATION SYSTEMS APPLICATIONS**]: Communications Applications; H.5.1 [**INFORMATION INTERFACES AND PRESENTATION**]: Multimedia Information Systems—*Video*

General Terms

Design, Human Factors, Experimentation, Measurement

Keywords

Zoomable Video, Tiled Video, Perception, Psychophysics

NOSSDAV'14, March 19-21 2014, Singapore, Singapore.

ACM 978-1-4503-2706-0/14/03 ...$15.00.
http://dx.doi.org/10.1145/2578260.2578267.

1. INTRODUCTION

While consumer video resolution has increased from HD to 4K, with 8K resolution arriving in foreseeable future, the physical screen size of mobile devices is normally constrained to ensure portability and ease of use. To allow better viewing of video details on such small screens, zoomable video streaming has been proposed [6, 5, 7, 10, 3] to allow viewers to zoom into the video and view selected regions in higher details. One implementation of zoomable video is to encode the video frames of the original video streams into fixed size tiles at different resolutions. Such tiling supports random access into selected region of interest (RoI) within the video.

Figure 1 illustrates the idea of tiled video. Each video frame is divided into a grid of small areas (tiles). The video can be considered as a three dimensional matrix of tiles. Tiles at the same y-x position are temporally grouped and coded along z axis. Tiles from one y-x position could be encoded/decoded independently with tiles from a different y-x position.

In existing works [7, 8, 3], at the server side, an original video is normally encoded into different versions (streams): frames of a low-resolution stream are constructed from a smaller number of tiles; and frames of higher-resolution streams are constructed from a larger number of tiles. At the client side, the number of tiles required to cover the physical screen resolution is fixed, therefore, the bandwidth consumption for each user will be mostly constant. Initially, a low resolution version of the video will be sent to users. When a user zooms into a RoI within the video, the server will first determine a suitable high-resolution stream based on the requested RoI size (zoom level). It then selects tiles covering the requested RoI from this stream. This mechanism allows users to see their regions of interest in detail without consuming more bandwidth.

The afore-mentioned RoI cropping technique performs well in small scale networks by unicasting video stream. In one of the use case we consider, the video stream is consumed by a large number of users within one location (e.g., in a concert hall or stadium). To overcome the scalability issues with such a large number of users and RoI requests, wireless multicast scheme is employed. When the RoI regions from multiple users partially overlap, tiles from the overlapped regions could be potentially multicasted to all interested users to save bandwidth consumption. In zoomable video, different users, however, may have different zoom levels (i.e., different RoI sizes) and will need tiles from different versions encoded at different resolutions, which prevents the potential benefits of wireless multicast.

Instead of fixing tile size, using a fixed number of tiles to encode and decode videos could be more effective. At the server side, an original video will be encoded into different resolution versions, but all versions consist of the same number of tiles. The same amount of tiles is required at the client side to decode each

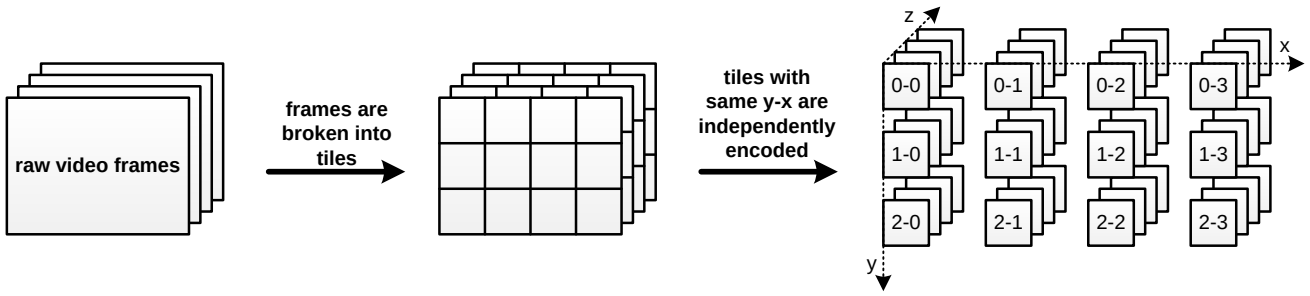


Figure 1: Tiled video

video frame. Within a frame, however, different tiles could come from different resolution streams. If a tile comes from a stream with resolution lower/higher than requested level, it will be scaled up/down accordingly. In zoomable video, when a user zooms into a region of interest (RoI) within the video, the server will first determine the tiles covering this RoI, and then associate each tile with an appropriate stream version, depending on their popularity and the resource constraints.

The proposed *mixed resolutions tiling* scheme has the following two essential advantages in tiled video streaming. First, benefiting from the scaling up/down operations for each tile, the multicast transmissions are considerably reduced. Next, by intelligently allocating resolution version to each tile, the mixing resolutions approach may considerably reduce bandwidth consumption without impairing much perceived video quality. For instance, the popular regions/tiles requested by many users could come from high-resolution streams; while tiles requested by one or few users could come from a low-resolution stream under limited bandwidth condition.

Although this proposed scheme saves bandwidth, the impairment to perceived quality is still unclear. Thus, in order to understand if, and at what thresholds, users could notice and/or accept the difference between original video and tiled video with mixed resolutions, we conducted a psychophysical study with 50 participants. Using the *method of limits* from psychophysics [4], we measure two perceptual thresholds – Just Noticeable Difference (JND) and Just Unacceptable Difference (JUD) – to understand the user perception about the quality of mixed-resolution tiled video.

In our experiments, we evaluate the quality difference of three well-known test sequences with different motion densities. The evaluation results demonstrate that in many cases, tiles from HD (1920×1080p) stream and 1600×900p stream could be mixed together without being noticed. Even when participants notice quality degradation in videos combined with tiles from HD stream and tiles from 960×540p stream, the majority of participants still accept the quality difference for low and medium motion videos; and greater than 40% participants accept the quality difference for the dense motion video.

The rest of the paper is structured as follows. Section 2 discusses existing works that closely related to our study. Section 3 presents our experiment setup and procedure. The evaluation results are shown in Section 4 and the conclusion and future work are discussed in Section 5.

2. RELATED WORK

Our work is closely related to the following three areas: (i) tiled streaming, (ii) RoI video coding, and (iii) psychophysical assessment.

Tiled streaming: To support HD video RoI cropping and streaming, Feng et al. propose tile-based solution [2], where each video frame is broken into a grid of tiles. Instead of transmitting the whole frame, a minimum set of tiles covering the selected RoI region is delivered. Based on tiling, zoomable video is proposed [7, 8], where video is encoded into multiple resolutions to reduce bandwidth consumption when users zooms out and to allow users to view higher resolution videos when zooming in.

RoI video coding: A rich body of work on real time RoI-based adaptive video coding has been proposed [9, 11, 12]. Applying the auto foreground (RoI region) and background regions identification, these studies encode some particular regions with a higher data rate. Further, the network bandwidth consumption could be reduced by decreasing the background quality. Apart from depending on the precise background identification techniques, the adaptive video coding solutions suffer from the dependency issues of the video coding, resulting in unnecessary transmissions without RoI cropping.

Psychophysics: As subjective quality assessment is the most accurate and reliable way to measure perceived quality of the given content, psychophysical methods are widely employed to assess the perceptual quality of the video. Recently, the techniques have been used to study user perception in multimedia systems. De Silva et al. investigate the Just Noticeable Difference in Depth in 3D video [1]. Wu et al. [13] identify the Just Noticeable Degradation and Just Unacceptable Degradation in 3D tele-immersive video. In our paper, we apply the similar method to measure the Just Noticeable Difference and Just Unacceptable Difference in mixed-resolution tiled video.

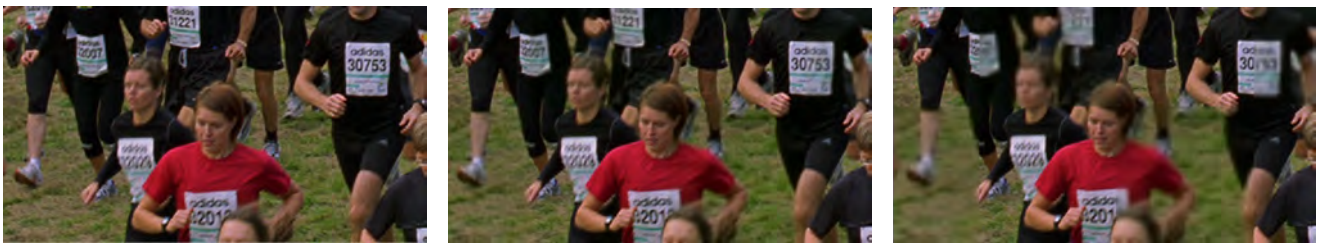
3. PSYCHOPHYSICAL EXPERIMENT

The purpose of the psychophysical experiment is to estimate two perceptual thresholds of video quality difference: Just Noticeable Difference (JND) and Just Unacceptable Difference (JUD). The two identified difference thresholds partition the quality degradation level (introduced by mixing tile resolutions) into the following three intervals: without noticeable quality degradation, with noticeable (but acceptable) quality degradation, and with unacceptable quality degradation.

3.1 Setup

Our experiments assess the quality of mixed-resolution tiled video using three standard HD (1920×1080p) test video files, *Crowd-Run* (dense motion, 50fps), *Old-Town-Cross* (medium motion, 50fps), and *Rush-Hour* (low motion, 25fps)[1]. The configurations for constructing the mixed-resolution tiled videos are detailed in the fol-

[1]Available at `http://media.xiph.org/video/derf/`



(a) Tiles from HD stream (b) Mixing tiles from stream levels 5 and 3 (c) Mixing tiles from stream levels 5 and 1

Figure 2: Mixing tile resolutions of Crowd-Run

(a) Tiles from HD stream (b) Mixing tiles from stream levels 5 and 3 (c) Mixing tiles from stream levels 5 and 1

Figure 3: Mixing tile resolutions of Old-Town-Cross

(a) Tiles from HD stream (b) Mixing tiles from stream levels 5 and 3 (c) Mixing tiles from stream levels 5 and 1

Figure 4: Mixing tile resolutions of Rush-Hour

lowing two subsections.

Table 1: The number of pixels in each frame and each tile at different resolution levels.

level	frame	16×9 tiles	80×45 tiles
5	1920×1080	120×120	24×24
4	1600×900	100×100	20×20
3	1280×720	80×80	16×16
2	960×540	60×60	12×12
1	640×360	40×40	8×8

3.1.1 Mixing Resolution Levels

We have five resolution levels for each video file, these levels are labeled from 5 to 1 (Table 1). The pixels of the original video frame at five resolution levels are: 1920×1080, 1600×900, 1280×720, 960×540, and 640×360.

In the experiments, we construct mixed-resolution tiled video by mixing two resolution levels, where the higher resolution level is denoted as R_H and the lower resolution level is denoted as R_L. Specifically, given a pair of R_H and R_L, we randomly allocate resolution level R_H or R_L to each tile with equal probability. For any particular pair of R_H and R_L, we restrict the range of R_H as $3 \leq R_H \leq 5$ and the range of R_L as $1 \leq R_L \leq R_H$. Figures 2, 3, and 4 show the screenshots of mixed-resolution tiled video.

3.1.2 Tile Size

Since the aspect ratio of the test HD video frame sequences is 16:9, we break the video frames into 16×9 tiles by default. As a result, each tile size (view region size) is $\frac{1}{16\times9}$ of the entire view region. To evaluate the impact of tile size, in addition to the default configuration, we generate another set of videos where each video frame is broken into 80×45 tiles. The number of pixels for a tile at each resolution level is shown in Table 1.

3.1.3 Data Rate

The average data rate (Mbps) of mixing resolution levels 5 and R_L in tiled video with 16×9 tiles and 80×45 tiles are represented in Table 2 and Table 3, respectively. Since the video data rate closely depends on its motion density, the full HD version of Crowd-Run experiences the highest data rate and the test sequence Rush-Hour has the lowest data rate. Besides the motion density, the tile

Table 2: Video data rates (*Mbps*) for configurations (5, R_L) with 16×9 tiles.

	5-5	5-4	5-3	5-2	5-1
Crowd-Run	26.44	22.32	19.91	17.78	15.53
Old-Town-Cross	7.15	5.81	5.06	4.68	4.2
Rush-Hour	5.12	4.40	3.83	3.37	3.01

Table 3: Video data rates (*Mbps*) for configurations (5, R_L) with 80×45 tiles.

	5-5	5-4	5-3	5-2	5-1
Crowd-Run	29.02	24.34	21.98	19.47	17.19
Old-Town-Cross	9.67	7.83	6.96	6.39	5.74
Rush-Hour	5.62	4.76	4.28	3.70	3.33

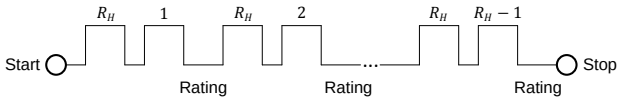
size is another important factor determining the video encoding efficiency. Due to the number of pixels in a finer-grained (80×45) tile is considerably smaller than in a coarse-grained (16×9) tile, encoding coarse-grained tiled video is more efficient than finer-grained tiled video (data rate), which is verified in the tables.

Regarding the bandwidth efficiency of mixing tile resolutions, the tables demonstrate that mixing tiles from resolution levels 5 and 4, the video consumes 14%-20% less bandwidth, compared to the video consisting of tiles from only level 5. Moreover, the bandwidth consumption can be reduced further with a smaller R_L value. For instance, around 35% bandwidth will be saved by mixing tiles from resolution levels 5 and 2 in all test sequences.

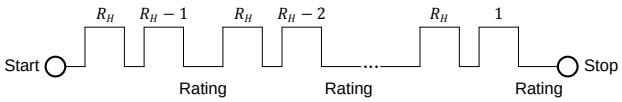
3.2 Procedures

Fifty adult participants were invited to participate in our assessment, primarily graduate students and research staffs from National University of Singapore. The sample consisted of 16 women and 34 men; all had normal vision. They were asked to watch the mixed-resolution tiled videos online[2] using a monitor with full HD display resolution.

For configurations with 16×9 tiles, we vary the high resolution level R_H from 5 to 3, 9 stimuli series are generated over three test videos. For configurations with 80×45 tiles, we generate stimuli series with $R_H = 5$. As a result, we have 12 stimuli series in total, which are shuffled in a random order and played.



(a) Ascending stimuli series



(b) Decending stimuli series

Figure 5: Experiment procedure. The video is composed by tiles with resolution level R_H and R_L. The numbers above represent the value of R_L, the first video in each pair is a standard tiled video where $R_L = R_H$, and the second video is a mixed-resolution tiled video.

[2]Online website is available at:
`http://liubei.ddns.comp.nus.edu.sg/resMix`

For each series, the stimuli is randomly manipulated in either an ascending or a descending order, the procedures are depicted in Figure 5. In a stimuli series, we fix the high resolution level R_H and vary the low resolution level R_L. As shown in the figure, each pair presents a standard video where $R_L = R_H$ and a mixed-resolution tiled video. After watching the videos in a pair ($10s$ per video), the participant is asked to rate the level of the difference between two videos. In particular, two questions are asked: (i) *is the quality difference noticeable* and (ii) *is the quality difference unacceptable*. In the case of ascending series, we increase R_L from 1. On each successive trial, we increase R_L by 1 util the participant eventually reports the difference is unnoticeable or $R_L = R_H - 1$. If the series is descending, the stimuli operates in an opposite direction. We start from $R_L = R_H - 1$ and gradually decrease R_L util the participant reports the difference is unacceptable or $R_L = 1$.

Using the above procedure, the obtained results fall into the following three categories: (i) The noticeable difference threshold and unacceptable difference threshold are both detected; (ii) Only the noticeable difference threshold is detected; and (iii) Neither noticeable difference threshold nor unacceptable difference threshold can be detected. Assuming that we have detected the noticeable difference threshold and unacceptable threshold, denoted by T_{ND} and T_{UD}, respectively, then according to *the method of limits* [4], we estimate the Just Noticeable Difference threshold as $\big(T_{ND} + (T_{ND} + 1)\big)/2 = T_{ND} + 0.5$. Similarly, we express Just Unacceptable Difference threshold as $\big(T_{UD} + (T_{UD} + 1)\big)/2 = T_{UD} + 0.5$. For the cases where we failed to detect the difference threshold, we set the corresponding Just Noticeable/Unacceptable Difference threshold to 0.

4. RESULTS

We first examine the configuration with 16×9 tiles. Figure 6 depicts the CDF distribution of participants that cannot notice any difference between mixed-resolution tiled video (5, R_L) and standard tiled HD video (5, 5). The CDF distribution of participants that accept the quality difference is present in Figure 7. The average measured thresholds of Just Noticeable Difference and Just Unacceptable Difference for R_H in the range from 5 to 3 are shown in Table 4 and Table 5, respectively. From the results, we can draw the following observations.

Feasibility of Mixing Tile Resolutions. The measured thresholds confirm the feasibility of mixed-resolution tiled video. The CDF distribution from Figure 6 implies that we can mix tiles with resolution levels 5 and 4 without being noticed in most cases. Further, the depicted result from Figure 7 indicates that more than 85% participants accept the quality difference with configurations where $3 \leq R_L \leq R_H = 5$; under these configurations, up to 30% bandwidth can be saved by mixing tile resolutions (Table 2). When we construct video from tiles at resolution level 5 and 2, almost all participants noticed the difference for video *Crowd-Run* and *Old-Town-Cross*. 40% to 65% of the participants, however, still accept the quality difference.

Impact of R_H. As expected, both the average JND threshold and the JUD threshold are positively correlated with the high resolution level R_H (Tables 4 and 5). The similar relationship is observed for the measured variations as well. The average thresholds, however, are not proportional to R_H. For instance, the average JND threshold gap of video *Crowd-Run* between $R_H = 5$ and $R_H = 4$ is 0.94, while the gap between $R_H = 4$ and $R_H = 3$ is only 0.65.

Impact of Content. With the same configuration, the results from Tables 4 and 5 show a great disparity in the measured average JND and JUD across three test videos. Overall, video *Crowd-Run*, which has the highest amount of motion among the three test

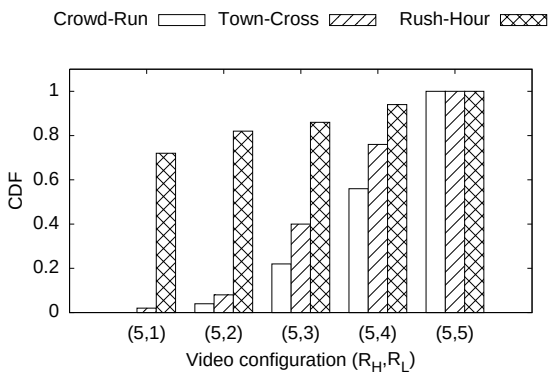


Figure 6: CDF distribution of participants that cannot notice any difference between mixed-resolution tiled video (5, R_L) and standard HD tiled video (5, 5).

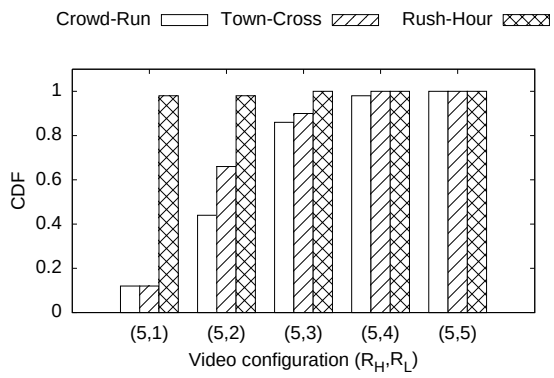


Figure 7: CDF distribution of participants that accept the quality difference between mixed-resolution tiled video (5, R_L) and standard HD tiled video (5, 5).

Table 4: The average Just Noticeable Difference threshold (number within parenthesis is the 95% Confidence Interval value).

R_H	Crowd-Run	Old-Town-Cross	Rush-Hour
5	3.68 (±0.52)	3.25 (±0.47)	0.81 (±0.23)
4	2.74 (±0.39)	2.31 (±0.34)	0.24 (±0.10)
3	2.09 (±0.30)	1.73 (±0.26)	0.11 (±0.06)

Table 5: The average Just Unacceptable Difference threshold (number within parenthesis is the 95% Confidence Interval value).

R_H	Crowd-Run	Old-Town-Cross	Rush-Hour
5	2.03 (±0.31)	1.76 (±0.27)	0(0)
4	1.64 (±0.26)	1.28 (±0.21)	0(0)
3	1.28 (±0.21)	0.69 (±0.14)	0(0)

videos, is most sensitive to the resolution mixing, as the highest average threshold and the greatest variation are detected. Interestingly, video *Rush-Hour*, which has the lowest amount of motion among the three test videos, performs remarkably different from others. It is difficult to notice the quality difference between the mixed-resolution tiled video and the standard version, thus the average measured thresholds and the variations are much smaller compared with other test videos.

Gap between JND and JUD Thresholds. For many cases, although participants could notice the difference, it is still acceptable. Generally, a greater gap value indicates a higher video quality tolerance degree when the quality difference is noticeable. From the Tables 4 and 5, we observe a significant gap between the average measured JND and JUD thresholds, especially for $R_H = 5$. In particular, the average gap quantities for video *Crowd-Run* and *Old-Town-Cross* with $R_H = 5$ are 1.65 and 1.49, respectively. As the tolerance space is reduced with smaller R_H value, the quantity of the threshold gap between JND and JUD will be reduced as well, as can be seen in both tables.

Impact of Tile Size. The comparison between the configurations with 16×9 tiles and 80×45 tiles is present in Tables 6 and 7. The threshold values with 80×45 tiles is slightly smaller than the corresponding threshold values with 16×9 tiles, which indicates that

Table 6: The average Just Noticeable Difference threshold where $R_H = 5$ (number within parenthesis is the 95% Confidence Interval value).

	Crowd-Run	Old-Town-Cross	Rush-Hour
16×9	3.68 (±0.52)	3.25 (±0.47)	0.81 (±0.23)
80×45	3.30 (±0.48)	3.04 (±0.44)	0.76 (±0.20)

Table 7: The average Just Unacceptable Difference threshold where $R_H = 5$ (number within parenthesis is the 95% Confidence Interval value).

	Crowd-Run	Old-Town-Cross	Rush-Hour
16×9	2.03 (±0.31)	1.76 (±0.27)	0(0)
80×45	1.76 (±0.29)	1.63 (±0.25)	0(0)

the quality degradation introduced by mixing resolutions is slightly less obvious for the finer-grained tile size (80×45) compared with the coarse-grained tile size (16×9). The finer-grained tiles, however, are generally less efficient in terms of encoding and transmission bandwidth. Therefore, we need to balance the trade-off between the video quality and the efficiency to obtain an appropriate configuration.

5. CONCLUSION

In this paper, we evaluate the impact of mixing tiles with different resolutions when streaming tiled video. Investigating two paramount factors, our subjective assessment confirmed the feasibility of this approach, especially for video with low to medium motion. Taking identified Just Noticeable Difference threshold as guiding parameter, we are able to save bandwidth consumption by mixing resolutions without being noticed; and the Just Unacceptable Difference threshold can be utilized to further reduce bandwidth consumption while the video quality is still acceptable.

From the evaluation results, we conclude that in most cases, by mixing tiles from HD (1920×1080p) stream and 1600×900p stream, we can save 14%-20% bandwidth without any noticeable perceptual quality loss. A video consisting of mixed tiles from HD stream and 1600×900p stream, consumes about 35% less bandwidth, compared to the video consisting of tiles from only HD stream. We can therefore save 35% bandwidth with acceptable per-

ceptual quality loss when viewing videos with low and medium motion. When viewing video with high dense motion, by mixing tiles from HD stream and 1280×720p stream, the tiled video saves around 25% bandwidth consumption with acceptable quality degradation.

Future Work. Our future work includes the following categories: (i) Currently, we only assess the perceptual thresholds by randomly mixing resolutions of tiles. If we can allocate tile resolutions more intelligently (e.g., lower resolutions on regions with low motion), we can further reduce the bandwidth with acceptable quality degradation. (ii) We are applying the results obtained to tile schedule for wireless multicast of tiled video streams. We are looking into how to optimally allocate resolution versions to each tile to better utilize the wireless bandwidth and improve perceptual video quality of the users. (iii) We also plan to study how mixing tile resolutions can be applied to other applications, such as error protection/recovery of video and DASH video streaming.

Acknowledgment

This research is supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office.

6. REFERENCES

[1] D Varuna SX De Silva, Warnakulasuriya Anil Chandana Fernando, Gokce Nur, Erhan Ekmekcioglu, and Stewart T Worrall. 3d video assessment with just noticeable difference in depth evaluation. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 4013–4016. IEEE, 2010.

[2] Wu-chi Feng, Thanh Dang, John Kassebaum, and Tim Bauman. Supporting region-of-interest cropping through constrained compression. In *Proceedings of the 16th ACM international conference on Multimedia*, pages 745–748. ACM, 2008.

[3] Wu-chi Feng, Thanh Dang, John Kassebaum, and Tim Bauman. Supporting region-of-interest cropping through constrained compression. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, 7(3):17, 2011.

[4] George A Gescheider. *Psychophysics: the fundamentals*. Psychology Press, 1997.

[5] Aditya Mavlankar, Pierpaolo Baccichet, David Varodayan, and Bernd Girod. Optimal slice size for streaming regions of high resolution video with virtual pan/tilt/zoom functionality. In *EUSIPCO*, 2007.

[6] Aditya Mavlankar, David Varodayan, and Bernd Girod. Region-of-interest prediction for interactively streaming regions of high resolution video. In *Packet Video*. IEEE, 2007.

[7] Ngo Quang Minh Khiem, Guntur Ravindra, Axel Carlier, and Wei Tsang Ooi. Supporting zoomable video streams with dynamic region-of-interest cropping. In *MMSys*. ACM, 2010.

[8] Ngo Quang Minh Khiem, Guntur Ravindra, and Wei Tsang Ooi. Adaptive encoding of zoomable video streams based on user access pattern. In *MMSys*. ACM, 2011.

[9] Aniruddha Sinha, Gaurav Agarwal, and Alwin Anbu. Region-of-interest based compressed domain video transcoding scheme. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, volume 3, pages iii–161. IEEE, 2004.

[10] Ray van Brandenburg, Omar Niamut, Martin Prins, and Hans Stokking. Spatial segmentation for immersive media delivery. In *ICIN*. IEEE, 2011.

[11] Haohong Wang and Khaled El-Maleh. Joint adaptive background skipping and weighted bit allocation for wireless video telephony. In *Wireless Networks, Communications and Mobile Computing, 2005 International Conference on*, volume 2, pages 1243–1248. IEEE, 2005.

[12] Haohong Wang, Yi Liang, and Khaled El-Maleh. Real-time region-of-interest video coding using content-adaptive background skipping with dynamic bit reallocation. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 2, pages II–II. IEEE, 2006.

[13] Wanmin Wu, Ahsan Arefin, Gregorij Kurillo, Pooja Agarwal, Klara Nahrstedt, and Ruzena Bajcsy. Color-plus-depth level-of-detail in 3d tele-immersive video: a psychophysical approach. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 13–22. ACM, 2011.