Context-aware Frame Rate Adaption for Video Chat on Smartphones

Xin Qi
Department of Computer Science
College of William and Mary
Williamsburg, VA 23185 USA
xqi@cs.wm.edu

Qing Yang
Department of Computer Science
College of William and Mary
Williamsburg, VA 23185 USA
qyang@cs.wm.edu

David T. Nguyen
Department of Computer Science
College of William and Mary
Williamsburg, VA 23185 USA
dnguyen@cs.wm.edu

Gang Zhou
Department of Computer Science
College of William and Mary
Williamsburg, VA 23185 USA
gzhou@cs.wm.edu

UbiComp'13 Adjunct, September 8–12, 2013, Zurich, Switzerland.
ACM 978-1-4503-2215-7/13/09.
http://dx.doi.org/10.1145/2494091.2494122

- This work is supported by NSF grant CNS-1250180

Abstract
As mobile video traffic is becoming dominant, balancing the mobile video quality and bandwidth usage is a relevant but hard problem. Particularly, in this project, we attempt to reduce the bandwidth usage of video chats through frame rate adaption.

The key idea of this project is to save bandwidth through reducing frame rate at the sender and interpolate the 'missing' frames at the receiver for a video chat. Additionally, the sender dynamically adapts the frame rate with respect to inertial sensor readings in order to keep the scene change between consecutive frames small and prevent strong artifacts from the frame interpolation.

Author Keywords
Video Chat; Smartphones; Frame Interpolation; Frame Rate Adaption

ACM Classification Keywords
C.4 [Performance of Systems]: Design Studies

General Terms
Measurement, Design, Performance

Introduction
In recent years, video chat apps have become highly popular among smartphone users. They allow us to stay

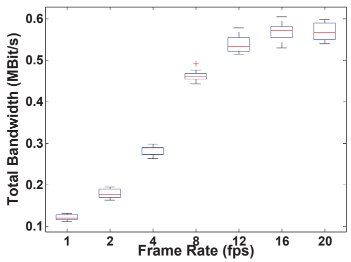

Figure 1: Total Bandwidth vs. Frame Rate.

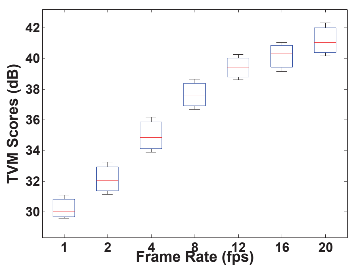

Figure 2: Video Quality vs Frame Rate.

in touch with family, friends, and colleagues anywhere we go. It is expected that there will be 29 million smartphone video chat users in 2015 [1].

However, streaming video chats results in high bandwidth usage. For example, the upload and download bandwidth requirements for low quality video chats over Skype are both 300Kbps [3]. Taking T-Mobile [4] as a sample carrier, its 500MB monthly data plan, with the cost of $50 per phone, can support the use of such low quality video chats over Skype only for 1.85 hours. Thus, reducing the bandwidth usage of video chats apps on smartphones is imperative. This project attempts to achieve this goal.

Related Work

It has been demonstrated that video streaming apps are among the most popular smartphone apps [9]. However, "*they consume much more bandwidth than other apps*" [6]. To lower the bandwidth requirement of video streaming, many video compression techniques [2][5][14] have been proposed. Rather than a new compression technique, our work introduces a novel context-aware frame rate adaption framework that reduces bandwidth usage of mobile video chats. Other works [8][10] propose miscellaneous solutions to lower the bandwidth usage of downloading *existing* videos. In contrast, our work aims to lower the bandwidth usage of streaming *live* video chats.

Measurement

In this section, we investigate through measurements how do the bandwidth usage and video quality vary under different frame rates.

Figure 1 summarizes the bandwidth measurement results. In the figure, when the frame rate of video chats is below 12 fps, we observe that the total bandwidth usage rapidly increases as the frame rate increases. Above 12 fps, the bandwidth usage increases slowly since video compression techniques begin taking effects. The observations indicate that below the default frame rate range ($12 \sim 20$ fps) adopted by smartphone video chat apps, increased frame rate introduces obviously more data to be transmitted per time unit. *Thus, it is possible to reduce the bandwidth usage of performing video chats on smartphones through reducing frame rate.*

To quantify video quality, we use a state-of-the-art no-reference metric, TVM (Temporal Variation Metric) [7]. Figure 2 summarizes the video quality measurement results. In the figure, we observe that the objective video quality decreases as the frame rate decreases. For example, the TVM scores indicate that the mean square difference between consecutive frames at 4 fps increases by about 214% compared to that at 12 fps. The larger the difference between consecutive frames, the jerkier the video. It is also reported that the subjectively perceptual video quality also decreases as the frame rate decreases [12]. *Thus, we cannot simply reduce the frame rate of video chats to save bandwidth, since the video quality deteriorates as the frame rate decreases.*.

Design

In our design, users select a target frame rate, which is lower than the default one, for video chats based on their needs. During a video chat, the sender adopts the target frame rate for most of the time and only adopts the default frame rate when it detects severe smartphone vibrations with respect to inertial sensors (accelerometer and gyroscope) readings. The receiver interpolates the 'missing' frames when the instant frame rate (calculated based on frame interval) is smaller than the default one. The purpose of the frame rate adaption at the sender is to

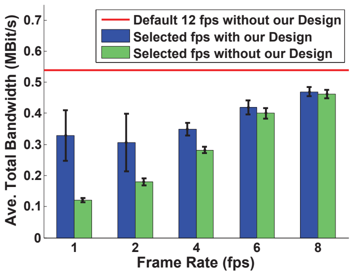

Figure 3: Average Bandwidth usage vs. Frame Rate.

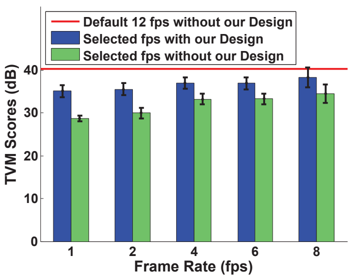

Figure 4: TVM Score vs. Frame Rate with and without our Design.

keep the scene change between consecutive frames small in order to prevent strong artifacts from the frame interpolation algorithm.

Frame Rate Adaption
In our design, the following metrics are used to quantify smartphone vibrations.

$$m_A = \sum_{t=start}^{end} [|\partial A_x(t)| + |\partial A_y(t)| + |\partial A_z(t)|] \quad (1)$$

$$m_G = \sum_{t=start}^{end} [|\partial G_x(t)| + |\partial G_y(t)| + |\partial G_z(t)|] \quad (2)$$

where $\partial A_x(t)$, $\partial A_y(t)$, and $\partial A_z(t)$ are acceleration changes in three measured dimensions; while $\partial G_x(t)$, $\partial G_y(t)$, and $\partial G_z(t)$ are angular velocity changes in three measured dimensions. In addition, $start$ and end are the starting and ending time stamps of each measurement period. We choose 60 milliseconds as the measurement period, which is small enough to quickly reflect the vibration in real-time.

In run-time, if the value of either m_A or m_G is larger than the corresponding threshold, the smartphone vibration is considered *severe*. Otherwise, the vibration is considered *light*.

Frame Interpolation
In our design, we choose the cross dissolve algorithm [11] for frame interpolation. Its effectiveness has been demonstrated in [11]. It is also computationally efficient since it only involves additions of two frames and multiplications of a frame by a constant.

Since the algorithm utilizes two frames as input, the current frame is not displayed until the next frame arrives. Thus, we add a common delay to synchronize the audio and video at the beginning of each video chat.

Results

We implement our design on the Android-based version of Linphone, a popular open-source video chat on smartphones. We recruit 10 pairs of subjects. Each pair performs video chats at different frame rates (1, 2, 4, 6, 8, 12 fps) for bandwidth measurements. Each video chat session lasts for 6 minutes. The subjects are instructed to perform video chats as usual under typical video chat scenarios such as sitting and standing. There is no other physical constraints imposed on the subjects.

Figure 3 illustrates the average total bandwidth usage. The red line depicts the average bandwidth usage at the default 12 fps of Linphone without our design. The blue bars illustrate the average bandwidth usage at the selected frame rates with our design. The results demonstrate that our design is able to reduce the average bandwidth usage under typical video chat scenarios. The bandwidth saving ranges from 13.0% to 43.2%.

We also execute a user study, in which we recruit 21 pairs of subjects from the College of William and Mary campus to perform video chats with and without our design under each selected frame rate. We implement a recorder in Linphone to automatically record all video chats, each of which lasts for 6 minutes. We calculate the TVM scores for all recorded videos and summarize them in Figure 4. The red line depicts the average TVM score at the default 12 fps of Linphone without our design. The blue bars illustrate the TVM scores at the selected frame rates with our design; while the green bars illustrate the TVM scores at the selected frame rates without our design.

Figure 4 statistically demonstrates that our design is able to alleviate the video quality degradation resulted from frame rate reduction. For instance, the TVM scores at 4 fps indicate that the mean square difference between consecutive frames with our design is decreased by 58% compared to that without our design. This significant decrease in the mean square difference between consecutive frames results in smoother videos.

Work in Progress

- Design a user-friendly interface to help users with varying video quality requirements set an appropriate frame rate to save bandwidth with great confidences.
- Investigate the relationship between power consumption and frame rate, and the impact of our design on power consumption.
- Investigate bandwidth savings of our design in other mobility cases such as walking and sitting in a vehicle.
- Collect subjective scores to investigate whether our design can maintain subjects' perceptual experience.
- Investigate other frame interpolation algorithms, such as the optical-flow based algorithm in [13].

References

[1] Juniper research report. http://juniperresearch.com/viewpressrelease.php?pr=209.

[2] Rfc6386 - vp8 data format and decoding guide. http://tools.ietf.org/html/rfc6386.

[3] Skype Bandwidth Requirement. https://support.skype.com/en/faq/FA1417/how-much-bandwidth-does-skype-need.

[4] T-mobile plans:. http://www.t-mobile.com/shop/plans/family-plans.aspx.

[5] H.264: Advanced video coding for generic audiovisual service. International Telecommunication Union (2013).

[6] White paper: Live and on-demand streaming video with lifesize uvc video center (March, 2012).

[7] Chan, A. J., Pande, A., Baik, E., and Mohapatra, P. Temporal quality assessment for mobile videos. In *Proc. of Mobicom*, ACM Press (2012).

[8] Chen, J., Ghosh, A., Magutt, J., and Chiang, M. Qava: quota aware video adaptation. In *Proc. of CoNEXT*, ACM Press (2012).

[9] Falaki, H., Mahajan, R., and et al. Diversity in smartphone usage. In *Proc. Mobisys*, ACM Press (2010).

[10] Keller, L., Le, A., Cici, B., Seferoglu, H., Fragouli, C., and Markopoulou, A. Microcast: cooperative video streaming on smartphones. In *Proc. of Mobisys*, ACM Press (2012).

[11] Kemelmacher-Shlizerman, I., Shechtman, E., Garg, R., and Seitz, S. M. Exploring photobios. In *Proc. of SIGGRAPH*, ACM Press (2011).

[12] Ou, Y.-F., Liu, T., and et al. Modeling the impact of frame rate on perceptual quality of video. In *Proc. of ICIP* (20108).

[13] Rakêt, L. L., Roholm, L., and et al. Motion compensated frame interpolation with a symmetric optical flow constraint. In *Proc. of ISVC* (2012).

[14] Wang, Y., Zhang, Y.-q., and Ostermann, J. *Video Processing and Communications*. 2001.