User-adaptive mobile video streaming using MPEG-DASH

Yuriy A. Reznik

InterDigital Communications, Inc.
9710 Scranton Road, San Diego, CA 92122

ABSTRACT

We describe an implementation of DASH streaming client for mobile devices which uses adaptation to user behavior and viewing conditions as means for improving efficiency of streaming delivery. Proposed design relies on sensors in a mobile device to detect presence of the user, his proximity to the screen, and other factors such as motion, brightness of the screen and ambient lighting conditions. This information is subsequently used to select stream that delivers adequate resolution implied by viewing conditions and natural limits of human vision. We show that in a mobile environment such adaptation can result in significant reduction of bandwidth usage compared to traditional streaming systems.

1. INTRODUCTION

During last two decades Internet streaming has experienced a dramatic growth and transformation from an early concept into a mainstream technology used for delivery of multi-media content.[1–3] A recently issued MPEG-DASH standard[4] consolidates many advances achieved in the design of streaming media delivery systems, including full use of the existing HTTP infrastructure, bandwidth adaptation mechanisms, latest audio and video codecs, etc. Yet, some challenges in implementation and deployment of streaming systems still exist. In particular, they arise in delivery of streaming video content to mobile devices, such as smartphones and tablets.

On one hand, many mobile devices are already matching and surpassing HDTV sets in terms of graphics capabilities. They often feature high-density "retina" screens with 720p, 1080p, and even higher resolutions. They also come equipped with powerful processors, making it possible to receive, decode and play HD-resolution videos. On the other hand, network and battery/power resources in mobile devices remain limited. Wireless networks, including latest 4G/LTE networks, are fundamentally constrained by capacities of their cells. Each cell's capacity is shared between its users, and it can be saturated by as few as $5 - 10$ users simultaneously watching high-quality videos.[5] High data rates used to transmit video also cause high power consumption by the receiving devices, draining their batteries rapidly.

All these factors suggest that technologies for reducing bandwidth and power use in mobile video streaming are very much needed. In this paper we describe one such technology. It is based on an observation that in many cases, mobile phone users can see only a fraction of information projected on the screen.

2. FACTORS AFFECTING USER ABILITY TO DISCERN VISUAL INFORMATION

We illustrate some factors affecting user ability to discern visual information Figures 1 and 2. For instance, the user may hold a phone close to his eyes, or at arm's length.[11,12] This affects viewing angle and density of information seen on the screen. Ambient illuminance may also change significantly. The user may be in the office, outside under direct sunlight, in a shadow, or in a completely dark area. Reflection of ambient light from the screen lowers the contrast of video or images seen by the user.[6] Finally, the user may pay full attention to visual content on the screen, or he could be distracted.

Together with characteristics of mobile display and user vision, all these factors affect the capacity of the "visual channel", serving as the last link in a communication system delivering information to the user. The main idea of this paper, as well as several of our related publications[7–10] is to show that characteristics of this last link can also be effectively measured and utilized in optimizing streaming video delivery. The recently developed MPEG-DASH standard[4] offers an excellent framework using which this idea can be realized.

Author's e-mail address: yuriy.reznik@interdigital.com

Applications of Digital Image Processing XXXVI, edited by Andrew G. Tescher,
Proc. of SPIE Vol. 8856, 88560J · © 2013 SPIE · CCC code: 0277-786X/13/$18
doi: 10.1117/12.2026911

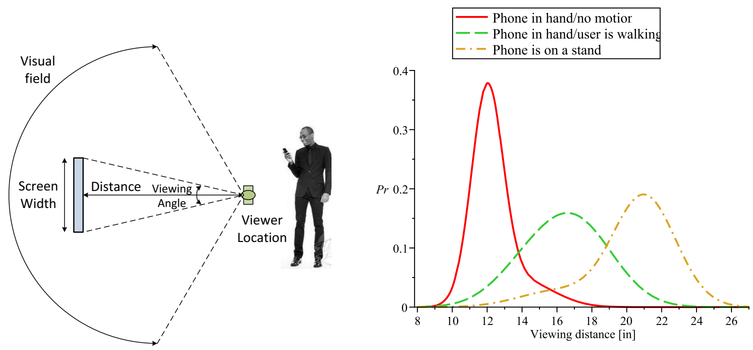

Figure 1. Characteristics of mobile viewing setup. The right sub-figure shows how distribution of viewing distances can be affected by user activity.

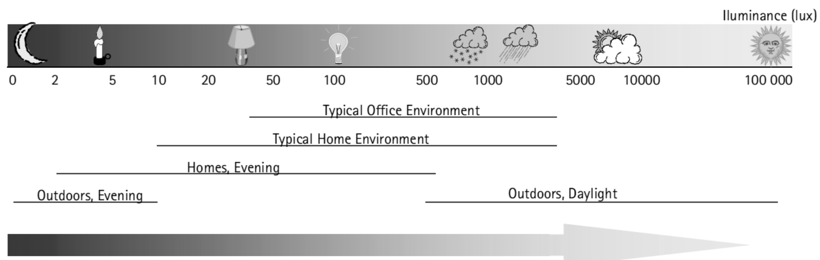

Figure 2. Ambient illuminance in different environments.[6]

3. MPEG DASH-STANDARD

We present a conceptual model of a DASH-based mobile video streaming system in Figure 3. The original video content is captured, encoded, and placed on an HTTP server. To scale distribution, the content may also be pushed to many servers forming a Content Distribution Network (CDN). It is typically the web browser or a streaming client application running on a mobile phone (UE) that discovers this content, retrieves it, and shows it on a mobile device.

3.1 Content preparation

In order to support bandwidth adaptive streaming, the content is usually encoded at a plurality of bit rates. Such encodings are also prepared such that they consist of multiple segments with time-aligned boundaries, allowing switches between encodings at different rates. In MPEG DASH standard, points at which switching is allowed are called stream access points (SAP). In the simplest case, SAP may correspond to an I- or IDR- video frame, allowing sequential decoding of all frames that follow. In addition to producing encoded media streams the encoder also produces a file containing information about parameters of each of the encodings and URL links to them. This file is called media presentation description (.mpd) file.

3.2 Adaptation to bandwidth changes

The streaming session is controlled entirely by the DASH streaming client. It opens an HTTP connection to the server, retrieves the .mpd file, and learns about different encodings (representations) that are available on the server. Then it picks representation with most suitable bitrate, and start retrieving its segments by issuing HTTP GET requests. As bandwidth changes, the streaming client may request segments encoded at different bit rates, allowing uninterrupted playback of the content. We illustrate this in Figure 3.

3.3 Communication of encoding parameters

MPEG-DASH media presentation description file allows encoders to share specific parameters of each encoded version of the content. In case of video, these parameters include resolution (width $\times$ height), pixel aspect ratio, frame rate, and required bandwidth. When the content is prepared, the encoder may choose to use different combination of these parameters to produce encodings for each target bitrate. The encoder may

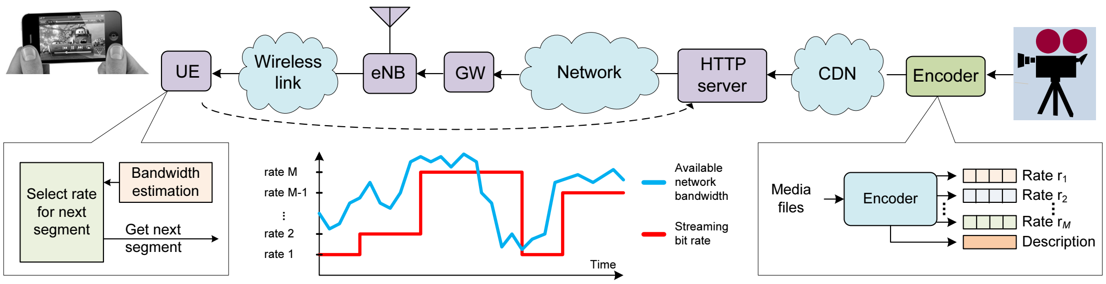

Figure 3. Illustration of functionality of mobile DASH-based streaming system. The multimedia content is encoded at encoded at multiple rates, and segmented in chunks allowing client to select portions that can be delivered in real-time, while also adopting to changing network bandwidth.

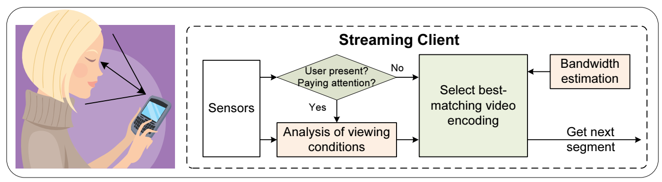

Figure 4. Illustration of functionality of DASH streaming client incorporating adaptation to user behavior and viewing conditions.

also produce multiple encodings considering different screen resolutions and other specific capabilities of target devices, allowing streaming clients to pick versions that are optimized for each particular device.

4. ENABLING ADAPTATION TO USER BEHAVIOR AND VIEWING CONDITIONS

We provide conceptual illustration of user-adaptive design of DASH streaming client in Figure 4. In order to adapt to viewing conditions, the client uses sensors of a mobile device, such as front-facing camera, accelerometer and gyroscope to detect the presence of the user, his proximity, and viewing angle. The client also uses ambient illuminance sensor and information about brightness settings of the screen to estimate effective contrast ration of the screen.

Using these estimates, the client obtains minimum characteristics of encoded video, such as spatial resolution, framerate, and bitrate that are sufficient to achieve high level of visual quality. In finding such characteristics

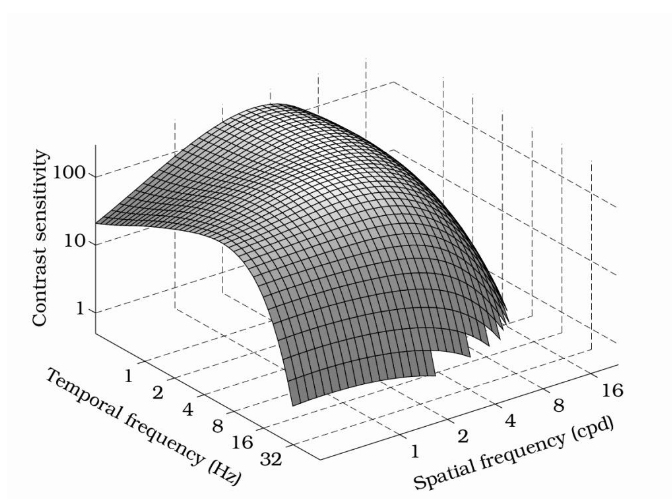

Figure 5. Model of statio-temporal contrast sensitivity function of human vision.[13] The x/y- axes represent the spatial and and temporal frequencies of a contrast-reversing pattern. The vertical axis represents the observer contrast sensitivity thresholds obtained for each contrast-reversing pattern.

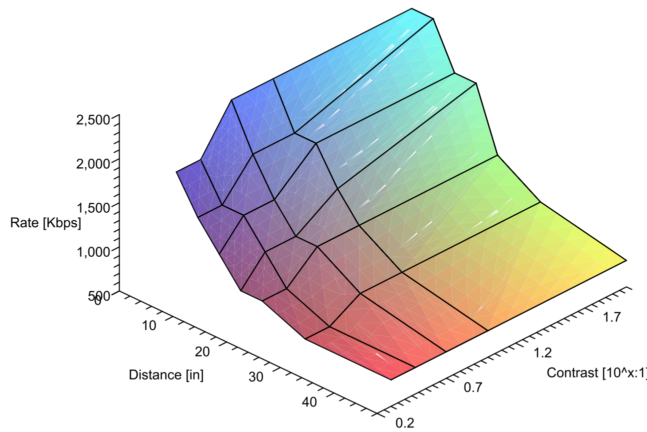

Figure 6. Example of rate allocation achieving approximately the same level of perceived quality under different viewing distances and contrast rates. This particular allocation was obtained assuming reproduction on a mobile device with 720p-resolution screen, 340 dpi pixel density, and when using the H.264 (Main profile) video encoder.

the client can use spatio-temporal contrast sensitivity functions[13] (see Figure 5) or other related results from studies on human vision and video coding. Once such characteristics are obtained, the client searches through a list of available video representations and selects one that is best suited for delivery.

As illustrated in Figure 6, adaptation to viewing distance and contrast can result in lowering bandwidth required to receive video. Increase in viewing distance lowers our ability to discern individual pixels and hence it becomes possible to select representations encoded using lower resolution and bitrate. Likewise increase of ambient illuminance lowers effective contrast of the screen and range of spatial frequencies that we can see. This also opens opportunity for lowering the resolution and required bitrate. For additional details the reader is referred to our related publications.[7–10]

In cases when client detects that user is not present next to the device, even more significant bandwidth savings are possible. For example, the client may stop receiving video while continuing playing only audio track.

As bandwidth usage is directly related to power consumption in mobile phones, the above described optimizations can also result in increased battery life. Additional benefits may include reduced congestion and re-buffering probability and improved quality of user experience.

5. CONCLUSION

We have shown that MPEG-DASH standard enables design of intelligent streaming systems adapting not only to bandwidth but also to factors affecting user ability to see visual information. Such adaptation can result in reduced bandwidth usage, increased battery life, and improved quality of user experience.

REFERENCES

[1] D. Wu, Y.T. Hou, W. Zhu, Y-Q. Zhang, and J.M. Peha, "Streaming video over the Internet: approaches and directions," *IEEE Trans. Cir. Syst. Video Tech.*, Mar 2001, vol. 11, no. 3, pp. 282-300.

[2] G. J. Conklin, G. S. Greenbaum, K. O. Lillevold, A. F. Lippman, and Y. A. Reznik, "Video Coding for Streaming Media Delivery on the Internet," *IEEE Trans. Cir. Syst. Video Tech.*, Mar 2001, vol. 11, no. 3, pp. 20-34.

[3] I. Sodagar, "The MPEG-DASH Standard for Multimedia Streaming Over the Internet," *IEEE Multimedia*, Oct-Nov, 2011.

[4] ISO/IEC 23009-1 *Information Technology – Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats*, ISO/IEC, January 5, 2012.

[5] A. Talukdar, M. Cudak, and A. Ghosh, "Streaming Video Capacities of LTE Air Interface," *Proc. IEEE Int. Conf. Comm. (ICC)*, 2010, pp. 1-5.

[6] J. Bergquist, “Resolution and contrast requirements on mobile displays for different applications in varying luminous environments,” *Proc. 2nd Int. Symp. Nanovision Science.*, 2005, pp. 143-145.

[7] Y. Reznik, et al., “User-adaptive mobile video streaming,” *Proc. of IEEE Visual Communication and Image Processing*, Aug 2012.

[8] R.Vanam, Y.Reznik, “Improving the Efficiency of Video Coding by using Perceptual Preprocessing Filter,” *Proc. Data Compression Conference*, March 2013.

[9] R.Vanam, Y.Reznik, “Perceptual pre-processing filter for user-adaptive coding and delivery of visual information”, *Proc. Picture Coding Symposium*, Dec 2013.

[10] Y.Reznik, R.Vanam, “Improving coding and delivery of video by exploiting the oblique effect”, *Proc. IEEE Global Conf. Sig. Inf. Processing*, Dec 2013.

[11] Y. Bababekova, M. Rosenfield, J. Hue, and R. Huang, “Font Size and Viewing Distance of Handheld Smart Phones,” *Optometry and Vision Science*, July 2011, vol. 88, no. 7, pp. 795-797.

[12] J. Young, M. Trudeau, D. Odell, K. Marinelli, and J. Dennerlein, ”Touch-screen tablet user configurations and case-supported tilt affect head and neck flexion angles,” Work: A Journal of Prevention, Assessment and Rehabilitation, Volume 41, Number 1, 2012, pp. 81-91.

[13] D. H. Kelly, “Motion and vision. II Stabilized spatio-temporal threshold surface,” *Journal of the Optical Society of America*, 1969, vol. 69, pp. 1340-1349.