Signal Processing: *Image Communication* 33 (2015) 41–50

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Towards quality of experience determination for video in augmented binocular vision scenarios

Patrick Seeling

Department of Computer Science, Central Michigan University, Mount Pleasant, MI 48859, USA

A R T I C L E I N F O

Article history:
Received 5 June 2014
Received in revised form
25 February 2015
Accepted 25 February 2015
Available online 5 March 2015

Keywords:
Augmented reality
Multimedia systems
Perceptual quality
Quality of experience
Quality of service

A B S T R A C T

With the continuous growth in the consumer markets of mobile smartphones and increasingly in augmented binocular vision wearable devices, several avenues of research investigate the relationships between the quality perceived by mobile users and the delivery mechanisms at play to support a high quality of experience for mobile users. In this paper, we present the first study that evaluates the relationships of mobile movie quality and the viewer-perceived quality thereof in an augmented binocular vision setting employing commercially available head-mounted see-through devices. We find that participants tend to overestimate the video quality when compared to a scaled representation and exhibit a significant variation of accuracy that leans onto the movie content and its dynamics. Our findings, thus, can broadly impact future media adaptation and delivery mechanisms for this new display format of mobile multimedia and spur follow-up research in this increasingly popular domain.

1. Introduction

In recent years, the amount of connected devices that are carried by mobile users has increased drastically and will become one of the dominant drivers for future mobile networking, as described by Cisco, Inc. [1]. A secondary forecasted trend is the continuously large fraction of mobile data that is required due to multimedia consumption while users are "on-the-go." While currently, smartphones and tablet computers are the dominant form of media consumption and display, the prospect of reality-augmenting wearable devices will likely account for a significant portion of the interaction with mobile multimedia content in future immersive communication systems [2]. Augmented reality has been an area of research in ubiquitous computing for some time [3] and is subject to ongoing research efforts [4]. Several application scenarios were evaluated in recent years in different domains, such as smartphone-based information overlay systems [5,6], outdoor systems with multiple elements [7,8], navigation [9], or general information systems combining both [10].

Several industry-based solutions were developed recently in parallel to augmented reality devices, which target the multimedia playback application scenario in the predominantly consumer market space. Augmented reality devices that are performing in a heads-up-display (HUD) or head-mounted display (HMD) manner are increasingly targeting the professional and consumer application space alike, indicating future broad adaptation. While these types of devices are available in a broad variety of implementations (see, e.g., [11] for an overview of different types), a slow convergence of systems has begun, especially in the consumer space. Examples for current commercial off-the-shelf (COTS) devices available include the Occulus Rift, Sony HMZ-T1 Personal 3D Viewer, Epson Moverio BT-100, or Google Glasses. We note that only the latter two are optical see-through devices and thus similar to the one presented in [12], showcasing how these device types have undergone additional improvements and are now consumer-graded.

The evaluation of these types of systems and related issues has attracted different research groups and a recent survey [13] highlights ongoing issues for the various system

E-mail address: patrick.seeling@cmich.edu
URL: http://patrick.seeling.org

http://dx.doi.org/10.1016/j.image.2015.02.006

types. Evaluations performed also target the user-perception of augmentation for daily life scenarios, such as in [14], or how to limit the amount of additional information, as in, e.g., [15]. Perceptual evaluations oftentimes consider the segmentation of virtualized/augmented items, such as in, e.g., [16]. There are several constraints that have to be considered in this particular scenario, especially from a communications point of view, when targeting the delivery of video data to these types of systems, as the replication of video content with natural features might behave significantly different from overlaid computer-generated information. In [17], the authors evaluate an industrial system that consists of an opaque HMD that displays video sequences at different target bit rates (and resulting imperfection or distortion levels) and user select different encoder properties at the target rate-points, resulting in a combination of frame rate and compression. Our evaluation is significantly different in that we provide participating users with a see-trough COTS HMD at prescribed video frame rates. Significant differences can be expected for the perceived video quality based on the type of the visual display [13].

In this paper, we investigate the applicability of existing video quality metrics, such as the frequently used peak signal-to-noise ratio (PSNR), structural similarity index metric (SSIM), and video quality metric (VQM), in the augmented reality space and correlate encoded video quality with subjectively rated perceived video quality levels. The perceptual video quality is measured using mean opinion scores (MOS) obtained from multiple human test subjects according to [18]. We note that currently, no specific testing standard has been established for the determination or evaluation of audio-visual objective or perceptual qualities employing COTS see-through devices or for augmented reality settings. In turn, we consider the existing standard as outlined in recommendation ITU-T BT.500-13 [18] as a general guideline for the experiments we perform here. In this seminal work, we provide a high-level overview of subjective quality ratings for longer movie segments performing an initial view at the underlying characteristics at play.

The broad potential for implementation in future systems that contain augmented single view or binocular vision display modalities is manifold, as media adaptations for specific video material might benefit content and network providers while maintaining a sufficiently high level of quality of experience [21]. Here, we focus on the evaluation of video compression quality without potentially lossy network transport when viewed in a scenario where binocular vision is augmented.

In the remainder of this contribution, we initially describe the measurement setup (including the wearable device, the developed mobile application, and the encoded video sequence characteristics) in greater detail. We continue with a detailed description of the obtained results in Section 3 and evaluate the participating users' video quality selection performance in Section 4. We conclude with an outlook on future activities in Section 5.

2. Measurement setup

In this section, we initially describe the employed wearable head-mounted optical see-through display and the application we developed for the experimentation. We continue with a description of media characteristics and performed experimentation with volunteering participants.

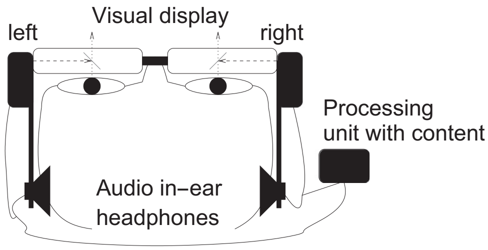

Fig. 1. Schematic view of the head-wearable binocular see-through glasses with in-ear headphones and the main processing unit that additionally stores the audio-visual content for local playout.

2.1. Binocular augmented vision heads-up display

We employed the Epson Moverio BT-100 mobile viewer, which consists of a wearable 3D-capable heads-up display unit and a central processing unit. The processing unit features both a directional pad and a touch pad and employs the Android Operating System version 2.2 (“Froyo”). The unit is connected via wires to deliver video signals and power to the see-through display unit, with a display control being integrated into the wired connection. We illustrate the overall system configuration in Fig. 1, noting that no networking is involved for the display of the content, as it is contained on the processing unit.

The display unit has a resolution of 960×540 pixels of 24-bit color with LED light sources and a 23 degree field of view. Without the additionally available shade, a maximum of 70% transparency is realized for the display. The images are projected from a display panel built into the device's sides, from which light is reflected through a lens, and in turn hits a half-mirror layer in the light guide material. As we consider an evaluation of a commercially available off-the-shelf augmented binocular vision HMD, no specific calibration was performed. Factory settings were applied, which allow for 24-bit color reproduction at 60 Hz and the built-in LED light intensity was set to maximum for highest contrast.

2.2. Mobile application

We developed a mobile Android application that can be executed on the wearable display's control unit. The application displays a movie sequence (video and audio content), followed by a Likert-scale question to rate the quality of the last displayed sequence. We illustrate this approach in Fig. 2.

Initially, a random quality for a movie sequence is selected and its value stored in a text file on-device before starting the audio/video play-out from the on-device storage, disabling potential network transmission impacts. After play-out, the user is asked to select a quality level from a presented Likert-scale, with each selection of a quality level captured in the same text file on-device. We do not enforce a time limit for the rating procedure, as users need to interface with the mobile application through the processing unit's touchpad. Adding a time limit at this stage would increase the potential stress on participants as they make their selections, in turn potentially influencing their ratings. This