Computer Communications 125 (2018) 56–64

Contents lists available at ScienceDirect

Computer Communications

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

Performance assessment of three-dimensional video codecs in mobile terminals

Almudena Sánchez*, Matías Toril, Marta Solera, Salvador Luna-Ramírez, Gerardo Gómez

Departamento de Ingeniería de Comunicaciones, E.T.S.I.Telecomunicación, Universidad de Málaga, Bulevar Louis Pasteur, S/N, Málaga 29010, Spain

ARTICLE INFO

Keywords:
3D video
Codecs
Mobile terminals
Subjective
Objective

ABSTRACT

Understanding the most relevant factors influencing the end user experience is key to design and manage multimedia services in mobile networks. Thus, user demand can be dealt with efficient strategies of resource management. In this paper, a comprehensive analysis is carried out to evaluate the Quality of Experience (QoE) obtained by three-dimensional (3D) video codecs in mobile terminals. The analysis covers the most widespread 3D video coding formats, namely the frame-compatible Side-by-Side (SbS) and Multiview Video Coding (MVC) schemes. The analysis considers both subjective and objective measurements over compression strategies that combine bit rate reduction and frame-rate dropping. Subjective tests are done by presenting 3D video sequences with different coding parameters to users who judge image quality, depth perception and visual comfort. For this purpose, a mobile phone with autostereoscopic screen is used. Then, mean opinion scores are compared with objective measurements obtained by a video quality measurement tool. Results show that MVC outperforms SbS in terms of image quality. However, when internal parameter settings are set to very restrictive configurations, the impairment of the original image are similar for both codecs. Likewise, results show that the reduction of the video bit rate is the key parameter controlling video quality, and that the use of frame-rate dropping is not enough to counteract the impairment introduced by bit rate reduction. Finally, it is confirmed that a simple QoE indicator for 3D video can be obtained from objective measurements.

1. Introduction

In the last decade, advances in consumer electronics have led to an unprecedented growth in the number of applications and devices in telecommunication networks. Youtube, Vine, and social media applications, such as Facebook, Instagram or Snapchat, allow users to upload, watch and share videos. As a result, video content has become a major contributor to Internet traffic, and it is expected that video traffic will be 80% of all consumer Internet traffic by 2019 [1].

In parallel, the introduction of tablets and smartphones has paved the way for new multimedia mobile services. New mobile terminals make it easier to watch and share videos, which has become a daily activity for many users. Mobile video is forecast to grow from 50% in 2015 to 70% in 2021 of all mobile data traffic [2]. Thus, mobile networks have to cope with the huge amount of data and strict performance requirements associated to video services.

The blooming of video applications in the entertainment area has also favoured the return of on-and-off technologies, such as three-dimensional (3D) video, which re-started in 2009 with the Avatar film. Thereafter, a large number of productions have been distributed in 3D format. 3D video has not only found its way into films, but it has also jumped into other areas, such as medicine (e.g., to help in surgery [3] or diagnosis [4]), sports (e.g., for advanced player performance analysis [5] or for investigation purposes (e.g., to monitor animal behaviour [6]). Furthermore, 3D devices have evolved from the first 3D television sets at home to intelligent 3D-capable mobile terminals with autostereoscopic screens [7,8].

To provide a successful 3D video service, it is necessary to somehow measure user's satisfaction or Quality of Experience (QoE), understood as how well a service meets customers goals and expectations, rather than focusing only on network performance, which was the goal of Quality of Service (QoS) [9].

In this paper, a thorough study of the Quality of Experience (QoE) provided by 3D video codecs in mobile terminals is presented. Two of the most well-known standards for 3D video coding, namely frame-compatible Side-by-Side (SbS) format and Multiview Coding (MVC), are analysed by performing both subjective and objective measurements. Subjective assessment is carried out by means of a smartphone with an autostereoscopic screen. In that terminal, observers are provided with a comprehensive set of 3D video sequences corresponding to different

---

* Corresponding author.
E-mail address: almudena@ic.uma.es (A. Sánchez).

https://doi.org/10.1016/j.comcom.2018.04.016
Received 3 December 2017; Received in revised form 22 March 2018; Accepted 24 April 2018

coding parameter configurations that reduce both bit rate and frames. The aim is to evaluate several 3D video aspects, such as image quality, depth quality and visual comfort. Subjective measurements, measured as mean opinion scores (MOS), are compared with classical objective measurements, such as Peak Signal-to-Noise Ratio (PSNR) and Multi-Scale Structural Similarity (MS-SSIM), obtained by a publicly-available video quality measurement tool.

1.1. Related work

Distributing high-quality 3D video content requires a large bandwidth. Thus, researchers have focused their efforts on the development of efficient 3D coding techniques to reduce the amount of transmitted data [10]. A first approach, referred to as Simulcast, consist of distributing a full-resolution stereo video by two independent views of maximum quality in parallel. Thus, computational load is kept to a minimum, although data rate is doubled compared to conventional 2D video.

Alternatively, frame-compatible formats were born to restrict the data flow of the video representation by combining sub-sampled versions of the left and right views into a single frame, which can then be coded with conventional 2D video codecs, such as MPEG2, H.264/MPEG4 Part 10 Advanced Video Coding (AVC) [11] or High Efficiency Video Coding (HEVC) [12]. Thus, operators can still use the existing infrastructure to introduce stereoscopic services at the expense of slightly deteriorating 3D video quality. An intermediate approach is Mixed Resolution Stereo Coding (MRSC), where the resolution of one view is reduced to exploit that the perceived quality of stereoscopic video with views of different sharpness is rated by the observer close to the sharper view [13].

More advanced schemes specifically designed for 3D video, such as H.262/MPEG2 Multi-View Profile [14] and H.264/MPEG4 Multiview Video Coding (MVC) [11], make use of differential coding between left and right view to exploit inter-view redundancy (stereo video). Nonetheless, the resulting coding rate is still 50% larger than that of a single view [15] because the information related to view's differences is also sent in order to restore the secondary view in the decoder. Alternatively, depth-based coding schemes (V + D, for video and depth) [16,17], add depth information (usually, a greys map) to a classical 2D video stream, enabling depth-image-based rendering (DIBR) of additional viewpoints in the decoder to generate the 3D video representation with full resolution. The resulting coding rate is only 10–20% higher than conventional video.

Another important issue of 3D video is the need for new metrics to assess video quality. Subjective metrics for stereoscopic video have significant differences compared to 2D video metrics, due to the clever way the human visual system handles similarities and dissimilarities between views. A first difference is the inclusion of depth information in video sequences, for which several video databases exist in the public domain [18]. Moreover, subjective 3D video assessment requires evaluating new indicators, in addition to picture quality, such as depth quality, naturalness, emotion, sense of presence or even the price of the network service, as recommended in ITU-BT.2021-1 [19]

Likewise, objective 3D video quality metrics also differ from those of 2D, since 3D video presents significantly different quality issues that are not encountered or do not have their equivalent in 2D. A naive approach to build 3D metrics from already existing 2D measurements is to average the value of any objective metric (e.g., Peak Signal-to-Noise Ratio [20], Structural Similarity [21], Moving Picture Quality Metric [22], or Video Quality Model [23]) for the left and right views [24]. More refined metrics take into account the properties of 3D vision (e.g., disparity [25] or binocular depth [26]). However, the latter are not currently available in open access measurement tools. In this work, the selected objective measurements are PSNR and MS-SSIM (described in detail in later sections), the latter being a multi-scale variant of Structural Similarity index. These measurements commonly used in video assessment are available in several publicly available video measurement tools, allowing the reader to replicate the study presented here. Furthermore, these tools allow the use of common video file formats (e.g., mp4, yuv, avi ...). Thus, there is no need of a preprocessing stage where the original video format is modified to compute the desired objective metric.

Regarding the novelties that this article provides, multiple references have evaluated the performance of 3D video coding schemes on large screen formats with objective [15,27], or subjective tests [28–30]. However, few studies have extended the analysis to mobile phones with small autostereoscopic screens, where not only screen size impacts MOS scores, but also other factors such as the maturity of the 3D technology or user familiarity with autostereoscopic screens. In [31,32], several stereoscopic coding schemes suitable for mobile devices are compared based on objective metrics, such as rate-distortion curves and computational complexity. These studies aim to find the best codec parameter settings depending on processing power and memory of the mobile device. However, these studies do not take into account the user subjective perspective. Nasralla et al. [33] presents subjective measurements of 3D video over a Long Term Evolution (LTE) network under packet losses following a Gilbert–Elliot Channel Model. However, it does not consider bit rate variations in the network and source, which is an important parameter affecting video quality.

A more comprehensive analysis of the experience provided by 3D codecs in mobile phones is presented in [34,35]. It was shown there that the use of the high coding profile (i.e., hierarchical B-frames and context-adaptive binary arithmetic coding) can provide the same quality as baseline profile with lower bit rates. Our study extends the results in [34,35] by also including frame dropping as a rate reduction scheme to check if a combination of both bit rate and frame dropping can obtain a higher QoE.

The rest of the paper is organized as follows. Section 2 reviews 3D video compression techniques. Section 3 describes the assessment methodology used in this work. Section 4 presents the results of the subjective assessment of 3D video coding approaches carried out with real observers. Section 5 presents the results of objective measurements. Section 6 compares the results of subjective and objective tests. Finally, Section 7 presents the main conclusions of the study.

2. 3D video coding

This section describes the 3D video coding schemes used in this work. The selected coding formats include a scheme that makes use of existing 2D video codecs (frame-compatible Side-by-Side, SbS) and a scheme based on a base view (Multiview Video Coding, MVC). Both schemes have been standardized and provide backward support to legacy devices. For this reason, SbS and MVC have become the most widespread 3D video schemes.

2.1. Side-by-Side

In frame-compatible formats, the stereo video signal is a multiplex of the two views into a single frame (or a sequence of frames). To keep the same sequence size, the left and right views are sub-sampled and merged. Decimation can be done in the spatial or temporal domain. In the former case, the two views are decimated horizontally or vertically and stored in a side-by-side (SbS), top-and-bottom (TaB) or line/column-interleaved format [36]. In the latter case, the left and right views are interleaved as alternating frames or fields. Both options are referred to as frame sequential and field sequential schemes, respectively [37].

Fig. 1 shows how SbS exploits spatial redundancy to reduce the horizontal resolution into a single frame. The process begins at the encoder by low-pass filtering the image of each view to reduce aliasing before sub-sampling. Then, the left and right views are decimated. Decimation can be performed by taking the same position of pixels in