Computer Communications 35 (2012) 1838–1845

Contents lists available at SciVerse ScienceDirect

Computer Communications

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

Depth-color based 3D image transmission over wireless networks with QoE provisions

Zhe Song [a], Honggang Wang [a,*], Yonggang Wen [b], Dalei Wu [c], Ken Lee [a]

[a] University of Massachusetts Dartmouth, North Dartmouth, USA
[b] Nanyang Technological University, Nanyang Avenue, Singapore
[c] Massachusetts Institute Technology, Cambridge, MA, USA

ARTICLE INFO

Article history:
Available online 24 March 2012

Keywords:
QoE-driven
3D image
Wireless transmission
Depth map
Patch map

ABSTRACT

The deployment of 3-D image techniques is one of the most promising fields among the development of new applications for natural image scenes. Driven by urgent demands from industry and users, 3-D image technology has received significant research attention in recent years. 3D image streaming gives users an extra dimension of visual sense that greatly improves the liveliness and joyfulness of user experience. However, they also raise new challenges (e.g., bandwidth and energy efficiency) especially in achieving satisfactory Quality of Experience (QoE) performances. QoE has become critical multimedia quality metrics determining whether a potential multimedia application or service is successful or not. In this paper, we propose a QoE-driven wireless 3D image transmission scheme with depth-color source coding adaptations according to the wireless network conditions. Specifically, our contribution includes: (1) developing a patch-pixel based source coding scheme for 3D image transmission; (2) proposing a 3D image quality model; (3) and developing a quality-driven 3D image transmission approach based on the quality model. Experimental results demonstrate that the proposed techniques can significantly improve the QoE of 3D image over wireless networks.

1. Introduction

With the rapid growth of the 3DTV and 3D mobile applications in the world, the research on 3D images and video has been drawn much attention from both industry and academia. As a new type of media, multiview imaging (MVI) [1] has attracted increasing attention, thanks to the rapidly dropping cost of digital cameras. This opens a wide variety of interesting new research topics and applications, such as virtual view synthesis, high performance imaging, image/video segmentation, object tracking/recognition, environmental surveillance, remote education, industrial inspection, 3DTV, and free viewpoint TV (FTV) [2]. While some of these tasks can be handled with conventional single view images/video, the availability of multiple views of the scene significantly broadens the field of applications, enhancing the resulting performance and user experience. 3DTV and FTV are some of the most important applications of MVI and are new types of media that expand the user experience beyond what is offered by traditional media. They have been developed by the convergence of new technologies from computer graphics, computer vision, multimedia, and related fields. 3DTV, also referred to as stereo TV, offers a three-dimensional (3D) depth impression of the observed scene, while FTV allows for an interactive selection of viewpoint and direction within a certain operating range. To enable the use of 3DTV and FTV in real-world applications, the entire processing chain, including multiview image capture [3], 3D scene representation [4], coding [5], transmission [6], rendering [7], and display [8], needs to be considered. There are numerous challenges to implement such a processing chain and further develop its mature systems for wide applications. To overcome these challenges, a variety of research work has been carried out on each component of the processing chain.

3D scene representation formats integrate various types of data, such as multiview video, and geometry data in form of depth or 3D meshes. In general, these result in a tremendous amount of data that needs to be transmitted or stored. Therefore, efficient compression is a key condition for the success of such applications. The compression of 3D data has recently received much attention in research and development. Technology has reached a good level of maturation. However, since the field is still very young compared for instance to classical 2D video coding, there is still a lot of room for improvement and optimization. 3D geometry can also be represented by per-pixel depth data associated with the color image. Depth needs to be clipped in between 2 extremes Znear and Zfar, and the range in between is most often scaled nonlinearly. Investigations have shown that such data can be encoded very efficiently, e.g. at

* Corresponding author.
E-mail address: hwang1@umassd.edu (H. Wang).

0140-3664/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.comcom.2012.03.008

5–10% of the bit rate that is needed to encode the associated color image at a good quality. This means that the extension from 2D image to 3D image comes at a little limited overhead. However, this is only true for a limited navigation range.

Further, some handheld devices such as smartphones have been started to be used to support 3D image streaming over cellular networks. As shown in Fig. 1, a typical 3D application is that users capture 3D scene by using a multi-camera based smartphone and upload the 3D image to the base station over wireless networks (i.e., cellular networks). In such networks, image applications do not only require intensive bandwidth, but also demand high QoE. The QoE has become a new measurement method to replace the traditional Quality of Services (QoS) for multimedia communication applications. There have been significant research efforts in improving QoE-driven systems. With the resolution of multimedia picture reaching a high level that is so hardly to have some big improvement, the expectation of greatly enhanced user experience raises 3D image technology to be one of the hottest topics within the area of image technology.

3D image presents a series of more lively images to users than traditional two-dimension (2D) one, and will increase the level of QoE. The QoE acts as the extension of QoS (Quality of Service) by taking into account the subjective measurement of a customer's experience, which gives people a comfortable and flexible user-environment. The challenge of this technique mainly lies on the largeness of data in contrast with the limited bandwidth of wireless communications and the quality of contents, and these two factors give a huge impact on the QoE of 3D image.

There are several stereo formats, coding schemes, and display technologies coexisting nowadays. Among them, a new technique, known as depth image-based rendering (DIBR) [9] is highly explored. It represents a 3D image based on a monoscopic image and associated per-pixel depth information (simply called color and depth maps). The color map refers to three components—Y, U, and V in the image frames while the depth map uses only one component to store the depth information of each pixel according to the real position of cameras as well as objects being filmed. With this technique, it can capture the stereoscopic sequences more easily compared to the traditional left and right view techniques, and save bandwidth as well as storage requirement. However, the technique cannot handle occlusion. Some part of information is missing due to the shift of objects horizontally when changing the viewpoint from left to right or vice versa. This results in distortion and may cause a big drop down of picture quality when objects are within a low range of depth (very close to the viewpoint).

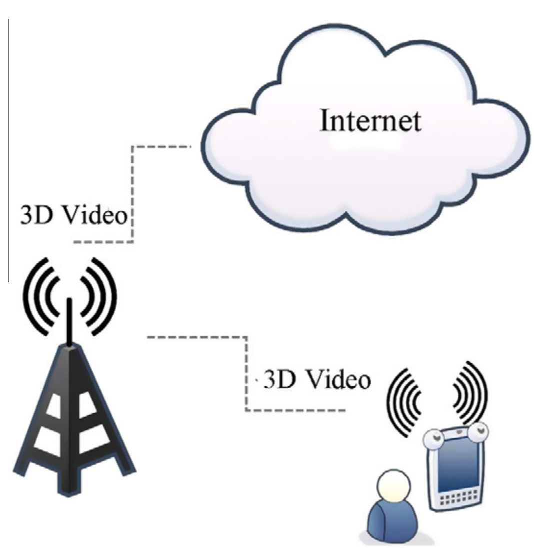

Fig. 1. A typical 3D application.

In summary, error-prone wireless channels, largeness of the multimedia data, and limited bandwidth are three major challenges for 3D multimedia transmission, which significantly impact the QoE of 3D image. In this paper, we deal with these challenges and propose a QoE-driven coding and transmission scheme based on the 2D-plus-depth technology for image applications over wireless networks [10]. The scheme includes space-domain coding, patch map based error detection on depth map and image rendering, and transmission adaptation. This paper especially gives a relatively detailed discussion on the patch map generation and patch map based error detection on depth map, scalable patch transmission strategy. The experimental results shows the feasibility and superior performance of the proposed approach.

2. Related works

A significant amount of works have been conducted in the areas of QoS–QoE analysis, 3D image and wireless multimedia transmission. We outline some of these works in this section and emphasize the innovation of our work for enabling QoE-driven 3D image through wireless networks.

In recent years, many objective and subjective image quality evaluation metrics are proposed and some of them can be found in [17–19]. In [10], a QoE-driven adaption scheme for image applications is proposed and discussed. The major idea of this scheme is to achieve bit rate adaptive control by image quality measurements based on the (Mean of score) MOS values, which are obtained by an objective and non-intrusive QoE prediction model derived from those measurements. The 3D image differentiates itself from regular image by including image from two or more viewpoints that have certain correlations. The DIBR technique takes advantage of high correspondence of the image pairs by using depth map to improve the coding efficiency and reduce the amount of data. Some of the related researches can be found in [9,11–13] and [14–16,20,23–27]. Most of these research focuses on depth map generation and distortion analysis, but few of them raises the discussion about the occlusion problem and its effect on the QoE of 3D image as well as the possibility of improving the wireless transmission scheme using the two-way relationship between the depth map and the occlusion pixels (referred to as patch pixels or patch map in the rest of the paper).

In [21], a image quality prediction model is proposed to predict the quality of a set of image frames at the sender side given the packet loss rate of the wireless channel. This prediction considers the motion prediction in image coding and error concealment in image decoding, and can predict the received image frames that are coded in both inter-mode and intra-mode. The model proposed in our paper is based on the intra-mode prediction in [21]. With the combination of DIBR technology, we propose a QoE-driven scheme for 3D image transmission over wireless channel which includes patch map generation, 3D-image quality prediction, and patch map based depth map error detection. The study presented in this paper is to improve QoE performance of wireless 3D image transmission.

3. QoE-driven coding and transmission for 3D image

Fig. 2 shows a QoE-driven transmission design for 3D image. Its major components include disparity calculator, pace domain encoder, and packetizer, image/image quality modeling, patch map based depth map error detection, and retransmission and channel coding. In this design, a disparity calculator is a generator that generates depth map with a stereo image pair. Based on depth