



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 of 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.

© 2012 Elsevier B.V. All rights reserved.

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 Z_{near} and Z_{far} , 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).

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

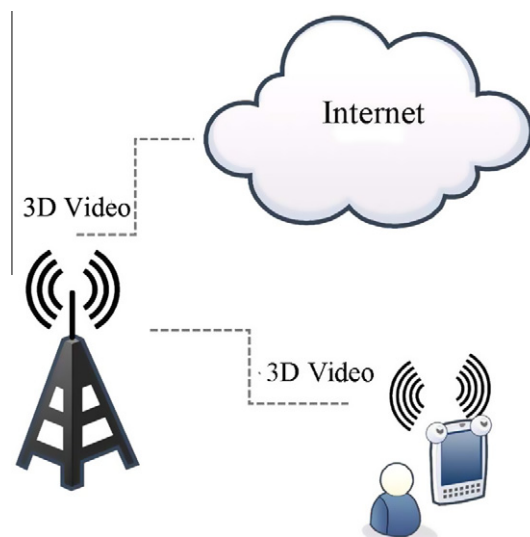


Fig. 1. A typical 3D application.

and may cause a big drop down of picture quality when objects are within a low range of depth (very close to the viewpoint).

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

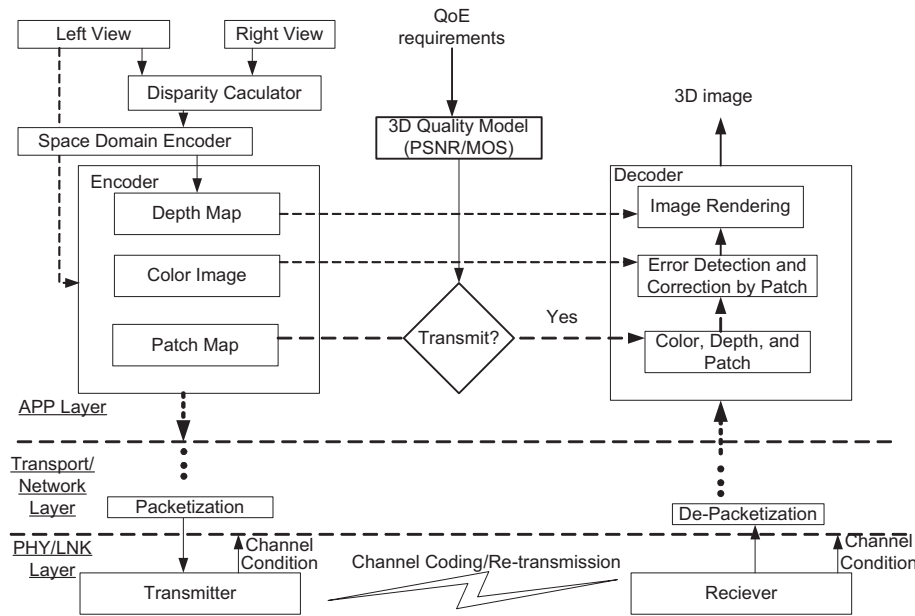


Fig. 2. QoE-driven transmission design for 3D images.

map, a space domain encoder calculates the specific ranges of pixels or information that appear in one of the image pair while getting lost in the other due to the horizontal shift of object for the change of viewpoint. At the transmitter side, color, patch and depth are encoded for the transmission. The expected quality is estimated by our proposed 3D image quality model. We compare the predicted quality with the users' QoE requirement to determine if the patch pixel should be transmitted. The encoder and packetizer are used for sender to do the source coding and packetization. At the receiver side, decoding and de-packetization are performed for constructing 3D images. An image rendering module uses the decoded information, basically, image with its depth map and a patch map, to render the stereo image pair. The 3D image quality prediction model is derived from the intra-frame image quality estimation by taking into account error concealment techniques at the receiver's site. The model is used to predict image quality from network QoS parameters such as packet error rate or application QoS parameters such as source coding rate. These predicted metrics give feedback information to the sender side. In the paper, we measure the image quality in terms of PSNR (Peak Signal to Noise Ratio) by comparing the reference image pairs with the rendered ones.

In this paper, we mainly consider 3D image quality measured by PSNR which closely related to QoE performance. However, the QoE results need to take the end users' perception into the consideration. The end users' perception of service quality is critical for quality evaluation of multimedia applications. User's perception of quality is widely evaluated by the Mean Opinion Score (MOS). As described in [22], a typical conversion between PSNR and MOS is that the values of PSNR (dB) ">37, 31–36.9, 25–30.9, 20–24.9, <19.9" corresponds to MOS values "5, 4, 3, 2, 1", respectively. Therefore, we assume that the MOS values are obtained from PSNR by the mapping conversion as described in [22]. The challenges of QoE measurement still remain and further comprehensive studies will be conducted in our future works. One major innovation in the design is the development of a patch map based depth map error detection/correction scheme. In contrast to the common mechanism using channel code to protect data, our method achieves the error detection and error correction capability by using patch map which is part of the information data and only introduces a little

extra overheads (described in Section 5). While error on depth map is detected, a retransmission request is being sent to the sender side for the retransmission. In our proposed scheme, the patch map generation, which distinguishes ours from other methods or technologies, is critical for the rendered 3D image transmission. In the following sections, we give our evaluation and experiment results to demonstrate the proposed scheme.

4. Patch map generation

Stereoscopic impression is formed by views from two parallel viewpoints. Since distance between two viewpoints is relatively short, the modification between two views is not very significant, especially when objects are much away from the viewpoints. Generally, when generating the right image based on the left image and its depth map, we need to shift the pixels of the left image at different extent according to the depth map. Here the depth map shows luminance in proportion to the distance from the camera. Nearer surfaces are darker; further surfaces are lighter. So the pixels within a darker part will be moved a longer distance in comparison with those within a lighter area in order to construct the right image. In this procedure, some part of the right image can be constructed from the left image directly while some others cannot. Some parts of viewed objects are blocked when seen from the left but show up when seen from the right. So here comes a limitation of signal channel depth map. We raise a method of selecting some pixels directly from the right image which cannot be derived from the left image. In the following we will discuss the algorithm of selecting pixels as well as corresponding efficiency and effectiveness issues.

Suppose the depth map has 8 bit depth, which means that it can represent a range of up to 256 different distances—from 0 (all white, furthest) to 255 (all black, nearest). The '0' part does not move while the '255' part moves over the longest distance when view moves from left to right. A simple scenario is that, according to the depth distance and the distance between the two viewpoints, we assign a degree of pixel shifting to each of the depth level and use this to shifted pixels. In this process, the left image

points at locations (x_L, y) are transferred to new locations (x_R, y) for the right view. This process is defined with:

$$x_R = x_L - \alpha_x \times t_c \times \left(\frac{1}{Z} - \frac{1}{Z_c} \right), \quad (1)$$

where α_x is the focal length of the reference camera expressed in multiples of the pixel width and t_c is the distance between the left and right cameras. Z_c is the convergence distance located at the zero parallax setting (ZPS) plane and Z denotes the depth value of each pixel in the reference view. Note that the y component is constant since the virtual cameras used to capture the virtual views (left-right) are assumed to be located at the same horizontal plane.

Depth map sequences have characteristics that are quite different from those of standard color image or image. Since depth map rarely contains any texture and is predominantly flat with sharp edges marking the boundary between objects at different depth, compensation pixels are needed across these edges between different levels of offset.

To explore the compensation pixels quantitatively, we assume that at a specific range of edge between two depths (Z_a, Z_b) , the length along axis Y is Y_{ab} . To notify that the offset of pixels only occurs along axis X , so the quantities of compensation pixels are determined by Y_{ab} , Z_a and Z_b .

As defined above, the scenario is that given the left image and its depth map, we need to synthesize the right image, so pixels of objects at different depth should all together be shifted towards left. When scanning the depth map from left to right, compensation pixels are needed when depth increases, which is $\Delta Z(x, y) = Z(x+1, y) - Z(x, y) > 0$. As described in [5], under the assumption of a parallel camera setting, the relationship between pixel position offset $\Delta P(x, y)$ and the depth of pixel $Z(x, y)$ can be written as a horizontal translation:

$$\Delta P(x, y) = \alpha \times t_c \times \frac{Z(x, y)}{255} \times \left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}} \right), \quad (2)$$

where α is the focal length of the camera in the horizontal direction with the unit of pixels, t_c is the distance between two cameras (horizontal), and Z_{near} and Z_{far} are the nearest and the farthest depth values, which correspond to the values of 255 and 0 in the depth map, respectively. This reveals that there is a linear relationship between the depth $Z(x, y)$ and the pixel position offset,

$$\Delta P(x, y) = k_1 \times Z(x, y), \quad (3)$$

where k_1 can be calculated as

$$k_1 = \alpha \times t_c \times \frac{1}{255} \times \left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}} \right). \quad (4)$$

Table 1

The algorithm for the formation of patch pixel.

<p>Input: left image (image to be transmitted), right image (image to be synthesized when rendering), depth map of left image k_1; Output: patch map $P[][]$; for each row i of the depth map matrix D</p> <pre> { for each column j of the depth map matrix D { if $D[i][j+1] < D[i][j]$; ** it has depth information Define variable $edgeleft=j - D[i][j] * k_1$; Define variable $edgeright=j + 1 - D[i][j+1] * k_1$; Fill three components (red, green and blue) from the right image matrix, starting from $R[i][edgeleft][1]$, $R[i][edgeleft][2]$, $R[i][edgeleft][3]$ to $R[i][edgeright][1]$, $R[i][edgeright][2]$, $R[i][edgeright][3]$ into the patch map $P[][]$; } } </pre>
--

Now, given the pixel position offset, we propose an algorithm described in Table 1 to choose the pixels in the right image, and all the images are stored in three-dimensional matrices $(n * n * 3)$ and depth map is stored in two-dimensional matrix $(n * n)$:

5. 3D image quality prediction model

As shown in Fig. 2, our scheme requires one color frame and its corresponding depth map and patch pixel to be transmitted together. At the receiver side, another color frame is reconstructed from these three components and a stereo frame pair is formed. Notice that distortions are both on transmitted color frame and depth map. So to predict the overall quality of the 3D image, we need to evaluate the transmitted frame and the reconstructed frame separately. We assume that the packet loss rate p is available at the sender side. This can either be specified from the initial negotiations, or calculated based on the transmission protocol. The original image frame n is denoted by f_n , and the corresponding depth map is denoted by d_n . The bits are packetized before transmission. The packets are constructed such that the loss of one packet does not affect the decoding of other received packets.

When a packet is lost, an error concealment technique is applied for estimating the missing segment. The reconstructed value of frame n at decoder is denoted by F_n , while its depth map is D_n . For pixel i in frame n , f_n^i denotes its original value and F_n^i denotes its reconstructed value at decoder. The original depth value and the reconstructed depth value are denoted by d_n^i and D_n^i respectively. The overall expected distortion for this pixel in the color frame is

$$E(s_n^i) = E(f_n^i - F_n^i)^2 \quad (5)$$

We observe that both color frame and depth map experience packet loss, so the distortion of a transmitted color frame is different from that of a reconstructed color frame. We consider these two cases separately.

5.1. Transmitted frame distortion estimation

First, we assume that the pixel i in the current frame is correctly received, which means that $F_n^i = f_n^i$, and the probability of this event is $(1 - p)$. If the packet that this pixel i belongs to is lost, then the previous frame $(n - 1)$ is checked. If the previous GOB (refer to [21]) is received, then the median motion vector is calculated and associate pixel i with pixel k in the previous frame. Thus we can have $F_n^i = f_{n-1}^k$, and the probability is $p(1 - p)$. If the previous GOB is lost, then the motion vector is set to be 0 and we have $F_n^i = F_{n-1}^i$ with the probability of p^2 . Combining the three cases discussed above, the first and second moments of F_n^i are calculated as follows (also given in [21]):

$$E\{F_n^i\} = (1 - p)(f_n^i) + p(1 - p)E\{F_{n-1}^k\} + p^2E\{F_{n-1}^i\} \quad (6)$$

$$E\{(F_n^i)^2\} = (1 - p)(f_n^i)^2 + p(1 - p)E\{(F_{n-1}^k)^2\} + p^2E\{(F_{n-1}^i)^2\} \quad (7)$$

5.2. Reconstructed frame distortion estimation

The prediction of reconstructed frame includes the prediction of both the transmitted color frame and depth map. We apply the same estimation strategy to the depth map, thus the D_n^i is given by

$$E\{D_n^i\} = (1 - p)(d_n^i) + p(1 - p)E\{D_{n-1}^k\} + p^2E\{D_{n-1}^i\} \quad (8)$$

We define $M_{ij}(\cdot)$ as a mapping from pixel j in a transmitted color frame to pixel i in the corresponding reconstructed color frame which means that i and j are the same 'pixel' from two different

point of view. If the transmitted frame is the left frame, then we use r_n to indicate the original value of the right frame (frame to be reconstructed at the receiver side), and R_n as the corresponding reconstructed value. Thus in the original stereo image frame pair, we have

$$r_n^i = M_{ij}(f_n^i, d_n^i) \quad (9)$$

Then the first and second moment of R_n^i can be calculated as

$$E\{R_n^i\} = M_{ij}(E\{F_n^i\}, E\{D_n^i\}) \quad (10)$$

$$E\{(R_n^i)^2\} = M_{ij}(E\{(F_n^i)^2\}, E\{(D_n^i)^2\}) \quad (11)$$

Then, the expected distortion D_n^i of the reconstructed pixel by considering both the constructed color map and depth map can be expressed as a function of $E\{R_n^i\}$ and $E\{(R_n^i)^2\}$, i.e.,

$$D_n^i = f(E\{R_n^i\}, \{(R_n^i)^2\}), \quad (12)$$

where function $f()$ can be obtained by performing data fitting.

6. Patch map based error detection on depth map

Intuitively, both the quality of color map and depth map affects the quality of 3D image. In our proposed scheme, whether or not the patch map is transmitted with color and depth maps depends on the required QoE as shown in Fig. 2. In order to achieve a satisfactory image quality, any of these three parts needs to be protected for the transmission. Let $\Delta Depth(x, y)$ be a depth change in a horizontal scan like the scenario above, and this change leads to a generation of a specific range of patch pixels. Also in the same way, we can define the other kind of depth changes as $\Delta Depth'(x, y)$. In this case, no pixels are missed, but some are discarded. We let 'D' indicate the ranges of pixels associate with $\Delta Depth'(x, y)$. In fact, we don't need the actual pixels for D, so D just consists of several pairs of starting and ending points of pixel ranges, which is in an very small data quantity. Therefore, the transmission overheads even can be ignored.

The main idea of the patch map based error detection on depth map is that, at the receiver side, when $\Delta Depth(x, y)$ or $\Delta Depth'(x, y)$ occurs, we can detect whether it's valid or an error by looking up the 'Patch' and 'D' according to the relationships $\Delta Depth(x, y)$ and $\Delta Depth'(x, y)$, which are sufficient and necessary conditions. This conclusion lies on the premise that 'Patch' and 'D' are under a very high level protection and generally error free. So under this condition, if there is an error detected, a retransmission request will be sent to the sender side to resend the related data. Intuitively, the depth map contains far less data than the color map, so it will be more efficient to just deal with the depth map than the whole 3D image. We perform some experiments about this mechanism and show the results in Section 6. In the 3D image wireless transmission system discussed above, the decoder side will perform estimation on those lost packets or pixels. This estimation would result in inaccurate results comparing to real values, which lead to distortions on both the color frames and depth frames. We perform the error detection in our simulation under the assumption that all the patch pixels are highly protected and perfectly received. The simulation results are shown in Section 7.

7. Experiment results

The experiment is based on a pair of intrusive stereo images (i.e., left and right images) as shown in Fig. 3, which uses color segments to stand for items at different depths. The difference of segment positions corresponds to the predefined parameters and the depth values of each one.

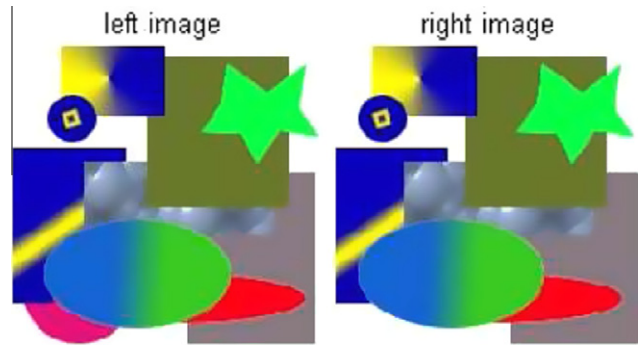


Fig. 3. Left image and right image.

The depth map for the left image is shown in Fig. 4. Using our algorithm, the patch map is calculated and the result is shown in Fig. 4(b). Fig. 4(c) shows the rendered image with the patch map. As shown in Fig. 4(c), the distortion of the rendered image is mainly because of the inaccuracy of the depth map which as a result causes distortions on the patch map. Intuitively, if the patch map is fully accurate, rendered image should be exactly the same as the original one. So no more comparisons are made in an experimental way. The graphical display of pixel range set D is displayed in Fig. 4(d). In the second experiment, we discuss the accuracy of the proposed image quality prediction model. In [21], the performance of this method has been discussed under a 2D image condition. In our experiment, we explore the prediction on both frames of the stereo pair. One example of the original image frame pair and the corresponding depth map is shown in Fig. 5.

There are ten objects in a frame, and each has a different depth value. Objects are moving through the video. No errors are considered in this experiment, and the channel has a packet loss rate of p . In Fig. 6, the sequence with 17 color frames and 17 depth frames is encoded, packetized and transmitted through a wireless channel with a packet loss rate of 0.3.

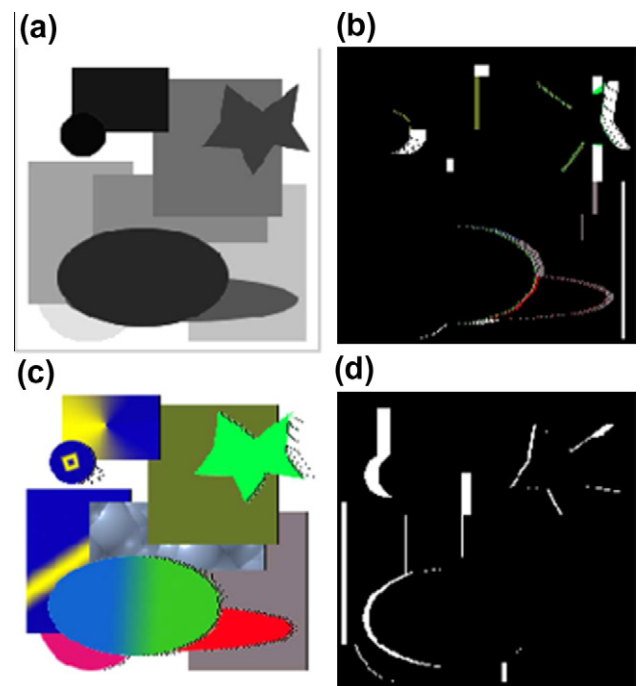


Fig. 4. Depth map.

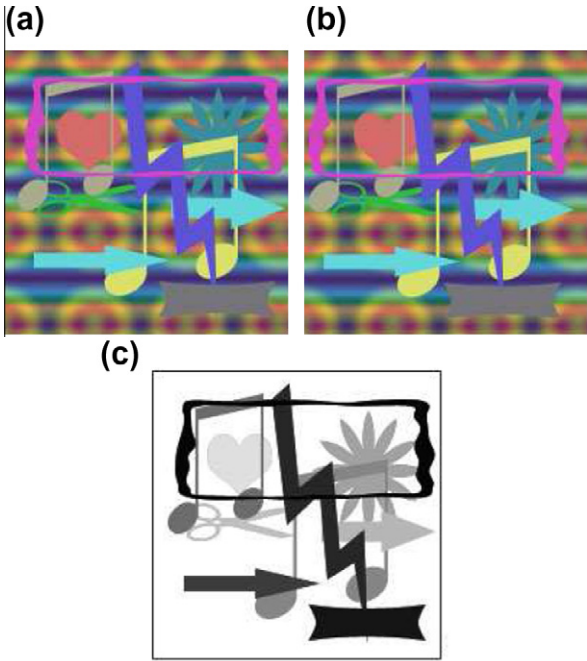


Fig. 5. Frame pair and depth map.

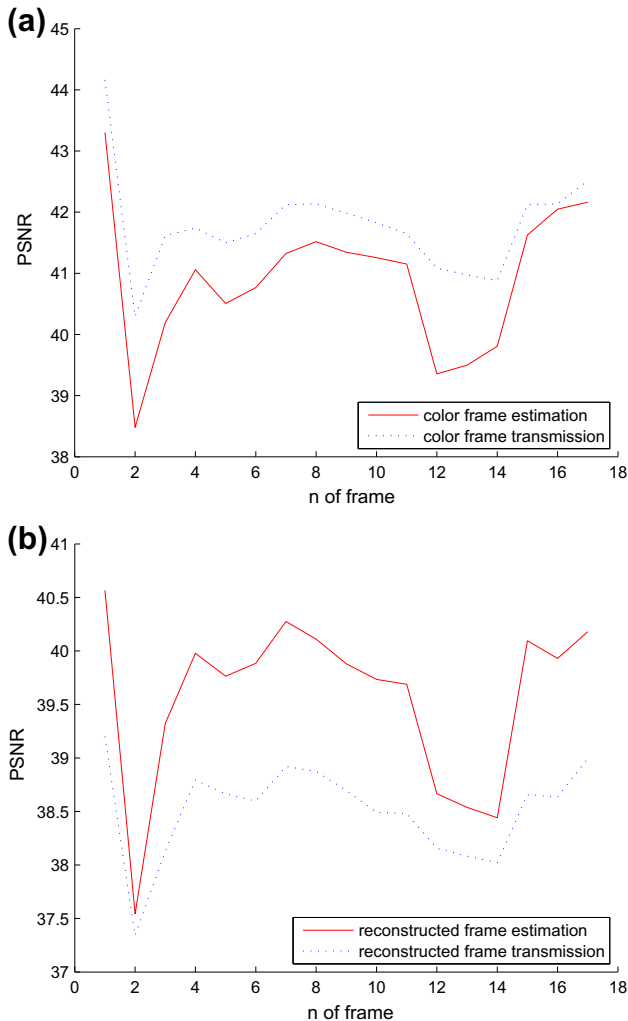


Fig. 6. Estimated PSNR and actual PSNR for transmitted color frame and reconstructed image frame with $p = 30\%$.

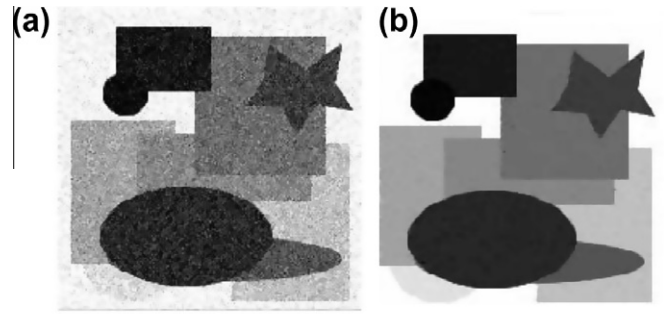


Fig. 7. (a) Depth map with additive white Gaussian noise; (b) A refined depth map.

Fig. 6(a) shows the comparison between the estimated PSNR using our proposed model described in Section 5 in and the actual PSNR of the transmitted color frames. Fig. 6(b) shows the comparison between the estimated PSNR and the actual PSNR of the reconstructed 3D image frames. Intuitively, the prediction of transmitted color frames is generally more accurate than the prediction of reconstructed image frames. Furthermore, it is interesting to note that by using this model, the transmitted color frames keep getting lower than actual quality estimation while the reconstructed color frames keep getting higher than actual quality estimation. It is observed that the depth estimation is also critical for the accuracy of the model. In general, our model can be used to estimate the transmitted 3D image quality in an accurate manner. In our experiment, we assume that the wireless connection is through AWGN channel which is represented by a series of outputs Y_i at discrete time event index i . Y_i is the sum of the input X_i and noise, Z_i , where Z_i is independent and identically distributed and drawn from a zero-mean normal distribution with variance n (the noise). Z_i is further assumed to not be correlated with the X_i

$$Z_i \sim N(0, n) \tag{13}$$

$$Y_i = X_i + Z_i \tag{14}$$

Fig. 7(a) shows the received depth map without any protection through AWGN channel. After applying the patch map based error detection and retransmission for error correction, we can get a refined depth map as shown in Fig. 7(b):

Another experiment takes ten different BER values of error depth map into consideration. We show the values in Table 2:

From the results shown in Fig. 7 and Table 2, we can draw the following conclusions: (1) The BER of depth map can be highly reduced, but not totally error free using the algorithm of this experiment; (2) There is no linear relationship between the error depth map BER and the error corrected depth map BER; (3) The experiment results vary randomly within the scope, which means that even we have the same error depth map BER, the value of error corrected depth map BER may be different; Further experiments are done by taking 10000 samples of variance of the AWGN from 0.001 to 0.1001. Table 3 shows three values of the ratio of BER of corrected depth map to BER of error depth map:

In these three scenarios as shown in Table 3, the patch map based error detection on depth map performs nearly perfect error detection. When applying this error detection to our proposed mechanism, the distortions are considered to be caused by packet loss through the wireless channel. Fig. 8 shows the PSNR comparison between the original depth map at the decoder side and that after the patch pixel based error detection (assuming the retransmission is successful). The quality is improved by 7.5232 dB on average.

In Fig. 9, we compare the PSNRs of the reconstructed color frame before and after performing the error correction and retransmission

Table 2
Error corrected depth map BER.

Error depth map BER	0.8462	0.8487	0.8512	0.8513	0.8487
Error corrected depth map BER	0.1340	0.1267	0.1222	0.1120	0.1104
Error depth map BER	0.8521	0.8522	0.8495	0.8530	0.8531
Error corrected depth map BER	0.1076	0.1013	0.1000	0.0943	0.0907

Table 3
Upper Bound.

	1	2	3
Upper bound	3.943802e-002	3.882281e-002	3.922083e-002

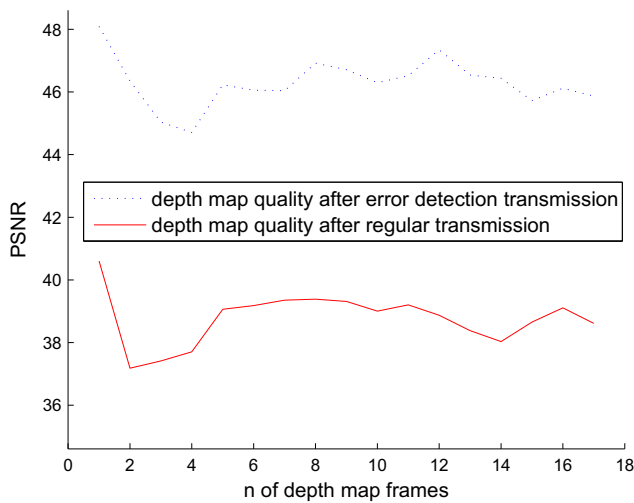


Fig. 8. PSNR comparison between the original depth map at the decoder side and the one after performing the patch pixel based error detection.

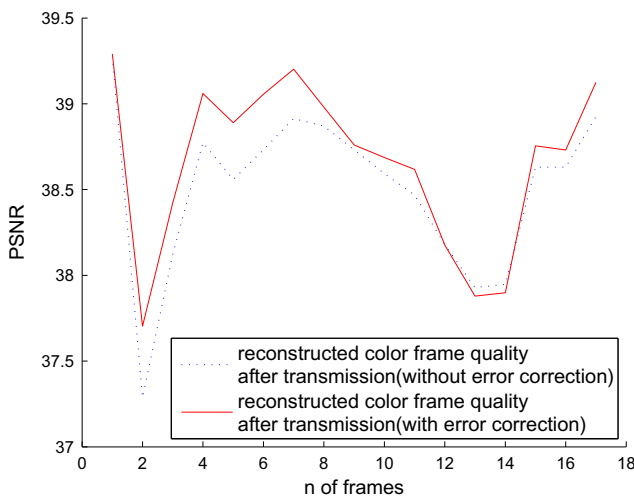


Fig. 9. PSNRs of the reconstructed color frame before and after performing the error detection and retransmission to the depth map.

to the depth map. The results vary within different quality conditions. In our simulation, when the reconstructed color frame PSNR is higher than 38.2 dB, the error detection improves the whole quality, and this improvement is in proportion to the original quality. But when the PSNR is lower than 38.2 dB, the quality improvement on depth map is not significant or even sometime has negative effect on the whole quality. This is because that almost perfect depth maps are used to reconstruct the color frames from 12 to 18 as shown in

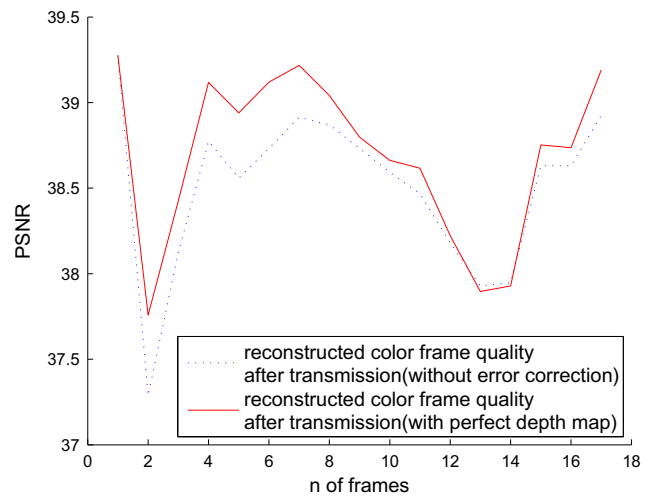


Fig. 10. Perfect construction.

Fig. 9. The fact can be justified by the similar trend (from 12 to 18) shown in **Fig. 10** where the received frame are constructed with perfect depth map.

8. Conclusion

In this paper, we proposed a wireless 3D image transmission scheme including image quality prediction model, quality improvement strategy, and its applicable scope. The major innovations in this paper include the techniques of patch map and patch map based error correction on the depth map. The techniques can be employed to not only detect the transmission errors and but also improve the quality of 3D image. It is a fundamental work to further 3D image transmissions, especially over the resource-constrained wireless networks. Based on the proposed depth-color source coding scheme, we firstly derive a depth-color quality prediction model that can estimates the transmitted 3D image quality over error-prone wireless environments. With the quality prediction model, an adaptive transmission 3D image transmission strategy has been developed. The experimental results demonstrate that our approach can achieve good multimedia quality under various channel conditions. Compared with traditional color-depth based 3D image compression and transmission approach, our approach can achieve higher multimedia quality over error-prone wireless environment with small communication overheads.

References

- [1] O. Schreer, P. Kauff, T. Sikora, 3D Video Communication: Algorithms, Concepts and Real-time Systems in Human Centred Communication, John Wiley & Sons Ltd, 2005.
- [2] A. Smolic, K. Muller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, T. Wiegand, 3D video and free viewpoint video – technologies, applications and MPEG Standards, in: Proceedings of International Conference on Multimedia and Expo (ICME 2006), Toronto, Canada, July 2006, pp. 2161–2164.
- [3] E. Stoykova, A.A. Alatan, P. Benzie, N. Grammalidis, S. Malassiotis, J. Ostermann, S. Piekh, V. Sainov, C. Theobalt, T. Thevar, X. Zabulis, 3D time-varying scene capture technologies – a survey, IEEE Transactions on Circuits Systems Video Technology 17 (11) (2007) 1587–1605.

- [4] A.A. Alatan, Y. Yemez, U. Gudukbay, X. Zabulis, K. Muller, C. Erdem, C. Weigel, A. Smolic, Scene representation technologies for 3DTV – a survey, *IEEE Transactions on Circuits and Systems for Video Technology* 17 (11) (2007) 1587–1605.
- [5] A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G. Akar, G. Triantafyllidis, A. Koz, Coding algorithms for 3DTV – a survey, *IEEE Transactions on Circuits and Systems for Video Technology* 17 (11) (2007) 1606–1621.
- [6] G.B. Akar, A.M. Tekalp, C. Fehn, M.R. Civanlar, Transport methods in 3DTV1 a survey, *IEEE Transactions on Circuits and Systems for Video Technology* 17 (11) (2007) 1622–1630.
- [7] S.C. Chan, H. Shum, K. Ng, Image-based rendering and synthesis, *IEEE Signal Processing Magazine* 24 (7) (2007) 22–33.
- [8] J. Konrad, M. Halle, 3-D displays and signal processing, *IEEE Signal Processing Magazine* 24 (7) (2007) 97–111.
- [9] C. Fehn, A 3D-TV system based on video plus depth information, *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 2, 9–12 Nov. 2003, pp. 1529–1533.
- [10] A. Khan, L. Sun, E. Jammeh, E. Ifeachor, Quality of experience-driven adaptation scheme for video applications over wireless networks, *Communications IET* 4 (11) (2010) 1337–1347.
- [11] W.J. Krzanowski, *Principles of Multivariate Analysis*, Clarendon press, Oxford, 1988.
- [12] NS2, <<http://www.isi.edu/nsnam/ns/>>.
- [13] P. Lai, A. Ortega, C. Dorea, P. Yin, C. Gomila, Improving view rendering quality and coding efficiency by suppressing compression artifacts in depth-image coding, in: *Proceedings of Visual Communication and Image Processing, VCIP '09*, San Jose, CA, USA, Jan. 2009.
- [14] W. Kim, A. Ortega, P. Lai, D. Tian, C. Gomila, Depth map distortion analysis for view rendering and depth coding, in: *Proceedings ICIP, 2009*, pp.721–724.
- [15] A. Khan, L. Sun, E. Ifeachor, Content clustering based video quality prediction model for MPEG4 video streaming over wireless networks, *IEEE ICC Conference, Dresden, Germany, 14–18 June 2009*.
- [16] N. Cranley, L. Murphy, P. Perry, Content-based adaptation of streamed multimedia, in: *IEEE International Conference on Management of Multimedia Networks and Services*, No. 7, San Diego, CA, 3–6 Oct 2004.
- [17] VQEG, Final report from the video quality experts group on the validation of objective models of video quality assessment, PHASE I, <<http://www.vqeg.org/>,2000>.
- [18] S. Jumisko-Pyykkö, J. Häkkinen, Evaluation of subjective video quality of mobile devices, in: *MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia*, pp. 535–538, 2005.
- [19] H. Knoche, J. McCarthy, M. Sasse, Can small be beautiful?: assessing image resolution requirements for mobile TV, *MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia*, 2005.
- [20] R. Krishnamurthy, B.-B. Chai, H. Tao, S. Sethuraman, Compression and transmission of depth maps for image-based rendering, in: *Proceedings of IEEE International Conference on Image Processing, ICIP 2001*, Thessaloniki, Greece, Oct. 2001.
- [21] R. Zhang, S.L. Regunathan, K. Rose, Video coding with optimal inter/intra-mode switching for packet loss resilience, *IEEE Journal on Selected Areas in Communications* 18 (6) (2000) 966–976, <http://dx.doi.org/10.1109/49.848250>.
- [22] A. Khan, L. Sun, E. Jammeh, E. Ifeachor, Quality of experience-driven adaptation scheme for video applications over wireless networks, *IET Communications Special Issue on Video Communications over Wireless Networks* 4 (11) (2010) 1337–1347.
- [23] L. Zhou, X. Wang, W. Tu, G.-M. Muntean, B. Geller, Distributed scheduling scheme for video streaming over multichannel multi-radio multi-hop wireless networks, *IEEE Journal on Selected Areas in Communications* 28 (3) (2010) 409C19.
- [24] L. Zhou, H.-C. Chao, A.V. Vasilakos, Joint forensics-scheduling strategy for delay-sensitive multimedia applications over heterogeneous networks, *IEEE Journal on Selected Areas in Communications* 29 (7) (2011). 1358C67.
- [25] C. Lai, Y. Huang, J. Chen, W. Ji, M. Chen, Design and integration of the opencore based mobile tv framework for DVB-H/ T wireless network, *ACM/Springer Multimedia Systems* 17 (4) (2011) 299–311.
- [26] J. Chen, Q. Yu, Y. Zhang, H.-H. Chen, Y. Sun, Feedback based clock synchronization in wireless sensor networks: a control theoretic approach, *IEEE Transactions on Vehicular Technology* 59 (6) (2010) 2963–2973.
- [27] S. He, J. Chen, Y. Sun, David K.Y. Yau, N.K. Yip, On optimal information capture by energy-constrained mobile sensors, *IEEE Transactions on Vehicular Technology* 59 (5) (2010) 2472–2484.