INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG2006/M12849
January 2006, Bangkok

Title: Reconstruction of Multi-view Images from Layered Depth Images
Source: GIST and ETRI
Authors: Yo-Sung Ho, Seung-Uk Yoon, Sung-Yeol Kim, and Eun-Kyung Lee
(Gwangju Institute of Science and Technology)
Kugjin Yun, Sukhee Cho, Namho Hur, and Soo-In Lee
(Electronics and Telecommunications Research Institute)
Status: Proposal

1 Introduction

Layered depth image (LDI) is an efficient approach to represent three-dimensional (3-D) objects with complex geometry for image-based rendering (IBR). We have been proposed a framework for multi-view video coding using the concept of LDI as a 3-D approach unlike other 2-D based video coding techniques [1]. In this document, we describe the reconstruction of multi-view images from the decoded LDI as the successive work of MVC using LDI [2][3].

2 Reconstruction of Multi-view Images from the Decoded LDI

In our previous works [2][3], we have generated LDIs from the natural multi-view video sequence, “Breakdancers”. The first eight color and depth frames of the sequence for camera zero are used to generate the first LDI frame; the second 16 images are used to make the second LDI frame; and so on.

Figure 2 shows the reconstruction procedure of multi-view images from the decoded LDI data.

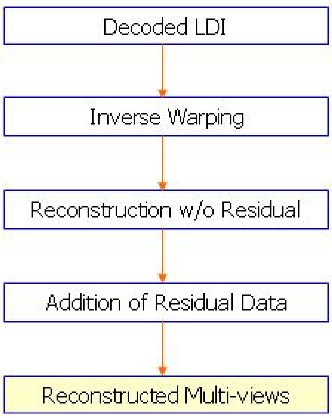

Fig. 2. Reconstruction of multi-view images from the decoded LDI

We start from the decoded LDI because the reconstruction process can be an inverse procedure of LDI generation. After the inverse warping to the world coordinate for whole pixels in LDI, we perform the incremental 3-D warping again. At this time, the target of the warping is to each camera location. Figure 3 shows the result of the inverse warping.

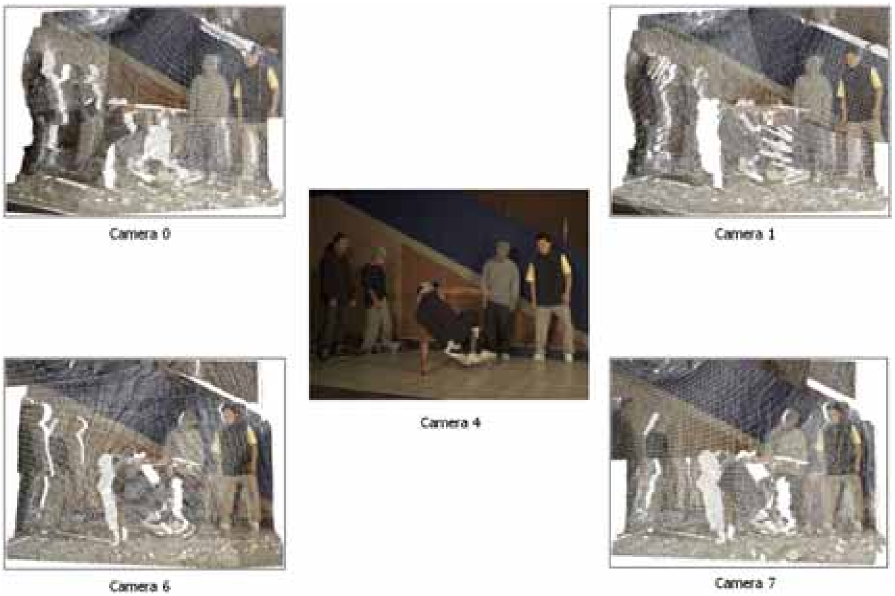

Fig. 3. Characteristics of the each layer of the generated LDI frame

As we can observe in Fig. 3, there are several holes in the resultant images because information is lost when pixels move to the reference camera location during the generation of LDI. In order to fill the holes, we can use color and depth information contained in depth pixels (DPs) of the back layer. Each layered depth pixel (LDP) contains different number of DPs and each DP has color and depth information. By using those DPs from other camera locations, we can restore some empty pixels. The reconstruction result using the additional DPs hanging in the back layer of LDI is depicted in Fig. 4.

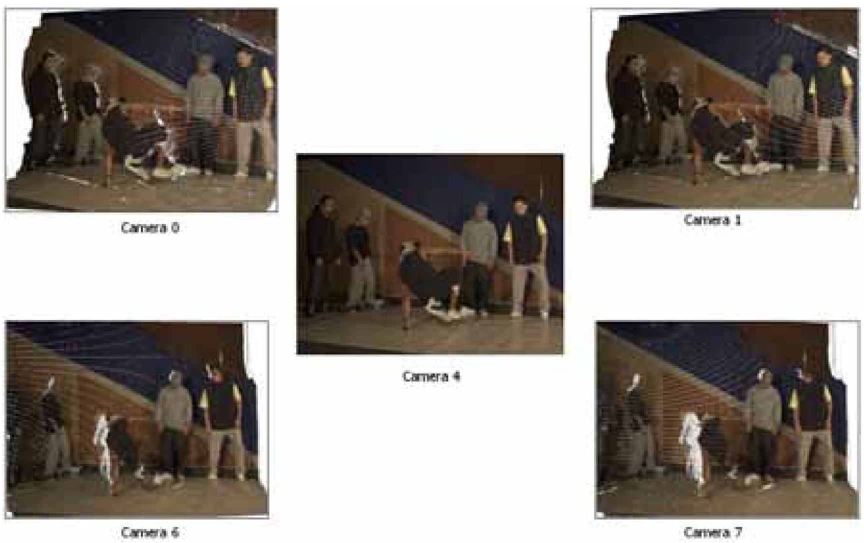

Fig. 4. Reconstruction without residual data

In Fig. 4, we have used no residual data from the original images, thus the left-most side and right-most side of the reconstructed images have large holes. The major reason of these results is the long distance among cameras. Actually, the camera arrangement of MSR data is 1-D arc of eight cameras with 20cm spacing. Thus, it is natural that we cannot find sufficient information to restore both sides of the reconstructed images. For example, there are lots of unfolded regions in camera 0, which means that other cameras cannot capture certain regions in camera 0.

We therefore need to fill the holes using the compensation module, as described in our overall framework [1]. During the compensation process is performed, we have used the residual data extracted from the original multi-view images. We can calculate residual information by subtracting empty pixel locations from the original multi-view images. Figure 5 represents the final reconstruction results using the residual information. Although there are some uncomfortable regions in the final result, it could be improved by exploiting more accurate method of selecting proper depth pixels to fill the holes.

Fig. 5. Reconstruction with residual data

In Table 1, we have compared the data size between sum of frames of the test sequence and the generated LDI frame. In each table, sum of frames means that the summation of eight color and depth images of the test sequence.

We have exploited two kinds of encoding methods, one is data aggregation with horizontal direction [2] and the other is the layer filling. The latter means that the empty pixels are filled with the pixels of the first layer. We can remove filled pixels in the decoding step because we know the number of layers in LDI. Based on the reconstruction results, we can determine the threshold, how much we can allow the difference among depth values. The table shows the result using the threshold value of 3.0, but the data size did not decrease much as the threshold value is over 3.0 as our experiments.

Table 1 does not list the data size of residual information. The residual information mainly depends on the distance among cameras and the actual viewing range. In our experiments, the size of residual data has changed based on the depth thresholding.

Table 1. Comparison of data size for the "Breakdanders" sequence

	1st 8 Frames	2nd 8 Frames
Sum of frames (color + depth) [kbytes]	25,165.9	25,165.9
LDI frame generated from 16 images [kbytes] (Threshold = 3.0, bit allocation for depth = 8 bits)	13,924	13,803
Encoded LDI frame using the data aggregation (with horizontal direction) [kbytes]	131.7	133.8
Encoded LDI frame using layer filling [kbytes]	48.4	48.2
Simulcast using AVC (color + depth) [kbytes]	137.7	132.5
Simulcast using AVC (color only) [kbytes]	97.4	96.3

3 Conclusion

In this document, we have described the reconstruction procedure for multi-view images from the decoded LDI. Although we could not test with other sequences without depth information, we believe that our approach could be helpful to encode sequences with depth information. Because the existence of depth information is the basic assumption of our framework, there are several obstacles to estimate depth values using post-processing for sequences without depth at present. However, the proposed encoding methods result in the smaller data size than the simulcast case in terms of depth coding. In addition, if we consider 2-D+depth representations to provide 3-D depth impression, the proposed framework would be appropriate for processing of those kinds of data.

4 References

[1] ISO/IEC JTC1/SC29/WG11 m11582, "A Framework for Multi-view Video Coding using Layered Depth Image," January 2005.

[2] ISO/IEC JTC1/SC29/WG11 m12278, "Intermediate Result on Multi-view Video Coding using Layered Depth Images," July 2005.

[3] ISO/IEC JTC1/SC29/WG11 m12485, "Generation and Coding of Layered Depth Images for Multi-view Video," October 2005.