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

ISO/IEC JTC1/SC29/WG11
MPEG2004/M11278
October 2004, Mallorca

Title: Multi-view Video Coding using Layered Depth Image

Source: GIST and ETRI
Authors: Yo-Sung Ho, Seung-Uk Yoon, and Sung-Yeol Kim
(Gwangju Institute of Science and Technology)
Daehee Kim, Sukhee Cho, Kugjin Yun, Chunghyun Ahn, and Sooin Lee
(Electronics and Telecommunications Research Institute)

Status: Proposal

1 Introduction

Layered depth image (LDI) is an efficient approach to represent three-dimensional objects with complex geometry for image-based rendering (IBR). LDI has already presented as a tool for multi-texturing and IBR in MPEG-4 AFX CE A8 [1]. In AFX, the functionality of LDI is mainly focused on texturing and rendering using depth. In this document, we describe how to encode multi-view video sequences with depth by using the concept of LDI. Before explaining our idea, we will explain the concept and characteristics of LDI.

2 Layered Depth Image

LDI contains potentially multiple depth pixels at each pixel location. Each depth pixel contains depth value along with its color. The farther depth pixels play a role in filling the disoccluded regions that occur as the viewpoint moves away from an LDI camera. Figure 1 represents the conceptual diagram of LDI [2].

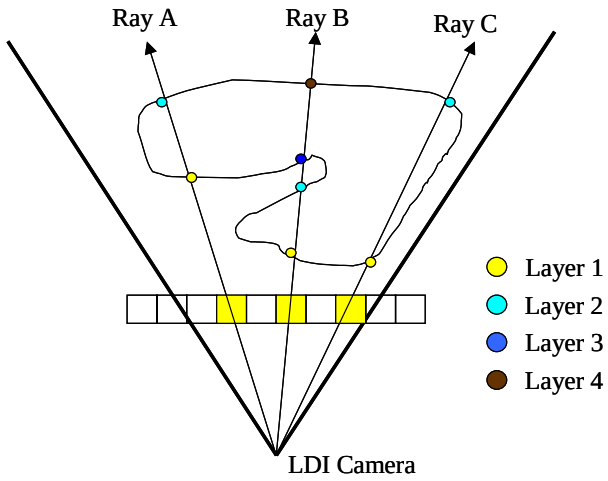

Fig. 1. Layered Depth Image

LDI can be generated from multiple depth images by means of a 3-D warping technique. As shown in Fig. 2 [2], the LDI scene viewed from C1, an LDI view, is constructed by warping pixels in other camera locations, such as C2 and C3. When the warped points, c and d, are placed in the same pixel location, their depth values are compared. If the difference between depth values is greater than a predefined threshold, a new layer is created; otherwise, the warped pixels are merged. This procedure is called pixel ordering.

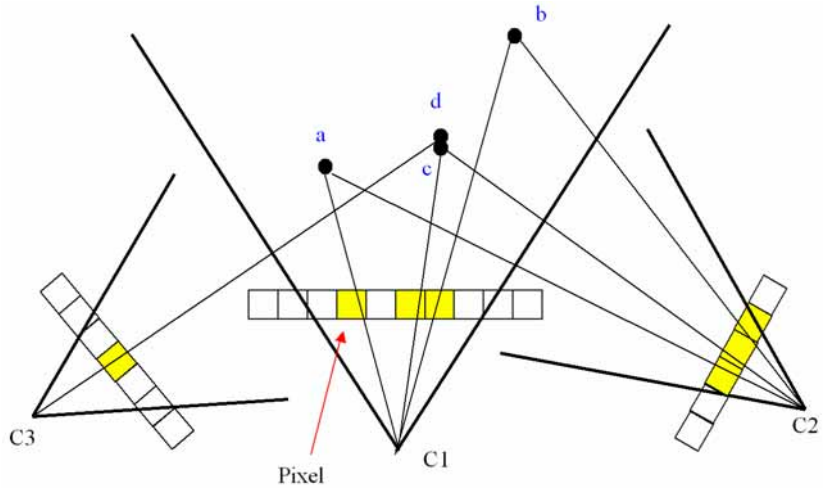

Fig. 2. Generation of LDI from Two Depth Images

LDI contains several attribute data together with multiple layers at each pixel location. A single LDI pixel consists of color, depth, and splat table index that support rendering of LDI. In detail, each LDI pixel contains 63 bit information: 8 bits each for the R, G, and B color components, 8 bits for the alpha channel, 20 bits for the depth of the object, and 11 bits for splat table index. The splat table index is in turn divided into 5 bits for the distance, 3 bits for the x norm, and 3 bits for the y norm. It is used to support various pixel sizes in rendering of LDI. The overall data structure of LDI is shown in Fig. 3.

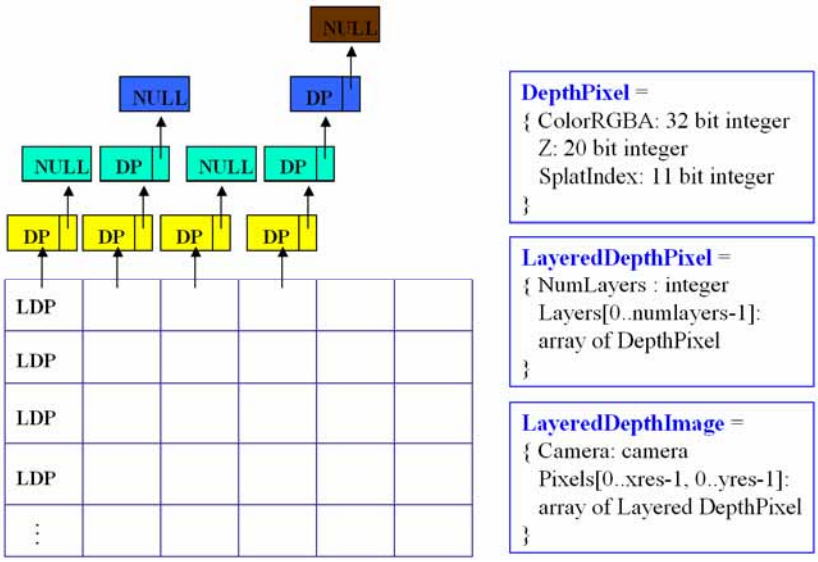

Fig. 3. Data Structure of LDI

Three key characteristics of LDI are as follows: (1) it contains multiple layers at each pixel location, (2) the distribution of pixels is sparse in the back layer, and (3) each pixel has multiple attribute values. Because of these special properties, LDI enables us to render arbitrary views of the scene at new camera positions. In other words, if LDI can be constructed from multi-view images with depth, it can easily regenerate new views with less data. Consequently, we can compress multi-view video sequences by using LDI coding algorithms [3].

3 Conversion between Multi-view Video and Layered Depth Images

ETH provides test sequences with depth [4]. Since we need depth images of multi-view video sequences, we adopted ETH-Zurich test sequence as our experimental data. ETH-Zurich data consists of three different sequences of color, shape, and depth. It contains camera parameters also. Three types of data provided by ETH-Zurich sequences are shown in Fig. 4.

(a) arctic1.col.yuv (b) arctic1.shape.yuv (c) arctic1.z.yuv

Fig. 4. ETH-Zurich Test Sequence

There are 16 cameras and each camera provides three types of sequences with the 640x480 resolution. Figure 5 represents the configuration of 16 cameras.

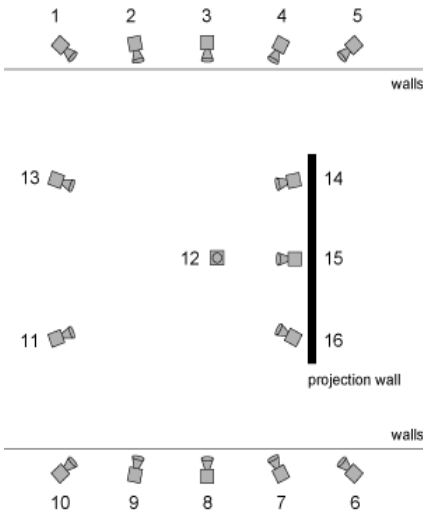

Fig. 5. The Configuration of 16 Cameras

First, we perform 3-D warping from 16 camera locations to a single LDI view. We use the following warping equation expressed by [5]

$$u_2 = \frac{\bar{a}_1 \cdot (\bar{b}_2 \times \bar{c}_2) u_1 + \bar{b}_1 \cdot (\bar{b}_2 \times \bar{c}_2) v_1 + \bar{c}_1 (\bar{b}_2 \times \bar{c}_2) + (\dot{C}_1 - \dot{C}_2) \cdot (\bar{b}_2 \times \bar{c}_2) \delta(u_1, v_1)}{\bar{a}_1 \cdot (\bar{a}_2 \times \bar{b}_2) u_1 + \bar{b}_1 \cdot (\bar{a}_2 \times \bar{b}_2) v_1 + \bar{c}_1 (\bar{a}_2 \times \bar{b}_2) + (\dot{C}_1 - \dot{C}_2) \cdot (\bar{a}_2 \times \bar{b}_2) \delta(u_1, v_1)}$$

$$v_2 = \frac{\bar{a}_1 \cdot (\bar{c}_2 \times \bar{a}_2) u_1 + \bar{b}_1 \cdot (\bar{c}_2 \times \bar{a}_2) v_1 + \bar{c}_1 (\bar{c}_2 \times \bar{a}_2) + (\dot{C}_1 - \dot{C}_2) \cdot (\bar{c}_2 \times \bar{a}_2) \delta(u_1, v_1)}{\bar{a}_1 \cdot (\bar{a}_2 \times \bar{b}_2) u_1 + \bar{b}_1 \cdot (\bar{a}_2 \times \bar{b}_2) v_1 + \bar{c}_1 (\bar{a}_2 \times \bar{b}_2) + (\dot{C}_1 - \dot{C}_2) \cdot (\bar{a}_2 \times \bar{b}_2) \delta(u_1, v_1)} \quad (1)$$

where $\dot{C}_1$, $\dot{C}_2$ are camera positions, $(\bar{a}_1, \bar{b}_1, \bar{c}_1)$, $(\bar{a}_2, \bar{b}_2, \bar{c}_2)$ are basis vectors, and δ is a disparity; (u_1, v_1) is a pixel coordinate of an image plane at $\dot{C}_1$ and (u_2, v_2) is that at $\dot{C}_2$. We can generate a new view at $\dot{C}_2$ by means of warping an image from the camera location $\dot{C}_1$ to $\dot{C}_2$.

The LDI view can be selected from one of the 16 positions of cameras, or it can be an arbitrary viewpoint. Color and depth sequences are used to construct LDI frames. As shown in Fig. 6, the first frames of color and depth sequences are collected and warped to the selected LDI view by using 3-D warping and pixel ordering. Consequently, 16 depth images with color construct the first frame of LDI sequence. In this document, the LDI sequence and LDI frames have the same meaning as the collection of layered depth images. Once we obtain the LDI frames from the above procedure, then the reconstruction of multi-view is a basic function of LDI. Since LDI contains all necessary information to generate an arbitrary view, it can reproduce any viewpoints. In addition, we can apply LDI coding algorithms to compress the converted LDI frames.

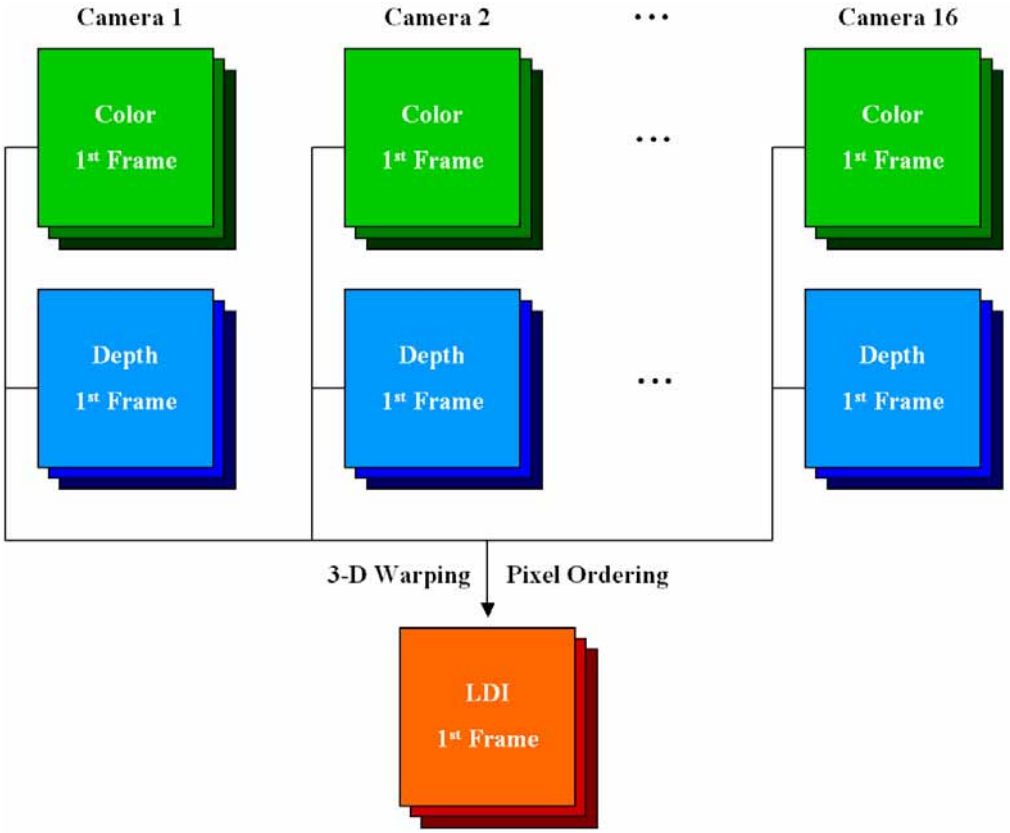

Fig. 6. Conversion from Multi-view Video to LDI Frames

4 Conclusion

In this document, we introduced the concept and characteristics of LDI. Furthermore, we explained the conversion procedure between multi-view video and LDI. Our main idea is that multi-view sequences with depth can be efficiently coded by using the concept of LDI. We can apply efficient LDI coding schemes by constructing LDI frames from multi-view video with depth. For the next meeting, we will compare our proposed approach with previous multi-view video coding algorithms for various multi-view sequences.

5 References

[1] ISO/IEC JTC 1/SC 29/WG 11/N4220, "Animation Framework eXtension Core Experiments Description," July 2001.

[2] J. Shade, S. Gortler, L. He, and R. Szeliski, "Layered Depth Images," ACM SIGGRAPH, pp. 231-242, July 1998.

[3] ISO/IEC JTC 1/SC 29/WG 11/M11279, "Coding of Layered Depth Image using Coherency between Point Samples," October 2004.

[4] 3-D Video at ETH Zurich, http://graphics.ethz.ch/3dvideo/main.php

[5] L. McMillan, "An Image-based Approach to Three-Dimensional Computer Graphics," Ph.D. Dissertation, 1997.