REAL TIME 3D AVATAR TRANSMISSION USING CYLINDER MAPPING

Shin-ichiro Takahashi, Masaaki Ikehara

Dep. of EEE, Keio Univ.
Yokohama, Kanagawa 223-8522, Japan
{takahashi, ikehara}@tkhm.elec.keio.ac.jp

Isthiaq Rasool Khan, Masahiro Okuda

Dep. of IMS, The Univ. of Kitakyushu
Kitakyushu, Fukuoka 808-0135, Japan
{khan, okuda-m}@env.kitakyu-u.ac.jp

ABSTRACT

In this paper, we propose the client-server system which accomplishes acquisition, transmission and reconstruction of the 3D shape at a remote location in real time. The server consists of stereo cameras and a PC cluster, in which each PC creates a range image using multiple color images obtained from the stereo cameras connected to it. All PCs on the server side work in synchronization and generate range images in parallel with each other, which are then transmitted to the client as a point cloud. On the client side, the 3D model is reconstructed using a received point cloud by the point-based rendering method. We propose the cylindrical mapping method, which removes overlaps and annoying artifacts efficiently before the rendering process. All the procedure of generation, transmission, and reconstruction of successive range images is done in real time.

***Index Terms*—** Real Time Systems, Stereo Vision

1. INTRODUCTION

In recent years, there has been much interest in the acquisition of 3D image sequences of non-static objects to generate 3D movies having multiple viewpoints. In [1], high-definition 3D image sequences were reconstructed using multiple views obtained by 49 cameras, however, the modeling process works offline. Although many other 3D imaging systems have been proposed as well, the systems which work in real time like our approach are practically useful. In [2]–[5], the reconstruction of 3D image sequences is accomplished in real time using the visual hull method by using a PC cluster and the parallel processing. Unfortunately the visual hull method approximates a rough shape of the object, the high computational power is required to represent details and color textures. In reason, that is time-consuming. Therefore, our proposed system avoids the visual hull method at the modeling stage and employs the stereo matching method to create the range images. The 3D model is reconstructed based on integrated multiple range images. Some stereo matching methods to generate a range image in real time exist, and some of them have been commercialized. The Digiclops camera by Points Grey Research [7] uses this technique and obtains range images at 10-30 frames per second (fps).

The data reduction is extremely important to realize a real time 3D imaging system. In [2]–[4], the surface of the reconstructed 3D models is represented by the polygonal mesh. Generally, the polygonal mesh consists of spatial coordinates, colors , and associated data of vertices positions. Because the positions data must be encoded in the lossless way, it affects the data reduction and the compression ratio. In our proposed system, pixels in the integrated range images are assumed as 3D spatial points. Therefore additional associated position data are not needed, and each pixel is transmitted directly as the point cloud. Our rendering method is based on point based geometry which is the well known method of point cloud rendering.

In certain situations, when the camera-calibration or ROI extraction is failed, these cause annoying artifacts and overlaps between different range images during the 3D model reconstruction. To address this problem, the cylindrical mapping method is employed before the rendering process, and that is similar to the approach presented in [9]. A point cloud is segmented by mapping on the inner wall of the cylinder. Invalid points are determined by comparison with visual axes of cameras and the direction of the reconstructed surface by segmented points. And invalid points are removed. The procedures also works in real time.

2. REAL TIME 3D IMAGING SYSTEM

2.1. Overview

Our proposed client-server system is depicted in Fig. 1, in which the client reconstructs 3D models and the server obtains, integrates and transmits the 3D shape of a target object. The client and the server are connected through the Gigabit network each other. The server consists of N stereo cameras and N+1 PCs. The stereo cameras surround a target object, and each camera is connected to one of the PCs through IEEE1394. The N PCs connected with each camera capture multiple color images and create N range images by the stereo matching method. And then the N PCs transmit the N range images to the PC in the server called *Parent PC*. After that, the bit streams are packetized and transmitted to the client. Relative positions of cameras are calibrated and informed to the client in advance. The parent PC sends the synchronization signal to the other PCs. The client receives a bit stream from the parent PC and reconstructs a 3D model based on it.

2.2. Tasks in Server

At the beginning, the parent PC sends a command to other N PCs. That command makes them to obtain color images from the connected stereo cameras. The N PCs create a range image based on these color images by the stereo matching. The Region of Interest(ROI) extraction method is a simple background subtraction. Here, a background is an image taken previously without any objects. The background subtraction process extracts an object by removing pixels corresponding to the background image. Those pixels are assessed by comparing color and depth of the background image and the obtained image. The pixels in ROI are quantized by the linear quantization and transmitted to the client. After receiving N bit streams, the parent PC retransmits a command to the N PCs so that they start to obtain color images from the connected camera. The

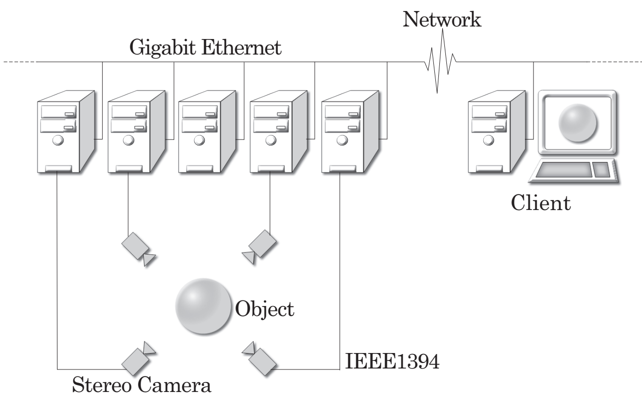

Fig. 1. **Real Time 3D Imaging System:** The Client-Server System. The client reconstructs the shape of the 3D object at the remote location based on range images in real time.

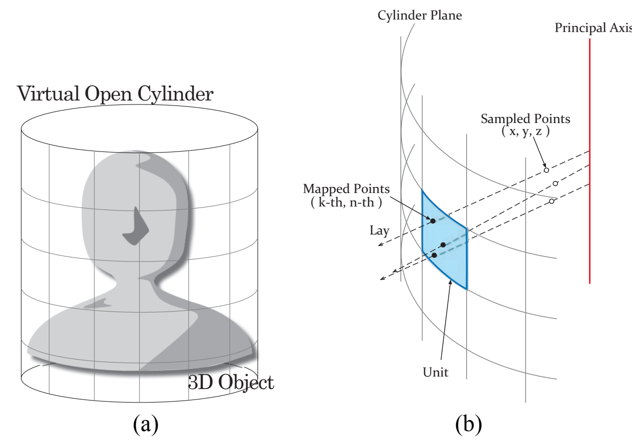

Fig. 2. **Mapping on the Wall of the Cylinder:** 3D object with the virtual open cylinder (a). Image of mapping on the inner wall of the cylinder (b).

parent PC integrates N bit streams and packetizes it with little headers, after that, transmits each packet to the client sequentially.

2.3. Tasks in Client

The client consists of one PC and a display. Bit streams are transmitted to this PC from the parent PC. A received bit stream consists of spatial coordinates and color data of points. The point based geometry is employed in rendering the 3D object. The 3D model is reconstructed by colored primitives which are located in the three-dimensional space.

When the position of each range image is not perfectly aligned due to failed camera calibration, overlaps are caused between different range images. The overlaps are usually annoying. Moreover, unexpected pixels are left around the boundary of a object since the stereo matching and the background subtraction often fail in that area. These remained pixels cause image degradation. To remove overlaps and annoying artifacts which make image degradation, the cylindrical method is adopted before the rendering process. The cylindrical method can remove overlaps and unexpected pixels efficiently and quickly.

2.3.1. Cylindrical Method

Let's assume that a object is surrounded by a virtual open cylinder, as shown in Fig. 2(a). Each point on the object can be mapped on the inner wall of the cylinder by using some suitable methods. The employed mapping method is explained in detail in this section. After the whole object is mapped on the cylinder, a vertical cut in the surface of the cylinder can warp it to a plane. Each mapped point is allocated to a small domain. Those domain called *"Unit"*, and taken by dividing the plane by a regular grid. In order to assess whether a point is valid or invalid, it is compared to other points contained in the same Unit.

Angle Based Mapping

The method based on angle is employed for mapping the object to the surrounding cylinder. Due to its simplicity and low computational complexity, the method can be implemented in real time and is suitable for our real time imaging system.

It is assumed that a whole object is surrounded by the virtual open cylinder and the axis of the cylinder is along the principal axis of the object. When the object and the surrounding cylinder are cut into thin horizontal slices, each slice can be considered that a set of 2D points surrounded by a circle (shown in Fig 3(a)). A ray originating from the center of a circle and passing through a 2D point is used to map it on the circle at the intersection of the ray and the circle. By repeating this procedure for each point in each slice, the whole object can be mapped on the inner wall of the cylinder.

The mapping of a point to a surrounding circle can be determined by 2-parameters, x and z components. We transform them to the cylindrical coordinates

$$r = \sqrt{x^2 + z^2}, \tag{1}$$

$$\phi = \tan^{-1}\left(\frac{z}{x}\right) \tag{2}$$

The angle ϕ, gives the angle that a ray passing through the point and the center of the circle makes with the x-axis. The range of ϕ is 0 to 2π. And both of 0 and 2π refer to the same direction, i.e.,along the x-axis.

The circumference of the outer circle in each slice makes a row of the plane. If each row has M columns (or Units), a point is located in k-th Unit when the angle ϕ of that point satisfies the following equation (3).

$$\frac{2\pi(k-1)}{M} \leq \phi < \frac{2\pi k}{M} \tag{3}$$

The location of a Unit on the plane describes the average y and ϕ parameters of the points associated to it; therefore, a Unit has to store only one parameter, r, to describe the approximate 3D position of these points.

Unit Process

After the mapping is done, each Unit stores the allocated points. When there are points from different cameras in one Unit, we take it as the overlap. In order to remove this overlap, we consider the inner product of the visual axis of a camera and the tangent direction of the surface generated from points in the Unit.

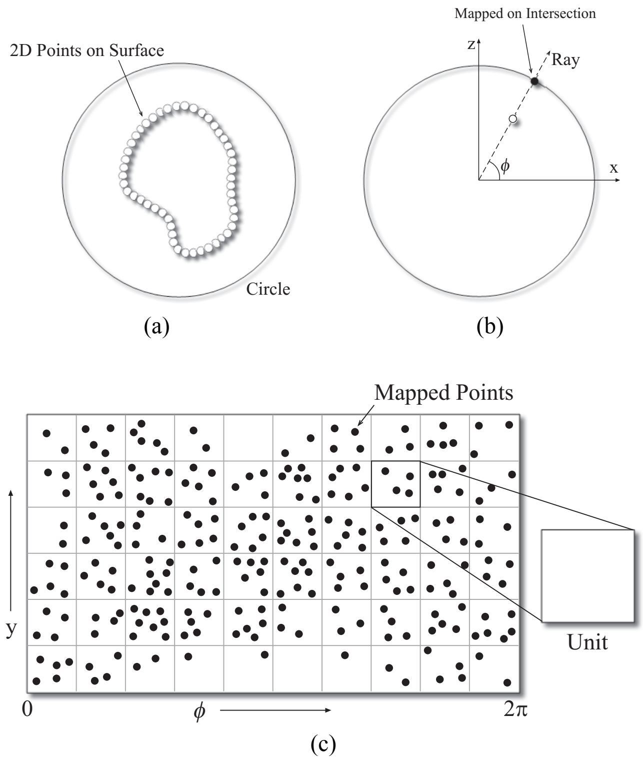

Fig. 3. **Angle based Mapping:** A Horizontal Slice of the 3D object with the surrounding cylinder (a). Image of mapping based on angle (b). A sampled point is mapped on a intersection point of circle by a ray. Unit Plane warped from the wall of the cylinder (c).

Firstly, if there are some points in the Unit, the number of points stored in the Unit is counted. When several points obtained from different cameras are in same unit, the tangent vectors of each surface are computed (Fig 4.I). And then these vectors are normalized. After each tangent vector is computed, the vectors are merged to one simply by averaging them with some weights and then set it as a tangent vector of the Unit (Fig 4.II). The inner product of the tangent vector and an axis of each camera is calculated. We decide that points obtained from the camera with the minimum inner product are valid. And others are invalid (Fig 4.III). By discarding the invalid points, only the valid points, which are obtained from one camera, remain. By repeating this procedure, the whole object can be handled.

2.3.2. *Rendering*

After the Cylinder mapping process is done, only valid points are rendered by the point based rendering method. The shape of primitives is a rectangle. The normal vector of each primitive is estimated by the average of differences of the current one and neighbors. A direction and a size of the primitives can be calculated dynamically by the estimated normal vector and the user-decided view point. The 3D model is reconstructed by locating these primitives on the three-dimensional space.

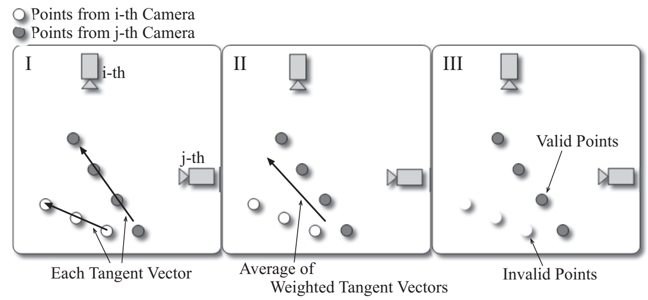

Fig. 4. **Procedure of Unit Process:** Gray points are 2D points obtained from i-th camera. White points form j-th. The inner products of the average of weighted tangent vectors and each axis of the cameras can be calculated in II. In this case, points from j-th camera is valid since the surface consists of entire points in this Unit is faced for j-th camera.

3. RESULTS

The server consists of five PCs(Pentium IV) and four stereo cameras. The target object is a human upper body, and a avatar is reconstructed in the client side. The cameras are located around the target object. Two cameras are at the front and back of the object, and the rest two cameras are at the front-right and front-left. All the PCs and the cameras are connected through Gigabit Ethernet, and the client is located in the same LAN. All data are transmitted using TCP. The resolution of both color images obtained from each stereo camera and range images is 240×320. The Camera-Calibration is done, and then the relative position of each stereo camera was informed to the client in advance.

The user can see the 3D model from arbitrary view point using the implemented viewer in the client. Resulting images from the viewer are shown in Fig. 5. Our system works about 5-10 fps, and the size of data actually transmitted to the client about 2-7 Mbps. Fig. 6 proves the validity of the cylindrical method. Fig. 6(a) shows the result of integrated four range images simply, and Fig. 6(b) shows one with the cylindrical method. In Fig. 6(b), annoying artifacts are removed and the facial expression is finer. Fig. 7 shows the horizontal cross-section view in the nose portion. Fig. 7(b) is the result with the cylindrical method and Fig. 7(a) without one. The three types of lines show different range images. We can be seen that the overlap is removed in Fig. 7(b).

The resulting sample movie is available at [10].

4. CONCLUSION

In this paper, we introduced the real time 3D imaging system by handling point clouds directly which is integrated by multiple range images without the polygonal mesh. Furthermore, annoying artifacts were removed efficiently and quickly by applying the cylindrical method. The advantage of our system is the transmission stream specialized by handling the data of the shape and the color as point clouds. However, many problem which should be examined furthermore are left, for example improvements of the image quality and frame rates, and being ROI extraction robust. Therefore, there still is much room for improvements left.

5. REFERENCES

[1] H. Saito, S. Baba, M. Kimura, S. Vedula and T. Kaneda, "Appearance-based virtual view generation of temporally-varying events from multi-camera image in the 3D room," *Second International Conference on 3-D Digital Imaging and Modeling (3DIM99)*, pp. 516–525, Oct. 1999.

[2] X. Wu, T. Matsuyama, "Real-time active 3D shape reconstruction for 3D video," *Proc. of* 3^{rd} *International Symposium on Image and Signal Processing and Analysis*, pp. 186–191, Sep. 2003.

[3] M. Ueda, D. Arita, R. Taniguchi, "Real-time free-viewpoint video generation using multiple cameras and a PC-cluster," *Pacific-Rim Conference on Multimedia*, pp. 418–425, Dec. 2004.

[4] S. Würmlin, E. Lamboray, O. G. Staadt and M. H. Gross, "3D video recorder," *IEEE Computer Society Press, Proceedings of Pacific Graphics '02*, pages 325-334, 2002.

[5] S. Würmlin, E. Lamboray, M. H. Gross, "3D video fragments: dynamic point samples for real-time free-viewpoint video," *Computers & Graphics, Special Issue on Coding, Compression and Streaming Techniques for 3D and Multimedia Data*, 2004.

[6] T. Kanade, A. Yoshida, K. Oda, H. Kano and M. Tanaka, "A stereo machine for video-rate dense depth mapping and its new application," *the* 15^{th} *Computer Vision and Pattern Recognition Conference*, pp. 196–202, June. 1996.

[7] PointGrey Research, Digiclops, http://www.ptgrey.com/products/stereo.html

[8] Rusinkiewicz, S., and Levoy, M., "Qsplat: a multiresolution point rendering system for large meshes," *SIGGRAPH '00*, pp. 343–352.

[9] Isthiaq Rasool Khan, Masahiro Okuda and Shin-ichi Takahashi, "Regular 3D mesh reconstruction based on cylindrical mapping," *IEEE International Conference on Multimedia and Expo*, TP1-6-2, June. 2004.

[10] http://vig.is.env.kitakyu-u.ac.jp/English/research/realtime_demo.html

Fig. 5. **Example of Client Viewer:** Resulting multi view point movie can be seen at the client viewer.

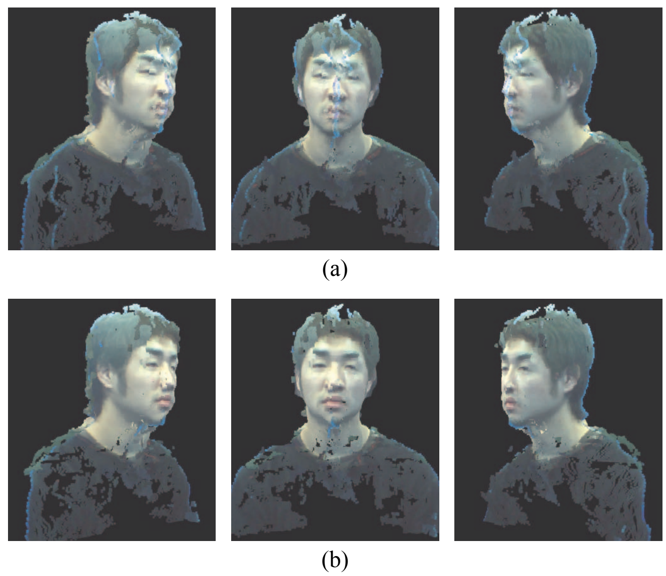

Fig. 6. **Validity of Our Proposed Method:** One frame of 3D image sequences without the Our Proposed Cylindrical Method (a). It can see that the facial expression of (b) is finer than (a).

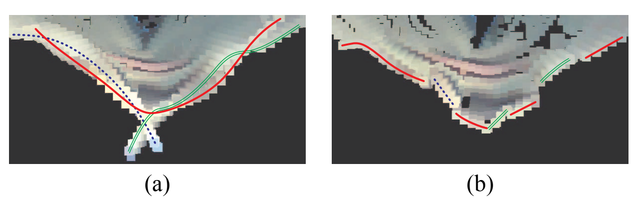

Fig. 7. **The Horizontal Cross-Section View in a Nose Portion:** Without Cylindrical Method (a). It can see that there are many overlapping area. With Cylindrical Method (b). Overlaps are removed especially front side.