4π Measurement System: A Complete Volume Reconstruction System for Freely-moving Objects

IIYAMA Masaaki†, KAMEDA Yoshinari‡ and MINOH Michihiko†

†Academic Center for Computing and Media Studies, Kyoto University
Yoshida-nihonmatsu, Sakyo-ku, Kyoto 606-8501 Japan
{iiyama,minoh}@media.kyoto-u.ac.jp

‡Institute of Engineering Mechanics and Systems, University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaragi 305-8573 Japan
kameda@image.esys.tsukuba.ac.jp

Abstract

A measurement system for reconstructing the volume of freely-moving objects is proposed. Previous systems are not taking into consideration for the objects that are not expected to stop their motion during measurement process, e.g. live insects. Our system can measure such objects.

Generally, the accuracy of reconstructed volume is improved by observing the object from as many directions as possible. In our system, the 20 cameras are arranged so that they can observe an object from 4π directions, while the previous systems can only observe from upper viewpoints.

Three issues of our system are discussed and evaluated in this paper; the accuracy of volume reconstruction, the spatial resolution and the influence of the object's motion on the resolution. Experimental results show that our system can reconstruct the kinds of freely-moving objects' volume.

1. Introduction

This paper describes a system that reconstructs objects' volume.

The reconstructed volume plays an important role in various applications. Displaying museum specimens in the virtual space is one of such applications. Several attempts to preserve the naturalistic and the archaeological materials as digital archives have been made recently. The digital archives contain the specimens' shape and allow us to display the specimens in the virtual space. In the virtual space, we can observe the specimens from arbitrary point of view, at all times. We call such space the virtual museum.

Previous researches digitized (measured) the volume of art objects such as sculptures and architecture[1][2]. However there is no attempt to digitizing creatures such as insects. Digitizing various creatures is significant for the entomological and the zoological field. The main reason that no attempt exists is that creatures are freely-moving objects. Many measurement systems have been proposed recently[3][4][5], however, these systems are designed for objects that stand still or for human that are expected to stop their motion during measurement process. These systems are not taking into consideration for measuring objects that are not expected to stand still (we call such objects freely-moving objects).

Our system aims at measuring the volume of small freely-moving objects, e.g. live insects. We focus on three issues in order to design our system.

The first one is reconstructing the complete volume, in other words, the volume reconstructed by 4π measurement. The complete volume, especially with high accuracy volume, is necessary for achieving the arbitrary points of view. In the previous measurement systems, the cameras were arranged upside of the reconstructed objects. These systems can reconstruct the upper surface of objects, while they can reconstruct the undersurface of objects. In order to reconstruct the complete volume, cameras should be arranged uniformly so that they observe the object from 4π directions.

The second one is the spatial resolution. In the virtual museum, digitized objects should have enough resolution to distinguish among the specimens. The sub-millimeter spatial resolution, which is enough to distinguish, was achieved in the previous systems. We also design our system that has sub-millimeter resolution.

The third one is measuring kinds of objects. The virtual museum should own as many specimens as possible. Each creature has various texture on its surface, so robustness to variation in texture of objects is required to our measurement system. There are two kinds of image-based method that measure the volume of objects: one is the method using the photometric information such as the multi-baseline stereo[6], and the other is the method using the geometric information such as the volume intersection method[7]. Compared with the multi-baseline stereo, the volume intersection method can deal with flat textured objects, because the volume intersection method only requires silhouettes on which the object is projected. Hence, in our system, the volume is reconstructed by the volume intersection method.

This paper is organized as follows: Section 2 describes a volume reconstruction method and discussion about the effectiveness of this method. Section 3 describes the overview of our system. The spatial resolution and the influence of the object's motion on the resolution are discussed in section 4. Experimental results and system evaluation are described in section 5. Conclusions are described in section 6.

2. Volume Reconstruction

We use the volume intersection method to reconstruct the volume of moving objects. The volume intersection method has two advantages over the multi-baseline stereo.

One is the robustness to flat textured objects because the volume intersection method only requires silhouettes on which the object is projected.

The other is the form of reconstructed results. The virtual museum requires the complete 3D data, in other words, 3D data that contains no hole. In the multi-baseline stereo, it is necessary to integrate the 3D surface data from a pair cameras in order to obtain complete 3D shape. Therefore, it is not guaranteed that the integrated data contains no hole. Compared with this fact, the obtained data from the volume intersection method is volume (voxel) data, and non-existence of hole is guaranteed.

In this section, we first describe the overview of volume intersection method, and next we describe the pseudo-volume where the object does not exist. Finally we describe the evaluation of the relation between the pseudo-volume and the number of cameras.

2.1. The volume intersection method

The volume intersection method needs the images observed from various directions. We fix the target space for reconstruction to a certain space, and several viewpoints are set to observe the space. The output data of the volume intersection method are in the form of voxels.

Let us denote cameras to observe the target space by $\boldsymbol{C_i}\,(i = 1, \ldots, m)$, where m is the number of cameras. The position and direction of each camera are calibrated as 3×4 perspective projection matrix $\boldsymbol{P}_i$ beforehand.

When an object is imaged by a viewpoint $\boldsymbol{C_i}$, they exist within the frustum that circumscribes its projected region on the image plane and whose apex corresponds to the focus point of the camera. We call this frustum the *Existence of Shadow Space* for $\boldsymbol{C_i}$ and denote $ESS_{\boldsymbol{C_i}}$. The projected region of the object is extracted by calculating the difference between the input image and the background image taken in advance. Let us denote the extracted object regions by R_i. In the case where the objects are image by m cameras, they exist within the product of all ESS_{C_i} (Figure.1). We denote it as ESS.

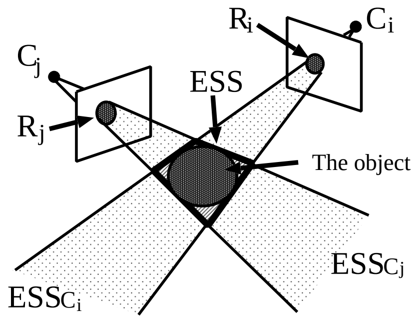

Figure 1. Volume intersection method

Reconstruction of object's volume in the voxel representation is performed as follows. We express the target space as the set of voxels. Let us denote the coordinate position of voxels in the target space as $\boldsymbol{v_k}$. a voxel $\boldsymbol{v_k}$ is projected onto the image plane of each camera C_i as

$$\boldsymbol{p}_i(\boldsymbol{v_k}) = \boldsymbol{P}_i \boldsymbol{v_k} \tag{1}$$

where $\boldsymbol{p}_i(\boldsymbol{v_k})$ is the projected pixel. $\boldsymbol{v_k}$ is an element of ESS if and only if $\boldsymbol{p}_i(\boldsymbol{v_k})$ is in R_i for every i.

$$ESS_{\boldsymbol{C}_i} = \{\boldsymbol{v_k} \mid \boldsymbol{p}_i(\boldsymbol{v_k}) \in R_i\} \tag{2}$$

$$ESS = \bigcap_i^m ESS_{\boldsymbol{C_i}} \tag{3}$$

As the number of cameras increases, the size of the ESS becomes smaller. In addition, the ESS always circumscribes the object, in other words, the ESS will never be smaller than the object. Accordingly, as the number of cameras increases, the volume of the ESS will be close to the volume of the object. However, in some case, the volume of an ESS will not be equal to the volume of the object, no matter how the number of cameras increases.

2.2. Pseudo-volume

There remains shape difference between the ESS and the object. In other words, an ESS contain the volume where the object does not exist. We call such volume the pseudo-volume. The pseudo-volume exists on the surface of the object, thus some surface of the object do not correspond to the surface of the ESS.

Let s be a surface point of the object. The necessary and sufficient condition that s corresponds to the ESS by the volume intersection method is following [10].

- There is at least one tangent line which only touches at the point of tangency s and does not intersect the object at the other points.

This condition indicates that if the infinite numbers of viewpoints are used, the surface of the reconstructed volume by the volume intersection method will be either a convex surface or a saddle like surface. Consequently, a pseudo-volume existing on such surfaces can be eliminated when we use enough cameras. On the other hand, a pseudo-volume on concave surfaces can not be eliminated even if we use infinite number of cameras. Thus, there are two kinds of pseudo-volume: one is able to eliminate by using more images, and the other is never eliminated. From now, we discuss the eliminatable pseudo-volume. We can not use infinite number of cameras, so we determine the reasonable number of cameras in order to eliminate the eliminatable pseudo-volume.

Relation between the amount of pseudo-volume and the number of cameras We first synthesized 3D objects, a cube, a sphere and a dinosaur model (Fig2 (a)). Next, we reconstructed the volume of each object by changing the number of virtual camera from 4 to 32 (4, 6, 8, 12, 20, 32(=12+20)). We arranged virtual cameras on the vertices of the regular n-hedron(s).

Fig.3 shows the relation between reconstruction ratio (which is calculated by reconstructed_volume/true_volume) and the number of camera. In Fig.2, the reconstructed volumes of the objects are shown. The reconstructed volume of the cube and the sphere must contain no pseudo-volume, if we use infinite number of cameras. As the number of cameras increases, these ratios close to 1, thus, reconstructed volume close to the synthesized object. The reconstruction ratios by using 20 cameras are 1.14 for the cube, 1.05 for the sphere, 1.20 for the dinosaur model. There is comparatively small ratio difference between 20-cameras and 32-cameras, and the ratio is close to 1. Thus, we used 20-cameras in our system.

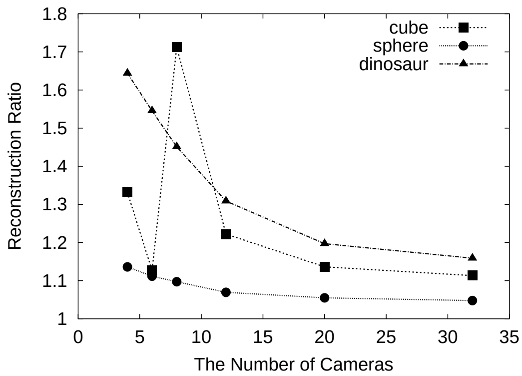

Figure 3. Reconstruction ratio and the number of cameras

3. System Overview

According to the discussion in section 2, the 20 cameras are arranged so that the system can observe an object from 4π directions, as shown in Fig.4. The undersurface of the object, which is unobservable by other systems, is observed by our system. We set the measurement region of our system ac-

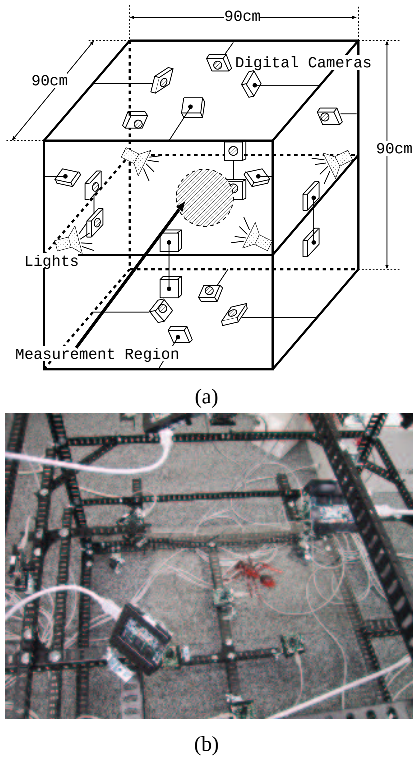

Figure 4. 4π **Measurement System**

cording to the size of target objects. It is the spherical region with a diameter of 20cm. Most insects fit in this region.

We calibrated the intrinsic and extrinsic camera parameters using Zhang's method[8] and Ueshiba's factorization-based method[9]. The average of calibration error on the images is about 0.7 pixels.

Our system uses IEEE-1394 board level cameras – Point Grey Research Dragonfly – with 6mm focal length. This camera has 640×480 progressive scan CCD sensor, and outputs 8-bit Bayer tiled image. Thus conversion from Bayer tiled image to RGB color image is needed by software processing.

The cameras are connected to 6 computers via SyncUnits, hence 3 to 4 cameras are connected to one computer. The computer sends capture control signals to the connected cameras, and stores the output image from the cameras. The capture control signal sent by each computer is buffered in SyncUnits, and transmitted from SyncUnit to the cameras simultaneously at 15Hz. Thus, our system can store the images at 15fps. The size of data for the camera is 640×480(pixels)$\times 8$(bit)$\times 15$(fps)≈ 4.4MBytes/sec, and the computers have 60GB RAID HDD. Hence, the measuring time of our system is about 78min.

Volume reconstruction needs silhouettes on which the object is projected. These silhouettes are extracted from observed images by the background subtraction method. In order to raise the accuracy of these subtracted silhouettes, the monochromatic (blue) screen covers the cubic frame.

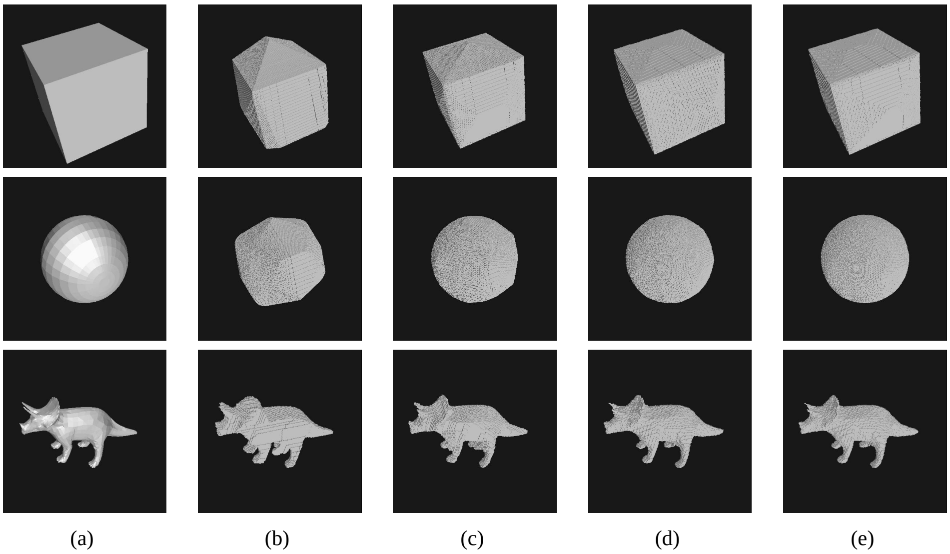

Figure 2. Reconstructed volume of synthesized data: (a) show the synthesized data, and (b)-(e) show reconstructed volume from 4, 12, 20 and 32 cameras.

4. The Spatial Resolution and the Influence of the Object's Motion

In order to evaluate our system, we discuss two issues in this section; the spatial resolution and the influence of the object's motion on it. The spatial resolution is influenced by the image resolution and shutter speed of the camera, in other words, they are influenced by equipment of our system.

The Spatial Resolution The reconstructed volume is described by the voxel representation. The size of each voxel depends on the resolution of observed image and the size of measurement region.

Now we focus on a pyramid region obtained by back projecting a pixel on an image in order to discuss the voxel size. The size of region where this pyramid and measuring region intersect gives the minimum size of each voxel.

In our system, each camera outputs 640×480 pixels image, and the measurement region is the spherical region with a diameter of 20cm. Thus, we set the size of each voxel to $20cm/480 \approx 0.5mm$ cube.

The Influence of the Object's Motion on the Spatial Resolution What we have to consider now is the influence of the object's motion on the spatial resolution. When the object moves while the camera's shutter is open, the motion blur is occurred on the observed image. Also, the difference of timing of the capture signal to the camera causes the difference of the position of projected object by the camera. This position difference should be less than the spatial resolution, in other words, this difference should be smaller than 1 pixel on each image.

The shutter speed of the camera is $1/16000$sec, and the size of each voxel is 0.5mm cube. When the target object moves less than $8m/s$, the size of motion blur on the observed image will be within one pixel on the image. Also, the error of the timing of transmitting the signal is about $1/8000$sec. When the target object moves at $4m/s$, the difference will be about 1 pixel. Hence, our system can measure objects that move up to $4m/s$. The walk speed of insects is about $1m/s$, so $4m/s$ speed limitation is sufficient for observing freely-moving live insects.

5. Experiment

In order to evaluate the performance of our system, we conducted experiments.

First, we reconstructed the volume of several objects placed on a transparent acrylic board (Fig.5). The position of cameras is shown in Fig.4. Some objects have flat texture and the lighting conditions generate the reflected region of the acrylic board on observed images. These reconstructed volumes are shown in Fig.5. These experimental results show the robustness of our system to flat texture and lighting conditions.

Next, we reconstructed the volume of a moving object, a human hand. Fig.6 show input images and the reconstructed volume. The tip of the middle finger moved at approximately 0.5m/s. This reconstruction results shows that our system can reconstruct the volume of moving objects.

Finally, in order to compare our system to simulated results, we reconstructed the volume of two objects whose geometry is known. The reconstructed volumes are shown in Table 1 and Fig.7. The reconstruction ratios by 20 cameras are close to the ratio of the simulated results, while the ratios by 12 upper cameras are distant from the simulated results. From these results, we consider that our 4π measurement system is almost equivalent to the simulated results. In addition, we reconstructed the volume of two moving objects. Each object moves at about 2m/s. The reconstruction ratios are shown in Table 1. There is small difference between the reconstruction ratio of the moving object and the reconstruction ratio of the standing still object. This difference is due to the noise on the silhouette images.

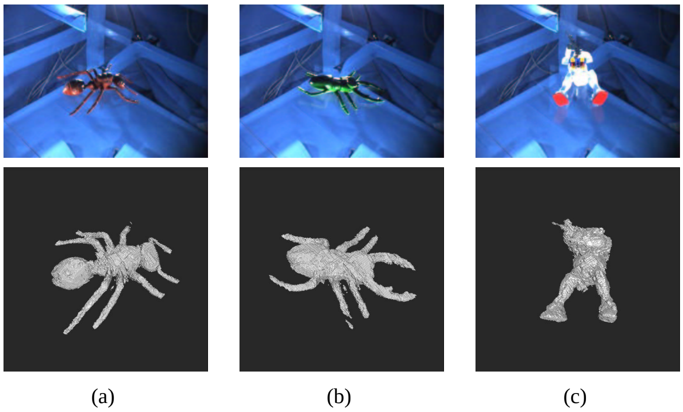

Figure 5. Observed images and reconstructed volume: upper row show one of the observed images, lower row show the reconstructed volume.

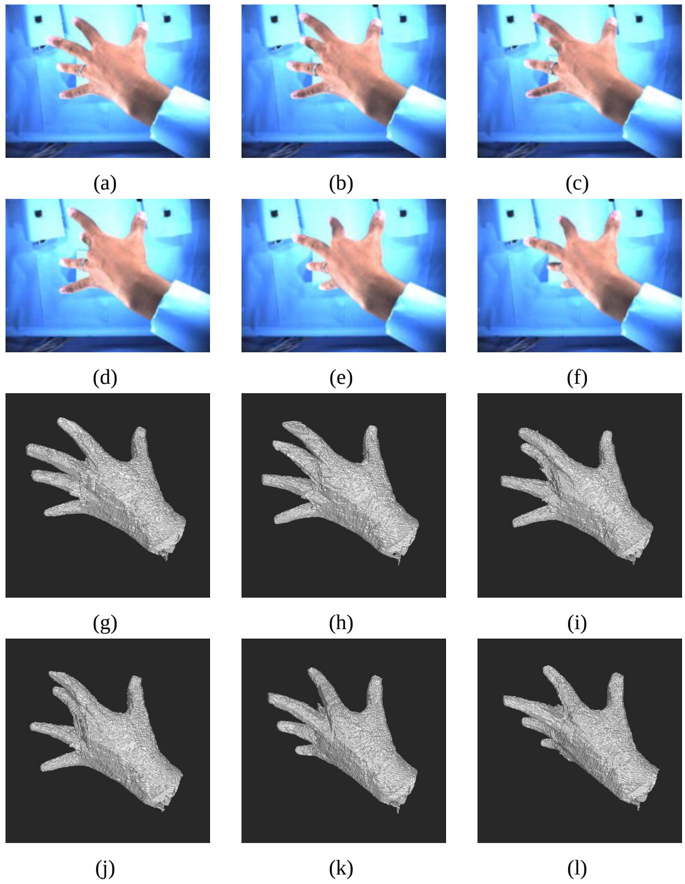

Figure 6. Reconstructed volume of moving objects: (a)-(f) show input images of a human hand motion, and (g)-(l) show reconstructed volume.

6. Conclusion

In this paper we proposed a system to reconstruct the volume of freely-moving objects. We arranged 20 cameras so that they observe the object from 4π directions. And we discussed and evaluated the accuracy of volume reconstruction, the spatial resolution and the influence of the object's motion on it. Our system has 0.5mm spatial resolution, and it can reconstruct objects moving at highest $4m/s$.

Data reduction remains as future work. The size of data for one camera is 640×480(pixels)$\times$8(bit)$\times$15(fps)$\approx$ 4.4MBytes/sec. We will reduce the acquired data by hardware processing and parallel computing[11].

The size of measurement region also remains as future work. Actually, it is necessary to determine the measurement region according to the object to measure. We set the measurement region to the spherical region with a diameter of 20cm. However our system can enlarge and expand the measurement region. The spatial resolution depends on the measurement region, so there is a tradeoff between the spatial resolution and the measurement region.

References

[1] Marc Levoy, et al., "The Digital Michelangelo Project: 3D Scanning of Large Statues," SIGGRAPH2000, pp.131–144, 2000

[2] Daisuke Miyazaki, Takeshi Ooishi, Taku Nishikawa, Ryusuke Sagawa, Ko Nishino, Takashi Tomomatsu, Yutaka Takase, Katsushi Ikeuchi, "The Great Buddha Project: Modelling Cultural Heritage through Observation," VSMM2000 (6th International Conference on Virtual Systems and Multimedia), pp.138–145, 2000

[3] Takeo Kanade, Peter Rander, P.J. Narayanan, "Virtualized Reality: Constructing Virtual Worlds from Real Scenes," IEEE MultiMedia, vol.4, no.1, pp.34-47, 1997

[4] R. Bajcsy, R. Enciso, G. Kamberova, L. Nocera, R. Sara, "3D reconstruction of environments for virtual collaboration," Proc. 4th IEEE Workshop on Applications of ComputerVision, p.160-167, 1998

[5] Wojciech Matusik, Chris Buehler, Ramesh Raskar, "Image-Based Visual Hulls," Proc. SIGGRAPH2000, pp.369-374, 2000

[6] M. Okutomi, T. Kanade, "A multiple-baseline stereo," IEEE Trans. Pattern Analysys and Machine Intelligence. PAMI-15(4).pp.353-363,1993

[7] W.N. Martin and J.K. Aggarwal, "Volumetric Descriptions of Objects from Multiple Views," IEEE Trans. on

Table 1. Statistics of reconstructed volume

object	camera	volume		reconstruction
		system	true	ratio
cube	12 (upper cameras)	$234.82cm^3$	$125cm^3$	1.879
cube	20	$146.62cm^3$	$125cm^3$	1.173
cube	20 (moving)	$150.24cm^3$	$125cm^3$	1.201
sphere	12 (upper cameras)	$79.12cm^3$	$65.45cm^3$	1.209
sphere	20	$69.29cm^3$	$65.45cm^3$	1.054
sphere	20 (moving)	$68.25cm^3$	$65.45cm^3$	1.043

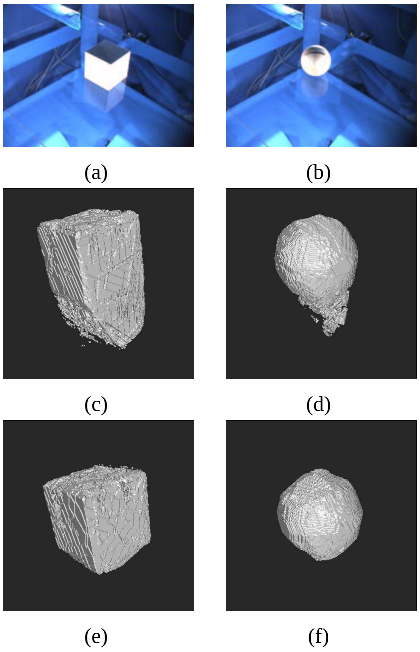

Figure 7. Accuracy evaluation: (a)(b) show one of the observed images, (c)(d) show the reconstructed volume from 12 cameras arranged at upper region of the object, (e)(f) show the reconstructed volume from 20 cameras(4π measurement).

Pattern Analysis and Machine Intelligence, Vol. 5, No. 2, pp.150-158, 1983

[8] Z. Zhang, “A flexible new technique for camera calibration,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22, No.11, pp.1330-1334, 2000.

[9] Ueshiba, T and Tomita, F, “A Factorization Method for Projective and Euclidean Reconstruction for Multiple Perspective views via iterative depth estimation,” Proc. European Conf. Computer Vision, vol.1, pp.196-310, 1998

[10] Aldo Laurentini, “How Far 3D Shapes Can Be Understood from 2D Silhouettes,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, No.2, pp.188-195, 1995

[11] Yoshinari Kameda, Takeo Taoda, Michihiko Minoh, “High Speed 3D Reconstruction by Video Image Pipeline Processing and Division of Spatio-Temporal Space,” Proceedings of MVA’98 IAPR Workshop on Machine Vision Applications, pp.406-409, 1998.