Proceedings of the 2006 IEEE International Conference on Robotics and Automation
Orlando, Florida - May 2006

# Robust Motion Capture System against Target Occlusion using Fast Level Set Method

Yumi Iwashita, Ryo Kurazume
and Tsutomu Hasegawa
*Graduate School of Information Science and Electrical Engineering*
*Kyushu University*
*6-10-1, Hakozaki, Higashi-ku, Fukuoka, Japan*
*yumi@irvs.is.kyushu-u.ac.jp*

Kenji Hara
*Graduate School of Design*
*Kyushu University*
*4-9-1,Shiobaru, Minami-ku, Fukuoka, Japan*

***Abstract*— This paper introduces a new motion capture system for recovering 3D models of multiple persons separately and robustly against occlusion in real-time. Various markerless motion capture systems using video cameras have been proposed so far. However, in case that there are multiple persons in the scene at the same time, it is quite difficult to reconstruct a precise 3D model of each person separately due to the occlusion between them. To deal with this problem, the Fast Level Set Method is utilized in the proposed system for integrating stereo range data which is captured by multiple stereo cameras located around the target people. To reconstruct precise 3D models in real-time, the proposed system is implemented on a PC cluster with seven PCs and four stereo cameras. Tracking experiment of multiple persons and real-time reconstruction of 3D human models using the proposed system are successfully carried out.**

## I. Introduction

Motion capture systems have been widely used for creating high-quality CG contents or 3D digital archive of professional skill performed by an expert. Several motion capture systems are available in the market and they can be classified in the following techniques; mechanical, magnetic, or optical systems. However, there are still many problems to be solved in these systems. For example, a user has to wear special equipments such as optical or electromagnetic markers. A special studio is also required for capturing body motion precisely.

Markerless motion capture technique using multiple video cameras is expected to overcome these limitations. This technique enables to recover not only motion of a human but also its appearance. The volumetric intersection technique and the multi-view stereo technique have been proposed so far.

The volumetric intersection technique is based on a silhouette constraint, that is, a 2D silhouette of an object constrains the object inside a frustum produced by back-projecting the silhouette from the corresponding viewpoint [1] [2]. Therefore, by extracting intersection of the frusta of all silhouettes, volumetric description of a 3D model which may be occupied by the object is obtained. Various algorithms have been proposed to construct a volumetric model in real-time, for example, the use of specialized hardware [3] [4], or a PC cluster [5] [6] [7].

However, a 3D model constructed by the volumetric intersection method is limited to a conservative approximation of the true shape of the target. This approximation is called the visual hull. For recovering a model including a concave shape more accurately, this method was extended to utilize photometric information [8], space carving method [9], and voxel occupancy algorithm [10]. In [11], a 3D voxel model was converted into a deformable mesh model, and the mesh model was refined using photometric constraint. However these methods are computationally expensive in general.

The multi-view stereo technique is based on the stereo correspondences between multiple images captured from different viewpoints. CMU's Virtualized Reality system [12] [13] arranged 51 video cameras to form a 3D dome. This system captured synchronized camera images and applied the multi-baseline stereo technique to construct a 3D model of an object. However, since these methods require off-line processing, they have difficulties in reconstruction of complex shapes such as human bodies in real-time. Voxel coloring method is also utilized [14], which checks the color consistency between the views. Though similar systems [15] [16] were also proposed to extract a 3D shape based on the multi-baseline stereo technique, off-line processing is indispensable.

Moreover, most conventional methods have been developed for tracking a single person. In case that there are multiple persons in the scene at the same time, optical occlusion occurs between multiple persons. This makes it quite difficult to reconstruct a 3D model of each person separately. In addition, since the visual hull represents all possible regions that may contain objects, some regions may be empty and the regions which contain the target cannot be distinguished clearly. Furthermore, a common problem for these conventional methods is that the calibration process of camera positions is considerably complex.

In this paper, a new 3D motion capture system is proposed for recovering 3D models of multiple persons separately and robustly against occlusion in real-time. We have developed a prototype system of the motion capture system using two stereo cameras. However, 3D models reconstructed in real-time were coarse because of the large calculation cost. To overcome this problem, the proposed system is implemented on a PC cluster with seven PCs for reconstruction of precise 3D models in real-time. In the proposed system, four stereo cameras are located around the target people, and the 3D model of each person is reconstructed separately by integrating stereo range data using the Fast Level Set Method (FLSM) [17]. The FLSM [17] was proposed as a high speed execution technique of the Level Set Method (LSM) [18] [19], which is used for several applications such as 3D modeling and motion tracking. The proposed system is able to estimate occluding regions and maintain separate and closed 3D surfaces of multiple persons using the FLSM. A camera calibration method is also proposed, which can be performed automatically using dense range data.

This paper is organized as follows; Section II introduces the Level Set Method proposed by Osher and Sethian, and its high speed execution algorithm named the Fast Level Set Method. A detail description of the new motion capture system is presented in Section III. Section IV shows some experimental results with human bodies. Finally, our conclusions are presented in Section V.

## II. Fast Level Set Method

The Level Set Method (LSM) [18] [19], introduced by Osher and Sethian, has attracted much attention as a method that realizes a topology free active contour model. Various applications based on the LSM have been presented so far including motion tracking, 3D geometrical modeling, and simulation of crystallization. However, the calculation complexity remains an open problem. To overcome this problem, the Fast Level Set Method (FLSM) [17] was proposed as a high speed execution technique, and it was applied to real-time applications, such as 2D real-time tracking of moving objects in video images [17]. In this section, basic ideas of the LSM and the FLSM are described.

### A. Level Set Method and its high speed execution algorithm

The LSM utilizes an implicit function $\Phi$ which is defined in a space one dimensional higher than that of where a contour (surface) of interest is described. This function $\Phi$, which is defined as a distance function from a current contour in general, is updated according to a next PDE (Partial Differential Equation).

$$\Phi_t = -F(\kappa) \mid \nabla\Phi \mid \tag{1}$$

where, $\kappa$ is a local curvature of $\Phi$, and $F$ is a speed function. The contour to be tracked is detected as the cells with a value of zero of the implicit function (zero level set), that is, the contour line of $\Phi = 0$. In the implementation of the LSM, the space is uniformly split by cells, and Eq. (1) is solved iteratively using numerical schemes such as the upwind scheme.

To solve Eq. (1), the speed function $F(\kappa)$ has to be determined at each cell for every update process of $\Phi$. The distribution of the $F(\kappa)$, which is known as the extension velocity field [20], is constructed as follows: i) at the current zero level set cell, $F$ is calculated according to the intensity of the current image at first; ii) next, at each cell except the zero level set cell, the speed function $F$ is copied from the nearest zero level set cell. However, finding the nearest zero level set cell needs large calculation cost.

In addition, since integral errors are accumulated during the calculation of updating, the reinitialization process is required in which the proper quantity (the distance from the current zero level set cell) at each cell is re-calculated periodically after several update processes. The constructed field is called the distance field. However, this reinitialization process is time consuming, because the nearest zero level set has to be determined at each cell. Therefore, this is also a major obstacle for the high speed implementation of the LSM.

To overcome these problems, several techniques have been proposed in the past, such as the Narrow Band Method [19], the Fast Marching Method [19], SFA (Sparse Field Algorithm) [21], and HERMES [22]. The most popular and efficient method is the one proposed by Adalsteinsson and Sethian [20] in 1999, which combines the Narrow Band Method and the Fast Marching Method. The calculation cost of this method is $O(n^2 \log n)$ (n is the number of cells along a side of the voxel space in 3D space).

### B. Fast Level Set Method

Though the Narrow Band Method is high speed execution algorithm compared with the conventional Level Set Method, the computational cost is still expensive. To overcome this problem, the Fast Level Set Method (FLSM) with the calculation cost of $O(n^2)$ was proposed [17].

The key idea of the FLSM is the use of a reference map (Figure 1 (a)). This map indicates classification in which each cell is categorized according to a distance from a center cell. For example, the class $R_r$ consists of cells which are located $\sqrt{r}$ away from the center cell.

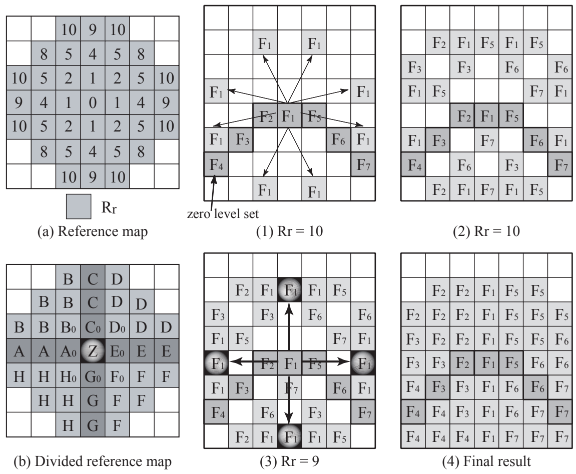



Fig. 1. Reference map and the construction process of the extension velocity field.

The extension velocity field is constructed efficiently using the reference map. Here, it is assumed that the speed function at the zero level set cell has already been determined. At first, one of the zero level set cells is chosen. Then, using the class $R_{\delta(\delta+1)}$ in the reference map, all cells which are located

$\sqrt{\delta(\delta+1)}$ away from this zero level set cell are selected, and the speed function at the zero level set cell is registered to these cells tentatively. This procedure is repeated for all the zero level set cells. Next, the same procedure is performed using the next class $R_{\delta(\delta+1)-1}$. This process is repeated until all the classes in the reference map are investigated. At the end of the registration processes for all the classes, the speed function of the nearest zero level set cell is registered at each cell, and the extension velocity field is constructed consequently (Figure 1 (c)).

In this process, the speed function is registered in order of a distance from the zero level set. Thus, by overwriting the distance stored in the reference map at the same time, the distance field can also be created. Though this reinitialization process is computationally expensive in general as explained above, additional cost of this process is quite small for the FLSM. Actually, it requires only memory access time and the total calculation cost is almost same as the one without the reinitialization process. Therefore, the reinitialization process can be performed frequently, even up to every update process, with trivial calculation cost.

Moreover, the method can be improved by using a divided reference map (Figure 1 (b)). For example, let us consider a zero level set cell $Z$. If a left cell $A_0$ adjacent to the cell $Z$ is also a zero level set cell, there must be no cells which are nearer to $Z$ than $A_0$ in the left side region of the cell $Z$ (A,B, and H). Therefore, by skipping the registration process in these regions, it is possible to reduce the number of useless overwriting and execute the construction process of the extension velocity field more efficiently.

The comparison of the calculation cost in 3D ($n \times n \times n$) is shown in Table I. The cost of the conventional level set method using the fast marching method [20] is $O(n^2 \log n)$. On the other hand, the calculation cost of this method is $O(n^2)$. Moreover, the additional cost of the reinitialization process is quite small in this method.

TABLE I

COMPARISON OF CALCULATION COSTS. $\delta(\ll n)$ IS THE WIDTH OF THE NARROW BAND.

| | Level Set Method | Fast Level Set Method |
|---|---|---|
| Construction of velocity field | $O(\delta n^2 logn)$ | $O(\delta^3 n^2)$ |
| Reinitialize | $O(\delta n^2 logn)$ | - |
| Updating process | $O(\delta n^2)$ | $O(\delta n^2)$ |
| Detection of zero level set | $O(\delta n^2)$ | $O(\delta n^2)$ |

## III. A New Motion Capture System for Recovering 3D Models Robustly against Occlusion

In this section, we propose a new prototype model of the 3D motion capture system using multiple stereo cameras and the FLSM. This system enables to reconstruct precise 3D models of multiple persons separately by applying the FLSM. To achieve real-time processing, the algorithm is implemented on a PC cluster with seven PCs. In the proposed system, each 3D models is recovered by the following procedure:

1) At first, multiple stereo cameras are installed as shown in Fig.2. The camera calibration to estimate mutual camera positions is performed by the procedure explained in Section III A. Internal camera parameters are assumed to be calibrated beforehand.
2) Synchronized depth images are captured from stereo cameras.
3) Occluding regions are estimated from the stereo range data and current positions of surfaces as described in Section III C.
4) The FLSM is applied and the isolated smooth surfaces of each person are reconstructed. The surface in the estimated occluding region remains stationary to maintain the closed and smooth surface.

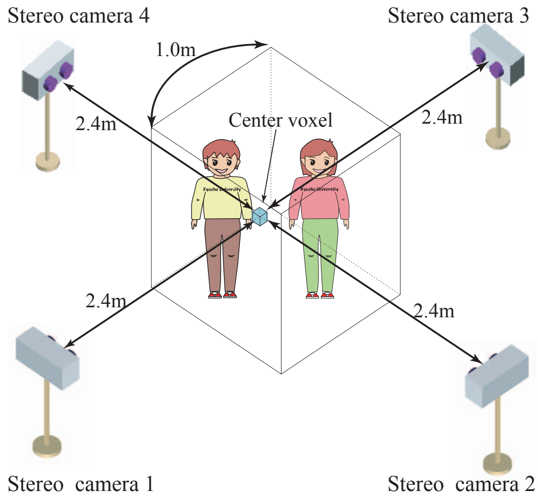



Fig. 2. 3D motion scanner using multiple stereo cameras and the camera settings.

### A. Camera calibration

Camera calibration to estimate mutual camera positions is indispensable when multiple cameras are used. Though quite complicated procedure is required for the precise camera calibration in general, it is possible for the proposed system to perform the precise camera calibration with simple procedure. Firstly, 3D stereo range data of an object with known shape is captured by all stereo cameras. Next, the iterative closest point (ICP) algorithm [23] is applied to the range data with respect to the object surface. Thus, all range data are aligned to the object surface. Finally, the relative camera position from the object is estimated. Since there is no need to extract specific feature points as conventional camera calibration procedures, the above procedure can be automated. Moreover, by utilizing dense 3D stereo range data, precise camera calibration result is performed.

### B. 3D shape reconstruction of human bodies using the FLSM

After calibrating the camera parameters, depth images of the target objects are captured from all stereo cameras, and the 3D models are reconstructed by the following procedure. At first, a 3D space is divided into small voxels ($N \times N \times N$, $N$ is the number of voxels along a side of the 3D space). Each voxel in the 3D space can hold one of two kinds of status, IN (internal human body) or OUT (external human body). As an initial state, IN is assigned to all voxels in the 3D space. Next, as shown in Fig.3, each pixel (x,y) of the depth image is backprojected onto the 3D space, and OUT is voted to voxels where the distance from the camera is equal or smaller than the depth value of (x,y). This process is repeated for all depth images, and a region where IN is assigned is extracted as a region-S.

Next, the FLSM is applied to the region-S for reconstruction of the 3D models. Here, voxels of the region-S which are adjacent to voxels voted OUT are defined as the stopping region, and other voxels of the region-S are defined as the internal region (Fig.4(c)). A closed initial surface including the whole 3D space is set at first, and it is shrunk and split by the FLSM. When the zero level set reaches to the stopping region, the propagation of the zero level set is stopped by setting a small value to the speed function at the corresponding voxel.

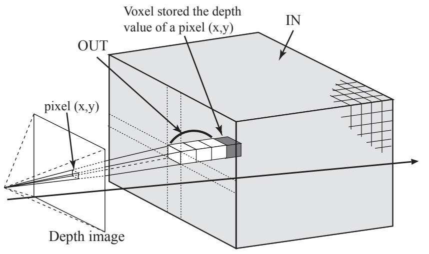



Fig. 3. Voxel voting.

### C. Robust reconstruction of 3D models against occlusion

Fig.4 (a) shows the cross section of the 3D space, where multiple human bodies and two stereo cameras are placed back and forth. Fig.4 (c) shows the stopping and internal regions. At this moment, all surfaces of the targets are not occluded and depth images are captured without occlusion. Therefore the separate 3D models can be reconstructed by applying the FLSM. However, when the target persons move as shown in Fig.4 (b), the mutual occlusion occurs. At this moment, since the stopping region is defined as shown in Fig.4 (d), a connected 3D model indicated with a dotted line in Fig.4 (b) is reconstructed along the occluding boundary. To reconstruct each 3D models separately, the zero level set in the occluding regions is not propagated by setting a small value to the speed function as shown in Fig.5. To evaluate whether a voxel is in the occluding regions or not, we utilize the change of the depth between consecutive periods. Here, $d1$ and $d2$ are distances from the zero level set A at time $t-1$ to the voxels in which depth value from each viewpoint is stored at time $t$ as shown in Fig.5. $D$ is the sum of $d1$ and $d2$, and $r_d$ $\left(0 \leq r_d \leq \frac{1}{2}\right)$ is the distance radio defined as the following equation.

$$r_d = \frac{min(d1, d2)}{D} \tag{2}$$

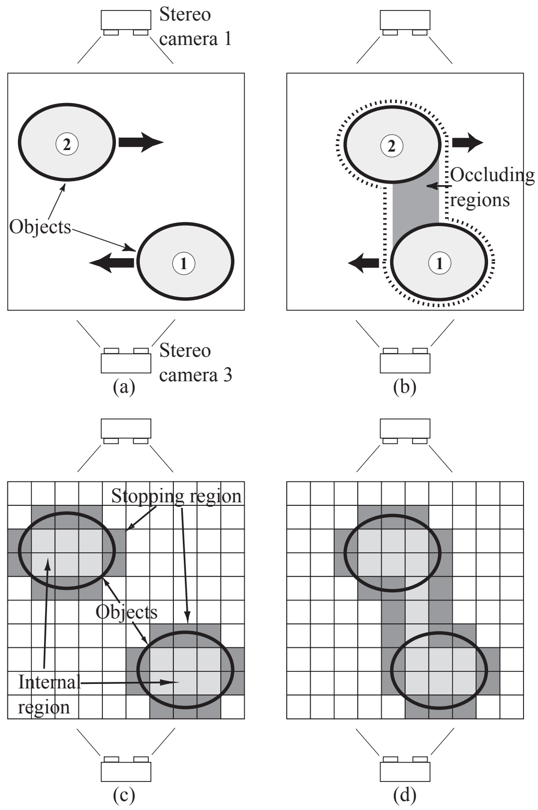



Fig. 4. Stopping region and inner region.

In our system, the speed function $F(\kappa)$ at time $t$ is defined as

$$F(\kappa) = k_I(a - b\kappa) \tag{3}$$

where $a = 1.0$ and $b = 0.1$ are used in following experiments, and $k_I$ is determined according to the distance ratio $r_d$ as follows.

1) In case that the zero level set A is in the stopping region, surface is not propagated from the zero level set at time $t-1$ . ($k_I = 0$)
2) In case that the zero level set A is in the internal region and $D > c$ (the region-S is thick enough along the depth direction).
   a) In case $r_d > d$ (the thickness is changed shortly), the zero level set is considered to be located in occluding regions, and the surface is not propagated from the zero level set at time $t-1$. ($k_I \approx 0$)
   b) In case $r_d \leq d$ (the thickness is almost same), the zero level set is considered to be located in outside of occluding regions, and the surface is propagated. ($k_I = C_1$)
3) In case that the zero level set A is in the internal region and $D \leq c$, the zero level set is considered to be located in outside of occluding regions, and the surface is propagated. ($k_I = C_2$)
4) In case that the zero level set A is in neither the internal region nor the stopping region, the surface is propagated. ($k_I = -C_3$)

In the following experiments, $c = \mathrm{N}/4$ and $d = 0.2$ are used. $C_1$, $C_2$, and $C_3$ are constants which depend on the speed and the size of moving targets. In the motion tracking experiment of multiple human bodies, we use $C_1 = 4.5$, $C_2 = 9.0$, and $C_3 = 18.0$, and in the rest of the experiments $C_1 = 7.5$, $C_2 = 15.0$, and $C_3 = 30.0$ are used. By above procedure, in case that the thickness of the target changes shortly, it is considered that the occlusion is occurred, and the position of the previous zero level set voxels in the regions is remained. Therefore, each 3D models of the multiple targets can be reconstructed even if the occluding regions appear.

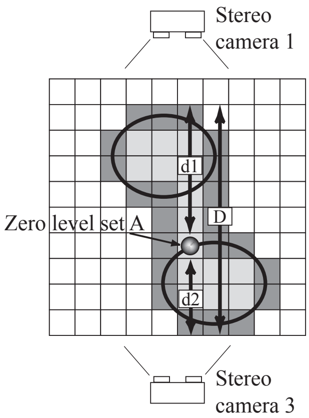



Fig. 5. Zero level set in occluding boundary.

### D. Implementation on a PC cluster system

For achieving real-time processing the above process is implemented on the PC cluster connected through Myrinetxp. The processing procedure for the system consisting two stereo cameras and seven PCs is as follows (Fig.6). At first, synchronized depth images are captured from two stereo cameras, which are connected to two PCs (PC1 and PC2, respectively). Next, the depth images are sent to four PCs which execute the FLSM calculation (PC3-PC6). Here, the 3D space is divided into four regions, and each region is assigned to one of four PCs, respectively. Each PC has two CPUs and calculates the FLSM in parallel. Finally, results of the FLSM are sent to a PC (PC7) for displaying 3D models.

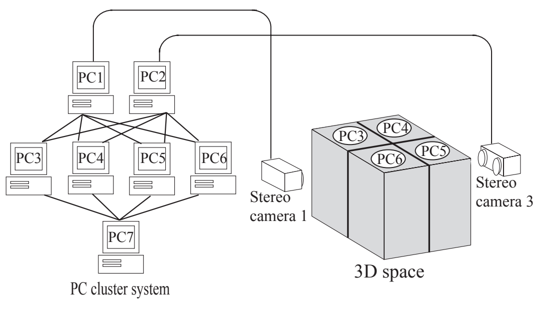



Fig. 6. PC cluster system.

### E. Characteristics of our system

Characteristics of our motion capture system are summarized as follows:

1) By recovering and tracking closed surfaces with the FLSM, smooth and isolated surfaces are recovered and maintained even if parts of depth images cannot be captured temporarily. Therefore, separate 3D models of multiple human bodies are reconstructed in case that there are multiple persons in the scene and mutual occlusions occur.
2) External camera parameters are estimated precisely with the simple procedure using the ICP algorithm [23] and dense stereo range data. This procedure can be performed automatically.
3) The proposed system is robust against spike noise within stereo range data by restricting the maximum value of the speed function and updating of an implicit function frequently.
4) Since the proposed system can easily distinguish between target regions and background regions using depth images, a special studio is not necessarily required.

In case that there are multiple persons in the scene, the occurrence of occlusion is a serious problem. Several target tracking techniques dealing with occlusion have been proposed so far, for example, tracking based on object contours using a monocular camera [24], [25] or appearances [26].

In the multiview-based tracking techniques, the weighted integration of image features across successive views has been proposed based on the visibility at each view [27]. Otsuka et al. proposed a framework for multiview occlusion analysis on a 2D plane using a recursive Bayesian estimation [28]. However, the targets are limited to be simple objects with known shape such as ellipsoids on the 2D plane. On the other hand, our proposed system can reconstruct 3D models with arbitrary shape.

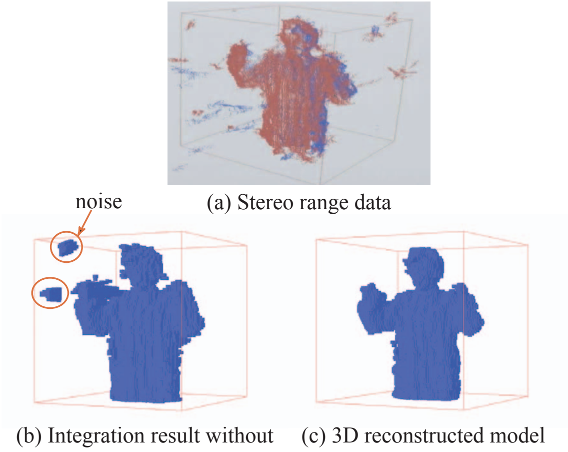
## IV. Experiments

In this section, we show some experimental results of the reconstruction and tracking of 3D models using the proposed system. The calculation is done by seven PCs with Pentium IV Xeon processor, 3.06 GHz, and the depth images are captured with Bumblebee stereo camera (PointGrey Inc.).

### A. Robust 3d model reconstruction against noise using FLSM

Firstly, Fig.7 shows the robustness of the proposed system against spike noise within stereo range data of a human body.

The multi-view stereo techniques compute correspondences across images captured from different viewpoint and recover the 3D model by triangulation. However, when the correspondence calculation fails, the depth image is corrupted by spike noise. On the other hand, the proposed system is robust against spike noise within stereo range data as shown in Fig.7. In this experiment, the 3D space is a cube 1.0m on a side, and 2 stereo cameras are placed at a distance of 2.4m from the center voxel of the 3D space as shown in Fig.2 (Stereo camera 1 and 3 are used). The resolution of the 3D space is $100 \times 100 \times 100$

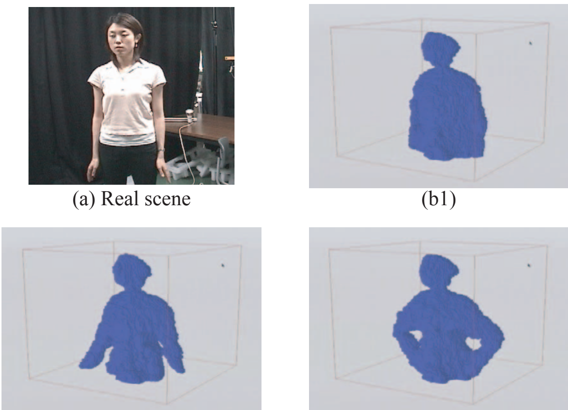



(a) Stereo range data

(b) Integration result without the FLSM

(c) 3D reconstructed model with the FLSM

Fig. 7. Effect of the FLSM.



(a) Real scene

(b1)

(b2)

(b3)

Fig. 8. 3D reconstruction of the human body.

(10mm voxel on a side). Fig.7(a) shows the raw stereo range data from two viewpoints. Fig.7(b) and Fig.7(c) show the integrated results of stereo range data with and without the FLSM, respectively. As seen in these results, it is verified that the FLSM has robustness against spike noise.

### B. Real-time 3D model reconstruction of a human body

We carried out experiments of tracking and reconstructing the 3D model of a human in real-time. Fig.8(a) and Fig.8(b1)-(b3) show the actual image of the target and the reconstructed 3D model, respectively. Stereo range data is acquired in every 37 [msec.]. Table II shows the calculation time of the FLSM with the PC cluster and single PC. The maximal processing time of the FLSM with the PC cluster is 24 [msec.] for one update period, and the maximal communication time between the PCs is 2 [msec.]. On the other hand, the processing time with single PC is 77 [msec.].

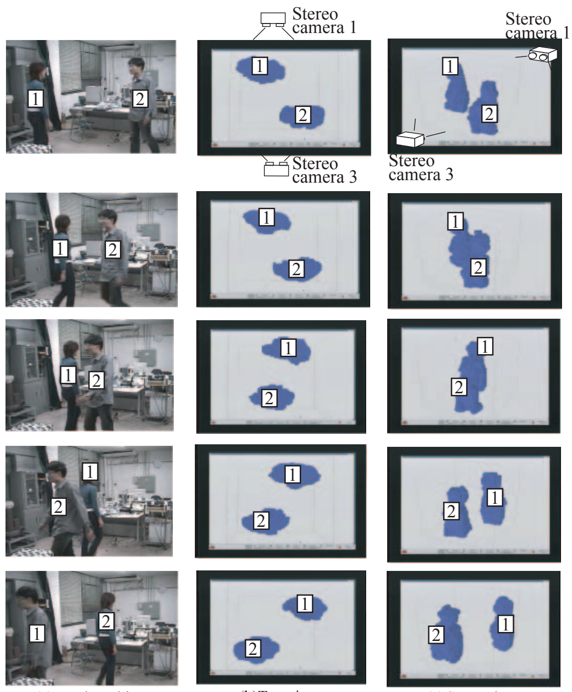
### C. Real-time 3D motion tracking of multiple human bodies

Fig.9 shows the 3D motion tracking experiment of multiple human bodies in real-time. Here, Fig.9(a) shows the actual images of the targets. Fig.9(b) and Fig.9(c) show the cross

TABLE II
CALCULATION TIME OF THE FLSM

| | PC cluster | | | Single PC |
|---|---|---|---|---|
| | Average | Max. | Min. | |
| Calculation time of the FLSM [msec.] | 17 | 24 | 9 | 77 |
| Number of the reconstructed 3D voxel | 5765 | 7418 | 3481 | 13357 |

section of the 3D space and the oblique section of the 3D space, respectively. The multiple targets walk across the 3D space as shown from Fig.4(a) to Fig.4(b). As seen in these results, each 3D models is tracked and reconstructed separately even if the mutual occlusion occurs. This result suggests that the proposed system is capable of detecting and tracking multiple human bodies in the 3D space individually even if parts of them are occluded.




(a)Moving objects (b)Top view (c)Cross view

Fig. 9. 3D reconstruction of multiple human bodies.

### D. Reconstruction of precise textured 3D model

Fig.10 shows experimental results of the precise 3D model reconstruction with texture image. The 3D voxel model is converted into triangular patches by the Marching Cubes algorithm, and then the texture images captured by four stereo cameras are mapped to the triangular patches.

## V. CONCLUSION

This paper described a new motion capture system, which reconstructs 3D models of multiple human bodies robustly against occlusion in real-time. In the proposed system, the Fast Level Set Method is applied to the stereo range data captured by multiple stereo cameras, and separate 3D models

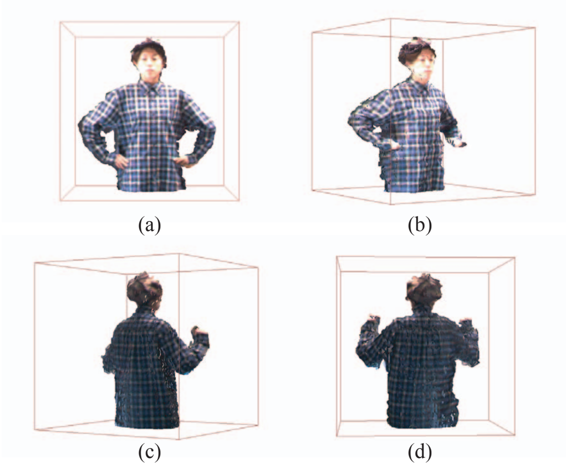


Fig. 10. Texture mapping result.

are obtained even if there are multiple persons in the scene. We implemented parallel programming on a PC cluster with seven PCs in order to perform the real-time 3D reconstruction. The camera calibration method is also proposed to be performed automatically using dense range data. The efficiency of the proposed system was verified through the experiments with multiple human bodies.

In the future work, we plan to install more stereo cameras around the targets, and develop a system which will be able to reconstruct a more number of 3D models in more details. Moreover, since the proposed camera calibration can be performed automatically with simple procedure, we plan to measure and conserve whole body motion of Japanese traditional folk dance in outdoor environment.

## REFERENCES


[1] W. Martin and J. Aggarwal, "Volumetric description of objects from multiple views," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, no. 2, pp. 150–158, 1983.

[2] A. Laurentini, "The visual hull concept for silhouette-based image understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 2, pp. 150–162, 1994.

[3] M. Li, M. Magnor, and H. Seidel, "Hardwareaccelerated visual hull reconstruction and rendering," in *In Proc. of Graphics Interface*, 2003.

[4] J. Hasenfratz, M. Lapierre, and F. Sillion, "A real-time system for full body interaction with virtual environments," in *Eurographics Symposium on Virtual Environments*, 2004, pp. 147–156.

[5] E. Borovikov and L. Davis, "A distributed system for real-time volume reconstruction," in *IEEE Proc. of CAMP-2000*, 2000.

[6] K. Cheung, T. Kanade, J. Bouguet, and M. Holler, "A real time system for robust 3d voxel reconstruction of human motions," in *Proceedings of the 2000 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2000, pp. 714–720.

[7] T. Matsuyama, X. Wu, T. Takai, and T. Wada, "Real-time dynamic 3d object shape reconstruction and high-fidelitytexture mapping for 3d video," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 14, no. 3, pp. 357–369, 2004.

[8] S. Seitz and C. Dyer, "Photorealistic scene reconstruction by voxel coloring," *International Journal of Computer Vision*, vol. 35, no. 2, pp. 1–23, 1999.

[9] K. Kutulakos and S. Seitz, "A theory of shape by space carving," *International Journal of Computer Vision*, vol. 38, no. 3, pp. 199–218, 2000.

[10] D. Snow, P. Viola, and R. Zabih, "Exact voxel occupancy with graph cuts," in *In Proc. Computer Vision and Pattern Recognition Conf.*, vol. 1, 2000, pp. 345–352.

[11] S. Nobuhara and T. Matsuyama, "Heterogeneous deformation model for 3d shape and motion recovery from multi-viewpoint images," in *Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission*, 2004, pp. 566–573.

[12] T. Kanade, P. Rander, and P. Narayanan, "Virtualized reality: Constructing virtual worlds from real scenes," *IEEE Multimedia, Immersive Telepresence*, vol. 4, no. 1, pp. 34–47, 1997.

[13] P. Narayanan, P. Rander, and T. Kanade, "Constructing virtual worlds using dense stereo," in *Proceedings of 6th IEEE International Conference on Computer Vision*, 1998, pp. 3–10.

[14] S. Vedula, S. Baker, and T. Kanade, "Image-based spatio-temporal modeling and view interpolation of dynamic events," *ACM Transactions on Graphics*, vol. 24, no. 2, April 2005.

[15] O. Faugeras and R. Keriven, "Complete dense stereovision using level set methods," in *In Fifth European Conference on Computer Vision*, vol. 1, 1998, pp. 379–393.

[16] P. Fua, "Regularized bundle-adjustment to model heads from image sequences without calibration data," *Int. Journal of Computer Vision*, vol. 38, no. 2, pp. 153–171, 2000.

[17] Y. Iwashita, R. Kurazume, T. Tsuji, K. Hara, and T. Hasegawa, "Fast implementation of level set method and its realtime applications," in *IEEE International Conference on Systems, Man and Cybernetics*, 2004, pp. 6302–6307.

[18] J. Sethian, "A fast marching level set method for monotonically advancing fronts," in *Proceedings of the National Academy of Science*, vol. 93, 1996, pp. 1591–1595.

[19] ——, *Level Set Methods and Fast Marching Methods, second edition.* UK: Cambridge University Press, 1999.

[20] D. Adalsteinsson and J. Sethian, "The fast construction of extension velocities in level set methods," *J. Computational Physics*, vol. 148, pp. 2–22, 1999.

[21] R. Whitaker, "A level-set approach to 3d reconstruction from range data," *International Journal of Computer Vision*, vol. 29, no. 3, pp. 203–231, 1998.

[22] N. Paragios and R. Deriche, "A pde-based level set approach for detection and tracking of moving objects," in *In International Conference on Computer Vision*, 1998, pp. 1139–1145.

[23] P. Besl and N. McKay, "A method for registration of 3-d shapes," *IEEE Trans. on PAMI*, vol. 14, no. 2, pp. 239–256, 1992.

[24] D. Koller, J. Weber, and J. Malik, "Robust multiple car tracking with occlusion reasoning," in *Proc. ECCV'94*, 1994, pp. 189–196.

[25] J. MacCormick and A. Blake, "A probabilistic exclusion principle for tracking multiple objects," *International Journal of Computer Vision*, vol. 39, no. 1, pp. 57–71, 2000.

[26] M. Isard and J. MacCormick, "Bramble: a bayesian multiple-blob tracker," in *International Conference on Computer Vision*, 2001, pp. 34–41.

[27] A. Mittal and L. Davis, "M2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene," *International Journal of Computer Vision*, vol. 51, no. 3, pp. 189–203, 2003.

[28] K. Otsuka and N. Mukawa, "Multiview occlusion analysis for tracking densely populated objects based on 2-d visual angles," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2004, pp. 90–97.