The Eighteenth International Symposium on Artificial Life and Robotics 2013 (AROB 18th '13),
Daejeon Convention Center, Daejeon, Korea, January 30-February 1, 2013

Real-time 3D-Shape Reconstruction System in Intelligent Space Based on Networked Vision Sensors

Takashi Ishimaru [1], Kazuyuki Morioka [2]

[1,2] *Dept. of Electrical Engineering, School of Science and Technology, Meiji University,* Japan

[1] ce11008@ meiji.ac.jp, [2] morioka@isc.meiji.ac.jp

Abstract: In recent years, intelligent spaces based multi-viewpoint cameras and range image sensors have been developed. Especially, real-time human 3D shape reconstruction system with multiple sensors in intelligent space is considered in this study. In order to achieve application of human shape reconstruction in intelligent space, it is necessary to integrate several kinds of sensors efficiently. This paper focuses on roles and functions required to each sensor node, and proposes systematic structures with RT-component for sensor integration.

Keywords: Intelligent Space, 3D-Shape Reconstruction, RT component

1 INTRODUCTION

Intelligent space is a platform, which organically integrates various sensors, computers and robots via network[1]. Especially, intelligent spaces, which multiple cameras and range image sensors are distributed, have been widely proposed. These systems aim to achieve various applications, such as supporting mobile robot navigation or human tracking with position estimation. In addition to human positions estimation, 3D shapes of humans can be reconstructed by cooperation of image sensors. Positioning and shape reconstruction of humans in intelligent space are promising for human-system interaction. In this study, real-time 3D shapes recognition systems in intelligent spaces are considered. This paper focuses on sensor nodes configuration in 3D-shape reconstruction systems for a target person by integrating image data captured by distributed sensors.

In order to acquire 3D shape of a target human in intelligent space, it is important how to process sensor data from sensors contributed to 3D shape reconstruction. However, intelligent space is a platform including various kinds of sensors. There are various kinds of sensors for contributing 3D shape reconstruction in intelligent spaces. Then, intelligent spaces require the system configuration for integrating various sensors effectively. This paper considers how to assign roles between intelligent space and application of 3D shape reconstruction. Especially, this paper proposes the system configuration that consists of hierarchical sensor nodes in intelligent space and application of 3D shape reconstruction independent of sensors. In the intelligent space side, various kinds of sensor data is integrated through a network as shown in Fig.1. This figure means that all sensor nodes provide same kinds of sensor data to application. In order to provide same kinds of sensor data, two or more sensors are hierarchically integrated for sensor nodes. On the other hand, the application side is independent to kinds of distributed sensors. Then, software configuration of application can be simple. Also, the configuration for accepting additional sensors or new sensors should be required for intelligent spaces as platform of various sensors integration. This paper describes followings. At first, integration methods for configuring sensor nodes from distributed sensors are described. And, an application of 3D shape recognition system for accepting sensor nodes is described. Finally, a total system including sensor nodes and application is shown.

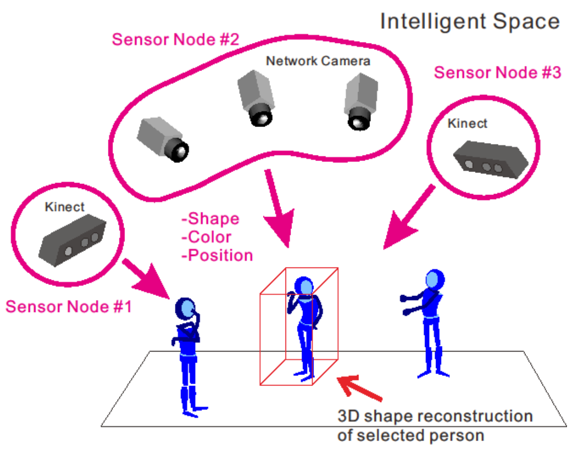


Fig 1. 3D shape reconstruction system

2 APPLICATION IN INTELLIGENT SPACE

As mentioned above, the intelligent space is a platform for integrating various sensors via network. Sensor data is integrated in order to achieve applications with networked sensors. The 3D shape reconstruction in this paper is also one of the applications in intelligent space. Sensors distributed in the intelligent space are not necessarily exploited only for the purpose of 3D shape reconstruction. These sensors should afford the other applications. This means that output of distributed sensors must not depend on the specific application. So, each sensor in intelligent space cannot provide suitable information for the application. On the other hand, the application should not include processes that depend on kinds of sensors, so that the application can accept different sensors added in intelligent space. The 3D shape reconstruction application needs following elements from distributed sensors. The first is 3D point cloud data of a target human. The second is color information corresponding to 3D points. The last is a target position in the intelligent space. Some sensors, such as Kinect sensors, can provide these elements by themselves. The other sensors, such as cameras or LRFs, cannot provide them. However, the sensors such as cameras or LRFs can provide same kinds of elements by integrating several sensors. In the proposed application, sensor nodes integrating several sensors are considered in order to provide information of same form to applications. Such same form information from sensor nodes will make the applications that flexibly accept several kinds of sensors in the intelligent space.

In this study, a system based on networked cameras and Kinect sensors is developed, because information required for 3D shape reconstruction can be easily acquired. Also, the system includes different kinds of sensors. Then, the system will show flexible acceptance of different sensors. As mentioned above, Kinect sensors can provide information required for 3D shape reconstruction. A standalone Kinect sensor can be regarded as a sensor node in 3D shape reconstruction application. On the other hand, raw data obtained by cameras cannot be directly used for 3D shape reconstruction. This raw data should be changed into the same data format in order to be accepted by 3D shape reconstruction application. In this case, a sensor group integrating several cameras can be regarded as a sensor node in the application. In this study, a system design based on RT component is performed.

3 SENSOR NODE COMPONENT

At first, every sensor data acquired in intelligent space should be generalized so that it becomes same data formats for 3D shape reconstruction application. For this purpose, processing in each sensor is integrated and wrapped in order to generalize data formats for 3D reconstruction application. Specifically, RT-component[2] based sensor nodes including several sensors are developed for 3D shape reconstruction application. In this application, three elements mentioned in Section.2 are same data formats for 3D data reconstruction. Therefore, same data formats from sensor nodes must be accepted and processed in 3D shape reconstruction application. Design and development of components of sensor nodes, which output same data formats, are explained below. Especially, sensor nodes based on cameras or Kinect sensors are described.

3.1 Component based on Kinect

A sensor node to output the same data format using a Kinect sensor is configured as follows.

Kinect sensors can acquire RGB image data and depth image data respectively. A target human on the image is tracked using depth data and human posture pattern. 3D point clouds of human can be also acquired from depth data around the target human position. Since Kinect sensors can RGB data in synchronization with the depth data acquired with the infrared camera, color information corresponding to 3D point clouds can be acquired. Then, data required for 3D shape reconstruction application is acquired by each standalone Kinect sensor. This means that each Kinect sensor can be regarded as a sensor node for 3D shape reconstruction application.

In addition, in order to integrate data between sensors nodes in intelligent space, each sensor node must be geometrically calibrated in common world coordinates. Each distributed Kinect has calibrated manually in advance in this system. Internal parameters and external parameters can be obtained by calibration in each sensor.

Although 3D point clouds obtained from a Kinect sensor are numerical values in a camera coordinate system of each Kinect sensor, 3D point clouds are transformed to world coordinates with internal and external parameters. Then, suitable data integrated in 3D shape reconstruction application is acquired by a sensor node by a Kinect sensor.

3.2 Component based on several cameras

Color information of a target human can be acquired with one camera. However, 3D point clouds and human positions in world coordinates cannot be acquired using one camera. Therefore, two or more cameras should be integrated for satisfying minimum requirements of 3D shape reconstruction application. Such integrated cameras can be treated as a sensor node to output the same data formats with a sensor node based on a Kinect sensor. Then, the application built on the platform of intelligent space should have following functions. That is a dynamical configuration of a sensor node according to kinds of applications.

A sensor node component including several cameras is explained below. First, this component obtains images from several cameras. Next, 3D point clouds of a target human are calculated using images from cameras connected to this component. Also, color information corresponding to the point is obtained from images.

Image processing methods in this component are as follows. At first, in order to obtain a human region in each camera image, background subtraction is performed between input images and a background image. Geometrical calibration of each camera is performed in advance in the similar way with Kinect sensors. Camera parameters obtained by calibration can be used for a volume intersection method integrating several camera images. Calculation results of the volume intersection method are regarded as 3D point clouds of a target human.

Details of the volume intersection method in the developed component are as follows. First, a 3D point in world coordinate system is transformed into a pixel position on image coordinate system of each camera. It is judged whether the transformed position is inside a human region in the image of each camera.

This judgment is applied to all cameras in a sensor node. When the transformed positions exist in human regions of all cameras, the 3D point is recorded as one of the points representing a target human. At this time, color information corresponding to that point is also recorded. This calculation is applied to points in a given cuboid around human position. These processes are implemented using GPU. Then, data formats for 3D shape reconstruction application can be obtained in real-time with a sensor node based on several cameras. These data formats are same with a sensor node based on a Kinect sensor. That is, this component based on cameras is acceptable as a distributed sensor node in 3D shape reconstruction application.

3.3 Data integration through a network

Two sensor node components mentioned above output data of same formats. A 3D shape reconstruction system is achieved by integrating them.
However, processing performance of sensor nodes was decreased when integrating them through a network. For example, a sensor node component based on Kinect can generate 3D point clouds and color information at about 100 milliseconds cycle. This loop cycle is sufficient performance for real-time 3D shape reconstruction. On the other hand, when integrating this component to the other component through a network for data sharing, a processing loop cycle decreases to 600 milliseconds. This degradation makes sensor node performance worse in the viewpoints of real-time 3D shape reconstruction.
In order to stabilize tracking human moving at natural speed, it is important to keep short processing cycle. Therefore, the following components are added for integration of sensor nodes through a network. First, the data compression components and decoding components are prepared among sensor node components and 3D shape reconstruction application. These components are efficient for short loop cycle of each sensor node component. In this study, several compression methods are evaluated toward executing a loop at about 100 milliseconds cycle in sensor node components.
As a result of evaluation, image data including color information is compressed with JPEG of high compression ratio because color information does not greatly affect the accuracy of shape reconstruction. Since 3D point clouds affect to reconstruction accuracy of human, PNG was adopted for them. PNG is lossless compression, so 3D shape reconstruction application can obtain data calculated in each sensor node without any changes.

4 FLEXIBILITY OF THE SYSTEM

4.1 Application integrating distributed sensor nodes

Each distributed sensor node component outputs following data formats as mentioned above. The first is 3D point cloud data of a target human. The second is color information corresponding to 3D points. The last is a target position in the intelligent space. 3D shape reconstruction application receives these from sensor node components through a network. This application execute following processes. First, this application integrates the 3D points on the world coordinate sent from sensor nodes. And, color information is put to a corresponding 3D point. Then, integrated 3D shapes are displayed using OpenGL. The range of reconstruction does not need to be the whole intelligent space. It is enough for this application to reconstruct around a target human. This makes processing time of integrating data short. So, this is efficient for real-time shape reconstruction.

4.2 Real time reconstruction through a network

The information acquired by each distributed sensor is unifi ed through a network. When the data volume transmitted is excessive, it has bad influence on processing of each compo nent. As an example, if a component based on Kinect passe d a network, it was investigated what bad influence it would have to the processing. The result of the influence is shown in Table 1.Table 1 expresses comparison of the processing ti me at one cycle in a component based on Kinect. The top of a table is a case where it processes independently without unifying data. The middle of a table is a case where the data through a network is unified. The lower berth of a table is a case where data integration which uses the compression co mponent explained in Section 3.3 is performed. Data was c ompared in this table. Time of the case not compressed and time of the case compressed is compared. The case of data c ompressed is 14 times as quick as the case of data not comp ressed. This has influence good for the performance of 3D s hape reconstruction application.

Table 1. Processing time in a component based on Kinect

	Time of one cycle
Case of only component	20[ms]
Case of not compressed	831[ms]
Case of compressed	60[ms]

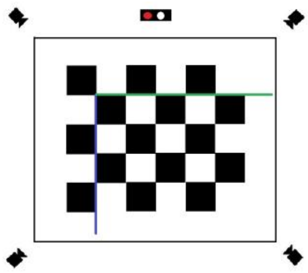


Fig 2. Sensors placements in an initial intelligent space

4.3 Flexible system and 3D shape reconstruction

In the proposed system, output of each distributed sensor node is the same format for the application.
This means that same frameworks to all sensor nodes are adopted in this system. The following advantages of the proposed system can be explained from this viewpoint. Even if a new sensor node is added to the system, the sensor node is easily integrated to the application. This makes the system more scalable. In order to show the advantages of the proposed system, addition of a new sensor node in this system is performed.
First, an initial intelligent space as shown in Fig. 2 was configured. In this space, four USB cameras and one Kinect sensor, which are geometrically calibrated in the common world coordinate, have been installed. Four USB cameras are integrated and it is regarded as one distributed sensor

node. One Kinect is also regarded as a distributed sensor node. Fig. 3 shows the experiment environment seen from the Kinect sensor installed in intelligent space. Each line on this figure expresses an axis of the world coordinate set in advance. The blue line is the X-axis on the world coordinate. The green line is the Y-axis. The red line is the Z-axis. Each voxel size in 3D shape reconstruction is set as 1cm x 1cm x 1cm.

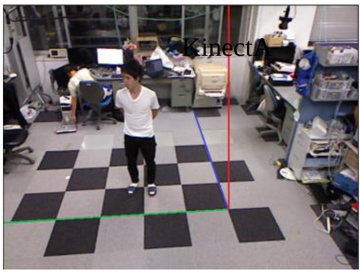


Fig 3. Intelligent space seen from KinectA

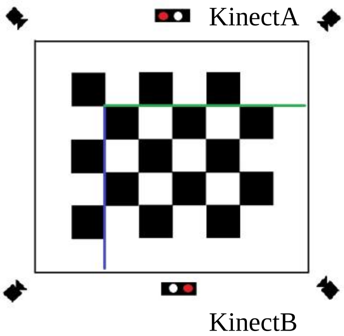
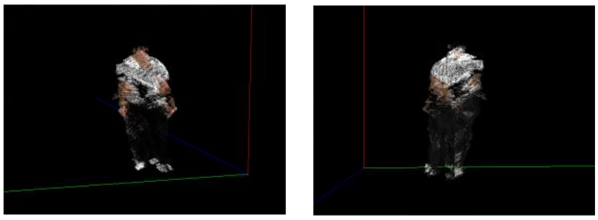


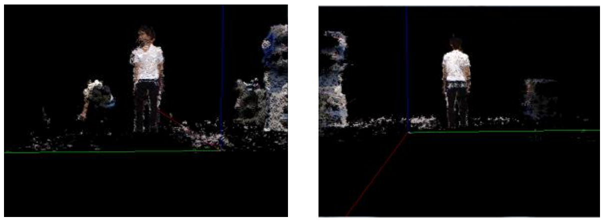
Fig 4. An intelligent space with an additional sensor

Next, the system as shown in Fig. 4 was configured. In this system, a sensor node based on one Kinect is added to the i nitial intelligent space as shown in Fig.2. Fig.5 shows 3D sh ape reconstruction results in the initial intelligent space of F ig. 2. Fig.5 (a) is the result of seeing a target human from a front viewpoint. Fig.5 (b) is the result of seeing a target hu man from a back viewpoint.Fig.6 shows 3D shape reconstru ction results in the improved intelligent space of Fig. 4.
The viewpoints of 3D shape reconstruction results are same with Fig. 5.These results show that shapes of Fig.6 are more accurate. Since the new sensor node was installed so that dead angles in intelligent space might be reduced, these results were previously expected. Generally, optimal placements of sensor nodes depend on scales or shapes of environments or purposes of the applications. When building such systems, developers usually configure the systems by trials-and-errors of sensor nodes placements. On the other hand, the proposed system configuration, which means integration of same format data from distributed sensor nodes, can contribute to flexibly adding new sensor nodes or new kinds of sensors to the systems. This data integration platform is required for intelligent spaces configured with various kinds of sensors. This also means that the intelligent spaces evolve according to addition of new sensors.



(a)Front viewpoint (b)Back viewpoint

Fig 5. 3D shape reconstruction at environment of Fig2



(a)Front viewpoint (b)Back viewpoint

Fig 6. 3D shape reconstruction at environment of Fig4

5 CONCLUSION

In this paper, a flexible system configuration for the 3D shape reconstruction application in intelligent space was proposed.
The proposed system was based on sensor nodes with same functions respectively. And each sensor node integrated several sensors in order to achieve the same function for the application.
Sensor nodes and the application were implemented as RT c omponents. Then, an experiment for adding a new sensor to the system was performed. Accuracy of 3D shapes of targe t humans depend on placements of sensors. The results sho w that a system for satisfying a purpose of the application is easily constructed with the proposed platform of sensor inte gration.

REFERENCES

[1] Joo-Ho Lee, Hideki Hashimoto, "Intelligent Space - concept and contents", Advanced robotics, Vol.16, No.3, pp.265-280, 2002
[2] Noriaki ANDO, Takashi SUEHIRO, Kosei KITAGAKI, Tetsuo KOTOKU, Woo-Keun Yoon, "RT-Middleware: Distributed Component Middleware for RT (Robot Technology)", 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2005), pp.3555-3560, 2005