Construction Progress Management and Interior Work Analysis Using Kinect 3D Image Sensors

Kosei Ishida[1]

[1] School of Architecture, Kogakuin University, Tokyo, Japan
E-mail : k_ishida@cc.kogakuin.ac.jp

Abstract –

Construction progress management and work analysis are often difficult in building interiors. In order to make progress management easier, the authors developed methods for recording the interior work environment. These methods record data using Microsoft Kinect sensors.

In this paper, the authors describe methods for recording work data utilizing several Kinect sensors. The authors recorded the following types of data:

1. **The work environment at the construction site**
2. **The shape of the building at the construction site**
3. **The work efficiency at the site derived from the motion capture data**

To record the three types of data, the authors developed a method for performing motion capture using several Kinects.

In addition, a shape recognition method for identifying building materials was developed. The shape recognition method utilizes point cloud data and camera image data. The authors also studied methods for analyzing the work efficiency data on the basis of skeletal tracking information.

Keywords –

Work efficiency; Point cloud; Kinect; Skeletal tracking; Construction management; Interior work

1 Introduction

Construction work is broadly divided into outdoor work such as building frame construction and indoor work such as interior finish work. The progress management and observation of outdoor work at a construction site are relatively easy to implement using methods such as employing the Global Positioning System (GPS) to identify worker positions and trajectories.

However, it is difficult to observe indoor work such as interior finish work because interior workers are located in buildings with room and floor divisions. It is generally not possible to use the GPS inside buildings to record worker positions and trajectories. Construction progress management and work analysis in indoor environments require a great deal of effort.

According to Japan's Ministry of Land, Infrastructure, Transport and Tourism, the cost of renovation and repair work in Japan was 11.02 trillion yen in 2013, and the cost is expected to increase in the future.

Moreover, if humans and robots work together on the construction site, we need the methods of collaborating safely with human workers in the same floor. For safety of construction workers, it is important to record pedestrian trajectories inside buildings.

In order to reduce construction costs with improved progress management, this paper focuses on methods for determining construction-worker positions and trajectories inside buildings. This requires shape recognition and motion capture devices to identify the 3D shape of the building and worker movements inside the building.

Several researchers have developed methods for recording pedestrian trajectories inside buildings. Some researchers have recently shown interest in 3D image sensors such as Kinect. A method for collecting trajectories with multiple Kinect sensors has been developed by Stefan Seera [1].

We also developed methods using Microsoft Kinect sensors [2]. The Kinect sensors collect camera data and calculate a 3D depth map. In addition, a Kinect sensor is able to track the skeletal images of one or two moving people.

In this paper, we describe methods for recording the work environment and worker trajectories using several Kinect sensors. The authors recorded the following types of data:

1. The work environment at the construction site
2. The shape of the building at the construction site
3. The work efficiency at the site derived from the motion capture data

2 Measurement Technique Using Multiple Kinect Sensors

Kinect sensors can maintain tracking through a range of a few meters; however, this is an insufficient range for measurements inside buildings. One Kinect sensor cannot measure all of the intricacies of building spaces, such as stairwells. In order to measure a wide area inside buildings, we have developed a measurement technique using multiple Kinect sensors.

2.1 Comparison of the Precision of 3D Point Clouds Acquired from a Kinect Sensor and 3D Laser Scanner

A Kinect collects camera data and calculates a 3D depth map. We developed a system to create 3D color point clouds from camera data and depth maps.

We then compared the precision of 3D point clouds acquired by a Kinect sensor and a Leica HDS3000 3D laser scanner. Figure 1 shows a Kinect sensor and 3D laser scanner setup in a measurement room with spherical targets. The distance between the Kinect sensor and the wall is 4 m. Figure 2 shows the 3D point clouds acquired by the Kinect (in red) and 3D laser scanner (in green and blue). As can be seen in Figure 2, the precisions of the 3D laser scanner and Kinect sensor are approximately ±3 and ±10 mm, respectively.

Figure 1. Kinect sensor and 3D laser scanner setup in a measurement room with spherical targets

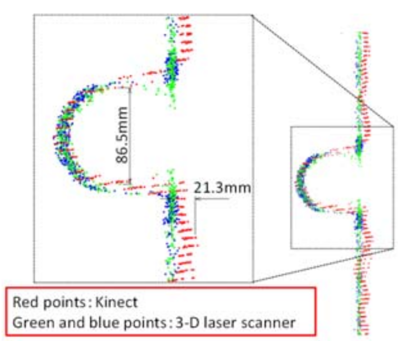

Figure 2. Comparison of the precision of 3D point clouds acquired from a Kinect sensor and 3D laser scanner

2.2 Shape Recognition of Circular and Spherical Targets

Two 3D depth maps with separate individual coordinate systems were used as inputs to perform the registration of point-cloud data based on multiple targets. We used spherical and circular targets to register multiple individual scans on a single coordinate system. Figure 3 shows the results for the shape recognition of circular targets based on 3D images and the measurement of the 3D coordinates of the circle center. Figure 4 shows the point clouds on spherical targets acquired by a Kinect sensor.

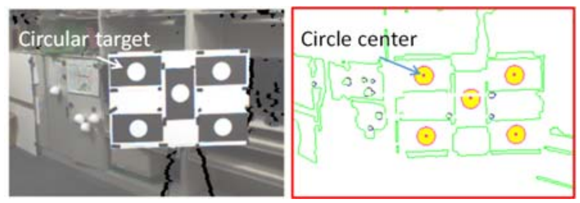

Figure 3. Shape recognition of circular targets based on 3D images and the measurement of the 3D coordinates of the circle center

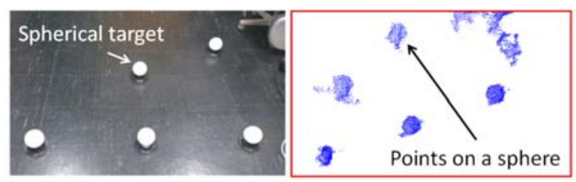

Figure 4. Point clouds of the spherical targets

2.3 Procedure for Registering Multiple Individual Point Clouds on a Single Coordinate System

In order to register multiple individual point clouds acquired by Kinect sensors on a single coordinate system, we developed a procedure for Kinect registration. This procedure consists of the following five steps:

1. Install more than four spherical targets for each Kinect.
2. Map the depth information from each Kinect.
3. Import point-cloud data and extract the points on the spherical targets.
4. Calculate the coordinates of the centers of the spherical targets using the least-squares method.
5. Register the target center coordinates into a CAD coordinate system.

Figure 5 shows spherical targets and the importation of point-cloud data from two Kinect sensors. Figure 6 shows multiple 3D depth images in a complete 3D model.

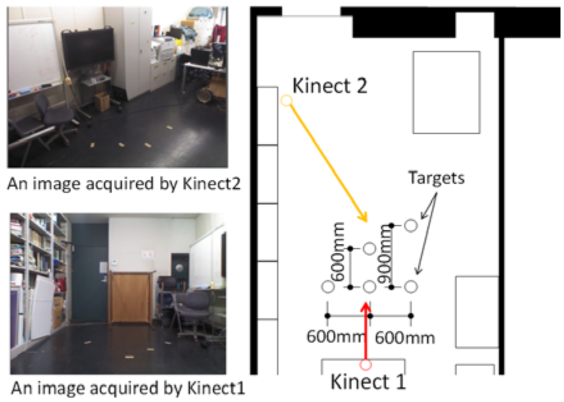

Figure 5. Measuring the position of the Kinect based on the spherical targets

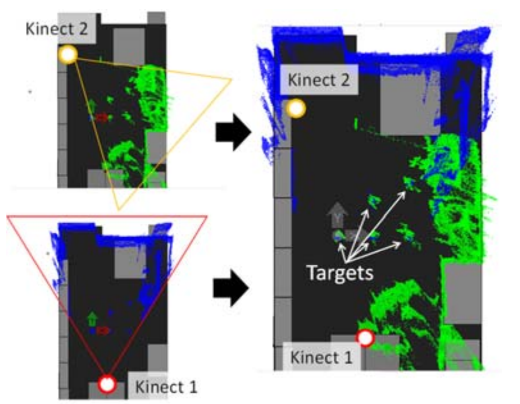

Figure 6. Multiple images in a complete 3D model

2.4 Evaluation of Pedestrian Trajectories Acquired by Kinect Sensors

We evaluate the pedestrian trajectories acquired by two Kinects. We collected skeletal tracking data in the room using two Kinect sensors.

Our evaluation data consist of two trajectory sets: the first data set was acquired by Kinect1 shown in Figures 6 and 8. The second data set was acquired by Kinect2. Two Kinect sensors measure the same pedestrian for 7668 frames ($\approx$ 259 s). The pedestrian walked around the perimeter of a triangle with a side length of 1300 mm. We analyzed the distance of the two trajectories.

Figure 7 shows the triangle on the floor and the positions of Kinect sensors. Figure 8 shows the point clouds and student trajectories acquired by the two Kinect sensors. Figure 9 shows the evaluation results of the pedestrian trajectories acquired by the Kinect sensors, and Table 1 summarizes the statistics of the measurement precision of the pedestrian trajectories.

Figure 9 shows the distribution of the Euclidean distance between the Kinect1 and Kinect2 trajectories acquired simultaneously. The mean and median distances are 82.9 and 70.5 mm, respectively. The deviation in the two trajectories is a few centimeters. The reason for the deviation is considered to be the deviation due to registering multiple individual point clouds on a single coordinate system. In addition, the reason for the deviation is due to the student's orientation in the camera image. Owing to the differences in the acquisition conditions such as the image in the pedestrian facing forward or backward, it is assumed that the two skeleton positions have different positioning of approximately 70 mm.

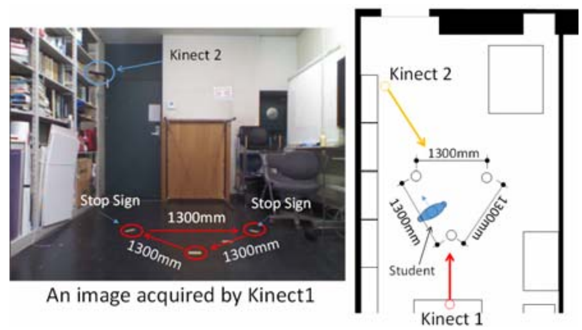

Figure 7. Collecting the skeletal tracking data

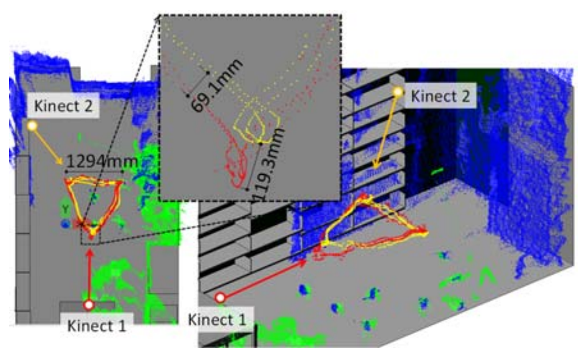

Figure 8. Point clouds acquired by Kinect1 (in blue) and Kinect2 (in green) and the student trajectories obtained by Kinect1 (in red) and Kinect2 (in yellow)

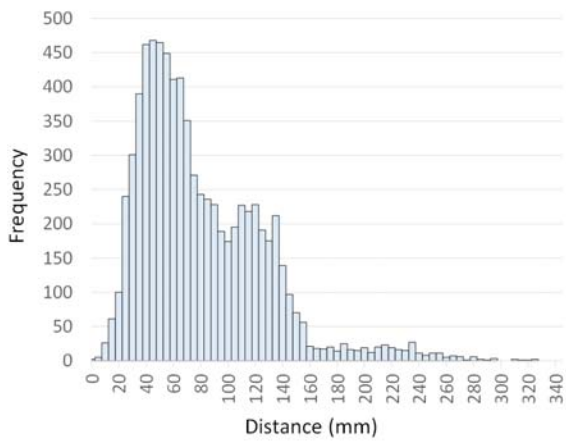

Figure 9. Evaluation results of the pedestrian trajectories acquired by the Kinect sensors.

Table 1 Statistics for the measurement precision of the pedestrian trajectories

Variable	Value
n	7668
Mean (mm)	82.9
SD (mm)	46.3
Maximum (mm)	325.7
Median (mm)	70.5

3 Method for Registering Skeletal Tracking Data from Multiple Kinect Sensors on a Single Coordinate System

The registration procedure discussed in the previous section is used with point-cloud data and 3D depth maps. The registration procedure for skeletal tracking data from multiple Kinect sensors is different because the skeletal tracking data include a time axis. We developed a method of registering skeletal tracking data from multiple Kinect sensors on a single coordinate system.

3.1 Procedure for Registering Skeletal Tracking Data from Multiple Kinect Sensors on a Single Coordinate System

In order to register skeletal tracking data from multiple Kinect sensors on a single coordinate system, we developed the following five-step procedure:

1. Register skeletal tracking data from each Kinect on individual coordinate systems (using the procedure discussed in the previous section).
2. Set up the measurement space for each Kinect.
3. Integrate the skeletal tracking data from all Kinect sensors into one dataset according to the data acquisition time.
4. Perform noise reduction on the skeletal tracking data.
5. Perform smoothing of the skeletal tracking data using a simple moving-average technique.

3.2 Measurement Space Setup for Kinect Sensors

Skeletal tracking data are generated from camera images and 3D depth maps obtained from the Kinect sensor. The shape of the skeleton is different in images obtained from the front and back camera positions. The data precision deteriorates at the point of camera switching. In order to prevent precision deterioration, the measurement space should be set up for each Kinect. Figure 10 shows the setup for four Kinect measurement spaces. The skeletal tracking data that are outside the measurement spaces need to be deleted.

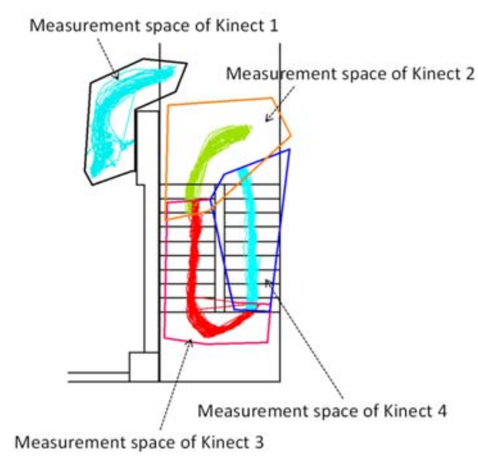

Figure 10. Setup of the Kinect measurement spaces

3.3 Integration of Skeletal Tracking Data into One Dataset According to the Acquisition Time

Skeletal tracking data have XYZ coordinates and Kinect acquisition times. It is necessary to integrate skeletal tracking data into one dataset according to the Kinect acquisition time. The method used to integrate the skeletal tracking data from multiple Kinect sensors into one dataset is shown in Figure 11.

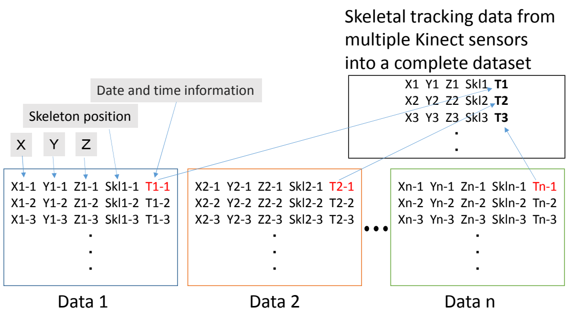

Figure 11. Method for integrating skeletal tracking data from multiple Kinect sensors into one dataset

3.4 Noise Reduction of Skeletal Tracking Data

The advantage of motion capture based on a variant of image-based 3D reconstruction is that it does not require a special marker for motion capture. Owing to the occasional erroneous recognition of skeletal positions, we have developed a noise reduction method for skeletal tracking data.

If the skeletal points exceed a speed of 6 m/s, noise is recognized, and points are deleted as errors. Figure 12 shows the noise reduction for skeletal tracking data.

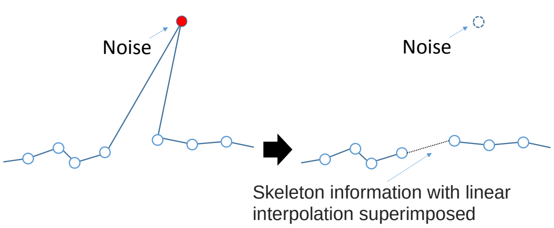

Figure 12. Noise reduction for skeletal tracking data

3.5 Data Smoothing of Skeletal Tracking Data Using a Simple Moving Average Technique

The skeletal tracking data obtained by a Kinect sensor are acquired at an approximate frame rate of 30 Hz. There is a large variability in the accuracy of the individual points. To reduce this variability, we have developed a method to smooth the skeletal tracking data using a simple moving average technique. Figure 13 shows the method for smoothing the skeletal tracking data. We normalize using two data points on both sides of a central point.

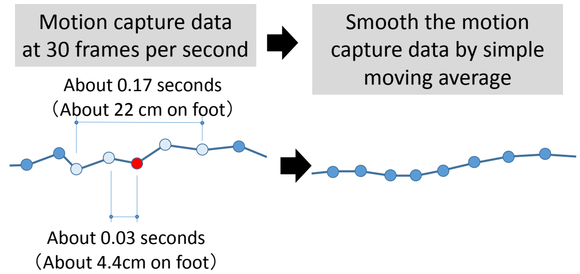

Figure 13. Method for smoothing the motion capture data using a simple moving average technique

4 Collection of Skeletal Tracking Data obtained by a Kinect Sensor in a Building Stairwell

We collected skeletal tracking data in a building stairwell using four Kinect sensors. In this experiment, a student (similar to a worker) hand-carried luggage on building stairs located between the eighth floor and the ninth floor. The student went up and down the stairs 20 times, acquiring data that was subsequently analyzed.

4.1 Kinect Setup in a Stairwell

To collect skeletal tracking data in a stairwell, we installed four Kinect sensors. Figure 14 shows the stairwell used in the experiment with the Kinect sensor set up on the wall. Using the spherical targets shown in Figure 14, we registered multiple Kinect sensors on a single coordinate system. The locations of the four Kinect sensors are shown in Figure 15.

Figure 14. Stairwell with Kinect sensors set up on the wall

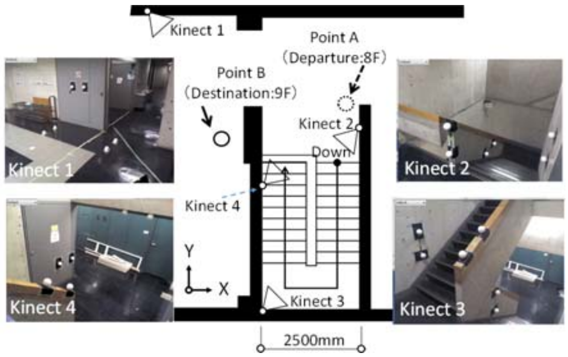

Figure 15. Layout of the stairs and the positions of the Kinect sensors

4.2 Collection of Point Cloud Data and Skeletal Tracking Data Obtained by Kinect Sensors in the Stairwell

We collected point cloud data and skeletal tracking data using Kinect sensors in a stairwell. Figures 16 and 17 show the results of this experiment. The raw 3D depth data (the point cloud data for the stairwell walls and stairs) are shown in gray, and the pedestrian trajectories of the student are shown in black. The pedestrian trajectories represent 20 round trips. Points A and B are on the eighth and ninth floors, respectively. The student traversed the stairs 20 times from Point A to Point B. We succeeded in measuring the shape of the stairs and collecting pedestrian trajectories.

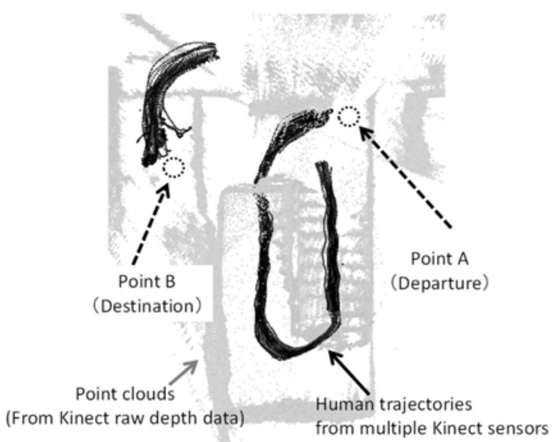

Figure 16. Point clouds on walls and stairs (in gray) and pedestrian trajectories (in black)

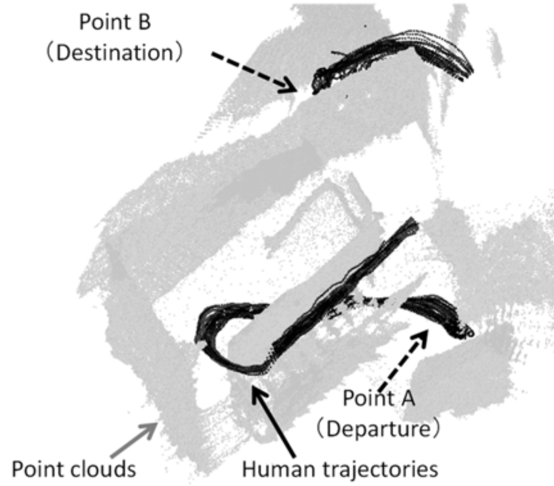

Figure 17. Isometric projection of the point clouds on the walls and stairs (in gray) and pedestrian trajectories (in black)

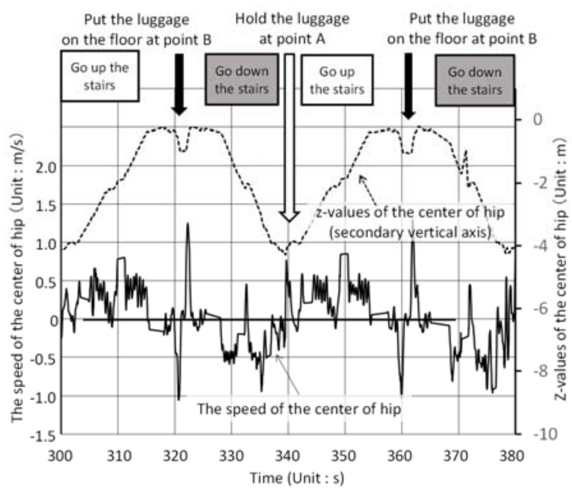

Figure 18. Results of the pedestrian position and velocity tracking

4.3 Analysis of the Skeletal Tracking Data in the Stairwell

The skeleton data consist of a set of 20 joints. We use the joint that is located at the center of the hip. The skeleton data encode a set of three-dimensional points $p_i = [x_i, y_i, z_i, t_i]$ including the acquisition time t_i. In this study, the Z-values of the center of the hip are z_i of the skeleton data.

The results of skeletal position and velocity tracking are shown in Figure 18. The horizontal axis represents the acquisition time of the skeletal tracking data. The left axis represents the speed (of the center) of the skeletal hip. The right axis of the graph represents the Z-values of the center of the hip.

In this experiment, the student pedestrian holds luggage at Point A and walks up the stairs. The pedestrian then places the luggage on the floor at Point B and walks down the stairs back to Point A.

From Point A to Point B, the Z-value of the center of the pedestrian's hip increased from −4.0 m in 300 s to −0.5 m in 315 s. From Point B, where the pedestrian places the luggage on the floor, back to point A, the Z-value of the hip center decreased from −0.5 m in 320 s to −1.0 m in 322 s. There is also a sudden decrease in the speed between 320 and 322 s.

4.4 Human Pose Recognition Using Skeletal Tracking Data

We developed a method for human pose recognition using skeletal tracking data. Human pose recognition is based on the pedestrian's speed and movement. This section describes the evaluation criteria used for human pose recognition.

The variables used in evaluation criteria for movement are the speed of the pedestrian in the X–Y plane, V_{xy} [m/s]; the total distance traveled after the state has changed, D_t [m]; and the duration of a human movement behavior, t [s].

The evaluation criteria for movement are defined in Table 2.

Table 2. Evaluation criteria for movement

State	Evaluation criteria
Stopped	$V_{xy} \leq 0.3$, $t \geq 1$, and $D_t \leq 0.4$
Walking slowly	$V_{xy} \leq 0.3$, $t \geq 1$, and $D_t \geq 0.4$
Walking	$V_{xy} \geq 0.5$ and $D_t \geq 0.4$
Transition state	None of the above conditions

The variables used in the evaluation criteria for standing and sitting are the Z component of the pedestrian's speed, S_z [m/s], and the height between the floor and the pedestrian's hip, H_{hip} [m].

The evaluation criteria used for standing and sitting are defined in Table 3. H_{hip} is calculated from the distance between the center of the pedestrian's hip and the floor, as shown in Figure 19.

Figure 20 shows the results of the height between the floor and the pedestrian's hip.

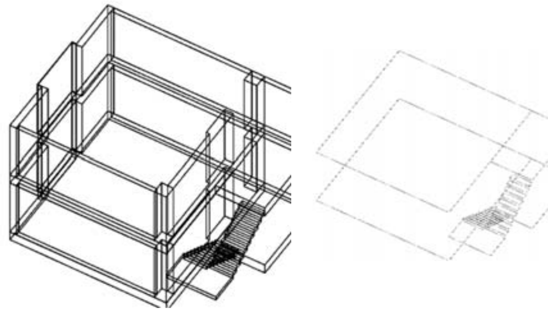

Figure 19. 3D CAD models of the stairwell and floor

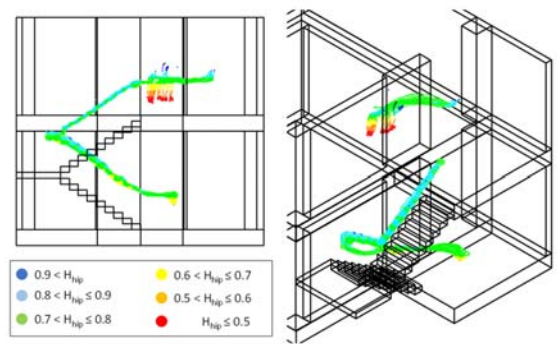

Figure 20. Height between the floor and the pedestrian's hip

Table 3. Evaluation criteria for standing and sitting

State	Evaluation criteria
Trying to stand up	$S_z \geq 0.8$
Trying to sit down	$S_z \leq -0.8$
Standing	$-0.8 < S_z < 0.8$ and $H_{hip} \geq 0.8$
Sitting	$-0.8 < S_z < 0.8$ and $H_{hip} \leq 0.6$
Transition state	None of the above conditions

Using the conditions in Table 2, we analyzed the skeletal tracking data with automatic human pose recognition. Figure 21 shows the results of the automatic evaluation of walking and stopping.

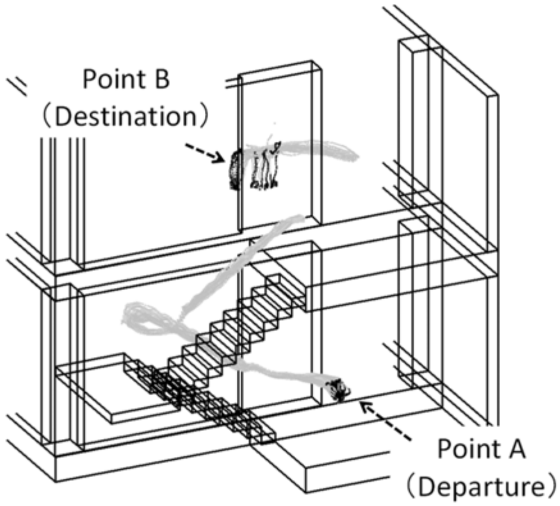

Figure 21. Results of the automatic evaluation of walking (in gray) or stopping (in black)

4.5 Automatic Acquisition of Repetitive Tasks

Construction work often includes repetitive tasks. In this experiment, a student pedestrian performed the repetitive task of climbing stairs 20 times from Point A to Point B. We developed a method for repetitive task recognition. This method is based on pedestrian positions and the evaluation criteria for walking and stopping. If the pedestrian worker has stopped, a task is likely finished.

Therefore, repetitive tasks are identified when a pedestrian has stopped by the time and location. Figure 22 shows the automatic acquisition of pedestrian trajectories for the first round-trip on the stairs.

Figure 23 shows the amount of time taken for each task repetition.

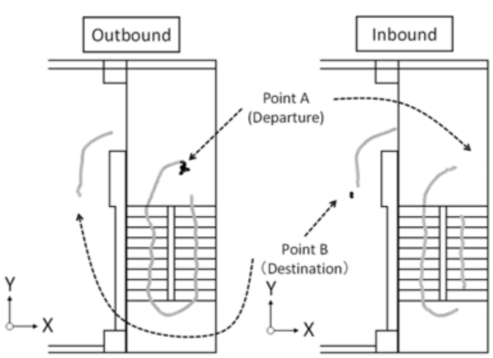

Figure 22. Automatic acquisition of pedestrian trajectories for the first round-trip

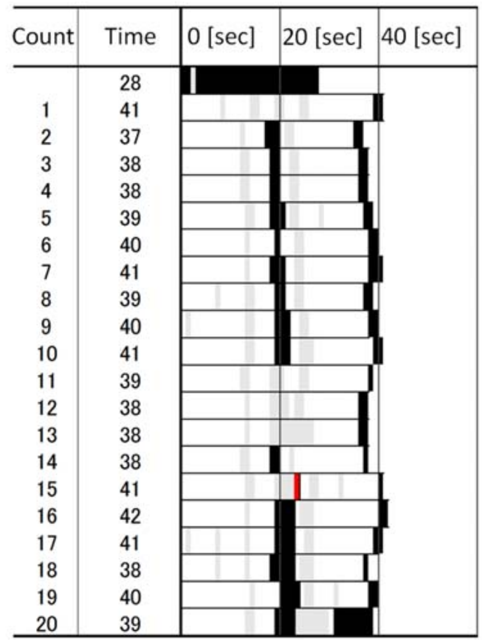

Figure 23. Time used for each task repetition

Black and white horizontal bars represent nonmoving and walking workers, respectively. Gray bars represent unmeasured time or transition states.

5 Conclusion

In this study, we have developed methods of registering skeletal tracking data from multiple Kinect sensors on a single coordinate system. We have analyzed the skeletal tracking data obtained in stairwells and have developed methods for human pose recognition using skeletal tracking data. We have also developed methods for the automatic acquisition of repetitive task data.

References

[1] Seera S., Brändlea N., and Rattib C. Kinects and human kinetics: A new approach for studying pedestrian behavior. *Transportation Research Part C: Emerging Technologies*, 48:212–228, 2014.

[2] Kosei I., and Yusuke O. Study on Analytical Method of the Workers Movements Monitor on Construction Site - Measurement Methods of Transport Route of the Worker -. *Proceedings of Thirty Symposium on Building Construction and Management of Project (Architectural Institute of Japan)*, 135-140, 2014.7.31