

Article

Implementation and Evaluation of a Wide-Range Human-Sensing System Based on Cooperating Multiple Range Image Sensors

Mikihiro Tokuoka ^{1,*}, Naoki Komiya ¹, Hiroshi Mizoguchi ¹, Ryohei Egusa ², Shigenori Inagaki ³ and Fusako Kusunoki ⁴

¹ Tokyo University of Science, 2641, Yamazaki, Noda, Chiba 278-0022, Japan; 7518523@ed.tus.ac.jp (N.K.); hm@rs.noda.tus.ac.jp (H.M.)

² Meiji Gakuin University, 1-2-37, Shirokanedai, Minato-ku, Tokyo 108-0071, Japan; egusa@psy.meijigakuin.ac.jp

³ Kobe University, 3-11, Tsurukabuto, Nada, Kobe, Hyogo 657-8501, Japan; inagakis@kobe-u.ac.jp

⁴ Tama Art University, 2-1723, Yarimizu, Hachioji, Tokyo 192-0375, Japan; kusunoki@tamabi.ac.jp

* Correspondence: tokmiki0708@gmail.com; Tel.: +81-80-8586-7197

Received: 30 December 2018; Accepted: 4 March 2019; Published: 7 March 2019



Abstract: A museum is an important place for science education for children. The learning method in the museum is reading exhibits and explanations. Museums are investing efforts to quantify interests using questionnaires and sensors to improve their exhibitions and explanations. Therefore, even in places where many people gather, such as in museums, it is necessary to quantify people's interest by sensing behavior of multiple people. However, this has not yet been realized. We aim to quantify the interest by sensing a wide range of human behavior for multiple people by coordinating multiple noncontact sensors. When coordinating multiple sensors, the coordinates and the time of each sensor differ. To solve these problems, coordinates were transformed using a simultaneous transformation matrix and time synchronization was performed using unified time. The effectiveness of this proposal was verified through experimental evaluation. Furthermore, we evaluated the actual museum content. In this paper, we describe the proposed method and the results of the evaluation experiment.

Keywords: three-dimensional range image sensor; time synchronization; coordinate transformation matrix; sensing interest; learning support system

1. Introduction

The museum is very important as a place for science education for children [1]. This is because the museum helps children gain knowledge by learning and through experience using content and learning materials [2]. The learning material and content in the museum are mainly about the exhibit and their explanations. Moreover, a panel and video tape recorder of explanation is used to complement the exhibits. In recent years, proposals have been drawn for further improving learning using these contents and learning materials. This proposal discriminates popular and unpopular exhibits using questionnaires and interviews, and changes the exhibits from time to time [3]. This is very important in further improving the efficiency of learning. In addition, research and development of museum learning support systems in science education is being researched and developed as a way to support children's learning [4–7]. As a method to evaluate these systems, subjective evaluations such as questionnaires and interviews, which are similar to evaluations of exhibits, are used frequently [4–9]. However, there are major problems in these evaluation methods; it is inefficient because it is time consuming. Furthermore, because it is a formal structured mechanism, it is difficult to obtain the natural opinion of the learner. Because we conduct interviews and questionnaires after the children

have experienced all content, we cannot quantitatively calculate the interest at a certain exhibit at a given time. Therefore, in essence, we cannot improve the content or learning materials. This is a problem that needs to be solved.

To solve these problems, techniques for sensing the interests of people have attracted attention and various research studies have been carried out. Some studies use a contact/noncontact sensor; studies that use contact sensors include measurements of physiological phenomena, electro dermal activity (EDA) [10] and electroencephalograms [11,12], estimation of a line of sight using eye gaze capturing devices, and blinking related to the most interests [13,14]. However, although these can be quantitatively measured, there is a problem that natural opinions cannot be obtained as the subjects may be stressed, and it takes time to wear these devices. Meanwhile, research using noncontact sensors includes measuring blinks using a web camera [15–20] and measuring the line-of-sight using an installation-type measuring instrument such as Tobii [21]. However, since these are noncontact methods, which allow measuring the natural body of the experimenter, there is a problem in that the measurement range is limited. For example, in a situation where there are many people like in a museum, there are cases where people cannot be tracked (Figure 1). Because people may overlap each other in the video, the sensor cannot recognize more than one person at a time. Therefore, it is necessary to measure these parameters within a range of 5–10 m. Therefore, there is a need for a system that measures quantitatively and in a noncontact manner the interest of “the place and the time” in a wide range. The learning effect in an actual museum has not been evaluated using our proposed technique.

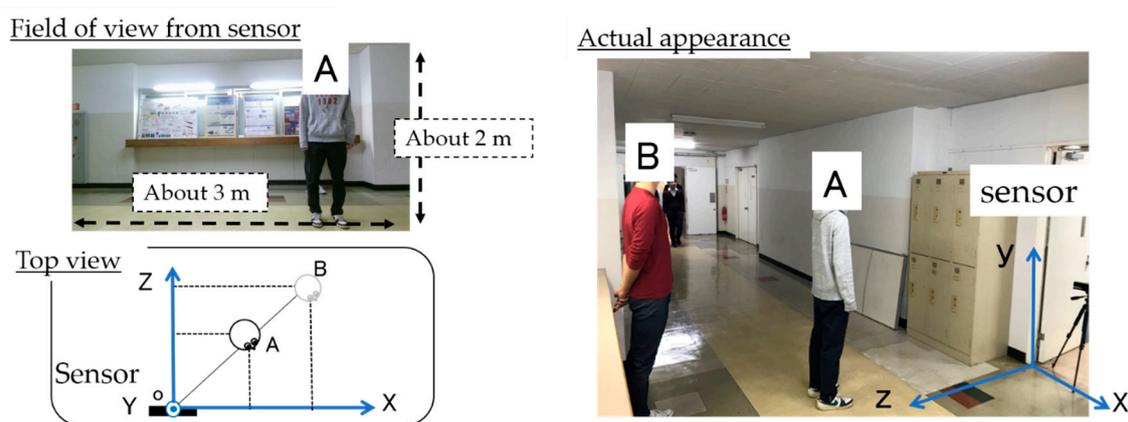


Figure 1. When people overlap each other in the sensor’s line of sight, the sensor cannot track these people. The sensor cannot recognize person B hidden behind person A.

We propose a system to solve these problems. By cooperating multiple noncontact sensors, quantitative interest in a wide range can be estimated in a noncontact manner. In particular, the proposed system observes behaviors such as a line-of-sight and eye blink for multiple learners at all times by using a large number of cooperating sensor groups arranged in the environment (for example, in the museum).

In this paper, we describe a method of coordinate transformation and time synchronization that can enable cooperation of multiple sensors and the results of the evaluation experiment. In addition, we describe the results of evaluating the contents and learning materials implemented in actual museums.

2. System

2.1. System Overview

We developed a system to estimate quantitative interests in a wide range by coordinating multiple noncontact sensors. The proposed system constantly observes the behavior of a learner by using a

large number of sensor groups arranged in a certain environment. Accordingly, we observed the interests of learners. Based on the obtained results, we realized the quantification of the interest “the place and the time” which could only be imagined fragmentarily through interviews and questionnaire-based surveys.

Figure 2 shows a model image of the system, and Figure 3 shows the system setup. The proposed system consists of a sensor group comprising multiple noncontact sensors and a data storage unit that accumulates all acquired data. The data storage unit includes elements, such as the direction of the learner’s face, detection of blinking, and gaze time in the gaze direction, related to the learner’s interest. In this paper, as the first step in realizing the system, we measure eye blinks, which are said to be most affected by human interest [22–24].

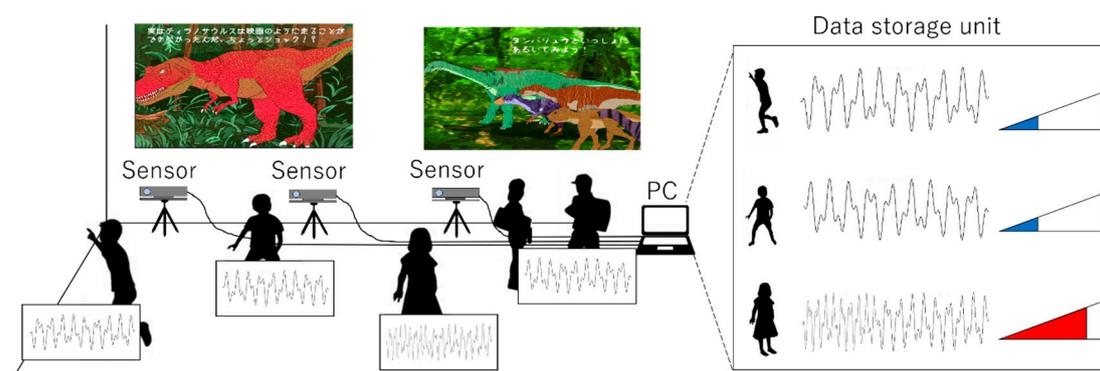


Figure 2. Model image of the proposed system. Simultaneously sensing the interests of multiple learners.

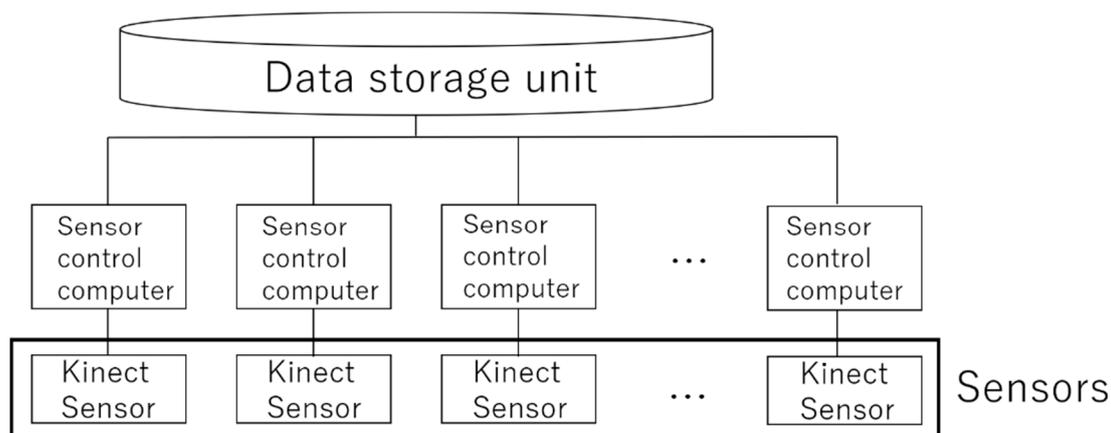


Figure 3. Setup of the proposed system.

To realize the system, we used a Kinect sensor [25] in this study. Microsoft’s Kinect sensor is a range-image sensor originally developed as part of an indoor video-gaming system. Although it is inexpensive, the sensor can obtain sophisticated measurements and adjudge the user’s location. In addition, this sensor can recognize humans and the human skeleton using the library in Kinect’s software development kit for Windows. The Kinect sensor can measure a three-dimensional skeletal location composed of 25 points on the human body, including the hands and the legs, and it can identify the user’s pose or status based on these functions. This skeletal information makes it possible to recognize various body movements.

Coordinate transformation and time synchronization for cooperation of multiple sensors are required for enlarging the measurement range by cooperating with multiple Kinect sensors. Coordinate transformation, time synchronization, and blink detection are all automatically initiated on our developed program. We describe our proposed methods below.

2.2. Coordinate Transformation Using Simultaneous Transformation Matrix

2.2.1. Coordinate Transformation

When the area requiring measurement is wide or there is a considerable amount of information to be measured, it is necessary to expand the measurement range by cooperating multiple sensors. However, when trying to cooperate multiple position measuring sensors, the coordinate system of each sensor and the measured value should be independent. As shown in Figure 4, the coordinate system of sensors 1, 2, and 3, and the measured value are independent. To measure the same object, it is necessary to unify it to one coordinate system, as shown in Figure 5.

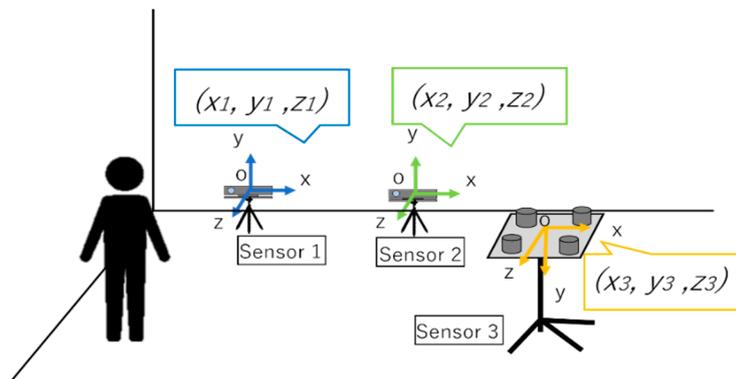


Figure 4. Issues in using multiple sensors. Even if the same target position is measured, the output values of the sensors do not match.

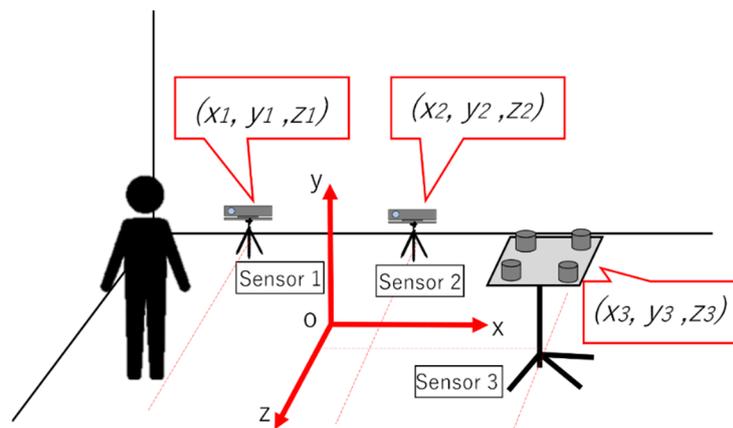


Figure 5. Unification of coordinate system.

In general, in many position measurement sensors such as the Kinect sensor, the measurement results in the coordinate system remain unique to each sensor. Therefore, simply by performing arbitrary multiple sensor placement and measurement, even if the same target position is measured, the output values of the sensors do not match and it is very difficult to coordinate the measured values. Various studies have been conducted to solve this problem. Unfortunately, conventional research also failed to accurately realize coordinate transformation through the cooperation of multiple sensors [26]. As the main method, those studies used a checker board for calibration [27]. However, although this method can unify the exact coordinate system, it takes a considerable amount of time to unify the coordinate system.

It takes a few hours for the checkerboard to move several centimeters at a time [27]. It is difficult to implement this in an environment such as a museum, where this has to be realized ad hoc.

We therefore propose a method of using the simultaneous coordinate transformation matrix to unify all coordinate systems in an arbitrary coordinate system. This is a method that can perform coordinate transformation in a short time and coordinate unification.

2.2.2. Method of Coordinate Transformation Using Simultaneous Transformation Matrix

We describe coordinate transformation using a simultaneous transformation matrix. We define $P(x, y, z)$ as the coordinates as seen from the coordinate system unique to Kinect sensor at the point P in space. Furthermore, we define $P'(x', y', z')$ as the coordinates in the unified coordinate system at the point P in the space. P is the coordinate system of Figure 4, P' is the coordinate system of Figure 5 after unification. At this time, by using the coordinate transformation matrix T , it is possible to convert to P' from P , using Equations (1) and (2).

$$\begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & q_x \\ r_{21} & r_{22} & r_{23} & q_y \\ r_{31} & r_{32} & r_{33} & q_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}, \quad (1)$$

$$P' = TP, \quad (2)$$

Coordinate representation such as that in Equation (2) is called a simultaneous coordinate representation. This can express the movement of the coordinate system by the expression of multiplication of one coordinate transformation matrix. We describe each component of the coordinate transformation matrix T as follows. Figure 6 shows the component of T .

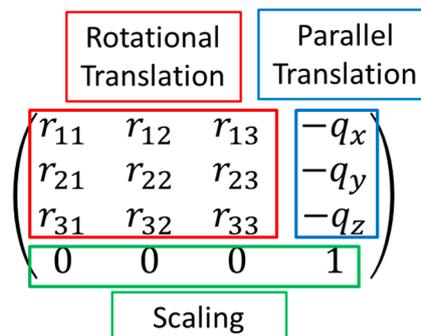


Figure 6. Translation matrix.

In Figure 6, the components from r_{11} to r_{33} represent the rotational movement of the coordinate system, and they are expressed by the multiplication of the equation of the rotational movement around each axis.

$$\begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} = R_x R_y R_z, \quad (3)$$

We describe the rotational movement R_x in the coordinate system around the X axis as an example of the rotational movement around the axis. As shown in Figure 7, when the coordinate system is rotated by θ_x about the X axis, the coordinates $P'(x', y', z')$ of the point in the coordinate system after movement are given by Equation (4). Equation (4) is expressed using coordinates $P(x, y, z)$ and θ_x .

$$\begin{aligned} x' &= x, \\ y' &= y \cos \theta_x + z \sin \theta_x, \\ z' &= -y \sin \theta_x + z \cos \theta_x, \end{aligned} \quad (4)$$

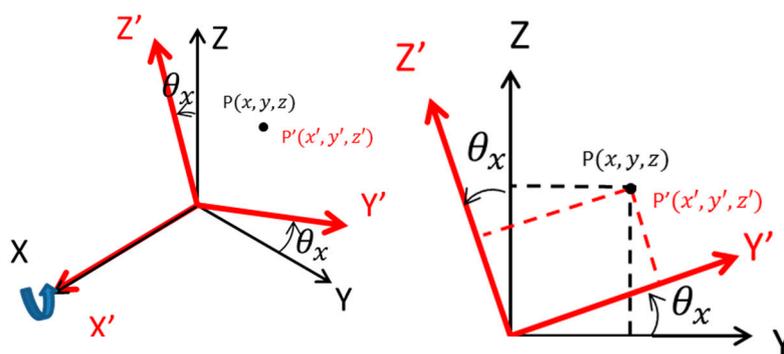


Figure 7. Rotational translation.

When Equation (4) is expressed as a matrix, Equation (5) is expressed as

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & \sin\theta_x \\ 0 & -\sin\theta_x & \cos\theta_x \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \tag{5}$$

From Equation (5), the rotational movement around the X axis is represented by a matrix of Equation (6).

$$R_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & \sin\theta_x \\ 0 & -\sin\theta_x & \cos\theta_x \end{pmatrix}, \tag{6}$$

Similarly, in the case of rotation about the Y and Z axes, the expressions for converting the coordinates on the original coordinate system to the coordinates in the coordinate system after rotation are expressed by Equations (7) and (8).

$$R_y = \begin{pmatrix} \cos\theta_y & 0 & -\sin\theta_y \\ 0 & 1 & 0 \\ \sin\theta_y & 0 & \cos\theta_y \end{pmatrix}, \tag{7}$$

$$R_z = \begin{pmatrix} \cos\theta_z & \sin\theta_z & 0 \\ -\sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{8}$$

Substituting Equations (7) and (8) into Equation (3) gets

$$\begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} = R_x R_y R_z = \begin{pmatrix} \cos\theta_y \cos\theta_z & \cos\theta_y \sin\theta_z & -\sin\theta_y \\ \sin\theta_x \sin\theta_y \cos\theta_z - \cos\theta_x \sin\theta_z & \sin\theta_x \sin\theta_y \sin\theta_z + \cos\theta_x \cos\theta_z & \sin\theta_x \cos\theta_y \\ \cos\theta_x \sin\theta_y \cos\theta_z + \sin\theta_x \sin\theta_z & \cos\theta_x \sin\theta_y \sin\theta_z - \sin\theta_x \cos\theta_z & \cos\theta_x \cos\theta_y \end{pmatrix}, \tag{9}$$

Next, q_x, q_y, q_z are components representing parallel movement in each axis direction. Combining the rotational movement and the parallel movement described above, the movement of the coordinates is expressed by

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} q_x \\ q_y \\ q_z \end{pmatrix}, \tag{10}$$

Equation (10) can thus be expressed as

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} q_x \\ q_y \\ q_z \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & q_x \\ r_{21} & r_{22} & r_{23} & q_y \\ r_{31} & r_{32} & r_{33} & q_z \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}, \quad (11)$$

Finally, the fourth line in Figure 6 represents the scaling of the coordinate system. In the case of the coordinate transformation used in this research, it is no need to enlarge/reduce; therefore, it is set to 1 (equal magnification), as shown in Equation (12).

$$1 = (0 \ 0 \ 0 \ 1) \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}, \quad (12)$$

Next, we describe how to calculate such a coordinate transformation matrix T. T can be calculated from the correspondence relationship between both coordinate systems if there is a point where the coordinates seen from the coordinate system of both the Kinect's coordinate system and the unified coordinate system are known.

When there are n points in space, as shown in Figure 8, the coordinates in the Kinect coordinate system of the n th point are defined as ${}^S P_n ({}^S x_n, {}^S y_n, {}^S z_n)$ and those in the unified coordinate system as ${}^F P_n ({}^F x_n, {}^F y_n, {}^F z_n)$. When the coordinates in the homogeneous coordinate system are represented by a matrix, the coordinates of the 1st to n th points as viewed from each coordinate system are expressed by

$${}^S P = \begin{pmatrix} {}^S x_1 & {}^S x_2 & \dots & {}^S x_n \\ {}^S y_1 & {}^S y_2 & \dots & {}^S y_n \\ {}^S z_1 & {}^S z_2 & \dots & {}^S z_n \\ 1 & 1 & \dots & 1 \end{pmatrix} {}^F P = \begin{pmatrix} {}^F x_1 & {}^F x_2 & \dots & {}^F x_n \\ {}^F y_1 & {}^F y_2 & \dots & {}^F y_n \\ {}^F z_1 & {}^F z_2 & \dots & {}^F z_n \\ 1 & 1 & \dots & 1 \end{pmatrix}, \quad (13)$$

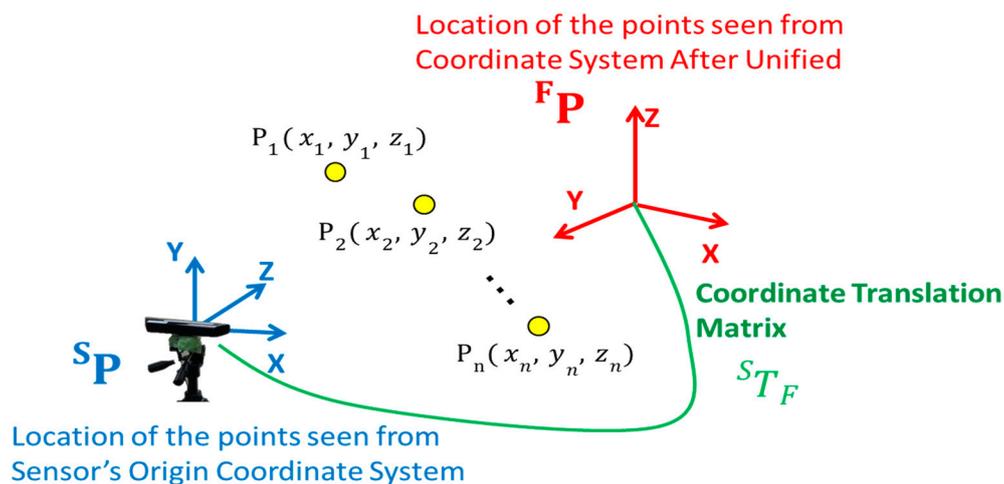


Figure 8. Coordinate system translation matrix.

Therefore, letting ${}^S T_F$ be the coordinate transformation matrix for transforming the coordinates of the points in the Kinect coordinate system into the coordinates in the unified coordinate system, this coordinate transformation can be expressed as

$$\begin{pmatrix} Fx_1 & Fx_2 & \cdots & Fx_n \\ Fy_1 & Fy_2 & \cdots & Fy_n \\ Fz_1 & Fz_2 & \cdots & Fz_n \\ 1 & 1 & \cdots & 1 \end{pmatrix} = {}^S T_F \begin{pmatrix} Sx_1 & Sx_2 & \cdots & Sx_n \\ Sy_1 & Sy_2 & \cdots & Sy_n \\ Sz_1 & Sz_2 & \cdots & Sz_n \\ 1 & 1 & \cdots & 1 \end{pmatrix} \quad (14)$$

$${}^F P = {}^S T_F {}^S P, \quad (15)$$

From Equation (15), a coordinate transformation matrix ${}^S T_F$ is calculated. When Equation (14) is $n = 4$ and these points are not on the same plane, the coordinate transformation matrix ${}^S T_F$ can be obtained as follows, using the inverse matrix of ${}^S P$ [28].

$${}^S T_F = {}^F P {}^S P^{-1}, \quad (16)$$

We calculate the coordinate transformation matrix by measuring with $n = 4$. To unify Kinect's coordinate system by this method, the point where unified coordinates are known on the real space and can be measured by Kinect (hereinafter, such a point is called a sample point) should be set. Therefore, in this research, the measurement result of the sample point measured by one Kinect (hereinafter referred to as the origin Kinect) is treated as a true value. Then, by substituting the true value and the results measured by other Kinect into Equations (14)–(16), the coordinate system of each sensor can be unified to the coordinates of the origin Kinect. After the calculation of the transformation matrix, each sensor is multiplied by a transformation matrix as shown below for the three-dimensional coordinate measurement result measured by other Kinect Sensor.

$$\begin{pmatrix} Fx \\ Fy \\ Fz \\ 1 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & -q_x \\ r_{21} & r_{22} & r_{23} & -q_y \\ r_{31} & r_{32} & r_{33} & -q_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} Sx \\ Sy \\ Sz \\ 1 \end{pmatrix}, \quad (17)$$

$$\begin{aligned} Fx &= r_{11} Sx + r_{12} Sy + r_{13} Sz, \\ Fy &= r_{21} Sx + r_{22} Sy + r_{23} Sz, \\ Fz &= r_{31} Sx + r_{32} Sy + r_{33} Sz, \end{aligned} \quad (18)$$

Through these calculations, it becomes possible to treat the three-dimensional coordinate measurement result measured by all Kinects as a unified value in the coordinate system of the origin Kinect. All these matrix calculations are performed on a program.

2.3. Time Synchronization

2.3.1. Summary

In the measurement by multiple sensors, the measurement start timing and the end timing are different because the time of each sensor is different. As shown in Figure 9, even if the same operation is recognized, a time lag occurs, and an accurate analysis cannot be performed; therefore, time synchronization of multiple sensors is indispensable. By performing time synchronization as shown in Figure 10, it is possible to perform an accurate analysis using multiple sensors. Nevertheless, with the conventional synchronization of multiple Kinect sensors, time synchronization has not been successful [29,30]. Conventional time-synchronization research is based on the establishment of a unified time and server. Previous studies that used a unified time did not state about the accuracy of time synchronization, assuming that the cycle of unified time does not fluctuate. However, there is

a possibility that the unified time deviates. In addition, any research that establishes a server and synchronizes time is accurate. However, to process and synchronize information with high capacity and geometric qualities like the Kinect V2 sensor with color images and depth information, it is necessary to establish an expensive server with a fast processing time. The establishment of an expensive server has no versatility if we consider the realization at our museum; this is the original purpose of our study. Therefore, we did not set up a server, and referenced the existing server, conducted time synchronization by considering the processing time of each personal computer, and evaluated the accuracy.

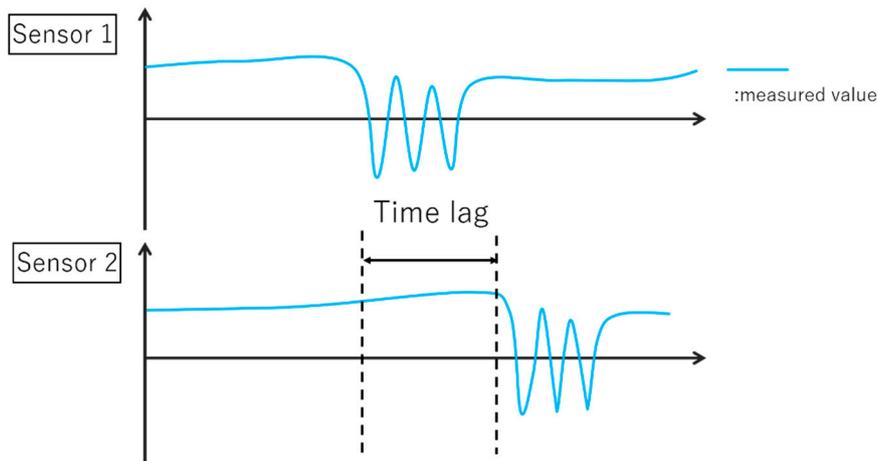


Figure 9. Time lag according to the difference of the time point of each sensor.

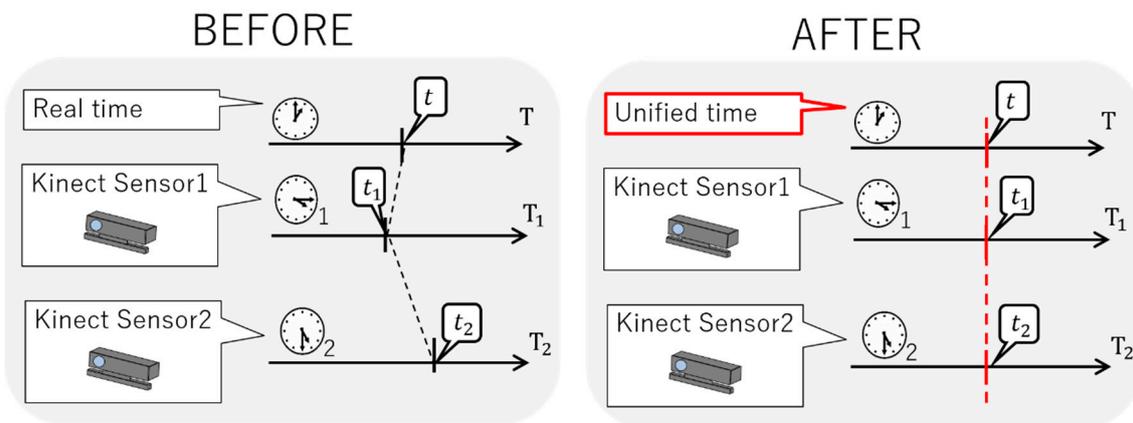


Figure 10. Time synchronization using unified time.

2.3.2. Time Synchronization Using Unified Clock

First, we created unified time and recorded the unified time on multiple sensors; then, we unified the time of multiple sensors to this unified time and realized time synchronization. Figure 11 shows the occurrence of this time synchronization. After the measurement, we unified the recorded time of all the sensors to the unified time. The frequency of the Kinect was 30 Hz, and the clock for the unified time was 100 Hz. By using this method, we ensured accurate time synchronization.

Next, we explain the algorithm of time synchronization. A personal computer was used to connect Kinect Sensors 1 and 2 and record the unified time on the Internet by using the DataTime class in the program. Time on the Internet refers directly to the time of the existing NTP server, and in this research, we referred to the NTP server of Tokyo Science University (nodarntp.rs.noda.tus.ac.jp); it can be used within our university. The time of the Internet was recorded at 30 fps, which is the sampling rate of a Kinect Sensor. However, depending on the performance of the personal computer, the recorded time

differs from the unified time by approximately 1 ms. Therefore, the error between the times of the NTP server and recording was constantly calculated in each personal computer, and a correction was made. As a result, time synchronization was performed.

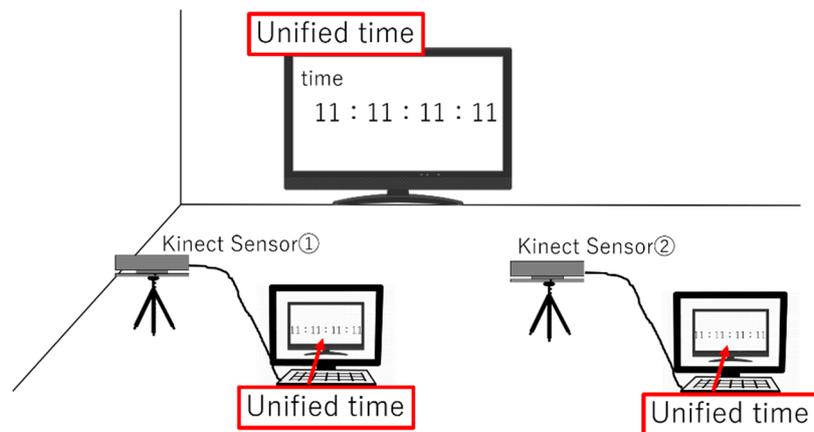


Figure 11. Time synchronization method.

2.4. Eye Blink Detection

It is said that people's interest can be gauged by the blinking of their eyes. We focused on the eye blink to quantify the interest of multiple people at the museum. Clearly, eye blinking is suppressed when an entity has caught people's interests. In other words, when people are more interested, the number of their eye blinks decreases. If we measure eye blink at all times, we quantitatively estimate the interest, engagement, and excitement "on the spot, at the time". However, in the conventional eye-blink measurement method, the measurement range is narrow because when learners look away, part of the eye-blink data is missed. The realization of the proposed wide-range eye-blink-measurement method could capture the eye-blink even when learners look away because it measures eye blinks using a coordinated sensor. In other words, the sensor always measures and records eye blinks. By measuring interest quantitatively, we clarified the learners' interest in the contents of the museums.

Eye-blink detection follows the flow of human detection, skeleton information detection, face recognition, and finally pupil detection, as detailed in the following text. First, the Kinect sensor performs human detection by using a database based on the learning data of skeletal information of a large number of people. Next, the sensor identifies a person based on the person's coordinate information. Accordingly, the same person was tracked in a multiple people environment. Then, the sensor recognizes a human face and extracts the 3D coordinates of 1347 points on the face as feature points [31]. Based on these information, the eye position, and then the eye blinking were determined. The two states of the pupil (open/closed eyes) were determined based on the ratio of the iris width to the maximum iris width obtained by counting the number of black pixels. The state was recognized as "OPEN" when opening eyes and "CLOSE" when closing eyes based on the set threshold value. It is said that people's interest can be gauged using their eye blink. Figure 12 shows the appearance of OPEN and CLOSE eyes. As shown in Figure 12, when the state changes from OPEN to CLOSE, and to OPEN again, we count it as one eye blink; the measurement cycle was 30 Hz.

In this way, the number of blinks is automatically measured.

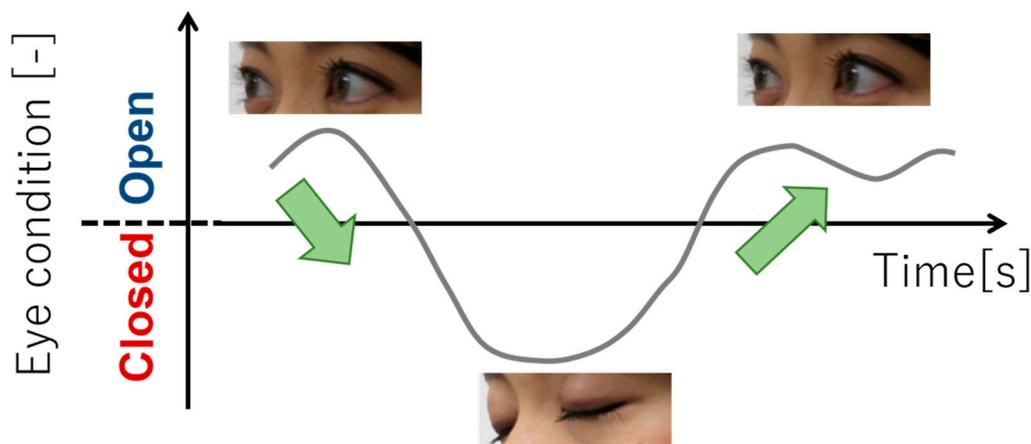


Figure 12. Recognition of eye condition.

3. Experiments

We describe the evaluation experiments and the results to evaluate the effectiveness of the coordinate transformation and time synchronization proposal method for multiple sensor cooperation, which is necessary for expanding the measurement range. In addition, we describe the results of evaluating the content of the museum using the proposed system.

3.1. Coordinate Transformation

We conducted an evaluation experiment on whether the coordinate system can be unified by coordinate transformation using simultaneous transformation matrix. As the first step in realizing the system, coordinate conversion was carried out using two Kinect Sensors to evaluate the coordinate system unification.

3.1.1. Evaluation Experiment of Coordinate Transformation

In this experiment, a jig with four sampling points that are not on the same plane for the unification of two Kinect sensors and coordinate system was used [28]. Figures 13 and 14 show the experiment setup. We unify the coordinate system of Kinect Sensor 1 and Kinect Sensor 2, as shown in Figure 13. First, as shown in Figure 13, a jig with sampling points is placed in front of two Kinect sensors, which measure the four-point coordinates of the jig. The simultaneous transformation matrix is calculated based on the coordinates of the four points measured by each Kinetic sensor. Next, one subject stands at a fixed measurement point, and each Kinect sensors measures the subject's coordinates simultaneously. The measurement result is coordinate-transformed by the calculated coordinate transformation matrix to unify the coordinate system. The total number of measurement points is 66 points. Figure 15 shows the measurement points. Next, we describe the method of processing persons and images programmed. First, each Kinect sensor recognizes the coordinates of sampling points for calculating the coordinate transformation matrix. Infrared rays were then irradiated into the field-of-view of the sensor and the position coordinates of the specified sampling point were measured according to the reflection time, depth information, and color-image information. Based on the coordinates measured using Kinect sensors 1 and 2, the matrix conversion from matrix to expression was performed on all the PCs connected to the two Kinect sensors on the program to calculate the coordinate transformation matrix. Based on the calculated coordinate transformation matrix, the coordinate system of Kinect sensor 2 was unified to the coordinate system of Kinect sensor 1. Therefore, it is possible to link all the information on the coordinate axis of Kinect sensor 1.

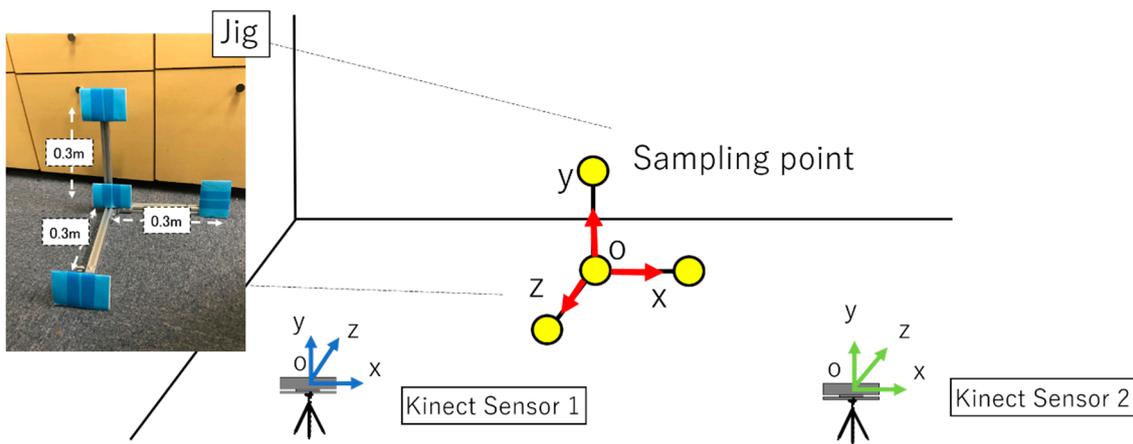


Figure 13. Experimental situation. The jig with four sampling points which are not on the same plane.

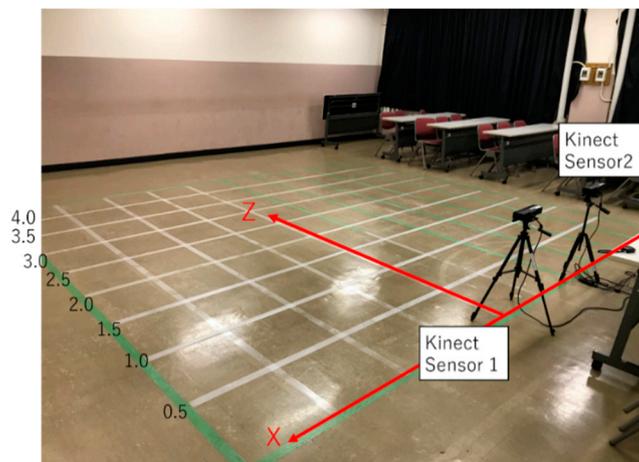


Figure 14. Experimental situation.

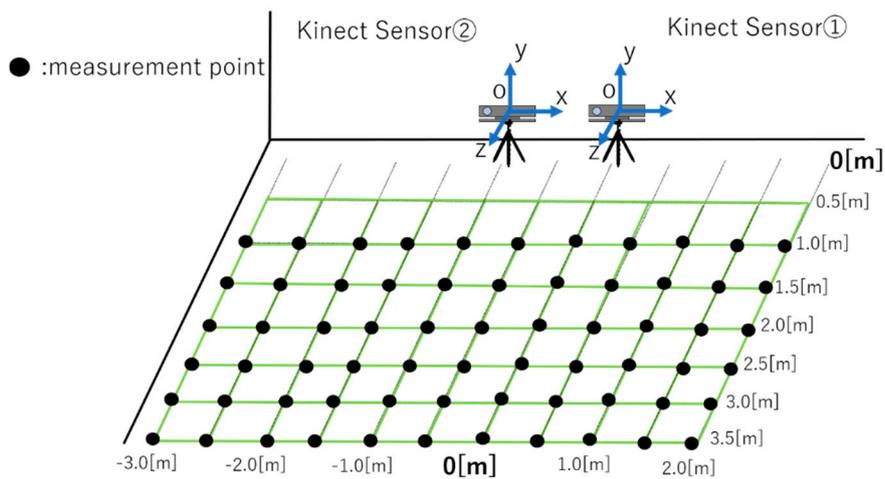


Figure 15. The total number of measurement points is 66 points.

3.1.2. Evaluation Experiment Result of Coordinate Transformation

We evaluate the error of the coordinate measurement result of Kinect Sensor 1 and the coordinate result obtained by coordinate conversion of the measurement value of Kinect Sensor 2 using the coordinate transformation matrix.

Figure 16 shows the coordinates of each measurement point after coordinate transformation; this is the measurement result of Kinect Sensor 1 and the measurement result of Kinect Sensor 2 that unified the coordinate system using coordinate transformation. In Figure 15, points that could not be measured outside the recognition range of the Kinect sensor are not plotted. From this result, it can be seen that the coordinate transformation is performed by the simultaneous transformation matrix. Next, we evaluate the error of coordinate transformation. The error is the difference in distance between the known coordinates and the distance difference between the known coordinates and the coordinates of person coordinates measured by Kinect Sensor 2; this is the distance between the known coordinates of the X-Z plane and the coordinates after coordinate system unification.

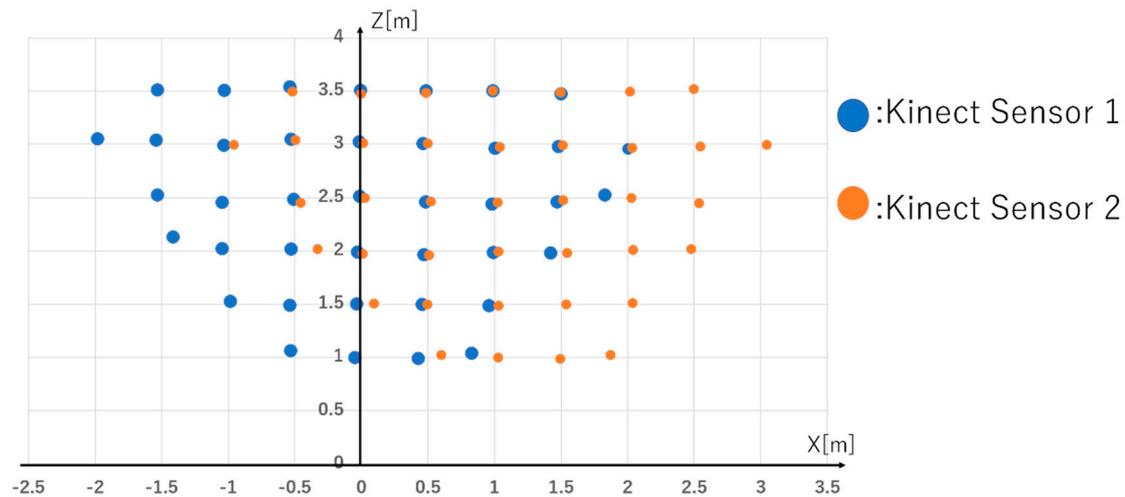


Figure 16. The total number of measurement points is 66 points.

The museum content we evaluate is for children. It is necessary to distinguish between two or more children, and therefore, we set the allowable error as half of the shoulder width of the child. Because the average shoulder width of the child is 33.86 cm, the allowable error is 16.93 cm [32]. Table 1 summarizes the errors of the coordinates obtained by coordinate conversion of the coordinates measured by Kinect sensor 2 in this experiment. All experimental results are within tolerance. The average error value of the coordinate transformation was 4.18 [cm], and the standard deviation was ± 2.98 [cm]. Thus, coordinate transformation by using the Kinect sensor can unify the coordinate system with an accuracy of 4.18 ± 2.98 [cm]. Therefore, the usefulness of the proposed method is proved.

Table 1. Error of the coordinates obtained by coordinate conversion of the coordinates measured by Kinect Sensor 2.

Z [cm]	X [cm]									
	-2.0	-1.5	-1.0	-0.5	0	0.5	1	1.5	2.0	
0	-	-	-	-	-	-	-	-	-	-
0.5	-	-	-	-	-	-	-	-	-	-
1	-	-	-	10.35	2.47	1.31	1.31	-	-	-
1.5	-	-	9.35	0.489	2.98	3.78	3.67	-	-	-
2.0	-	16.92	2.74	3.87	2.98	4.52	4.02	8.09	-	-
2.5	-	6.25	2.64	4.07	4.83	2.84	2.48	5.20	-	-
3.0	4.10	4.03	1.78	1.00	4.55	1.21	4.57	6.19	4.36	-
3.5	-	2.03	2.38	2.22	1.52	1.28	1.56	2.09	-	-

3.2. Evaluation Experiment on Time Synchronization

We conducted an evaluation experiment on whether the time synchronization of two Kinect sensors can be achieved using time synchronization with a clock for unified time.

3.2.1. Evaluation Experiment of Range of Measurement

In this experiment, we used a clock to unify time using two Kinect sensors. Figure 17 shows the experiment environment. We performed time synchronization between Kinect sensors 1 and 2, which are shown in Figure 17. We explain the experimental procedure below. First, we start measuring using Kinect sensor 1. Second, we start measuring Kinect sensor 2. Therefore, both Kinect sensors measure at and record for different arbitrary times. The measuring time is 53 s for each Kinect. After the end of the measurement, we standardized the Kinect sensors 1 and 2 with a clock for unified time. Through these experiments, we evaluate whether it is possible to synchronize time between Kinect sensors 1 and 2.

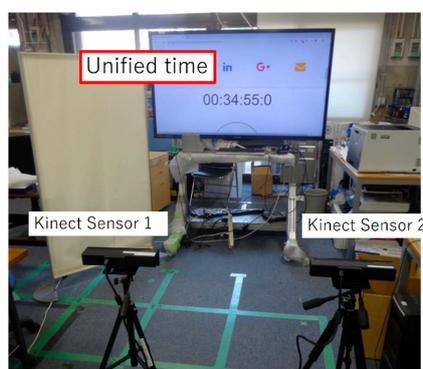


Figure 17. Experiment environment.

3.2.2. Experimental Result

The results of the verification experiment are shown below. Table 2 shows the experiment start time, experiment end time, and elapsed time for each of the unified time. Unified time started at 0 s and ended at 53 s. We evaluate the elapsed time of the unified time and Kinect sensors 1 and 2. 53 s have elapsed in unified time, 52.954 s have elapsed in Kinect 1, and 53.033 s have elapsed in Kinect 2. Therefore, the difference between the elapsed times of Kinect sensors 1 and 2 was 0.079 s. The average value of time-synchronization error was 0.0069 [s], and the standard deviation was ± 0.0024 [s]. As a result, time synchronization using the NTP server can be synchronized with an accuracy of 0.0069 ± 0.0024 [s]. Furthermore, we must accurately detect the eye blink, which occurs once, averaging 0.2 s [33]; therefore, we set 0.2 s as the allowable error. Since the difference in elapsed time is within the allowable error, the effectiveness of the proposed method for time synchronization was suggested.

Table 2. Experimental result of time synchronization.

	Start of Experiment		End of Experiment		Result	
	Unified Time [s]	Kinect Time [s]	Unified Time [s]	Kinect Time [s]	Elapsed Time (Unified Time) [s]	Elapsed Time (Kinect Time) [s]
Kinect Sensor 1	0	2.091	53.000	55.045	53.000	52.954
Kinect Sensor 2	0	3.979	53.000	57.012	53.000	53.033

3.3. Range of Measurement

We conduct an evaluation experiment on measure the range expansion enabled by linking two Kinect Sensors.

3.3.1. Evaluation Experiment of Range of Measurement

In this experiment, we used two Kinect Sensors and one subject. Figure 18 shows the state of the experiment. In the experiment, the subject walked in front of the two Kinect Sensors. The walking route is as shown in Figure 18. At that time, a comparison is made between the area of one Kinect Sensor person within the discovery range and the area of two Kinect Sensors person within the discovery ranges.

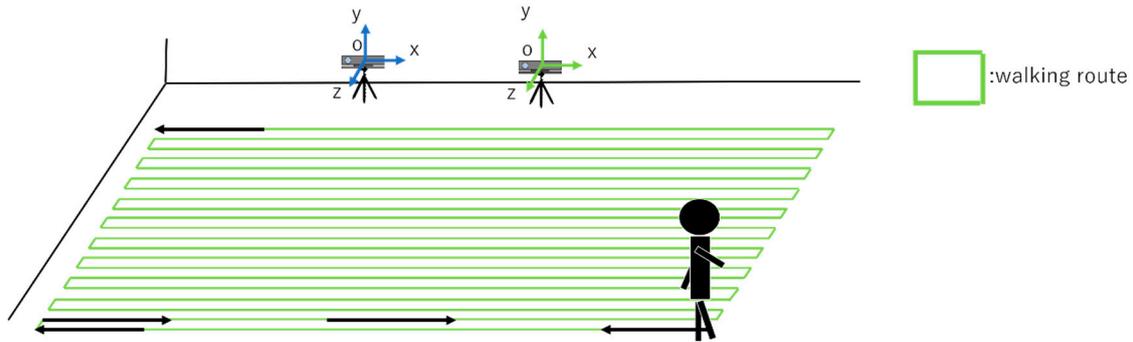


Figure 18. When multiple people overlap and the sensor cannot track multiple people.

3.3.2. Evaluation Experiment Result of Range of Measurement

We describe the experimental results. Figure 19a shows a person discovery range with one Kinect sensor. Figure 19b shows a person discovery range with two Kinect Sensors. The person tracking area using one Kinect Sensor is 7.028 m², and that using two Kinect Sensor is 11.23 m². The area was calculated using an approximation curve. From this result, it was found that the measurement area can be expanded by cooperating multiple Kinect sensors. By cooperating two Kinect sensors, it is possible to trace a person in the range of 5 m × 4 m. It was suggested that it is possible to further extend the measurement range by increasing the number of Kinect sensors.

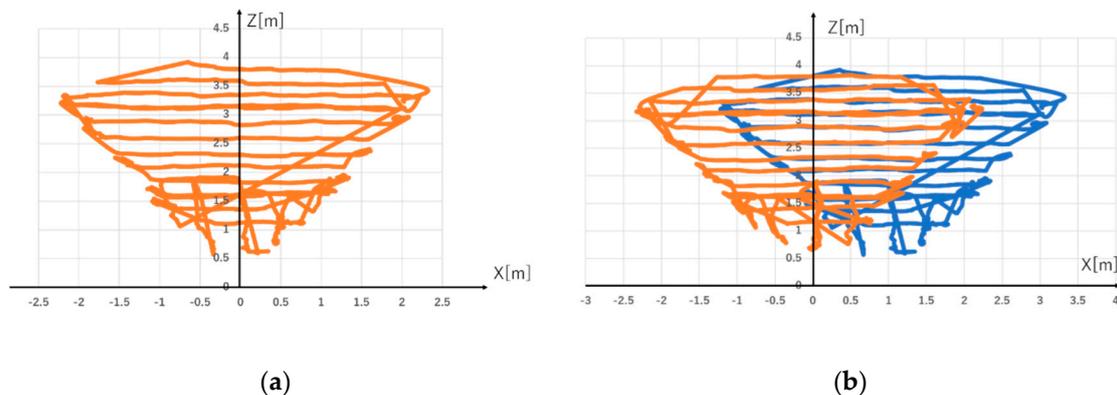


Figure 19. (a) shows that person tracking area of one Kinect Sensor. (b) shows that person tracking area of two Kinect Sensors.

3.4. Evaluation Experiment of Contents

From Sections 3.1–3.3, we conducted evaluation experiment on the elemental technologies to realize the proposed system. We evaluate contents (multiple movies) implemented in an actual museum by the proposed system using these element technologies. This museum learning support system includes what I develop [34].

3.4.1. Experimental Method

In this experiment, we evaluate contents and learning materials implemented in actual museums. We sensed four learners watching actual videos flowing in the museum using two connected Kinect sensors. During the experiment, we evaluate whether contents could be evaluated by an eye blink. The video is composed of different contents of four sections.

Figure 20 shows the experimental environment. By cooperating with two Kinect sensors that became clear in our experiment, it is possible for a person to track a range of 5 m × 4 m. Therefore, the museum contents to be evaluated in this experiment is done within 5 m × 4 m. We sense the interests of the four people who experience the content within this range.

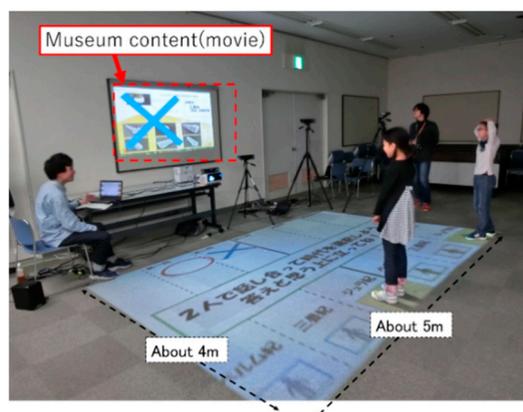


Figure 20. Evaluation experiment of museum contents [32]. We always sense four people within the measurement range.

3.4.2. Experimental Result

We describe the experimental results of one subject as an example. One eye blink is converted on the graph as shown in Figure 21. Conversion result are shown in Figure 21. In sections 1, 3, 5, and 7, subjects are taking a break. In sections 2, 4, 6, and 8, subjects are viewing content videos of museums. First, during the experiment, it is understood that the data is always acquired without data loss. As shown in Figure 22, there are many eye blinks during a break. We evaluate based on this result. First, the eye blink rate is calculated as

$$(\text{Eye blink rate}) [\text{times}/\text{min}] = \frac{(\text{Number of Eye Blink}) [\text{times}]}{(\text{Elapsed Time}) [\text{s}]} \times 60, \quad (19)$$

The results of calculating the blink rate are shown in Table 3. It is said that eye blink is suppressed when people has interest or attention [35]. In other words, the more subject have interest, the smaller subject's eye blink rate is. As shown in Table 3, eye blink rate increases during a break. In addition, when subjects watched content video, eye blink rate decreases and eye blink is suppressed. From this, it is confirmed that the subject is interested when subject is watching the content video. Next, the degree of interest is described. Eye blink rate is high in the order of 4, 8, 2, 6 from Table 3. From this, it turns out that the interest is high in the order of 4, 8, 2, 6. This result is supported by a questionnaire to the subjects. From these results, the effectiveness of the proposed method was suggested. It was possible to evaluate the content in museum by quantifying the degree of interest of the learner by the proposed method.

Table 3. Eye blink rate in each section.

	1	2	3	4	5	6	7	8
Eye blink rate	19.92	5.37	23.48	3.75	19.48	11.80	54.40	3.69

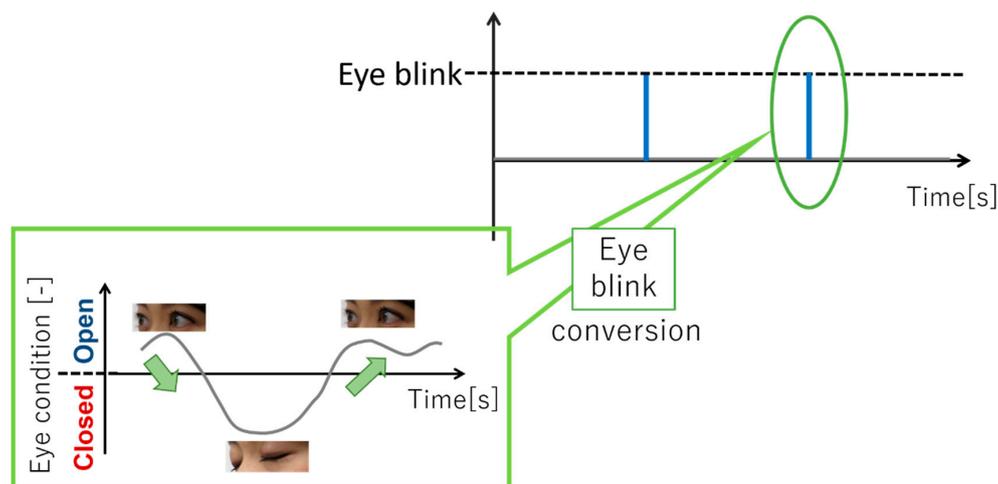


Figure 21. Eye blink conversion.

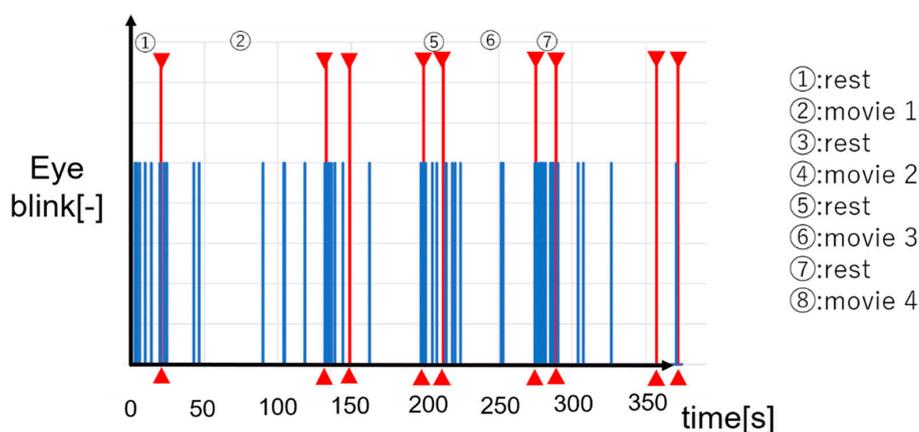


Figure 22. Experimental result after eye blink conversion.

4. Conclusions

In this paper, we described a method of coordinate transformation and time synchronization for the coordination of multiple sensors required for enlarging the measurement range, and discussed the evaluation results. To expand the measurement range by using multiple Kinect sensors, we proposed coordinate transformation using a simultaneous transformation matrix, ad hoc coordinate transformation, and time synchronization using unified time. In the experiment, we conducted experiments to evaluate the coordinate transformation using simultaneous transformation matrices and the usefulness of time synchronization using unified time. As a result, coordinate transformation was realized with accuracy, and it can identify multiple children. Furthermore, time synchronization could be realized with accuracy of detecting eye blinks. These results showed the effectiveness of the proposed method.

In the future work, we aim to add attention time or action in face orientation as an element to sense learners. By doing this, we aim to further quantify learner's interests.

Author Contributions: Conceptualization, M.T., H.M., R.E. and S.I.; methodology, M.T. and H.M.; software, M.T. and N.K.; data Curation, M.T.; writing—original draft preparation, M.T., N.K.; writing—review and editing Preparation, H.M., R.E., S.I. and F.K.; funding acquisition, H.M., S.I. and F.K.

Funding: This work was supported in part by Grants-in-Aid for Scientific Research (A), Grant Number JP16H01814, JP18H03660.

Acknowledgments: This work was supported in part by Grants-in-Aid for Scientific Research (A), Grant Number JP16H01814, JP18H03660. The evaluation was supported by the Museum of Nature and Human Activities, Hyogo, Japan.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Falk, J.H.; Dierking, L.D. *Museum Experience Revisited*, 2nd ed.; Left Coast Press: Walnut Creek, CA, USA, 2012.
2. Hein, G.E. *Learning in the Museum*; Routledge: Abingdon, UK, 2002.
3. Sheng, C.-W.; Chen, M.-C. A study of experience expectations of museum visitors. *Tour. Manag.* **2012**, *33*, 53–60. [[CrossRef](#)]
4. Nakayama, T.; Yoshida, R.; Nakadai, T.; Ogitsu, T.; Mizoguchi, H.; Izuishi, K.; Kusunoki, F.; Muratsu, K.; Inagaki, S. Novel Application of Kinect Sensor for Children to Learn Paleontological Environment -Learning Support System *based on* Body Experience and Sense of Immersion. In Proceedings of the 8th International Conference on Sensing Technology (ICST 2014), Liverpool, UK, 2–4 September 2014; p. #S7-8 (1)-(4), (USB Memory).
5. Yoshida, R.; Egusa, R.; Saito, M.; Namatame, M.; Sugimoto, M.; Kusunoki, F.; Yamaguchi, E.; Inagaki, S.; Takeda, Y.; Mizoguchi, H. BESIDE: Body Experience and Sense of Immersion in Digital paleontological Environment. In Proceedings of the International Conference on Human-Computer Interaction (CHI2015) Extended Abstracts, Seoul, Korea, 18–23 April 2015; pp. 1283–1288.
6. Tamaki, H.; Sakai, T.; Yoshida, R.; Egusa, R.; Inagaki, S.; Yamaguchi, E.; Kusunoki, F.; Namatame, M.; Sugimoto, M.; Mizoguchi, H. Science education enhancement within a museum using computer-human interaction technology. In Proceedings of the 8th International Conference on Computer Supported Education (CSEDU2016), Rome, Italy, 21–23 April 2016; Volume 2, pp. 181–185.
7. Tokuoka, M.; Tamaki, H.; Sakai, T.; Mizoguchi, H.; Egusa, R.; Inagaki, S.; Kawabata, M.; Kusunoki, F.; Sugimoto, M. BELONG: Body Experienced Learning Support System *based on* Gesture—Enhancing the Sense of Immersion in a Dinosaurian Environment. In Proceedings of the 9th International Conference on Computer Supported Education (CSEDU2017), Porto, Portugal, 21–23 April 2017; Volume 1, pp. 487–492.
8. Tan, T.-H.; Liu, T. The mobile-based interactive learning environment (MOBILE) and a case study for assisting elementary school English learning. In Proceedings of the IEEE International Conference on Advanced Learning Technologies, Joensuu, Finland, 30 August–1 September 2004; pp. 530–534.
9. Fusako, K.; Sugimoto, M.; Hashizume, H. Toward an interactive museum guide system with sensing and wireless network technologies. In Proceedings of the IEEE International Workshop on Wireless and Mobile Technologies in Education, Växjö, Sweden, 29–30 August 2002; pp. 99–102.
10. Yoshida, R.; Nakayama, T.; Ogitsu, T.; Takemura, H.; Mizoguchi, H.; Yamaguchi, E.; Inagaki, S.; Takeda, Y.; Namatame, M.; Sugimoto, M.; et al. Feasibility Study on Estimating Visual Attention using Electrodermal Activity. In Proceedings of the 8th International Conference on Sensing Technology (ICST 2014), Liverpool, UK, 2–4 September 2014; pp. 589–592.
11. Lewis, G.W.; David, L.R.-J. Evaluation of a Subject's Interest in Education, Training and other Materials Using Brain Activity Patterns. U.S. Patent No. 5,762,611, 9 June 1998.
12. Bang, J.W.; Lee, E.C.; Park, K.R. New computer interface combining gaze tracking and brainwave measurements. *IEEE Trans. Consum. Electron.* **2011**, *57*, 1646–1651. [[CrossRef](#)]
13. Zhan, Z.; Zhang, L.; Mei, H.; Fong, P.S. Online learners' reading ability detection *based on* eye-tracking sensors. *Sensors* **2016**, *16*, 1457. [[CrossRef](#)] [[PubMed](#)]
14. Lee, H.C.; Luong, D.T.; Cho, C.W.; Lee, E.C.; Park, K.R. Gaze tracking system at a distance for controlling IPTV. *IEEE Trans. Consum. Electron.* **2010**, *56*, 2577–2583. [[CrossRef](#)]
15. Wang, F.; Zhou, M.; Zhu, B. A Novel Feature Based Rapid Eye State Detection Method. In Proceedings of the 2009 IEEE International Conference on Robotics and Biomimetics, Guilin, China, 19–23 December 2009; pp. 1236–1240.
16. Harini, V.; Papanikolopoulos Nikolaos, P. *Detecting Driver Fatigue through the Use of Advanced Face Monitoring Techniques*; University of Minnesota-Twin Cities: Minneapolis, MN, USA, 2001.
17. Ryu, J.B.; Hyun, A.Y.; Seo, Y. Real Time Eye Blinking Detection Using Local Ternary Pattern and SVM. In Proceedings of the International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA), Compiegne, France, 28–30 October 2013; pp. 598–601.

18. Le, H.; Dang, T.; Liu, F. Eye Blink Detection for Smart Glasses. In Proceedings of the 2013 IEEE International Symposium on Multimedia (ISM'13), Anaheim, CA, USA, 9–11 December 2013; pp. 305–308.
19. Alioua, N.; Amine, A.; Rziza, M.; Aboutajdine, D. Eye state analysis using iris detection *based on* Circular Hough Transform. In Proceedings of the 2011 International Conference on Multimedia Computing and Systems (ICMCS), Ouarzazate, Morocco, 7–9 April 2011; pp. 1–5.
20. Lu, Y.; Li, C. Recognition of driver eyes' states *based on* variance projections function. In Proceedings of the IEEE Conference of 3rd International Congress on Image and, Signal Processing, Yantai, China, 16–18 October 2010; pp. 1919–1922.
21. Frutos-Pascual, M.; Begonya, G. Assessing visual attention using eye tracking sensors in intelligent cognitive therapies *based on* serious games. *Sensors* **2015**, *15*, 11092–11117. [[CrossRef](#)] [[PubMed](#)]
22. Holland, M.K.; Tarlow, G. Blinking and thinking. *Percept. Mot. Skills* **1975**, *41*, 403–406. [[CrossRef](#)]
23. Wood, J.; Hassett, J. Eyeblinking during problem solving: The effect of problem difficulty and internally vs externally directed attention. *Psycho-Physiology* **1983**, *21*, 18–20. [[CrossRef](#)]
24. Kim, D.; Choi, S.; Park, S.; Sohn, K. Stereoscopic Visual Fatigue Measurement Based on Fusional Response curve and Eye-blinks. In Proceedings of the 17th International Conference on Digital Signal Processing, Corfu, Greece, 6–8 July 2011; pp. 1–6.
25. Shotton, J.; Sharp, T.; Kipman, A.; Fitzgibbon, A.; Finocchio, M.; Blake, A.; Moore, R. Real-time human pose recognition in parts from single depth images. *Commun. ACM* **2013**, *56*, 116–124. [[CrossRef](#)]
26. Kaenchan, S.; Mongkolnam, P.; Watanapa, B.; Sathienpong, S. Automatic multiple kinect cameras setting for simple walking posture analysis. In Proceedings of the 2013 International Computer Science and Engineering Conference (ICSEC), Bangkok, Thailand, 4–6 September 2013; pp. 245–249.
27. Yang, R.S.; Chan, Y.H.; Gong, R.; Nguyen, M.; Strozzi, A.G.; Delmas, P.; Ababou, R. Multi-Kinect scene reconstruction: Calibration and depth inconsistencies. In Proceedings of the 2013 28th International Conference on Image and Vision Computing New Zealand (IVCNZ), Wellington, New Zealand, 27–29 November 2013; pp. 47–52.
28. Ganapathy, S. Decomposition of transformation matrices for robot vision. *Pattern Recognit. Lett.* **1984**, *2*, 401–412. [[CrossRef](#)]
29. Huang, Z.; Nagata, A.; Kanai-Pak, M.; Maeda, J.; Kitajima, Y.; Nakamura, M.; Ota, J. Automatic evaluation of trainee nurses' patient transfer skills using multiple kinect sensors. *IEICE Trans. Inf. Syst.* **2014**, *97*, 107–118. [[CrossRef](#)]
30. Saputra, M.R.; Guntur, D.P.; Paulus, I.S. Indoor human tracking application using multiple depth-cameras. In Proceedings of the 2012 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, Indonesia, 1–2 December 2012.
31. Cafaro, A.; Wagner, J.; Baur, T.; Dermouche, S.; Torres Torres, M.; Pelachaud, C.; Valstar, M. The NoXi database: Multimodal recordings of mediated novice-expert interactions. In Proceedings of the 19th ACM International Conference on Multimodal Interaction, Glasgow, Scotland, 13–17 November 2017; pp. 350–359.
32. Ishii, M.; Isogai, F.; Iizuka, S.; Amano, S. Fundamental Studies on Planning Clothes for Infants (Part 2). *J. Home Econ. Jpn.* **1975**, *26*, 402–407.
33. Królak, A.; Paweł, S. Eye-blink detection system for human–computer interaction. *Univers. Access Inf. Soc.* **2012**, *11*, 409–419. [[CrossRef](#)]
34. Tokuoka, M.; Komiya, N.; Mizoguchi, M.; Egusa, R.; Inagaki, S.; Kusunoki, F. Application of 3D Range Image Sensor to Body Movement Detection—Supporting Children's Collaborative Learning in Museums. In Proceedings of the 2018 Twelfth International Conference on Sensing Technology (ICST2018), Limerick, Ireland, 4–6 December 2018; pp. 394–398.
35. Tsugunosuke, S.; Ryuichi, Y.; Haruya, T.; Takeki, O.; Hiroshi, T.; Hiroshi, M.; Etsuji, Y.; Shigenori, I.; Yoshiaki, T.; Miki, N.; et al. Electrodermal Activity Based Study on the Relationship between Visual Attention and Eye Blink. In Proceedings of the 2015 Ninth International Conference on Sensing Technology (ICST2015), Auckland, New Zealand, 8–10 December 2015; pp. 639–642.

