

Simultaneous Children Recognition and Tracking for Childcare Assisting System by Using Kinect Sensors

Bin Zhang¹, Tomoaki Nakamura¹, Rena Ushioji², Takayuki Nagai¹, Kasumi Abe¹, Takashi Omori³, Natsuki Oka⁴, Masahide Kaneko¹

¹Department of Mechanical Engineering and Intelligent Systems, The University of Electro-Communications, Tokyo, Japan

²Faculty of Home Economics, Otsuma Women's University, Tokyo, Japan

³Department of Intelligent Mechanical Systems, Tamagawa University, Tokyo, Japan

⁴Faculty of Information and Human Sciences, Kyoto Institute of Technology, Kyoto, Japan

Email: zhang@radish.ee.uec.ac.jp

Received 30 June 2016; accepted 2 August 2016; published 5 August 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Developing a childcare assisting system is highly necessary due to the lack of nursery teachers, and it will make an important progress on effective utilization of nursery teacher resources. In this paper, we proposed simultaneous children recognition and tracking system by using Kinect sensors for the childcare assisting system to provide information for the nursery teachers. Each of the children is recognized by integrating his/her personal information of color, face and motion. The tracking problem is modeled as finding the MAP solution of a posterior probability, and is solved by using Markov Chain Monte Carlo (MCMC) particle filter. Our system can recognize and robustly track each child during class activities. Trajectories, motion ranges and relative distances information can be provided for the nursery teachers to assist their childcare work. The effectiveness of our system is proved through continuous monitoring of the children in a nursery school.

Keywords

Childcare Assisting System, Children Tracking, Children Recognition, MCMC Particle Filter

1. Introduction

In recent years, double-income households keep increasing, and more and more people want to send their child-

ren into nursery schools. Low birthrates in developed countries have also become a critical problem as these countries are rapidly becoming extremely aging societies. It is highly necessary to improve the quality of nursery schools to increase the birthrates. However, the number of qualified teachers is far from enough. Taking care of all the children and understanding all their behaviors is difficult and time intensive for the nursery teachers. At present, the teachers have to try their best to take care of all the children while leading the class activities, and to record their performances afterwards by memories. However, the teachers are difficult to pay attention to every child at the same time during the class activities, and they may not be able to remember all their behaviors. It is reported [1] that the nursery teachers are very interested in automatic childcare system because it would be helpful to their work. As a result, it is highly necessary to develop a childcare assisting system, by which all the children can be recognized and tracked automatically.

Researchers have developed some robotics systems for supporting childcare work. Hwang *et al.* [2] tried to use wearable sensors to recognize the motions of the children. However, the sensors are so obvious that the children may reject to wear them. Srivastava *et al.* [3] proposed to use sensor-based wireless networks to help the development in children behaviors. They explore wireless networking, middleware and data management technologies for targeting developmental problem-solving environments for early childhood education. They provide the idea of integrating multiple sensors information, but they did not talk about tracking method under complex environments with many children moving together. Sivalingam *et al.* [4] developed a multi-sensor based behavior monitoring system of at-risk children. They use a covariance based descriptor on the point clouds of Kinect sensors to track the subjects in the scene. However, their Kalman filter based tracking method will lose the individual that it is tracking if full occlusions last for a while and new trackers are initialized when they are seen again. Different from their work, we tackle the problem of recognizing and tracking children under complex environment during class activities by using multiple Kinect sensors. Each of the children is recognized and the recognition result is used for continuous tracking. It is more robust as the disappeared children can also be recognized when they show up again. Other researches about people recognition and tracking for visual surveillance have also been greatly developed. Detection and tracking people based on camera image analysis has gotten many impressive results [5] [6], even under some crowded environments [7] [8]. However, these image based tracking methods cannot show the localization of the people. Laser range finders, which use a laser to scan the distance to the objects, are also often used for people tracking [9] [10], and tracking methods based on particle filters have been successfully applied to public spaces [11] by setting multiple laser sensors. Although these methods work well in practice, they are limited to the estimation of people's positions. It is hard to identify the persons while tracking them, especially under crowded environments. The accuracies are also affected by the height of the sensors. Recently, 3D sensing has been noticed and researched due to the availability of 3D sensors, like Microsoft Kinect sensor. Human detection and tracking methods based on RGB-D information have been proposed [12]-[14]. These works use a single sensor information to track people, and are difficult to deal with occlusion problems. In some recent works [15] [16], researchers have proposed the combination of multiple 3D range sensors and successfully realized people tracking in a public space. However, their tracking methods are difficult for long time tracking, and also hard to apply for tracking children activities. We propose a children recognition and tracking system by using Kinect sensors. We can give the 3D motion trajectories even if the tracking targets are children. Moreover, all of the above-mentioned works did not think about the childcare problem from the viewpoint of nursery teachers. Shiomi *et al.* [1] [17] investigated the importance of developing childcare assisting systems, and showed some requirements from the viewpoint of nursery teachers. However, their system is still in a preliminary stage. Few researches were focused on childcare assisting with the purpose of helping the work of nursery teachers. The goal of our system is to provide necessary information for the nursery teachers.

In this paper, we propose simultaneous children recognition and tracking system by using Kinect sensors in classrooms of a nursery school. To solve the occlusion problem when multiple children move together and to get more personal information, we use multiple Kinect sensors, which are set from different views in different heights. Each of the children is recognized by integrating his/her personal information (color, face and motion). The tracking problem is modeled as finding the MAP solution of a posterior probability, and is solved by using Markov Chain Monte Carlo (MCMC) particle filter [18] [19]. Notice that the number of Kinect sensors can be adjusted according to the size of the classroom. It is a general system for simultaneous children recognition and tracking. **Figure 1** shows an example of the scenes where multiple children move together during a game.



Figure 1. The tracking scenes in a nursery school.

2. Robust Children Tracking System

In order to simultaneously recognize and track all the children in a classroom, we proposed an easy setting multi-sensor system by using Kinect sensors. Here, we show an example of using two Kinect sensors for a real classroom of nursery shown in **Figure 2**, and more sensors can be added to the system in the same way if the view of the two Kinect sensors is not enough. One Kinetsensor can almost cover 36 m^2 of space, just like the classroom shown in **Figure 2**. Another Kinect can be added for getting more personal information. The sensor positions in the classroom are also shown in **Figure 2**. Kinect 1 is set in the average height of children in front of the class to monitor the children with high qualified frontal face images, and Kinect 2 is set slanted in a higher height to monitor all of the children with less occlusions. In our system, only these two Kinect sensors are needed for monitoring the classroom. The accurate positions and slanted angles are not required in the sensor setting step, which ensures that the sensing equipment is easy to mount. The accurate coordinates of Kinect sensors will be estimated during the calibration process. The scene shown in **Figure 1** is taken from the slanted Kinect 2. The children can be tracked well when they are detected separately after projecting their position information to the ground. Too many children staying in a small space may cause overlapping of their trajectories. In our experiment, we successfully tracked around 20 children together in the classroom shown in **Figure 2**. The flowchart of our system is shown in **Figure 3**. It mainly contains four parts: initialization, children detection, children recognition and tracking. The recognition results are used as observation likelihood for the tracking process.

2.1. Children Detection

Firstly, we detect the positions of the children by the two sensors separately. Here, we take the slanted Kinect 2 as an example to explain the detection algorithm. As the sensors are set roughly, we firstly estimated the accurate height and slanted angle of the Kinect 2. We detected out all the planes using Point Cloud Library (PCL) [20] in the empty classroom, and segmented out the sub-horizontal plane with the lowest height as the floor plane. From the slanted angle of the floor plane, we can calculate the accurate Kinect slanted angle. This process needs to process only once as initialization every time after setting the sensors. We detect the children by projecting the transformed point cloud with the child height range (0.5 m - 1.2 m) on the ground, and find out projected points after deleting the background parts. Connected-component labeling [21] process is used to detect out the areas within children size by Equation (1).

$$l_{\min} \leq l_{\text{width}} \leq l_{\max}; l_{\min} \leq l_{\text{length}} \leq l_{\max} \quad (1)$$

Here, $l_{\text{width}}, l_{\text{length}}$ mean the width and length of a child candidate area, l_{\min}, l_{\max} mean the thresholds for a real child area. The range of l_{width} and l_{length} are same as the children may face to different directions.

The detection process is shown in **Figure 4**. For the scene shown in **Figure 1**, the projected result on the ground plane is shown in **Figure 4(a)**. The final children position detection result is shown in **Figure 4(b)**. We compared the detection results with the correct ones that we made manually from the point cloud one frame by

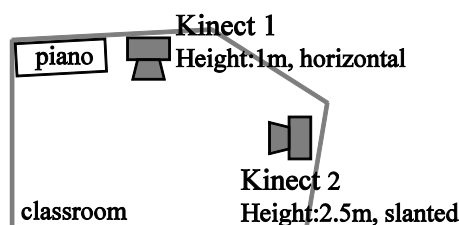


Figure 2. Sensors placement in the classroom.

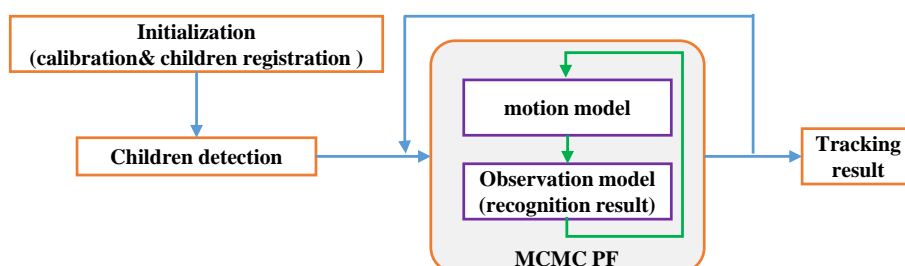


Figure 3. Flowchart of the system.

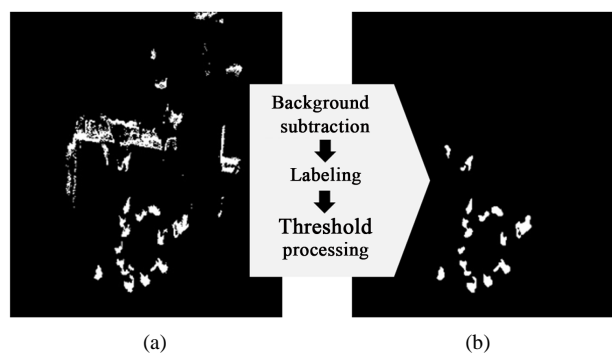


Figure 4. Children detection process. (a) xz image after projection; (b) Children areas.

one frame, and found that the average error for each child is around 10cm, with the standard deviation 8 cm. The detection results showed the position coordinates and their relative position relationships correctly. In this way, we can get the children position information from two Kinect sensors.

2.2. Calibration of Multiple Kinect Sensors

The 3D reconstruction resulting from multiple sensors strongly depends on a good calibration result. We address the problem by matching the same corresponding points of two Kinect sensors. We change the matching process into matching the corresponding points on 2D position maps to decrease the complexity and calculation. We try to match the points by affine transformation. The affine transformation matrix is calculated in advance during the initialization process. A series of corresponding points in 2 Kinect coordinate systems can be gotten easily by asking a person walking around in the classroom after setting the sensors. The single detected person by the two Kinect coordinate systems are surely be the same person. We choose 3 of them to calculate the affine transformation matrix. The remaining points are used to check the residual sum of squares (RSS). We repeat this process and find the best affine transformation matrix with the least RSS. Finally, we apply the best affine transformation matrix for calibration. Notice that we only process the calibration once after setting the sensors during the initialization process.

2.3. Simultaneous Children Recognition and Tracking

We model the children tracking problem using a sequential Bayesian framework. A child's state at time t can be

expressed as X_t (2 dimensional, location in 2D). When the observation information Z_t is gotten from sensors information at time t , we estimate the children states by finding the maximum-a-posteriori (MAP) solution of the joint probability. To find the most probable configuration, we estimate the MAP solution of $P(X_t | Z_t)$ by Equation (2).

$$P(X_t | Z_t) \propto P(Z_t | X_t) \int P(X_t | X_{t-1}) P(X_{t-1} | Z_{t-1}) dX_{t-1} \quad (2)$$

Here, $P(Z_t | X_t)$ represents the observation likelihood at time t , given the sensors input Z_t . It measures the confidence of a hypothetical configuration. $P(X_t | X_{t-1})$ is the motion model, which shows the smoothness of the trajectory over time. $P(X_{t-1} | Z_{t-1})$ is the posterior probability of time $t-1$. The posterior probability at arbitrary time t can be calculated from the probabilities from time 1 to $t-1$ sequentially if the posterior probability at initial time is given. The best configuration X_t is then the MAP solution. MCMC particle filter approximates the MAP solution as a set of discrete samples known as a Markov Chain.

2.3.1. Motion Model

The motion model $P(X_t | X_{t-1})$ can be modeled by giving the update rules as

$$X_t = X_{t-1} + X'_t dt; \quad X'_t = X'_{t-1} + X''_t dt; \quad X''_t = X''_{t-1} + \mu \quad (3)$$

Here, X'_t is the speed of the tracked child, X''_t is the acceleration of the tracked child, dt is the time between two continuous frames, and μ is a process noise for a child's motion getting from a Gaussian noise. We use the variable acceleration model to model the motion of the children as their motion are usually unpredictable.

2.3.2. Observation Likelihood

Given a hypothesized location of a child on the image, the observation likelihood measures the accuracy of the location. In order to track a particular child robustly, we use the particular child recognition result as the observation information. In our system, we proposed to use three detectors to recognize a particular child: a face detector, a color detector and a motion detector. Each single detector has its strength and weakness. The face detector is extremely reliable when frontal faces can be detected, but the face information may not be always available as the child may show his side/back to the sensor. The color detector is always available, but the accuracy is relatively low when different children wear similar clothes. The motion detector can effectively limit the motion range of the child as he/she cannot move a long distance in a single frame time. However, this detector is hard to distinguish the candidates that show up in the motion range. We propose to combine the detectors by using a weighted combination of detection responses as shown in Equation (4).

$$P(Z_t | X_t) \propto \exp\left(\sum_j w_j P_j(Z_t | X_t)\right) \quad (4)$$

Here, w_j is the weight of each detector, $P_j(Z_t | X_t)$ is the likelihood of each detector, and j stands for the types of the detector (face, motion or color).

Face detector is used to detect and recognize a particular child's frontal face. We employ the OKAOVISION [22] software in our system. The particular child face detector likelihood is calculated from the maximum recognition confidence score S_{face} .

$$P_{\text{face}}(Z_t | X_t) = \alpha(S_{\text{face}} - Th), \text{ if } S_{\text{face}} > Th; \quad P_{\text{face}}(Z_t | X_t) = 0, \text{ else} \quad (5)$$

Here, Th is the threshold of face identification confidence. α is the coefficient to adjust the range of the OKAOVISION recognition confidence to 0 - 1 so that it can be used as a probability.

The weight of face detector is influenced by the showing angle of the frontal face and its distance to the sensor. The angle of the face is changing from $(-\pi/2, \pi/2)$, so w_{face} can be calculated by

$$w_{\text{face}} = (1 - 2\theta/\pi) \exp(-L), \text{ if Face Detected; } w_{\text{face}} = 0, \text{ else} \quad (6)$$

Here, θ is the angle of detected face, and L is the distance from the Kinect sensor to the face.

Motion detector is a strong indicator of the presence of a person. The areas around the predicted position trend have a higher possibility to be the tracked target. The observation likelihood $P_{\text{motion}}(Z_t | X_t)$ is calculated from the distance D between the predicted position and detected children areas.

$$P_{\text{motion}}(Z_t | X_t) = \beta/D^2 \quad (7)$$

Here, β is the coefficient to adjust the range of the motion likelihood.

Color detector is used for searching out the child with similar color. As the Kinect 2 is set slanted in a relatively high height to decrease occlusions, and the other is set in front of the classroom, the color information of any child is almost available all the time. We find out the points in the Kinect point cloud that corresponds to the detected areas and match their histograms with the children that we are trying to track. The observation likelihood $P_{\text{color}}(Z_t | X_t)$ is calculated by the similarity S_{color} of the color, which is calculated from comparing the color histograms between the detected child and the registered ones. We use correlation of the Hue channel histograms in the HSV color space here.

$$P_{\text{color}}(Z_t | X_t) = S_{\text{color}} \quad (8)$$

The weight of motion and color detector are designed with the relationship as follows:

$$w_{\text{motion}} = \gamma w_{\text{color}}; w_{\text{face}} + w_{\text{motion}} + w_{\text{color}} = 1 \quad (9)$$

Here, γ is a constant value, showing that the weight of motion is γ times of the weight of color due to the motion information better reflects the real position of a child than the color. Children may wear similar clothes.

2.3.3. Tracking with MCMC Particle Filter

We have discussed the motion model and how to evaluate proposed tracking states through observation likelihood. Then we need to explore the space of these hypotheses to find the MAP solution. To efficiently explore the configuration space and obtain the MAP solution, we used the MCMC particle filter method [19]. Unlike [19], however, our goal is to recognize and track each of the child who ran during class activities, so we need to consider the status that children are close to each other, which shown as overlapped position on the projected ground area. To this end, an important contribution of our method is that we extended the method in three ways: 1) We use a group of independent particles to track one child and these particles will be never used for other children's tracking. Multiple children tracking can be realized by parallel running different groups of particles. It will make sure that the tracked person is fixed and never mixed with each other. 2) The conventional estimation result of MCMC particle filter is calculated as the mean value of all the re-sampled particles, but we proposed to use the center of the area that is closest to the mean value of the re-sampled particles as the real position of the child. In this way, the tracked child will be related to a detected child, and the tracking error will be reduced. 3) Multiple children can share one area gotten from the children position detection. When two or more children are close to each other, their projections on the ground may fuse with each other, and their detection result turns out to be a "big" area. Multiple children share this area in practice. In this way, our system can work even if the detected number of children keeps changing.

3. Experiments

In this section, we present the experimental tracking results of children during a rhythmic class in a nursery school. Then we try to analyze their behaviors for providing useful information to nursery teachers. The necessary information is investigated in advance by asking 8 nursery teachers and referring a social acceptance investigation research [1] about the childcare assisting system.

3.1. Children Recognition and Tracking Results

Our system not only can recognize and track children when they are static or move slowly, but also can work well when they move fast with kinds of occlusions and crossings. We took a very difficult scene of the drum game during a eurythmic class to show our tracking results. The teacher leded the children to walk or run along with drum rhythm. The drum game lasted 90 seconds, during which teacher leded the children walk in the clockwise direction for four circles. They moved slowly in the beginning and accelerated and decelerated many times during the game. They also stopped a while for talking during the game. **Figure 5** shows the tracking results of the teacher and two particular children. For each frame, 5 pictures are shown to express the status of the teacher or the children: "(a) color" shows the color image of the current frame and the persons that to be tracked; "(b) depth" shows the position information of the detected teachers and children. The Kinect 2 is set in the middle-bottom position in the picture; "(c) teacher" shows the trajectory of the teacher during the game; (d) and (e) show the trajectories of two children during the game. We can observe that the teacher walked in the clockwise

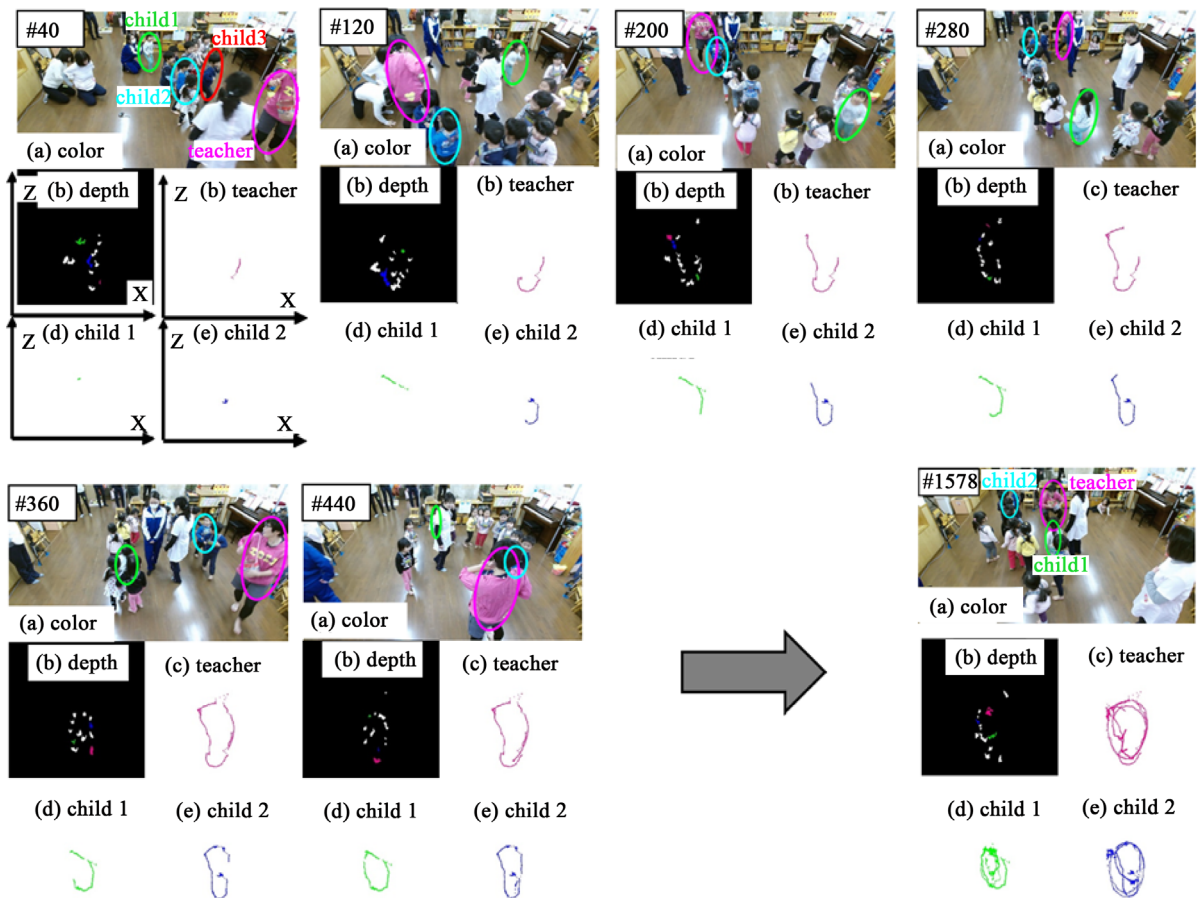


Figure 5. The tracking trajectories of the teacher and two children during the drum game.

direction for four circles. From the changing tendency of the teacher’s trajectory, we can also observe that the teacher walked slowly in the beginning, and stopped after walking two circles. After that, the teacher began to run fast for another two circles, after each of which she stopped and waited for the children. The child 1 and child 2 were also observed as moving in the clockwise direction for four circles, with the same speed changing tendency of the teacher except a time delay. We observed that the motion tendencies and trajectories of all the children are matched perfectly with that recorded by nursery teachers. In frame 120, we can see in (b) that the child 2 and the teacher share one same area in the depth image as they are close to each other. In this way, we solve the problem that less people are detected than the real number. This may leads to some errors of the position of the tracked child as the output will be the center of all the long area. However, this kind of error do not affect the motion tendency so much of the tracked children so that the trajectories can still show the motion of the children robustly. Besides, the teacher and the children can be continuously tracked without lacking any frame. We observed from our tracking result that the teacher led the children to move in the clockwise direction for four circles during the whole game. Their moving speed kept changing during the game. These results are almost the same with the results that the nursery teachers recorded, proving that our method is effective.

In order to show the validity of our system, we evaluated the tracking results by comparing with the correct ones. The correct results are generated manually by assigning the position from original information of Kinect sensors. We can calculate the tracking accuracies for each child. **Figure 6** shows the tracking error changing tendency of the teacher and the child 1. The average errors of the tracking result of them are 0.103 m and 0.122 m, with the standard divisions as 0.088 m and 0.112 m. This tracking accuracy is good enough for our purpose of analyzing the behavior of the children and providing necessary information to the nursery teachers. We observe that the teacher and the child 1 are successfully tracked by our system during the whole drum game with low error, although the distance errors becomes a little big for a few frames, as shown in **Figure 6**. These errors

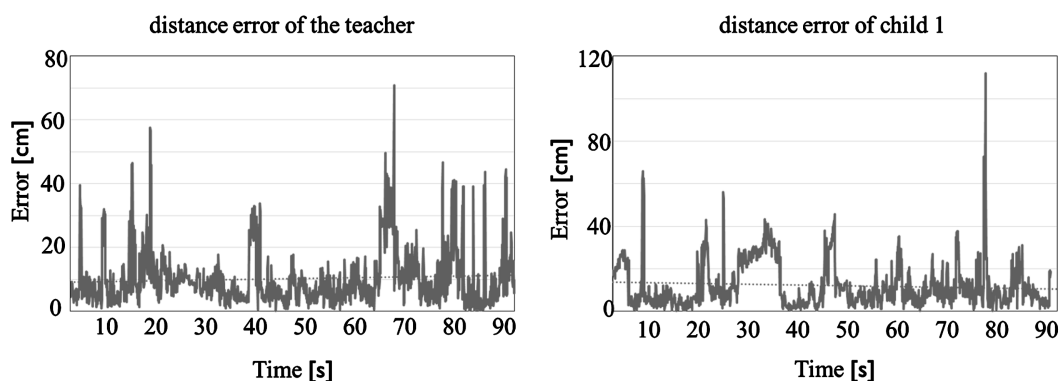


Figure 6. The tracking error changing tendency of the teacher and the child 1.

are caused by the overlapping of projected positions on the ground. As more than one child are closed to each other, they formed a “big” area in the human detection result. The center of this area is different from the real position of any single child. However, our method is more robust as the error would decrease after these children separated with each other, which is shown as breaking up of the “big” area in the detection result.

To show the effectiveness of our system, we also use the conventional multiple laser sensors based tracking method [14]. We observe that the conventional method cannot track the people well even if it can detect out the positions of them. The tracking result by conventional method is shown as **Figure 7**: “(a) color” shows the color image of the current frame and the persons that to be tracked; “(b) tracking result” shows the position information of the person to be tracked. The tracking result of the teacher shown as different colors in **Figure 7** mean that the teacher is tracked as different IDs. We observed that the teacher is only tracked for 46 s during the drum game (90s in total). After that (frame 948 in **Figure 7**), the teacher is also detected as a person, but recognized as another one, which led to the failure of continuous tracking (frame 948~frame 1578). During the drum game, the teacher is recognized as three different persons.

We observe that both the conventional method and our proposed system can work well when children are static or move slowly, but only our system can work well for the scenes that the children moving fast with kinds of occlusions and crossings with each other. We proved that our proposed system can be used for continuous monitoring and tracking each of the child during class activities.

3.2. Children Behavioral Analysis

The purpose of our work is to provide useful information for the nursery teachers to help their work. In this part, we show four different kinds of information that can support the work of the teachers. These results are designed based on the requirements of the nursery teachers and also are evaluated by them.

3.2.1. Motion Trajectories of the Children

As the nursery teachers need to record the activities of the children after class, they used to have to remember all the reactions or motions during the whole class. This is almost impossible as the amount of information is too huge. They can only remember some special reactions of a child and the performances of some special (very active or uncooperative) children. Our system can provide the motion trajectory of any child. This information can help the nursery teachers remember of the performance of any child. **Figure 4** shows the motion trajectories of some children. Our tracking results are almost the same with correct answers, which are recorded by the staffs, and we can provide the motion trajectory and tendency of any child. This information proved to be useful for the nursery teachers to help them remind the performances of the children.

3.2.2. Motion Range

The nursery teachers believe that the motion range and momentum of a child during the class can show the growth process and familiarity to the class. A younger child or a new member tends to be quiet. They will be more active with growing up. Motion range can be used as a quantitative index to show the growth of a child. Our system can provide accurate motion range information of the children. From the tracking results, we can calculate out the motion area of a child by finding the bounding rectangle of his/her trajectory. Their dynamical

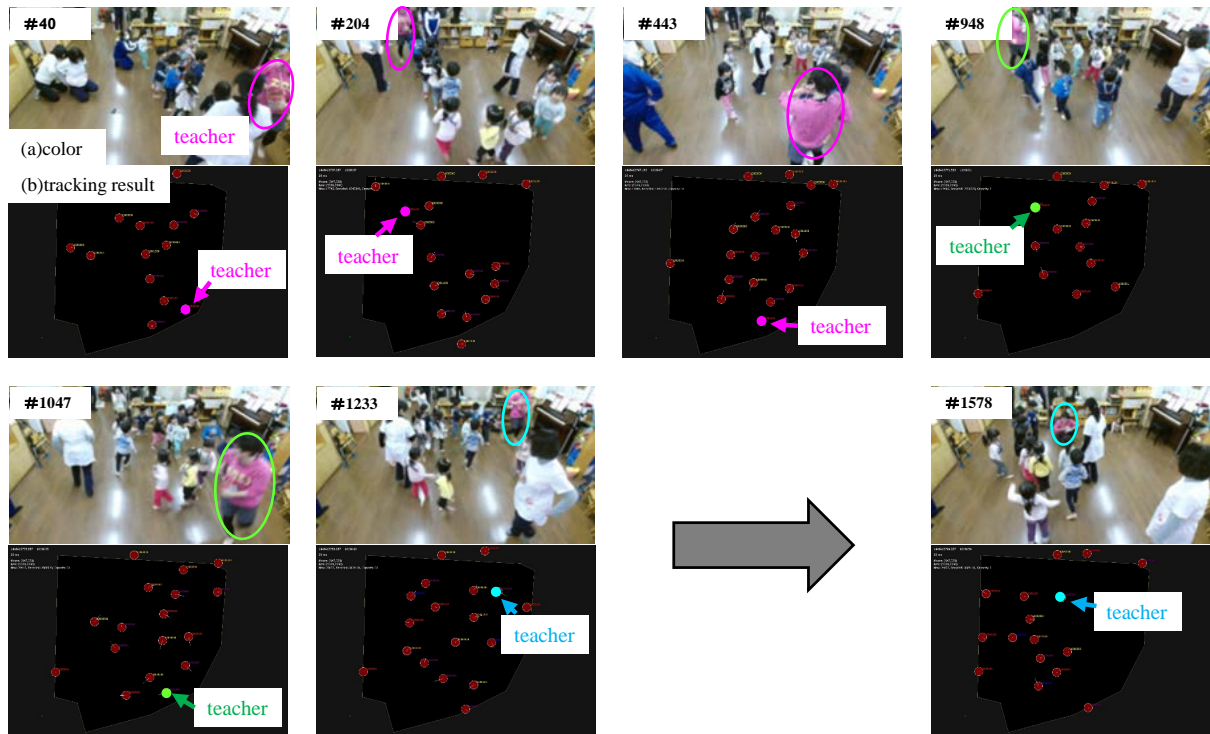


Figure 7. The tracking result based on multiple laser sensors during the drum game.

momentum can also be calculated from the length of the trajectories. The motion areas of the teacher and the two children during the drum game are shown in **Figure 8(a)** and their dynamical momentums are shown in **Figure 8(b)**. We observe that the motion range and momentum of the teacher are bigger than those of the children. This is because the teacher led the motions of the children. With these information, our system can be applied to monitor the developments of a child along with the growth of their ages after a long term observation. The contribution of our system is that we provide a quantitative way to analyze the growth of the children, which is very helpful for the nursery teachers.

3.2.3. Relative Distance

The nursery teachers need to know the relationship among the children for better leading their growth. Besides, the teachers also need to know how much a child rely on him/her in the daily life. This can be evaluated by the relative distances between two persons. From the tracking results, we can accurately calculate the relative distance between different persons. In **Figure 9(a)**, it shows the relative distances between the teacher and 3 different children. Their average relative distances during the game is 1.768 m, 0.797 m, 0.550 m with the standard deviation of 0.689 m, 0.416 m, and 0.267 m. We can see that child 2 and child 3 prefer to stay close to the teacher, and child 1 prefers to keep a small distance with the teacher. Similarly in **Figure 9(b)**, it shows the relative distances between child 3 and the other two children. Their average relative distances during the game is 1.623 m, and 0.307 m with the standard deviation of 0.592 m and 0.235 m. We can see that child 2 stays closer to child 3 at most of the time, and their distance is very small, even in touch with each other sometimes. On the other hand, child 1 usually keeps a small distance with them. We can infer their relationships that child 2 and child 3 are close friends and they like to play together. This information is especially useful for monitoring the children under natural status. By showing the children who like to play with each other, the teacher can understand the behaviors of the children better.

4. Conclusion

In this paper, we proposed a novel system of simultaneous children recognition and tracking system by using Kinect sensors, towards the goal of assisting the nursery teachers with the child care work. Each of the children

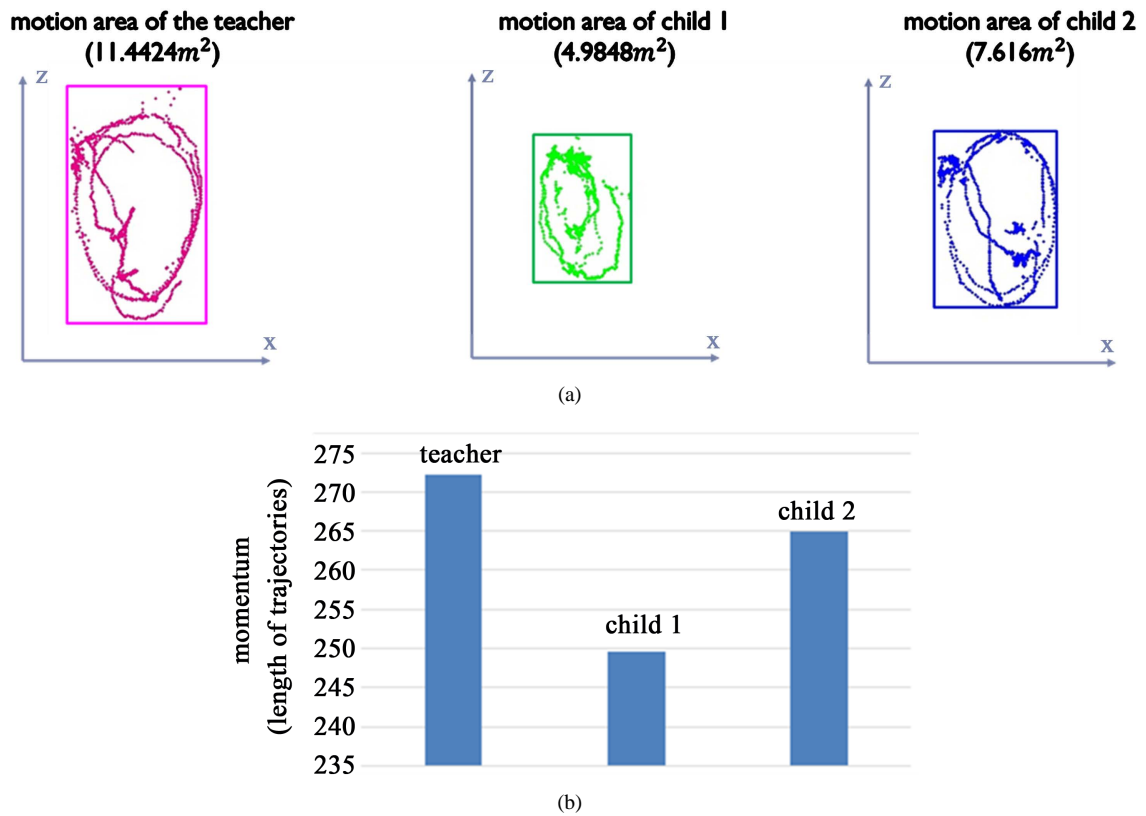


Figure 8. Motion area and momentums of different persons. (a) Motion areas of different persons; (b) Dynamical momentum of different persons.

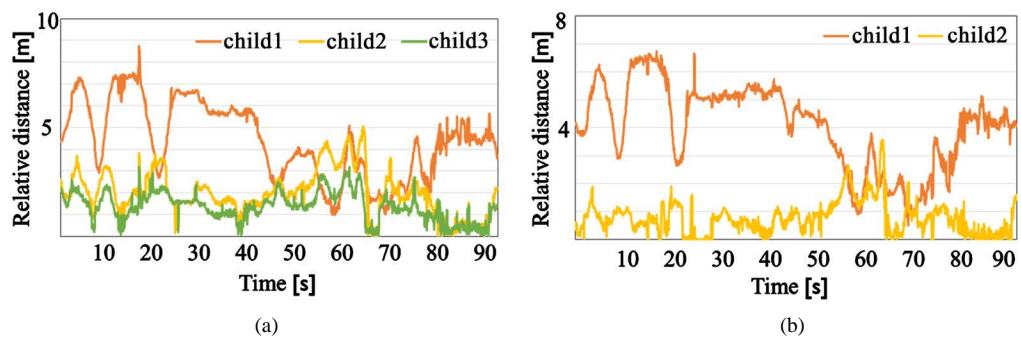


Figure 9. The relative distance relationships during the drum game. (a) Relative distance with the teacher; (b) Relative distance with child 1.

is recognized by integrating his/her personal information (color, face and motion). The tracking problem is modeled as finding the MAP solution of a posterior probability, and is solved by using Markov Chain Monte Carlo (MCMC) particle filter. We extended the tracking method by modifying the tracking result of each frame according to the detection results, and allowing different children share one human detected area for solving the problem of detected human number changing. By our system, we can recognize and robustly track each child during complex class activities. The effectiveness of our system is proved through comparing the tracking results with conventional laser sensors based method as our system can still work well when the children are moving with kinds of occlusions and crossings with each other. Trajectories, motion ranges and relative distances information can be provided for the nursery teachers to assist their childcare work. The information is designed according to the requirement of the nursery teachers and evaluated to be useful by them. However, the color information of each child cannot be repeatedly used as the children change their clothes every day. More

robust personal features need to be proposed for personal recognition. Future work will also be focused on understanding different scenes by the system and provide more information to the nursery teachers. This will be conducted by further communicating with the nursery teachers and understanding their needs.

Acknowledgements

This work was supported by Grant-in-Aid for Scientific Research on Innovative Areas 26118003. Special thanks to all the members in the nursery school.

References

- [1] Shiomi, M. and Hagita, N. (2015) Social Acceptance of a Childcare Support Robot System. *24th IEEE International Symposium on Robot and Human Interactive Communication*, 13-18. <http://dx.doi.org/10.1109/roman.2015.7333658>
- [2] Hwang, I., Jang, H., Nachman, L. and Song, J. (2010) Exploring Inter-Child Behavioral Relativity in a Shared Social Environment: A Field Study in a Kindergarten. *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, 271-280. <http://dx.doi.org/10.1145/1864349.1864373>
- [3] Srivastava, M., Muntz, R. and Potkonjak, M. (2001) Smart Kindergarten: Sensor-Based Wireless Networks for Smart Developmental Problem-Solving Environments. *The 7th Annual International Conference on Mobile Computing and Networking*, 132-138. <http://dx.doi.org/10.1145/381677.381690>
- [4] Sivalingam, R., Cherian, A., Fasching, J., Walczak, N., Bird, N., Morellas, V., Murphy, B., Cullen, K., Lim, K., Sapiro, G. and Papanikolopoulos, N. (2012) A Multi-Sensor Visual Tracking System for Behavior Monitoring of At-Risk Children. *IEEE International Conference on Robotic and Automation*, 1345-1350.
- [5] Shiomi, M. and Hagita, N. (2014) Preliminary Investigation of Supporting Child-Care at an Intelligent Playroom. *The Second International Conference on Human-Agent Interaction*, 157-160. <http://dx.doi.org/10.1145/2658861.2658911>
- [6] Moeslund, T., Hilton, A. and Kruger, V. (2006) A Survey of Advances in Vision Based Human Motion Capture and Analysis. *Computer Vision and Image Understanding*, **104**, 90-126. <http://dx.doi.org/10.1016/j.cviu.2006.08.002>
- [7] Tang, S., Andriluka, M., Milan, A., Schindler, K., Roth, S. and Schiele, B. (2013) Learning People Detectors for Tracking in Crowded Scenes. *International Conference of Computer Vision*, 1049-1056. <http://dx.doi.org/10.1109/iccv.2013.134>
- [8] Andriluka, M., Roth, S. and Schiele, B. (2008) People-Tracking-by-Detection and People-Detection-by-Tracking. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1-8. <http://dx.doi.org/10.1109/cvpr.2008.4587583>
- [9] Glas, D., Miyashita, T., Ishiguro, H. and Hagita, N. (2009) Laser-Based Tracking of Human Position and Orientation Using Parametric Shapemodelling. *Advanced Robotics*, **23**, 405-428. <http://dx.doi.org/10.1163/156855309X408754>
- [10] Chou, C., Li, J., Chang, M. and Fu, L. (2011) Multi-Robot Cooperation Based Human Tracking System Using Laser Range Finder. *IEEE International Conference on Robotics and Automation*, 532-537.
- [11] Glas, D., Miyashita, T., Ishiguro, H. and Hagita, N. (2009) Simultaneous People Tracking and Localization for Social Robots Using External Laser Range Finders. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 846-853. <http://dx.doi.org/10.1109/iros.2009.5354198>
- [12] Munaro, M., Basso, F. and Menegatti, E. (2012) Tracking People within Groups with RGB-D Data. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2101-2107. <http://dx.doi.org/10.1109/iros.2012.6385772>
- [13] Kirchner, N., Alempijevic, A. and Virgona, A. (2012) Head-to-Shoulder Signature for Person Recognition. *Proceedings of the IEEE International Conference on Robotics and Automation*, 1226-1231. <http://dx.doi.org/10.1109/icra.2012.6224901>
- [14] Hansen, D., Hansen, M., Kirschmeyer, M., Larsen, R. and Silvestre, D. (2008) Cluster Tracking with Time-of-Flight Cameras. *IEEE Conference Computer Vision and Pattern Recognition Workshop*, 1-6.
- [15] Brscic, D., Kanda, T., Ikeda, T. and Miyashita, T. (2013) Person Tracking in Large Public Spaces Using 3-D Range Sensors. *IEEE Transactions on Human-Machine Systems*, **43**, 522-534. <http://dx.doi.org/10.1109/THMS.2013.2283945>
- [16] Almazan, E.J. and Jones, G.A. (2013) Tracking People across Multiple Non-Overlapping RGB-D Sensors. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 831-837. <http://dx.doi.org/10.1109/cvprw.2013.124>
- [17] Ali, A. and Terada, K. (2009) A Framework for Human Tracking Using Kalman Filter and Fast Mean Shift Algorithms. *12th International Conference on Computer Vision Workshops*, Kyoto, 27 September-4 October 2009, 1028-1033. <http://dx.doi.org/10.1109/iccvw.2009.5457591>
- [18] Andrieu, C., Davy, M. and Doucet, A. (1999) Sequential MCMC for Bayesian Model Selection. *IEEE Signal Processing Workshop on Higher Order Statistics*, Caesarea, 14-16 June 1999, 130-134. <http://dx.doi.org/10.1109/HOST.1999.778709>

-
- [19] Choi, W., Pantofaru, C. and Savarese, S. (2013) A General Framework for Tracking Multiple People from a Moving Camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**, 1157-1591.
- [20] Rusu, R. and Cousins, S. (2011) 3D Is Here: Point Cloud Library (PCL). *IEEE International Conference on Robotics and Automation*, Shanghai, 9-13 May 2011, 1-4. <http://dx.doi.org/10.1109/icra.2011.5980567>
- [21] He, L., Chao, Y. and Suzuki, K. (2008) A Run-Based Two-Scan Labeling Algorithm. *IEEE Transactions on Image Processing*, **17**, 749-756. <http://dx.doi.org/10.1109/TIP.2008.919369>
- [22] Takigawa, E., Hosoi, S. and Kawade, M. (2004) Gender, Age, Human Race Identification and Estimation Technology of Facial Attributes—OKAO Vision. Image Lab, 15.4.



Scientific Research Publishing

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>