Noname manuscript No.
(will be inserted by the editor)

Estimating Children's Social Status through their Interaction Activities in Classrooms with a Social Robot

Tsuyoshi Komatsubara · **Masahiro Shiomi** · **Thomas Kaczmarek** · **Takayuki Kanda** · **Hiroshi Ishiguro**

Received: date / Accepted: date

Abstract We developed a technique to estimate children's social status in classrooms with a social robot. Our approach observed children's behaviors using a sensor network. We used depth cameras to track their positions and identified them with RGB cameras and exploited the presence of a social robot for the estimations. We specifically observed the children's behavior around the robot, expecting that their interactions with it would provide clues for estimating their social status. We collected data at an actual elementary school and observed 70 fifth graders from three different classes during six lectures for each class period. Our system tracked the positions of the children 93.4% of the time and correctly identified them 65.5% of the time in crowded classrooms that held 28 students. These results were used to estimate the children's social status. Our developed system successfully estimated the children's social status with 71.4% accuracy.

Keywords Estimation of social status · Sensor network · Robots for classroom · Human-Robot Interaction

1 INTRODUCTION

In the relatively near future, social robots are expected to interact with children in classrooms, including such situations as language education in classrooms [1, 2], language teaching by social supportive behavior [3], and encouraging students to pose questions to robot teachers in class because asking a robot is easier than asking a human teacher [4]. However, social robots in classroom environments with children need to understand the 'social status' of the children. Social status, which broadly refers to a person's rank within a particular social structure, can be estimated by occupation and education [5]. More specifically to classroom contexts, social status is established through social acceptance and social connections. Popular children rank high in the classroom social structure, and less popular or ignored children rank low [6] (Fig. 1).

Fig. 1 Children with high and low social status during a lesson

Social status is often studied in educational and developmental research because it is critically related to children's school life and academic performance. For instance, children with low social status often become the targets of bullying [7], and bullies tend to have high social status [8]. There are two bullying patterns: one involves a popular child who bullies unpopular individuals of the same gender; the other describes unpopular, aggressive boys who bully popular girls [9].

T. Komatsubara, M. Shiomi, T. Kaczmarek, T. Kanda
ATR, 2-2-2, Hikaridai, Keihanna Science City, Kyoto, 619-0288, Japan
Tel.: +81-774-95-1405
Fax: +81-774-95-1408
E-mail: {komatsubara.tsuyoshi, m-shiomi, thomask, kanda}@atr.jp

H. Ishiguro
Osaka University, 1-3, Machikaneyama, Toyonaka, Osaka, 560-8531, Japan
Tel. and Fax: +81-6-6850-6360
E-mail: ishiguro@sys.es.osaka-u.ac.jp

Social status also influences the academic outcomes of children. Children with low social status tend to have less successful academic performances [10]. The transformation of low social status into low academic performance is probably due to reduced motivation to succeed [11].

We speculate that future social robots will be used for interventions in the environments of children. The literature has identified and discussed the effect of intervention. For instance, the academic performance of children with low social status was improved by training their academic and social skills [12]. To reduce bullying, a few different interventions were considered, e.g., training non-bullying children with high-status to intervene in bullying situations with high-status bullies or helping low social status children acquire better social skills and encouraging them to build better social connections [8]. If a social robot were to understand each child's social status, it could be used during such interventions to improve the position of low social status children. The capability to understand social status is a critical step for such social robots.

2 RELATED WORKS

2.1 Social Robots for Children and Classrooms

Previous studies unveiled the possibility of using social robots to support the social lives of children. For instance, Woods et al. explored how differently bullied children tell stories featuring a robot than children who haven't been bullied [13]. Bethel et al. revealed that as an interviewer, a robot can be used to investigate sensitive events [14], and Tanaka et al. concluded that a robot can be accepted as a close peer of children [15]. Belpaeme et al. explored whether a robot's adaptation to children's characteristics effectively contributed to teaching mathematics [16]. Komatsubara et al. investigated whether a social robot increased science understanding in classrooms [17] and whether using pointing gestures encourage children to ask questions [18]. Shiomi et al. investigated whether a social robot stimulated the science curiosity of children through a long-term interaction [19].

However, until now, no research has revealed how to develop the capability of estimating social status, even though the literature suggests that social robots will eventually appear in classrooms.

2.2 Understanding People in Group Setting

To the best of our knowledge, no previous works have addressed techniques that estimate social status. Some techniques have estimated people's relationships (e.g., friends), although participants were required to use wearable sensors. For instance, Choudhury et al. developed a wearable device called a sociometer that records the proximity events of its carriers who were near each other from which they estimated people's social networks [20]. Similarly, children's friendships in a classroom were estimated by proximity information that was observed with RFID tags around an interactive robot [21]. Although these studies revealed the importance of observing proximity information, we found that it is difficult for children to constantly carry/wear a wearable device in school.

Without requiring that any device be carried, we can still identify individuals from their faces [22]. Nonetheless, face-based identification is limited because it can only identify people whose frontal or side face is observed by a camera. Nathan et al. proposed a shape-based identification method [23], but it would also probably perform poorly in such a crowded situation as a school environment. Thus, face-based identification alone does not adequately perceive people's behaviors. Instead, researchers have explored various techniques to combine tracking techniques with person identification (e.g., [24]).

Information from cameras has also been used for estimation. For instance, Aran et al. observed such nonverbal features as motion energy (motion regions on the upper torso including arm and head motions) and speaking turns in a small group meeting to estimate personality [25]. Hung et al. proposed a method to identify the dominant person in a group meeting from such features as speaking length and optical flow [26]. Even though many previous works addressed group-meeting situations, we focus on classroom situations. The difficulty is that children often move around; thus observing their gestures/motions is complicated (since they often move beyond the camera view and do not necessarily face the camera), and the room is too noisy (when they are allowed to talk) to apply techniques that analyze their auditory interactions.

Unfortunately, little is known about how to retrieve useful information from children's behavior in a classroom. In contrast, our work, which pioneers the challenge of retrieving information from classroom behavior, is built upon rather robust tracking and identification techniques. We describe the useful features we retrieved from an actual elementary school science classroom and the techniques that estimated children's social status.

3 SYSTEM DESIGN

3.1 Expected Relation between Behavior and Social Status

Observing children's social behaviors is critical to their social status. For example, children with high social status tend to be friendly, interactive, and helpful [6], but children with low social status tend to spend too much time alone [27];

children with high social status spend more time with other children. Since students with high social status often have high social power [28], they can more easily ask for help when facing a difficult problem during a lecture because they move around more, and they may more frequently use novel things like a social robot because they stay around it more.

3.2 Technical Requirements

Our approach is to understand children through observations. We considered the following requirements:

Non-wearable: Although previous studies indicated the possible use of wearable sensors [20, 21, 29–31], children are typically prohibited from bringing smartphones or other devices to elementary schools. Thus, observation must be done with non-wearable devices, i.e., with such sensors as cameras.

Tracking and identification: The behaviors described in Section 3.1 can be extracted if we can identify children and continuously estimate their positions. Without wearable sensors, a possible approach for person identification is with a camera, i.e., face identification (face recognition). However, face identification can only be done when a frontal/side face is visible to the camera; this is not always the case, since children often change their face direction or move around. Such cases require a good combination of tracking and identification systems.

Social robot in classrooms: We installed a robot that is designed to encourage self-learning. In our scheme, robots that help learning will be used in classrooms. Interaction with them will provide clues for an estimation system by a robot that actively influences the children's behavior.

3.3 Overall System Design

Figure 2 illustrates the architecture of our developed system whose design is based on the above considerations. Our observation system (Section 4) estimates children's positions by integrating both people-tracking and face-identification systems, while a social robot actively interacts with children to support self-learning. When children's faces are identified, their IDs are associated to the tracked entity in the tracking system. The estimation system (Section 5) judges children's social status based on the features extracted from their positions.

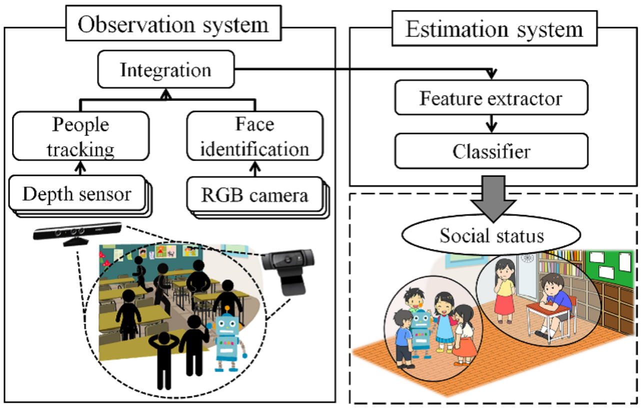

Fig. 2 System overview

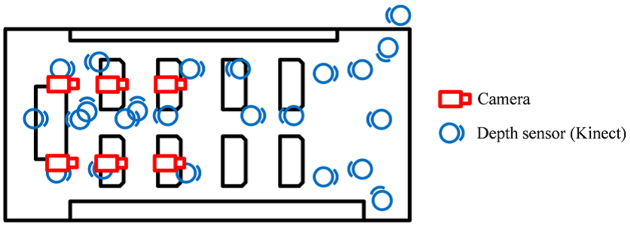

Fig. 3 Sensor arrangement in classroom

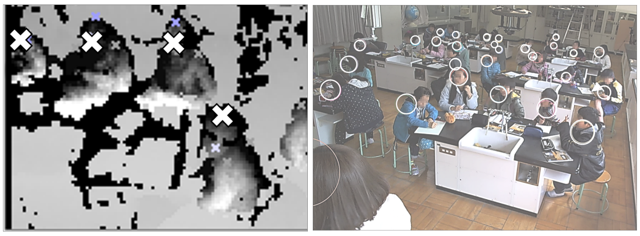

Fig. 4 Depth image with estimated head positions (left) and overall tracking results (right)

4 OBSERVATION SYSTEM

4.1 People-Tracking

We employed a people-tracking algorithm using depth sensors [32] and attached depth cameras to the ceiling to estimate the people's positions and heights based on head and shoulder shape detection. With our settings, the tracking system follows the positions of all the people in the area at 30 Hz with accuracy of about 30 cm. This system has robustness toward changes of color information due to different clothing because it uses depth information. Since the depth camera sees from the top down, it is also robust for crowded situations.

We arranged the locations of the depth sensors to efficiently cover the whole space. A sensor (Kinect, with a 57° horizontal and 43° vertical field of view (FOV)) covers approx. 4 m * 3 m of the space when attached to the ceiling at 2750 mm. 24 depth sensors covered an 8 by 16 m area of the room (Fig. 3). Fig. 4 shows a depth image from a sensor

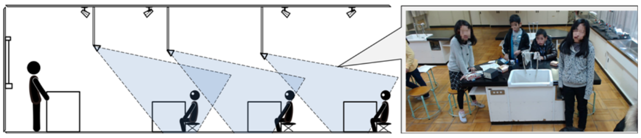

Fig. 5 Camera arrangement for face identification

and the classroom's tracking result. In this room four to five children are sitting around each desk, and the system tracked them well in such a crowded situation.

4.2 Person Identification

For person identification, we employed a face-identification approach using RGB cameras and commercially available software (Omron, OKAO Vision [22]).

We designed the camera configurations to improve the balance between the number of cameras and the chance of face identification. OKAO Vision requires a view of a face in a frontal direction within 20° (pitch) and 35° (yaw). The school requested that cameras not obstruct classroom activities, e.g., no cameras on the desks facing individuals. Here we assume that the children will at some point look at the front of the classroom, where the teacher usually stands, and so we only aimed one camera (Logicool, C920t, 70.5° horizontal and 43.6° vertical FOV) at each desk, which should capture all of the children at the desk (Fig. 5).

We also developed software to facilitate face registration that extracts the face regions from images and compares them with the faces registered in the database. Since face images vary depending on angles and facial expressions, registering many face images with different angles is a key factor for accurate face identification. Therefore, we implemented a function to semi-automatically find the best set of face images for registration. After processing the face detection and identification processes, a coder labeled the correct IDs for the face images and registered new individuals using the stored face images. After manual coding, the system optimized the face-identification performances using the new registered images by testing the influence of registering each image over a randomly selected subsample of images from the known yet still unrecognizable face images of the same individual. We continued these processes to optimize the face-identification system until the performance became saturated (Fig. 6).

4.3 Tracking and Identification Integration

We next describe the process that integrates the tracking and identification results. Head positions are tracked in 3D absolute coordinates, and faces are identified in the 2D camera view. Since each suffers from some positioning errors,

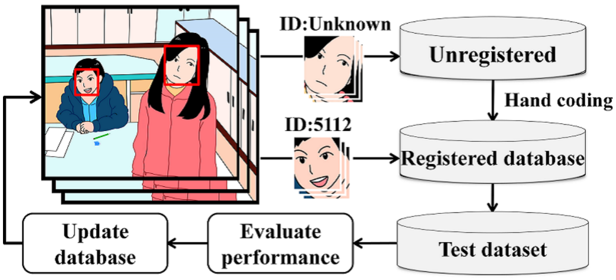

Fig. 6 Flow of finding best matches of registered images

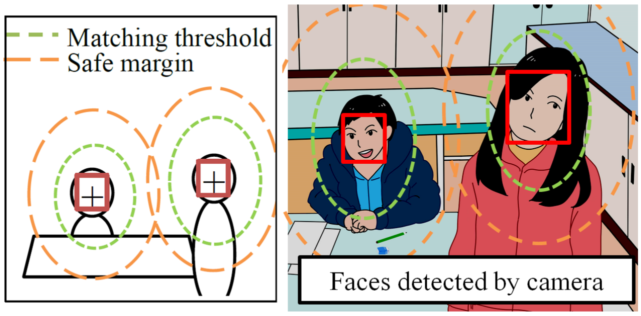

Fig. 7 Matching 3D people-tracking and face recognition

when multiple people are visible in a camera view, finding the correspondence of both positions is complicated. For our purpose we believe that failing to make an association is preferable to making an incorrect association. To prevent such associations, we defined a safe margin for integration. An association is only applied when an identified face and a tracked-head position are within a matching threshold and the head position is outside of the safe margin of all the other tracked-head positions (Fig. 7). We empirically configured the threshold and safe margin to be 20 and 30 cm horizontally, which is twice the vertical tolerance because the tracking errors were larger in that direction.

5 ESTIMATION SYSTEM

We used the tracking results (Section 4) to estimate the children's social status. From them, we computed various behavioral features (discussed in Section 3), which are used for machine learning with a Support Vector Machine (SVM).

5.1 Classroom Behavior Features

We computed the following features about children's classroom behaviors:

Time spent alone: Since less popular children (low social status) tend to spend more time alone than children with high social status (e.g., [27]), we measured the ratio of such times. For each child i and each time t, we computed whether any other children are within a threshold (D_{TH}) of this child and defined it as follows:

$$Time\,spent\,alone(i) = \frac{1}{Tracked(i)} \sum_{t=t_0}^{t_n} (isAlone(i,t) \cdot \Delta t) \quad (1)$$

$$isAlone(i,t) = \begin{cases} 1 & if(d(p(i,t), p(j,t)) > D_{TH} \;\; for\,all\; j(j \neq i)) \\ 0 & otherwise, \end{cases} \quad (2)$$

where $Tracked(i)$ returns the total tracking time of child i, Δt is the time step (33.3 msec) for this calculation, $p(i,t)$ is the x-y position of child i, and $d(a,b)$ is the Euclidean distance between x-y positions, a and b. Based on proxemics knowledge [33], we used multiple thresholds for D_{TH}: 500 mm as an intimate space to interact with friendly people and 1200 mm as the personal space to collaborative with others.

Number of surrounding people: Since children with high social status tend to be helpful, friendly, and interactive, they get many nominations [6], meaning that they have many friends or people who like them. Thus, they spend time with such individuals and well work with others in group-work activities. To capture this idea, we measured the average number of surrounding people. For each child and each moment, we computed the number of other people within a distance threshold (D_{TH}) from him/her and defined this feature as:

$$Number\,of\,surrounding\,people(i) = \frac{1}{Tracked(i)} \sum_{t=t_0}^{t_n} \sum_{j=\forall j} (if(d(p(i,t), p(j,t)) < D_{TH})). \quad (3)$$

Moving distance outside personal desk area: In classroom activities, children were split into groups and assigned to desks. While children often worked within the area of their desk, sometimes they moved around in the classroom. Since children with high social status interact with others more than children with low social status, they visited people outside their own desks more frequently (within their own desk area, they moved for the activity that was assigned to their group). We measured the average moving distance outside their desks. Here a child is judged to be outside of her own desk area if the distance from the desk exceeds a range threshold (R_{TH}) and defined it as follows:

$$Moving\,distance\,Outside(i) = \frac{1}{Tracked(i)} \sum_{t=t_0}^{t_n} (isOutside(i,t) \cdot d(p(i,t), p(i,t+\Delta s))) \quad (4)$$

$$isOutside(i,t) = \begin{cases} 1 & if(d_m(p(i,t), desk(i)) > R_{TH}) \\ 0 & otherwise, \end{cases} \quad (5)$$

where $desk(i)$ is the rectangular area of child i's assigned desk and d_m is the shortest Manhattan distance between the position and the rectangular area on the x-y plane (y is the classroom's long side). We set R_{TH} to 300 mm for situations where children change their own positions around the desk and 600 mm for situations where they go to other desks. We set Δs to 500 msec to eliminate noise effects in the tracking.

5.2 Robot-related Features

Since we believe that children's interaction with the robot will provide additional clues for an estimation system, we added the following features:

Time spent around the robot: Social status is closely connected to social hierarchy, in which a high rank denotes a higher priority access to resources [28]. When a robot is novel, since many children want access to it (e.g., [1]), a somewhat competitive situation exists where only a limited number of children can actually secure access to it. Children with high social status tend to gain more access than children with low social status. Therefore, we measured the ratio of the time that children spent around the robot. For each child and each moment, we computed whether the child was within a distance threshold (D_{R_TH}) from the robot and defined this feature as:

$$Spent\,time\,around\,the\,robot(i) = \frac{1}{Tracked(i)} \sum_{t=t_0}^{t_n} (if(d(p(i,t), p(r,t)) < D_{R\_TH}) \cdot \Delta t), \quad (6)$$

where $p(r,t)$ is the robot's x-y position. In the calculation of the robot-related features, based on proxemics knowledge [33], we used multiple thresholds for D_{R_TH}: 500 mm as the intimate space for friendly people, 1200 mm as the personal space for familiar people, and 3500 mm as the social space for acquaintances.

Number of surrounding people when around the robot: Since children with high social status interact with the robot and their friends or people who like them, we measured the average number of surrounding people who lingered around the robot. For each child and each moment, we computed the number of other people within a distance threshold (D_{R_TH}) from him/her and defined this feature as:

$$Num.\,of\,surrounding\,people\,when\,being\,around\,the\,robot(i) = \frac{\sum_{t=t_0}^{t_n} \sum_{j=\forall j} (aroundRobot(i,j,t) \cdot if(d(p(i,t), p(j,t)) < D_{R\_TH}))}{Tracked(i)} \quad (7)$$

$$aroundRobot(i,j,t) = nearRobot(i,t) \cdot nearRobot(j,t) \quad (8)$$

$$nearRobot(i,t) = \begin{cases} 1 & if(d(p(i,t), p(r,t)) < S_{TH}) \\ 0 & otherwise. \end{cases} \quad (9)$$

We set S_{TH} as the threshold distance to 3500 mm, which represents the social space from the robot.

5.3 Classification System

We used an SVM with a Radial Basis Function (RBF) kernel to classify social status as high or low. We trained the classifiers using the data obtained in the data collection (Section 6) and sought the best features from all the combinations of features and parameter C for SVM by a grid search using 10-fold cross validations. For implementation, we used the scikit-learn library [34].

6 DATA COLLECTION

6.1 Environments

Data collection was conducted in an elementary school's science room (Fig. 3) that is used for lectures about twice a week per class. Four to five children sit around each desk (six desks from the front were used). The lessons lasted 45 minutes, followed by a five to twenty minute free period.

6.2 Participants

The science rooms we observed were comprised of three classes of 5th graders with 84 students (14 girls and 14 boys in each class). Their average height was 147.2 cm (S.D. was 9.4). The experimental protocol was approved by our institutional review board (reference number 13-502-8) and the school administrators. All the children and their parents signed consent forms and agreed to have their behavior recorded.

6.3 Sociometry Questionnaires

We distributed sociometry questionnaires before the study and asked the children to list five friends. From their answers, we computed the index of sociometric status score (ISSS) as the social status of each child in each class with the following previous definition [35]:

$$ISSS = \frac{1}{2}\left(\frac{N_{nominated}}{N_p - 1} + \frac{N_{mutual}}{N_{max}}\right), \quad (10)$$

where $N_{nominated}$ is the number of nominations (i.e., other children listed this child as a friend), N_p is the number of children in the class, N_{mutual} is the number of mutual nominations (i.e., other children who listed this child, and those listed by this child), and N_{max} is the maximum number of nominations (i.e., five in our study). If this value is high, the child has high social status, i.e., the selection of low/high social status contains no subjective input from teachers or peers.

6.4 Procedure

Each class had six lessons during the study. The room remained available before and after the science lessons. Among the six lessons, five free-time sessions included the robot, which was available only during free-times before/after lessons and did not engage in interaction during lessons. A lesson usually included two parts: lecture and group-work. During lectures, the teacher usually spoke in the front of the class while students sat and listened quietly. In the group-work part, students formed groups based on their seat locations and conducted an experiment or worked with instruments. For instance, one time they changed the weight and the initial angle of a pendulum to study its characteristics.

(a) Lecture (b) Group-work (c) Free-time

Fig. 8 Children's behavior across different phases

6.5 Robot

We used a humanoid robot designed for human interaction to elicit anthropomorphic expectations. It has two arms (each with 4 degrees of freedom (DOF)), a head (3 DOFs), and is 120 cm tall. It has cameras and a speaker on its head with a Pioneer 3DX mobile base.

The robot interacted with children by quiz-style conversations about recent lessons. It started the interactions by greeting them by name and then asked a multiple choice question that was related to their science lessons: *What does a fetus receive through its umbilical cord? Please choose two answers: 1) oxygen, 2) blood, 3) nutrition, or 4) water.* In addition, it answered science questions related to recent lectures. The contents of the 40 quizzes were prepared from lecture materials. Through quizzes and answers to science questions, the robot was designed to help children review recent lessons and deepen their comprehension.

The robot, which generally operated autonomously, identified the locations of the children from the people-tracking infrastructure (Section 4.1) and oriented its gaze direction to each child individually while it spoke. It identified children by its own camera, which enabled it to refer to the children by name. However, since speech recognition remains too difficult in such noisy environments, an operator assumed that function's control. When a relevant question to the lecture topics was asked, the operator controlled the robot to provide information. When typical questions were asked, he selected from pre-implemented behaviors. Otherwise, he directly typed utterances to answer the questions.

Further analysis of a robot's effect on children's learning was previously reported [17]. Some of the children who actively interacted successfully learned through interaction with the robot.

6.6 Obtained Dataset

Behavioral Data

The observed children's behaviors were quite different across the three phases: lecture, group-work, and free-time. During lectures (Fig. 8(a)), the children sat at their desks, listened to the teachers, and stayed still. Since the lecture phase provided little information, we did not use any data from it.

During group-work (Fig. 8(b)), the children were usually at their desks, engaging in experiments. Some worked alone; others worked together. Some children visited other desks to check the progress of different groups or to ask questions.

During free-time (Fig. 8(c)), children were often with their friends. Before the science lessons, many were talking with friends, but some just sat at their desks and waited for the lessons to start. After the science lessons, some children gathered around the robot and interacted with it, while others chatted with friends or returned to their homerooms.

We separated the data based on the above definitions of the phases and used both the group-work and free-time phases for our estimations. We obtained 235 minutes of group-work and 112 minutes of free-time data.

Questionnaire Data

We got 70 valid data samples; some children were absent when the questionnaires were given. We categorized their social status as *high* or *low* using the obtained ISSS scores and used the average scores as the cutoff point. The *low* class included children whose scores were below average, and the *high* class included children whose scores were above it. 34 children were categorized in the *low* class and 36 in the *high* class.

7 EVALUATION

7.1 Evaluation of Observation System

We evaluated the correct tracking and identification ratios of the children during the lessons. A human coder watched five minutes of video of the group-work phase result, identified the children, and checked whether each child had been monitored by the tracking system and whether his/her id was correctly associated.

The tracking system tracked children 93.3% of the time (on average, 279.8 sec out of 300). Each child was correctly identified 65.5% of the time (196.5 sec on average). A wrong ID was associated 4.9% of the time (14.8 sec). For the remaining 22.9%, no ID was associated (68.6 sec).

We consider this result good, since the classroom situation is very complex. For instance, Fig. 9 (left) shows scenes where the tracking system failed to track some of the children who sometimes gathered around a desk to observe materials for the group-work experiment. In such situations, they were huddled too closely together to be separated, and some were occluded by others. Note that once tracking is lost, even though the system recovers from the error, the identification remains lost until the face is seen again. This is the main source of the gap between the time being tracked (93.3%) and identified (65.5%).

Fig. 9 Difficult tracking and identification moments

Table 1 Performance of SVM classifiers

Features	Performance
Only-with-robot	61.4%
Without-robot	61.4%
Proposed method	71.4%

Figure 9 (right) shows another case where the tracking system failed. Here it recognized the faces of two boys, but their bodies were so close that it only identified one of them. In such moments, tracking was unstable, and identifications were likely to fail. Nevertheless, even though such situations often occurred, our system was able to track and identify them for our purpose: estimating social status.

7.2 Evaluation of Estimation System

Since we expect that the robot's presence will add estimation clues, we evaluated the contribution of the *classroom behavior features* (Section 5.1) and the *robot-related features* (Section 5.2) as well as the performances of the following three conditions:

Without-robot: The only feature vector for SVM was the *classroom behavior features* (Section 5.1) that estimate how the system worked without a robot in the classroom.

Only-with-robot: The only feature vector for SVM was the *robot-related features* (Section 5.2) that estimate how the system worked if the estimation was done only with the information observed from the robot (i.e., only the children's behavior around it).

Both (proposed): All the features reported in Sections 5.1 and 5.2 were used. This is our proposed system.

For all the conditions, except for the above features, we used the system reported in Section 5 with the same training and tuning procedures reported in Section 5.3.

Table 1 shows the result. Our proposed system, which successfully estimated children's social status with 71.4% accuracy, outperformed the other two conditions. The performances of the without-robot and only-with-robot methods were both 61.4% and 61.4%. We conducted a paired t-test to compare the performances of the cross validation results among the proposed method and each alternative method.

Table 2 Confusion matrix

		(a) Only-with-robot: Estimation high	(a) Only-with-robot: Estimation low	(b) Without-robot: Estimation high	(b) Without-robot: Estimation low	(c) Proposed method: Estimation high	(c) Proposed method: Estimation low
Ground-truth	high	28	8	23	13	29	7
	low	19	15	14	20	13	21

Significant differences were shown between the proposed and only-with-robot methods (p=.046) as well as the proposed and without-robot methods (p=.001).

Table 2 shows a confusion matrix whose details we further analyzed. If we compare the *proposed* and *only-with-robot* conditions, the main difference is the estimation of children with low social status, whose result resembled the result of children with high social status. That is, our proposed method more successfully estimated children with low social status than the *only-with-robot* condition. In fact, many of them (with whom the proposed model was successful but not the *only-with-robot*) were children who were alone during free-time relatively far from the robot.

If we compare the *proposed* and *without-robot* conditions, the main difference is their estimation of children with high social status; the result resembled the result of the children with low social status. In fact, many children with high social status, with whom the proposed model was successful but not the *without-robot* model, were children who frequently interacted with the robot. Because the proposed method used information about the children's behavior around the robot, it outperformed the *without-robot* condition, showing that interaction with the robot provided an additional clue for estimation.

7.3 Analysis of Contributing Features

We also investigated which features were important for the estimation by eliminating each one by one and checking the performances. We identified high contributions from *time spent alone* during free-time and *time spent around the robot*. When we removed them, the estimation performance decreased by more than 10%. We investigated whether children with low social status spent more time alone (avg. value from Eq. 1, high: 0.642, S.D. 0.141, low: 0.718, S.D. 0.144) and less time around the robot (avg. value from Eq. 9, high: 0.014, S.D. 0.030, low: 0.007, S.D. 0.014) than children with high social status. A t-test showed a significant main effect in *time spent alone* (t(1,68)=2.244,p=.028), but it did not show a significant main effect in *time spent around the robot* (t(1,68)=1.166, p=.247). Children with high social status tended to be with friends or others, and such observations as togetherness (or aloneness) were keys for successful estimation. On the other hand, we need to carefully interpret these results. Even though a child is alone, she hasn't necessarily been socially rejected; these results only identified that such tendencies of children's behaviors are related to social status. In the following subsections, we retrieved scenes based on these contributing features.

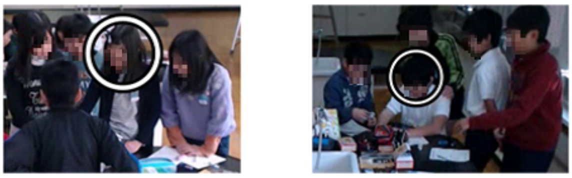

Fig. 10 High social status children spent free-time with others

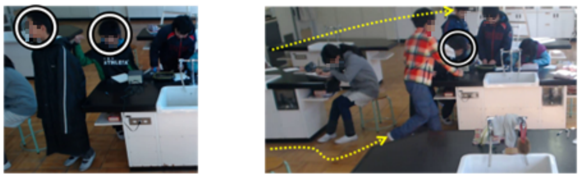

Fig. 11 High social status child formed a group

Fig. 12 Low social status child alone and watching others play

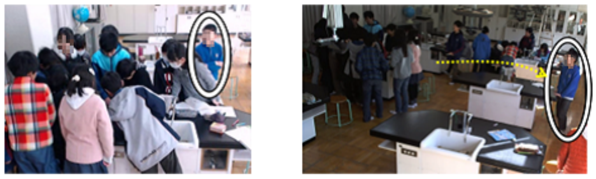

Fig. 13 Low social status child looking at a group from a distance

8 OBSERVATIONS

The evaluation in Section 7.2 revealed that both the *in-class behavior* and the *behavior around the robot* were important information for estimation. Here we further scrutinize the relevant behaviors of some of the children.

8.1 In-class Behavior Related to Social Status

Figure 10 shows children with high social status during free-time who often spent time with friends. After the end of each lecture, they gathered and talked. Fig. 11 shows another child with high social status. During his free-time, he invited his friends to his desk to play, where many gathered.

In contrast, children with low social status were less frequently with others and sometimes were alone. Fig. 1 shows

(a) t = 0 (sec) (b) t = 545 (sec)

Fig. 14 Children's behavior across different phases

(a) 2^{nd} lecture (b) 4^{th} lecture

Fig. 15 High social status child repeatedly interacted with robot

Fig. 16 Low social status child who seemingly longs to join a group around robot

Fig. 17 Low social status child who gave up joining a group

scenes where children with low social status were isolated and returning back to their seats alone, while other children had already left with friends or were talking with others. Fig. 12 shows another child with low social status. During free-time while other children were playing together, he stayed at his desk alone and watched them. Fig. 13 shows another child with low social status. Once during free-time, he approached a group of children who were playing together without joining them and stayed alone at a distance.

8.2 Behavior around the Robot

Children with high social status spent much more time around the robot than children with low social status. Fig. 14 shows a child with high social status who is interacting with the robot. In this situation, she and her friend interacted with it for 545 seconds. Fig. 15 shows another child with high social status. Every day she approached the robot and repeatedly interacted with it with her friends.

In contrast, children with low social status spent less time with the robot. Fig. 16 shows a child with low social status who seemed to want to interact with other children and the robot, but he only moved around a group of people. Fig. 17 also shows a child with low social status who was unable to participate. He looked at the group, approached it (Fig. 17, left), but gave up (Fig. 17, right).

9 DISCUSSION

9.1 Contribution of Observation System

Since most previous studies relied on human observations, children's classroom behavior has not been scrutinized. One reason is the difficulty of observing it in detail by sensors. Children often move around and stay very close to each other. Our study faced this difficulty using a sensor network that combined depth and RGB cameras. Our system tracked children 93.3% of the time and identified them 65.5% of the time. Even if room exists for better performance, our study identified one important step for understanding children in a classroom with sensor-based observation.

Perhaps the information from a system with only 65.5% accuracy is too noisy for estimation. However, note that 65.5% accuracy does not necessarily imply that the remaining 34.5% of the information was inaccurate. Only 4.9% of such identification was inaccurate, and situations where the target person was simply not identified comprised the remaining 29.6%, meaning that information is missing for the majority of these moments. If our observations continued for a long enough period, the loss of part of the observations would not be a big problem for the estimations. When a child was correctly identified, the features (Section 5) were correctly computed if other people were correctly tracked (with 93.3% accuracy), even without being identified. We just need such information as the number of other people. Nevertheless, we believe that our estimation is based on quite accurate features.

9.2 Contribution of Estimation of Social Status

Our method estimated social status with 71.4% accuracy. Given that the chance rate is 50%, perhaps this result is not very good. However, we believe that our achievement is solid because the estimation of social status is not simple. To the best of our knowledge, since no previous work has automatically estimated social status, we are unable to provide a simple comparison of performances. As a reference, looking at similar problems related to the estimations of human attributions might be fruitful. Although neither in a classroom nor about social status, some studies addressed the estimation of personality[1], where 2-class classifiers resulted in estimation results of personality with around 70% accuracy. For example, Mohammadi et al. developed a method that resulted in 60-72% accuracy [31]. Since observing classroom behavior is not easy, around 70% accuracy is a solid achievement for our initial challenge of the estimation of social status.

The comparison in Section 7 revealed that the observation of the children interactions with the robot provided useful clues for estimation. We found that the active interaction caused by the social robot increased its understanding of the children's social status. As far as observing a couple of early repeated interactions, children with high social status gathered around the robot with friends, while children with low social status less frequently participated in interactions. Note that we need to carefully consider how to treat observations from longer term interactions because the robot's novelty eventually wears off and the interactions of children with high/low social status will probably change over time.

If we conducted social status estimations by separating the data for each of the three classes, the performances became 65.7%, 63.7%, and 61.3%. This performance is relatively low compared to using all of the class data. We thought that separating the data for each class might allow us to use just a relatively small amount of data and lower the estimation performance.

9.3 Implications

This study describes a promising way of estimating children's social status. Thus, we should be able to develop a robot that supports children's activities. For example, since children with low social status are often alone, a social robot might encourage them to interact with others. We observed that when a social robot called a child by name, other children made space for him/her to join and they interacted together. Even with 71.4% estimation accuracy, a robot could encourage children who are estimated to have low social status to interact with it. With this approach, even though estimation failure does not cause serious negative results, the robot could encourage children with low social status to engage with other children.

[1] Personality and social status are clearly different constructs (concepts) in psychology. Social status is closely connected to social preference (i.e., whether everyone likes him/her), and hence it tends to be the consequence of individual capability, such as social competence [36] or the tendency to engage in aggressive behavior [6]. For example, a child with an extroverted personality is not necessarily popular and might be perceived as annoying or aggressive if he behaves badly; a shy introverted child might be liked if she is socially competent, e.g., helpful and cooperative.

9.4 Alternative Approaches

In this section, we compare the advantages and disadvantages of alternative approaches to estimate social status and tracking and identifying children in classrooms with our proposed method.

The first topic is an alternative approach to estimate social status. One possible method is to rate friendships and compute social status from them with Eq. 10, because we can accurately estimate friendships by observing proximity events (who is often with whom) [20, 21], when children's behaviors are accurately observed with observation systems. However, the accurate estimation of friendships or nominations (who likes whom) is not easy. For instance, we implemented a method [21] that identified proximity events as the moment when two children were within a certain distance, which we set to 2000 mm because it yielded the best performance. However, only 14.7% of the nominations were correctly identified, and 11.5% false-positive nominations were estimated. Because the estimated nominations were inaccurate, the estimated social status with Eq. 10 poorly matched the ground-truth. The 2-class classification accuracy was only 40.0%, which is also poor and below the chance rate, possibly because the information from the proximity events is rather incorrect in our system. When our system fails to identify a child's friend during proximity events, it mistakenly assumes that he/she is not with this friend.

The second topic is an alternative sensing approach. In this study we used an environmental sensor to estimate social status instead of such wearable devices as Bluetooth bracelets due to several technical requirements. Here we discuss the possibility of using wearable devices for this purpose. In the context of academic trials, even though asking children to wear devices might be allowed in some circumstances, such permission is unusual and often denied. Moreover, asking children to always wear such bracelets requires a significant amount of human resources because this is contrary to what children naturally do. Actually we did experience some difficulties in our school environment through discussions with the school principal. He (and his teachers) worried that the preparation time for such devices might erode into class time. During long-term experiments, the load of wearing/removing such devices during every class would obviously increase. Also, since children might forget to carry or wear them, opportunities to gather whole data would decrease in the experimental period. Another concern is that using so many wireless devices might deleteriously influence a school's Wi-Fi environments, which is also related to the management's view of the elementary school.

Furthermore, a sensor array approach provides more useful information than Bluetooth-based wearable sensors: the accuracy of positioning children. A wearable device enables estimates of relatively rough position relationships among children, but the accuracy of wireless-based positioning falls, especially in complex and crowded classroom environments. Unstable wireless signals might cause misunderstanding about estimating the relative position relationships among children. Also, it is difficult to estimate the absolute positions of many users accurately with such wearable sensors under about 10-m scale environments.

9.5 Limitations

Several limitations exist in our method that estimates the social status of children by observing their activities in classrooms. For estimations, we used SVM because it is a leading algorithm for classification problems and performs reasonably well. Using different classification methods might provide different performances. But in this study, we focused on whether the social status of children can be estimated through autonomously extracted features and combined the children's activities in both group-work/free-time to improve the performances rather than comparing the performance with different classification methods.

Even though we did not use any manually annotated identification data to estimate social status, such an approach might be useful when the number of features is limited. However, testing all combinations manually with many features would be difficult. Moreover, the system needs to automatically recognize informative features; if they are only recognized by human capabilities, it would be difficult to use the system. This approach might have difficulty identifying what features/combinations are best for this purpose, but building a system with current technologies would be appropriate.

We estimated social status with ISSS, which is a common metric in Japan to investigate social status in children-related researches [37–39]. To the best of our knowledge, ISSS is used more often than subjective reports from teachers, partly because reports have concluded that it provides more benefits. For example, it identified relationships among children who are shy and/or introverted [37]. Even if ISSS is mainly used in Japan and focuses on classroom situations, its definitions are reasonable to represent social status in a classroom because its simple calculation is based on the number of listed friends. In fact, observations of children behaviors indicated that children with high social status are more popular in classrooms.

In this study, since only the valid children data in the same grade were used, we could not add the data of more children from the same grade at the same school. We believe that the data of 70 children with around six hours for a binary classification problem (not multiple/complex classification) is adequate. Moreover, we conducted cross validation in the evaluation process to appropriately evaluate the performance of our developed model, even with a relatively small amount of data. However, additional data and analysis are needed to generalize the proposed method's validity.

We used proxemics knowledge to extract features such as personal distance, even though a previous study reported that children define and use personal space differently than adults [40]. On the other hand, several studies reported that the distances between children and robots are consistent with the social distance chosen by adults [41–43]. Therefore, we investigated the effects of children's personal space by changing the threshold (D_{TH}) for Eqs. (2) and (3). Since the children's ages were 11 to 12 in this study, we employed 110 cm as the interaction distance based on [40]. (about 90% from 120 cm). After recalculation of the performance, we found that the modified interaction distance did not change the performance.

Moreover, as an additional test, we employed 40 cm as an intimate distance; we note that the paper only investigated the difference of interaction distance with children, and therefore we used the same ratio for the modified intimate distance, which is 90% of 45 cm, based on the original intimate distance definition. But after recalculating the performance, we found that the modified intimate distance just slightly decreased the performance (68.3%). The modified intimate distance is not based on the collected data from a previous paper [40], as described above. One possible explanation is that in this grade the intimate distance between adults and children is similar. Since the accuracy of our human tracking system was about 30 cm, the recalculated performance was influenced by these factors. Thus, we must carefully interpret the achieved performance of this study, which is based on the proxemics knowledge of adults, but we believe that our proposed method's achieved performance has adequate impact even if we only use the proxemics knowledge of adults.

9.6 Future work

Several future directions exist for this study. If tracking and identification were improved, the system could provide more useful features for estimation. For this purpose, combinations of environmental and wearable sensors are promising, if we use wearable sensors in real environments. In fact, past research showed that integrating these different kinds of sensors achieved both precise person identification and tracking [44]. In this study, we focused on estimating social status. But we could of course use the estimation results in many other ways: providing such information to teachers and school counselors to help them identify potential problems in classrooms.

Moreover, the robot can be utilized to obtain more information for the estimation of social status in two ways: its sensors and interaction activities. For the former, the robot can integrate its own sensing data to the whole system. For instance, person identification with the robot's camera and the estimated distance relationship between robots and people using distance sensors such as laser range finders would improve the robust estimation of social status.

For the latter, the robot can change position relationships by moving around in the environment and actively interacting with children. Related to the above approach, changing positions would be helpful to identify children because the robot can move to an appropriate place based on the sensing areas of the environmental sensors. Conversation between children, in particular about their friends, would also be helpful to estimate social status, but ethical issues and a cautious conversation design must be considered for this approach.

10 CONCLUSION

We developed a social status estimation system that consists of a people-tracking system using depth sensors and a person-identification system using RGB cameras. The system extracts features from children's behaviors during their lessons and free-time and estimates their social status by SVM classifiers. We gathered children's behaviors at a science room in an elementary school. Our system tracked them in a classroom almost all of the time (93.3%) and correctly identified the tracked children 65.5% of the time. Our system estimated social status with 71.4% accuracy through tracking and identification.

11 ACKNOWLEDGMENTS

We thank the staff and the students of the elementary school for their support. This work was in part supported by JSPS KAKENHI Grant Numbers JP25240042, JP25280095 and JP15H05322.

References

1. Kanda T., Hirano T., Eaton D. and Ishiguro H. (2004) Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial. Human-Computer Interaction, vol. 19, pp. 61-84.
2. Han J., Jo M., Park S. and Kim S. (2005) The Educational Use of Home Robots for Children. IEEE Int. Workshop on Robot and Human Interactive Communication, pp. 378-383.
3. Saerbeck M., Schut T., Bartneck C. and Janse M. D. (2010) Expressive Robots in Education: Varying the Degree of Social Supportive Behavior of a Robotic Tutor. ACM Conf. on Human Factors in Computing Systems (CHI2010), pp. 1613-1622.
4. Howley I., Kanda T., Hayashi K. and Rosé C. (2014) Effects of Social Presence and Social Role on Help-Seeking and Learning. ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2014), pp. 415-422.
5. Hollingshead A. B. (1975) Four factor index of social status. (*It was initially unpublished work but extremely well cited. It was later published as: A. B. Hollingshead (2011) Four factor index of social status. Yale Journal of Sociology, vol. 8, pp. 21-52.)
6. Coie J. D., Dodge K. A. and Kupersmidt J. B. (1990) Peer Group Behavior and Social Status, in Peer Rejection in Childhood. S. R. Asher and J. D. Coie eds., Cambridge Univ. Press, pp. 17-59.
7. Salmivalli C., Lagerspetz K., Björkqvist K., Österman K. and Kaukiainen A. (1996) Bullying as a Group Process: Participant Roles and Their Relations to Social Status within the Group. Aggressive Behavior, vol. 22, pp. 1-15.
8. Crothers L. M. and Kolbert J. B. (2008) Tackling a Problematic Behavior Management Issue Teachers' Intervention in Childhood Bullying Problems. Intervention in School and Clinic, vol. 43, pp. 132-139.
9. Rodkin P. C. and Berger C. (2008) Who Bullies Whom? Social Status Asymmetries by Victim Gender. Int. Journal of Behavioral Development, vol. 32, pp. 473-485.
10. O'Neil R., Welsh M., Parke R. D., Wang S. and Strand C. (1997) A longitudinal assessment of the academic correlates of early peer acceptance and rejection. Journal of Clinical Child Psychology, vol. 26, pp. 290-303.
11. Van Laar C. and Sidanius J. (2001) Social status and the academic achievement gap: A social dominance perspective. Social Psychology of Education, vol. 4, pp. 235-258.
12. Coie J. D. and Krehbiel G. (1984) Effects of academic tutoring on the social status of low-achieving, socially rejected children. Child Development, pp. 1465-1478.
13. Woods S., Davis M., Dautenhahn K. and Schulz J. (2005) Can Robots Be Used as a Vehicle for the Projection of Socially Sensitive Issues? Exploring Children's Attitudes Towards Robots through Stories. IEEE Int. Workshop on Robot and Human Interactive Communication, pp. 384-389.
14. Bethel C. L., Eakin D., Anreddy S., Stuart J. K. and Carruth D. (2013) Eyewitnesses Are Misled by Human but Not Robot Interviewers. ACM/IEEE Int. Conf. on Human-Robot Interaction, pp. 25-32.
15. Tanaka F., Cicourel A. and Movellan J. R. (2007) Socialization between Toddlers and Robots at an Early Childhood Education Center. Proc. of the National Academy of Sciences of the USA (PNAS), pp. 17954-17958.
16. Belpaeme T. et al. (2012) Multimodal Child-Robot Interaction: Building Social Bonds. Journal of Human-Robot Interaction, pp. 33-53.
17. Komatsubara T., Shiomi M., Kanda T., Ishiguro H. and Hagita N. (2014) Can a social robot help children's understanding of science in classrooms? Proc. of Int. Conf. on Human-Agent Interaction, pp. 83-90.
18. Komatsubara, T., Shiomi, M., Kanda, T., and Ishiguro, H. (2017). Can Using Pointing Gestures Encourage Children to Ask Questions?. International Journal of Social Robotics, 1-13.
19. Shiomi M., Kanda T., Howley I., Hayashi K., and Hagita N. (2015) Can a Social Robot Stimulate Science Curiosity in Classrooms? International Journal of Social Robotics, Volume 7, Issue 5, pp. 641-652.
20. Choudhury T. and Pentland A. (2003) Sensing and Modeling Human Networks Using the Sociometer. IEEE Int. Symp. on Wearable Computers, pp. 216-222.
21. Kanda T. and Ishiguro H. (2006) An Approach for a Social Robot to Understand Human Relationships: Friendship Estimation through Interaction with Robots. Interaction Studies, vol. 7, pp. 369-403.
22. Lao S. and Kawade M. (2005) Vision-Based Face Understanding Technologies and Their Applications. Advances in Biometric Person Authentication, pp. 339-348.
23. Kirchner N., Alempijevic A. and Virgona A. (2012) Head-to-shoulder signature for person recognition. IEEE Int. Conf. On Robotics and Automation (ICRA), pp. 1226-1231.
24. Huang K. S. and Trivedi M. M. (2003) Distributed Video Arrays for Tracking, Human Identification, and Activity Analysis. IEEE Int. Conf. on Multimedia and Expo vol. 2, pp. 9-12.

25. Aran O. and Gatica-Perez D. (2013) One of a Kind: Inferring Personality Impressions in Meetings. Proc. of the ACM Int. Conf. on Multimodal Interaction, pp. 11-18.
26. Hung H., Jayagopi D., Yeo C., Friedland G., Ba S., Odobez J. M., Ramchandran K., Mirghafori N. and Gatica-Perez D. (2007) Using audio and video features to classify the most dominant person in a group meeting. Proc. of the ACM Int. Conf. on Multimedia, pp. 835-838.
27. Gottman J., Gonso J. and Rasmussen B. (1975) Social Interaction, Social Competence, and Friendship in Children. Child Development, Vol. 46, pp. 709-718.
28. Cummins D. (2005) Dominance, status, and social hierarchies. The handbook of evolutionary psychology, pp. 676-697.
29. Olguín D. O., Gloor P. A. and Pentland A. (2009) Wearable Sensors for Pervasive Healthcare Management. Int. Conf. on Pervasive Computing Technologies for Healthcare, pp. 1-4.
30. Kalimeri K. (2013) Towards a Dynamic View of Personality: Multimodal Classification of Personality States in Everyday Situations. Proc. of the ACM on Int. Conf. on Multimodal Interaction, pp. 325-328.
31. Mohammadi G. and Vinciarelli A. (2012) Automatic Personality Perception: Prediction of Trait Attribution Based on Prosodic Features. IEEE Trans. on Affective Computing, vol. 3, pp. 273-284.
32. Brscic D., Kanda T., Ikeda T. and Miyashita T. (2013) Person Tracking in Large Public Spaces Using 3d Range Sensors. IEEE Trans. on Human-Machine Systems, vol. 43, pp. 522-534.
33. E. T. Hall, (1966) The hidden dimension, Anchor Books.
34. Pedregosa F. et al. (2011) Scikit-learn: Machine Learning in Python. JMLR 12, pp. 2825-2830.
35. Tanaka K. (1979) Theory and Methods of Sociometry. Meiji Press (in Japanese).
36. Inderbitzen-Pisaruk H. and Foster S. L. (1990) Adolescent friendships and peer acceptance: Implications for social skills training. Clinical Psychology Review, vol. 10, pp. 425-439.
37. Shoji Sato, et al., "Interpersonal Relationship, Social Adjustment and Preventive Intervention for Children," The Annual Report of Educational Psychology in Japan, 38, pp. 51-63, 1999.
38. Mino, Yoshio, et al. "Sociometric status of children with attention deficit disorder." Nippon Eiseigaku Zasshi (Japanese Journal of Hygiene) 42.5 (1987): 913-921
39. Kazuaki Sugihara, et al., "A comparison between perpetrators and victims of ijime on sociometric status and personality traits," Tsukuba Psychological research, 8, pp. 63-72, 1986.
40. Aiello, John R., and Tyra De Carlo Aiello. "The development of personal space: Proxemic behavior of children 6 through 16." Human ecology 2.3 (1974): 177-189.
41. Yuto Yamaji, Taisuke Miyake, Yuta Yoshiike, P. Ravindra S. Silva, and Michio Okada. STB: Child-Dependent Sociable Trash Box. International Journal of Social Robotics, 2011; 3 (4): 359 DOI: 10.1007/s12369-011-0114-y
42. Vázquez, Marynel, et al. "Spatial and other social engagement cues in a child-robot interaction: Effects of a sidekick." Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction. ACM, 2014.
43. Walters, Michael L., et al. "Close encounters: Spatial distances between people and a robot of mechanistic appearance." Humanoid Robots, 2005, 5th IEEE-RAS International Conference on. IEEE, 2005.
44. Shiomi M., Kurumizawa K., Kanda T., Ishiguro H., and Hagita N. (2014). Finding a Person with a Wi-Fi Device in a Crowd of Pedestrians. Advanced Robotics, Volume 28, Issue 7, pp. 441-448.

Tsuyoshi Komatsubara received M.Eng. degree from Osaka University, Osaka, Japan, in 2016. He was an internship student at the Intelligent Robotics and Communication Laboratories (IRC), Advanced Telecommunications Research Institute International (ATR). His research interests included human-robot interaction and educational robots.

Masahiro Shiomi received Ph.D. degrees in engineering from Osaka University, Osaka, Japan, in 2007. He is currently a Group Leader with the ATR Intelligent Robotics and Communication Laboratories. His research interests include human-robot and human-agent interaction.

Thomas Kaczmarek is a research engineer at ATR Intelligent Robotics and Communication Laboratories since he graduated from Telecom SudParis in France in 2012 with a major in High-Tech Imaging. He develops sensor systems, perception algorithms, and other systems used for human-robot interaction research.

Takayuki Kanda received the B.Eng., M. Eng., and Ph.D. degrees in computer science from Kyoto University, Kyoto, Japan, in 1998, 2000, and 2003, respectively. He is currently a Group Leader with the ATR Intelligent Robotics and Communication Laboratories. His research interests include intelligent robotics and, human-robot interaction.

Hiroshi Ishiguro received a D. Eng. in systems engineering from the Osaka University, Japan in 1991. He is currently Professor of Department of Systems Innovation in the Graduate School of Engineering Science at Osaka University (2009-) and Distinguished Professor of Osaka University (2017-). He is also visiting Director (2014-) (group leader: 2002-2013) of Hiroshi Ishiguro Laboratories at the Advanced Telecommunications Research Institute and an ATR fellow. His research interests include sensor networks, interactive robotics, and android science. He received the Osaka Cultural Award in 2011. In 2015, he received the Prize for Science and Technology (Research Category) by the Minister of Education, Culture, Sports, Science and Technology (MEXT).