IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 1, 2018, accepted December 24, 2018, date of publication January 1, 2019, date of current version January 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2890438

Object Recognition-Based Second Language Learning Educational Robot System for Chinese Preschool Children

QIN WU[1,2], SIRUI WANG[1], JIASHUO CAO[1], BING HE[2], CHENMEI YU[1], AND JIANBO ZHENG[2,3]

[1]School of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China
[2]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[3]Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong

Corresponding author: Jianbo Zheng (jb.zheng@siat.ac.cn)

This work was supported in part by the Shenzhen–Hong Kong Innovative Project under Grant SGLH20161212140718841, in part by the Shenzhen Engineering Laboratory for 3D Content Generating Technologies under Grant [2017]476, in part by the Guangdong Technology Project under Grant 2016B010108010, Grant 2016B010125003, and Grant 2017B010110007, and in part by the National Basic Research Program of China (973 Program) under Grant 2014CB744600.

ABSTRACT Research in psychology, pedagogy, and physiology has shown that learning a second language will be a great benefit to preschool children. Compared with adults, children have more advantages in learning, such as better pronunciation and intonation. However, because of children's inattention, the effects of the second language education in many preschools are not ideal in China. To make children learn the second language effectively, it should be integrated into their daily lives and in line with their interests. In this paper, an educational robot system with object recognition technology is introduced, which aims to provide innovative second language learning services for preschool children in China. The proposed system combines object recognition and projection with English teaching and makes objects in daily life more interesting with projected animation to attract children's attention. On the projection screen, children can interact with a robot by touch or movement and can easily trigger more interactive effects. To evaluate the effectiveness of the proposed system, we conducted an experiment, and the results showed that the system can improve the language learning efficiency for the Chinese preschool children by interacting with objects under proper guidance.

INDEX TERMS Educational robot, intelligent robot, object recognition, preschool education, second language learning.

I. INTRODUCTION

Children are developing and changing very quickly [1]. Although they have lower levels of purpose, attention and comprehension in language learning than teenagers and adults, they have the basic capabilities for learning languages, and are even better than teenagers and adults in voice imitation and voice identification [2]. Therefore, second language education for preschool children is very important and should not be ignored. In recent years, more new technologies have been applied in the field of second language education, such as multimedia English learning through mobile devices [22], networks of remote teaching and built-in voice function of intelligent education robots [3]. However, most of these methods of second language learning are for teenagers and adults. For many preschool children, especially in China, the only way to learn English is in the English classroom. They are not exposed to the new technologies used in language education [4]. This is mainly because preschool children have various learning styles, and there is a huge gap between the attitudes of children in different ages towards to the same learning style.

Therefore, several effective education approaches that work for teens and adults cannot directly apply to preschool education [5]. The discovery of an effective language teaching method welcomed by preschool children is a widespread concern.

In recent years, the market of English education for child has grown rapidly in China. Many parents have recognized

the importance of English education to their children. Many English courses or institutions for children have emerged, and bilingual teaching programs have been introduced into some kindergartens [6]. Some well-educated parents choose to teach their children English themselves, while some parents cannot because of the lack of time or English skills. In the latter situation, they usually let the children learn English through their mobile devices by playing English programs, such as teaching video and English animation, or by learning words through apps. The major shortcoming of learning English on mobile devices is that most real entities are taught in 2D images or some animation effects. Considering children are still under cognition development, the mapping relationship between pictures and real items has not been completely formed in their minds; hence, they cannot link the learned English words to the corresponding real entities. Therefore, even if the English words have been taught, children still do not know their real meanings and how to use them in daily life; this results in a low learning efficiency, and the children gradually lose interests in English learning. Guiding children in object cognition during English teaching does not only stimulate their interests in learning, but also leads to a higher learning efficiency [24], [26].

In this paper, we propose a projection robot system with object recognition technology for language education to provide innovative language learning services for preschool children. The proposed robot system combines object recognition and projection with English teaching, and makes objects in daily life more interesting with projected animation to attract children's attention. On the projection screen, children can interact with the system by touch or movement, which easily triggers more interactive effects. The system consists of three main components:

1. Projector. It casts images and objects on the flat surface with which the child interacts.

2. Kinect. It takes pictures of objects on a fixed area to realize object recognition and finger tracking.

3. Main controller. It receives the captured object of the camera and identifies and controls the content of the projector playback.

II. BACKGROUND AND RELATED WORK

Language learning is an indispensable and important focus in educational robotics. Educational robotics is a new research field, spanning less than three decades, and novel ways and modes of interaction for learning are still being explored.

The unique feature about educational robots is that they combine traditional education and computers with robotics. They have physical bodies, share physical space with humans and use human cognition and behavior to communicate with humans in a more natural way [7]. Thus far, much has been achieved in language learning via educational robotics, and the importance of creating new teaching methods remains a highly relevant, active and growing research field in both academia and industry.

Several researchers have conducted studies on children interacting with educational robots in a learning environment and responding [8]. Breazeal *et al.* [9] studied on children treating robots as interlocutors. The children supplied information to the robots and retained what the robots told them. Hall *et al.* [10] studied on map reading with an empathic robot tutor; they showed this scenario with a Nao robot interacting with students reading a map. There are other works on the degree of the social supportive behavior of a robotic tutor [11]. These studies explore the effectiveness of a robot as a tutor or learning partner, and their results show that interaction with educational robots in a learning environment can improve children's activity.

Several research suggests that educational robots have potential as learning companions and tutors for young children in early language education. G. Gordon in his paper discusses the affective personalisation of a social robot tutor for developing a second language skill in children [12]. Tanaka and Matsuzoe [13] constructed a new educational framework to promote spontaneous learning in children by teaching through a robot. Some articles have examined the proposition that robots could form relationships with children, and that children might learn from robots as they learn from other children [14]. These findings prove that a successful vocabulary learning in children depends on their relating to the robots as interactive, social creatures [23], [25].

There is also research on the use of storytelling and reading skills to promote the oral language development and story comprehension ability of preschool children. Participating in storytelling can improve the verbal fluency, listening skills and vocabulary of children. Particularly, reading can be an effective method for expanding the vocabulary of young children, when they are encouraged to actively process the story materials. For example, *Storytelling in Early Childhood* is a captivating book which explores the multiple dimensions of storytelling and story acting and shows how they enrich language and literacy learning in the early years [15]. Kathryne McGrath's study provides a framework for implementing the art of storytelling in individual classrooms [16]. In addition, the effects of storytelling and story reading on the oral language complexity and story comprehension of young children have been studied [17]. These studies prove that conversational reading is an effective way to improve the vocabulary of children.

Some studies explore the past and present status of object recognition technology. The current leading approach for object detection is the region-based convolutional neural network (R-CNN) proposed by Girshick *et al.* [18], in which object detection is decomposed: first, object locations are generated, and then CNN classifiers are used to identify object categories at those locations. Most importantly, Google scientists proposed a deep CNN architecture codenamed Inception, which set the state of the art for classification and detection in the ImageNet Large-scale Visual Recognition Challenge (ILSVRC14), 2014 [19]. These research results all introduce reliable algorithms for object recognition

technology using neural networks [21] based on computer vision.

Some related work is about the interaction between embodied robots and virtual robots for preschool children. They compare the effects of an embodied robot-coacher and its virtual agent on a game-like exercise task performed by pre-schoolers through a controlled experiment. The study suggests that virtual and embodied robotic agent coaches might be active in promoting the engagement preschool children in motor activities [32]. Other studies have looked at using virtual robot-mediated play activities to assess cognitive skills. The study shows that virtual robots are a viable alternative to the use of physical robots for assessing children's cognitive skills, with the potential of overcoming limitations of physical robots [31]. However, it is unclear whether these functions can be applied to preschool children's learning in a second language.

The use of interactive screen media such as smartphones and tablets by young children is increasing rapidly. However, research regarding the impact of this portable and instantly accessible source of screen time on learning, behavior, and family dynamics has lagged considerably behind its rate of adoption [33]. In addition, the display screen of the mobile device is very harmful to the eyes of children. These high-brightness electronic screens can cause diseases such as myopia, amblyopia, and astigmatism in children's eyes.

In this paper, we explore an interactive projection technology based on object recognition for use in Chinese home-schooling robots for preschool children. Although object recognition technology has been an active research area for the past 20 years, its use in interactive systems has been intermittent. This may be due to the key technologies proposed by many research institutes to optimize the accuracy, advancement and stability of massive databases. However, research based on the interaction of object recognition in children's education is still insufficient. To date, almost all research on the use of educational robots as child learning partners use computer systems as an auxiliary tool, rather than a more natural form of interactive engagement. Therefore, the educational robot system designed in this study can fill this vacancy. Our system can provide more exciting and effective interactive interfaces in the field of robotics. In the current work, however, we focus on the human-computer interaction technology, while considering children's preschool education, second language learning and home education.

III. SYSTEM DESIGN

The system architecture is shown in Figure 1. Through the development of hardware and software modules, most of the equipment was assembled into a host that integrates Kinect [27], projectors and system controls. The functions of the Kinect camera are objects identification and motion capture in the system. The model of the projector used is XGIMI H1S [28]. The main function of the projector is to image the interactive process. The system control part consists of a computer, into which the developed software module is installed. In addition, the projector can be erected vertically on its stand, and its height and screen size can be adjusted according to the needs of children of different heights.

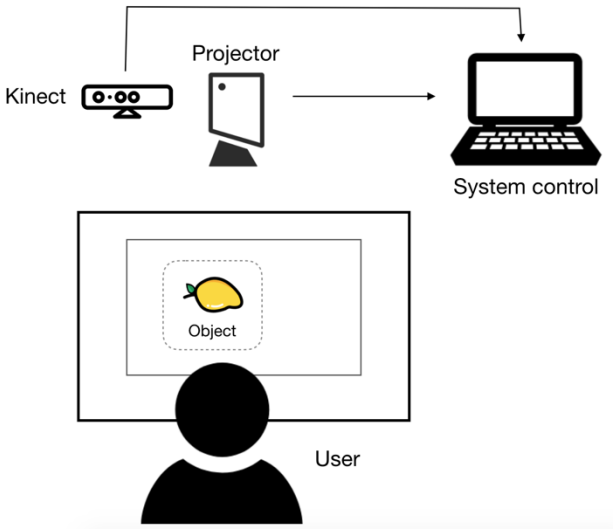


FIGURE 1. System architecture diagram.

A. DYNAMIC CAPTURE

The system dynamic capture device uses Kinect 2 for Windows. The system can use Kinect 2 to identify items, capture user operation gestures and perform picture control. The resolution of the TOF camera in the system is 512×424, which can capture the depth information in the scene, while the system uses a 1920×1080 high-definition camera capture screen. The depth information is processed using the Kinect for Windows SDK kit. The Kinect-related algorithm is also optimized in the system. Finally, by detecting whether the fingertip touches the object and the movement of the object, the position of the interactive trigger point is accurately determined, thereby providing the user with accurate and interactive operation feedback.

B. PROJECTION EQUIPMENT

The model of the projector in the system is XGIMI H1S, which is mainly composed of an LED light source and a highly transparent coated glass lens. The H1S lens has a high resolution of 1920 × 1080, a contrast ratio of 8001–10000:1, and a brightness of 1100 ANSI lumens. The H1S can project images in the range of 30–300 inches. Compared to ordinary projectors, the used projector can project a clearer picture under the same distance. The high-precision projector can make the picture color closer to reality, and the high-definition device can improve the sensitivity of the system control. In addition, H1S uses diffuse reflection imaging and is a non-direct light source. The user's eyes do not get tired even if the user watch the screen for a long time; this provides parents and children with a healthier viewing method.

FIGURE 2. The demonstration site of the robot system.

C. SYSTEM CONTROL

The system control module uses a Windows laptop. It performs data processing and ensures the normal operation of the software system and provides users with a stable interactive operation process (as shown in Figure 2). The software program is developed by Unity. The hardware system consists of three parts: computer, Kinect and projector. Parents can quickly and easily install, use, disassemble and store these three devices, which is beneficial for timely home education.

IV. INTERACTION DESIGN

Preschoolers' theory-of-mind development follows a similar age trajectory across many cultures [29]. In previous research, we found that preschool children in China are highly dependent on virtual characters, especially cartoon characters [29], [30]; therefore, we designed a mascot Pobi for the whole system to guide users to learn (as shown in Figure 3). The inspiration of the design for this chubby Pobi is from a fox. It is an orange fox which has a lively character and is friendly to children. We not only design the image of Pobi but also design several actions and expressions (including reading, watering, eating and being happy). Pobi is integrated into the system as a virtual learning partner of children to accompany parents in giving children a good interactive experience.

When users place physical objects into recognizable areas for an interactive operation, Pobi guides them in English to touch, drag, click and press to interact with objects. The whole interactive content does not only include knowledge of English, but also the related features under the guidance of Pobi. Pobi tells users how to use the items in daily life and other related features. The contents almost cover life encyclopedia, music appreciation, mathematics education, geography learning, life recollection, good habits development of children and emotional warmth [20]. During the process, the users can both learn English and expand their surface knowledge by interacting with the objects, thereby improving their cognitive ability.

The English word library includes more than 10 items (Table 1). The most common items are chosen, such as Mango, Bank card, Cookie, Key, Toothbrush and other items with which Chinese preschool children daily relate. The interaction of some items are introduced in detail in the following chapters.

A. MANGO

with children so that children can learn how to correctly eat mango. The interaction begins with an animation of Pobi plugging a straw into a mango but is unable to drink fruit juice (Figure 4a). Users are guided with a voice to think on the right way to eat a mango. Then, the user will need to slide on the mango to cut the mango, after which Pobi runs to eat the cut mango and thanks the user (as shown in Figure 4b). At the end of the interaction, mango-related words are displayed for extended learning; for example, yummy, delicious and cut. In this way, it becomes convenient to expand English vocabulary learning in children. At the same time, we hope that

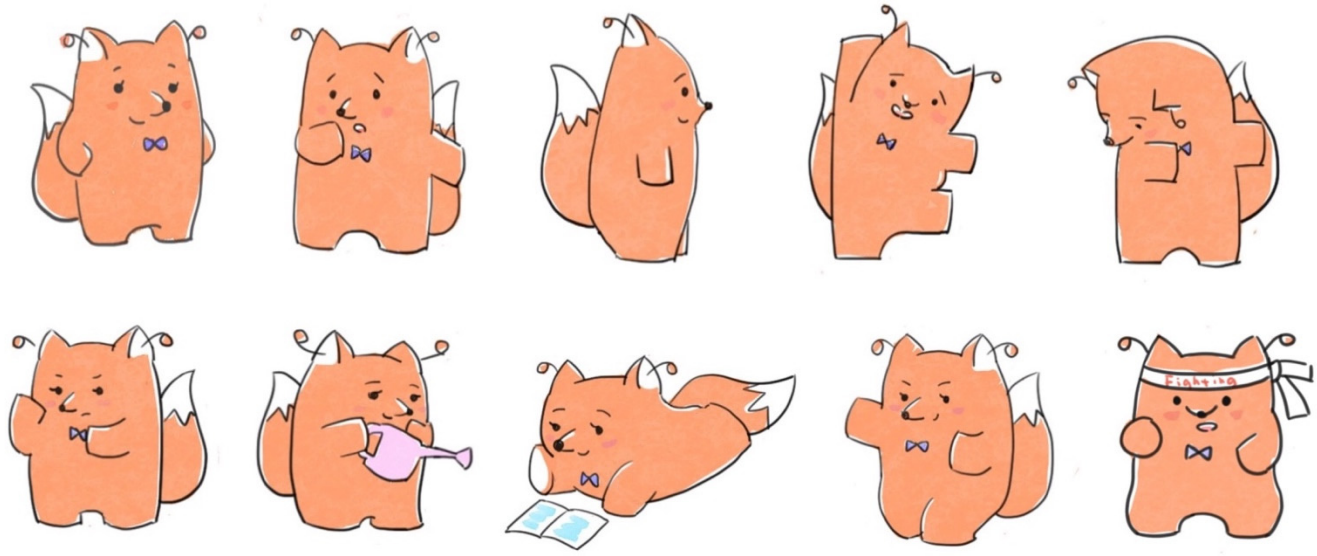
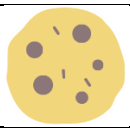


FIGURE 3. The design and expression of Pobi.

TABLE 1. Summary of english words.

	Mango	Touch, Slide	Slice, Cut, Delicious	Cut mango, how to eat mango
	Bank card	Touch, Press	One, Two, Three, Four, Five, Six, Seven, Eight, Nine, One hundred, One thousand	Saving money, Taking money
	Cookie	Touch and drag	Mother, Father, Forest	Family care, Maze game
	Map	Touch, Click on the map	China, America, Canada, Australia, Russia, Africa, South America, South Pole, Panda, Eagle, Koala, Beaver, Brown bear, Hippo, Sloth, Penguin	Know the world
	CD	Touch, Press	Music, Piano, Play	Playing the piano, Listening to music
	Straw	Touch, Move the straw	Water, Down, Put	Drink juice
	Photo	Touch, Click the photo	Memory, Toy, Newspaper, Cook Family	Happiness memory
	Key	Touch, Move the key	Lock, Door, Room	Unlocked and open the door
	Toothbrush	Touch and drag	Teeth, Toothpaste, Mouth	Dental care
	Lollipop	Touch	Honey, Confectionery, Sugar	Eating the lollipop
	Pencil case	Touch, Click on pencil case	Pencil, Pen, Eraser, Compasses, Stapler	Know Stationery

with this design, children can learn mango-related English vocabulary, the correct way to eat a mango and learn to help other partners.

B. KEY

Keys are daily necessities and the most commonly used tools for unlocking objects. We designed an interaction with keys to make children aware of how keys are used and the importance of security. In the interaction, Pobi first prompts the user to insert the key into the projected door lock (as shown in Figure 5a). Then, the door slowly opens, while Pobi tells the user to open the door in the animation. When the user pushes the door open, the robot behind the door says 'hello' (as shown in Figure 5b). Only if the key is properly held

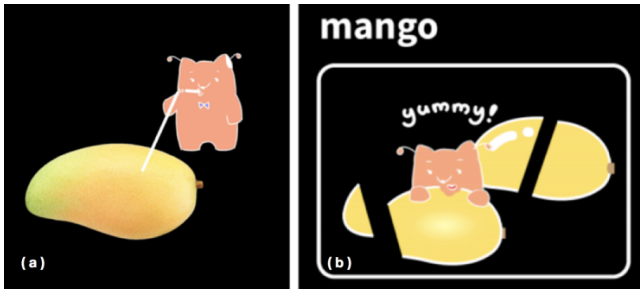


FIGURE 4. Interactive of mango. (Pobi dink the mango juice, Pobi feels satisfied and expresses gratitude).

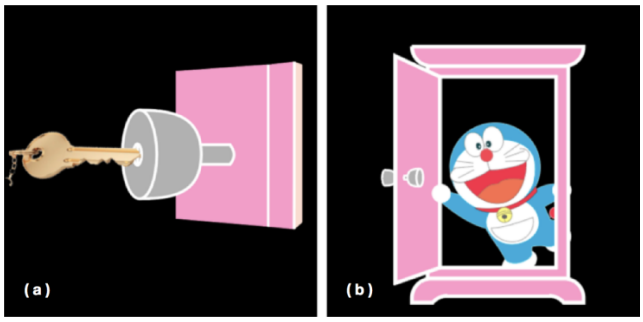


FIGURE 5. Interactive of key. (Insert key into door lock, the robot says 'hello').

in the application can the door be opened according to the specified gesture. With this, in addition to the words learning, the children can know the importance of the key to the family, and that it must be kept safe.

C. TOOTHBRUSH

Toothbrushes, as one of the most basic oral health cleaning tools, are necessities for daily life. We designed an interaction with toothbrushes to make children learn how to use toothbrushes and the importance of oral health. In the robot system, when the user places a toothbrush in the object recognition area, a projected image of a dirty tooth appears on the screen. It is covered with dental plaques and has a painful expression (Figure 6a). Then, Pobi painfully asks the users for help and guides them on how to use solid toothbrushes to remove dirt from their teeth. When the last piece of dirt is brushed off, the tooth becomes clean and Pobi becomes happy (Figure 6b). Through this interactive design, we hope that children can correctly brush their teeth with toothbrush, realize the importance of tooth cleaning and develop a good habit of brushing their teeth frequently.

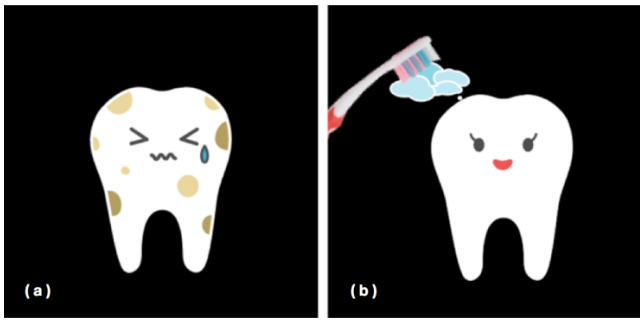


FIGURE 6. Interactive of toothbrush. (Dirty teeth and unpleasant expression, Clean teeth and happy smile).

D. BANK CARD

Bank cards are important to human daily life. The bank card is designed to inform children on how to use a bank card in the ATM. During the interaction, an input cipher interface appears (Figure 7a), and Pobi prompts the user to enter a six-digit password. When the password has been entered, a selection interface appears (Figure 7b). This interface has three instructional functions: deposit, withdraw, balance. At the end of the application, it is necessary to withdraw the card to complete the operation. Through the bank card interactive design, we hope that children can understand the ATM operation process and the importance of money deposit and withdrawal.

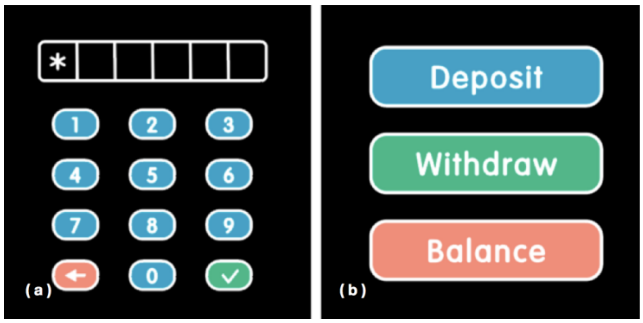


FIGURE 7. Interactive of bank card. (Password input interface, Functional selection interface).

E. COOKIE

Cookies are very common in daily life, and they are very popular among Chinese children. We designed the interaction with cookies to express the warmth of family. When the system recognizes the cookies, cookies with face and hands are projected, and a child cookie introduces its parents to the users (Figure 8a). After the introduction, Pobi guides the users into a maze game in a forest, in which the user needs

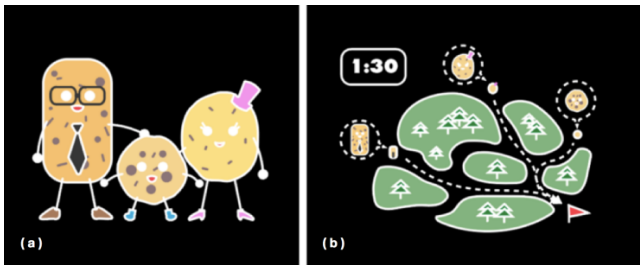


FIGURE 8. Interactive of Cookies. (Cookie family, The maze game for cookie family).

to drag the cookie rope and help the cookie family reunite within the specified time (Figure 8b). If the user successfully rescues the cookie family, the family will thank the user. If the rescue fails, the screen will show the cookie crying alone in the forest. We hope that through this interaction, the children will understand the importance of harmony and cohesion in the family, and that maintaining a warm family requires the efforts of every family member.

V. OBJECT RECOGNITION TECHNOLOGY

In this section, we mainly introduce the key technologies of the system's object recognition. This scheme involves identifying objects and then interactively projecting them on the desktop. As shown in the system architecture diagram. The Kinect camera over the desktop captures the object (as shown in Figure 2). When the objects in the designated area have been captured, our software system based on the OpenCV image recognition technology identifies the type of object. The PC enters the corresponding interactive scene according to the result of the recognition, and projects the related word and interaction mode of the item. Meanwhile, the Kinect fingertip tracking technology obtains the depth information of the desktop, as well as information on the position and touch or gesture operation.

Here, image recognition is based on OpenCV and Unity. Achieving a balance between recognition accuracy, detection delay and computational efficiency is not an easy task. In this section, we describe these details and show the basic modules of the proposed system function. These modules constitute the object recognition technology of our software. To evaluate our method, we implemented the speeded up robust features (SURF) algorithm provided in OpenCV. In addition, we tested the stability of the system when in use. The experimental results show that our object recognition technology is very stable during the interactive operations. In the next subsections, we introduce fingertip tracking, object recognition and location tracking technologies in this system.

A. FINGERTIP TRACKING MODULE

In our educational robot system, some interactions require physical contact with the objects; therefore, the system needs to locate the contact position and distance between the finger and the object. When a touch occurs, the finger covers an area of the desktop. It is difficult for the system to accurately determine the touch features. Therefore, our application first detects the state of the Kinect at run-time. When the Kinect starts to input data to the computer, the cost module records the output value as the initial value and obtains the initial state of the desktop. The height of the desktop pixels relative to the horizontal protrusions is as shown in Figure 9. The PC can record the initial information of the desktop by recording the elevation of the object, and complete the initialisation and then store the desktop information. When the initialisation is completed, the system determines the height of the fingertip by the difference between the height of the fingertip and the height of the desktop (as shown in Figure 9).

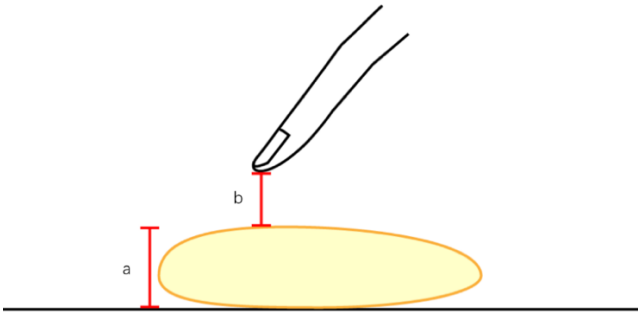


FIGURE 9. Schematic diagram of the fingertip tracking module. (a: The height of the surface of the object. b: The distance from the finger to the object.).

During the operation of the system, Kinect continuously detects and obtains the depth information of the current frame. This module compares information with the initial stored values and marks all pixels that generate differences (as shown in Figure 10a). Next, this module reduces the noise of the pixels individually and removes the pixels (as shown in Figure 10b) which are caused by the change of ambient light and data errors. Then, the PC finds out the isolated points in the region and removes the parts of the hand (as shown in Figure 10c), so that it can obtain the coordinate information of the hand. Finally, the vertex of the y-axis in the hand structure can be found and used as the fingertip coordinate (as shown in Figure 10d). In addition, to consider the change of gestures, we provide examples of gestures of different users. We use these data to train our machine learning model to keep the data as constant as possible.

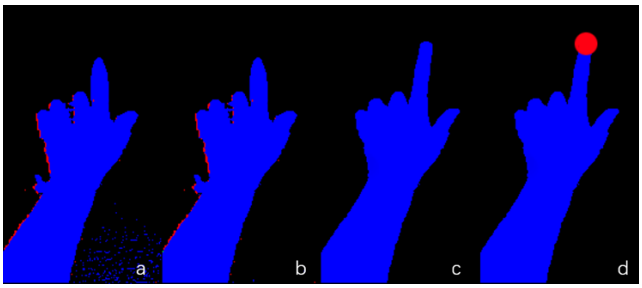


FIGURE 10. Schematic diagram of the fingertip tracking module.

B. OBJECT RECOGNITION MODULE

We use the SURF algorithm provided by OpenCV to obtain SURF features from the database. During the system operation, the recognition module tailors the information captured by the camera, and the system only reserves a part of the recognition area, thus effectively reducing the amount of computation and external interference. The PC compares the tailored picture with the feature points in the previously extracted items, and then, the system finds out the SURF feature sample that correlates most with this frame. When the PC terminal obtains the object information, it sends the sample type to the Unity 3D through a socket, and starts the

matching application in Unity 3D. If the correlation between the frame and all SURF feature samples is lower than a certain threshold, the system will be certain that there is no sample similar to the object in the database. In addition, if the PC recognizes that the object has been removed from the recognition area, a termination instruction will be sent to the scene in the Unity 3D through the socket, which means the system will interrupt the word-learning and exit the current application.

C. POSITION TRACKING MODULE

In our system, some applications require users to pick up objects and interact with the projected dynamic effects. For example, for the word 'toothbrush', users need to use physical toothbrushes to clean the dirty tooth in the picture. In this type of application, we used Python to develop a location tracking module based on color segmentation. We set it such that when the object recognition module recognizes the physical object in the scene, the system calls the module. The location tracking module acquires real-time color information on the desktop through the color camera. We set the module to convert the obtained color information to the color space, and convert the RGB color space to a more suitable color space for the color segmentation of the HSV.

We use image tools to extract color parameters of objects in the HSV space. Taking a red toothbrush as an example, the test value of the toothbrush is between HSV [156, 150, 46] and HSV [180, 255, 255]. Therefore, we search for the data belonging to the color region in the scene through the inRange method provided by the OpenCV visual library. Then, we can get the location of other entities in the picture by using the graphics obtained by mask. The PC gets the straight-line coordinates of the toothbrush on the y-axis, and the parameters obtained are the coordinates of the physical movement that we need to locate. Finally, the data are transferred in (x, y) format to Unity 3D through a socket, and then the position of objects and projection pictures can be tracked.

D. THE USER INTERFACE MODULE FOR GESTURE CONTROL

The interactive technology of physical recognition is built into the Unity 3D scene in the proposed system. The user's operation instructions are acquired and processed by the Kinect camera, and no physical mouse click or touch operation is generated in the interactive process. Therefore, the UGUI module provided by Unity 3D engine cannot be used. To solve this problem, we designed gesture control module to implement the user interface and layout. In this module, a rectangular collision body for the interactive object is set, and a cursor with a collision body to follow the motion of the finger is used. When a finger generates a click action, the system aligns the clicked object and activate its built-in click command. However, the fluctuation of the depth value returned by Kinect causes the problem of clicking on the edge and the fluctuation of Boolean value that indicates whether a touch has occurred. This phenomenon is similar to the click operation caused by the aging button inside the mouse and is misjudged as a double click operation. Fluctuation of the value near the boundary causes repeated operations (as shown in Figure 11). The figure shows 35 frames of tapping operations. The horizontal axis indicates the time unit of frame and the vertical axis indicates the depth which belongs to the blurry area where the touch occurs. At a threshold of 30, five touch functions are triggered. Among them, two triggers generated near the 11th and 30th frames correspond to a false touched. The two triggers near the 14th and 28th frames are the data oscillations generated when the touch occurred.

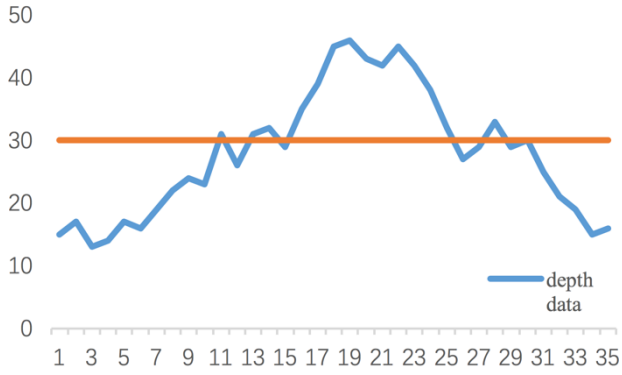


FIGURE 11. Multiple triggers caused by data fluctuation.

To reduce the misoperation caused by the border effect, we set a directional threshold for the click region. When the finger's depth value enters the tap range, the threshold of tap changes; therefore, the finger needs to be raised higher to leave the tap range. After leaving the tap area, the system needs to lower the finger to enter the tap range. Therefore, the clicked boundary value will become a dynamic range in this method and will change with the direction of fingertips. Thus, the misoperation caused by the numerical fluctuation will be reduced (as shown in Figure 12). The first four short triggers and one long 10-frame length trigger are integrated by the algorithm for a length of 16 frames and one long trigger. The improved long trigger coverage is more consistent with the best trigger range generated by click operations.

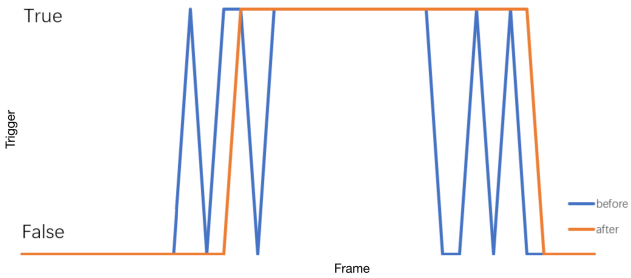


FIGURE 12. The comparison of trigger effect before and after the proposed method.

VI. EXPERIMENT & RESULTS

In order to evaluate the effectiveness of our system better, as well as to simulate the application environment, we designed a two-phase study. At the first phase, we invited

several children to learn some English words by our system, and record the reaction and other detail information (as shown in Figure 13). Then, we conducted three return visits for the learning result for each child, respectively 24 hours, 48 hours and one week after they finished the test. The feedback results show that our system can effectively help children learn English words.

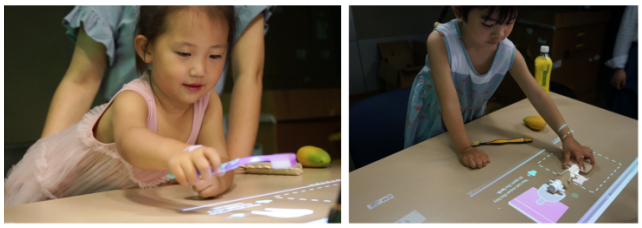


FIGURE 13. User test for robot system.

A. PARTICIPANTS

Participants were recruited from three preschool classrooms in a single school located in the Shuangliu Area, Chengdu City, China. Twenty-six children between the ages of 3-5 years old participated in the experiment (11 boys, 15 girls, age M = 4, SD = 0.5).

Participants were randomly assigned and counter-balanced across conditions with respect to their age and gender, and all participants' native language is not English.

Out of twenty-six participants, only twenty children who completed the whole test process. Among these twenty, seven children did not follow the verbal instructions of the system or experimenter during the test. Their records were thus excluded from the analysis. At last, there were thirteen children's test records contained in the analysis (5 boys, 8 girls, age M = 4, SD = 0.7).

B. INTRODUCTION TO THE TEST PROCESS

There were eight words be learned by participants in test. According to the order of learning, they are Mango, Cookie, Key, Toothbrush, Map, Bank Card, Photo, Lollipop.

At the beginning of the test, we asked each participant if he/she knows those eight words and recorded his/her answers. Then we demonstrated how to use the item identification of the system to them. Only after they have completed the operation correctly three times, they can start learning words. The time limitation of learning a single word is 5 minutes, and it takes an average of one to two minutes to complete a word interaction, which means the participant can keep learning a word for two or three times in five minutes, or they can choose to learn the next word once finished current word's interaction.

During the test, an experimenter, who could help participants to continue test, stands behind children. The experimenter would handle the item that our test needed and give it to participants when they want to learn the next word. Also, the experimenter would record the reactions and other details of each participant by taking pictures and writing down key words.

We repeated the word test with the participants in the same way after 24 hour later, and checked how well they remembered the interaction. And we did the return visit after 48 hours and a week. And we repeated the test again 48 hours and a week later

C. RESULT PRESENTATION

In order to evaluate the learning efficiency, an assessment of the words will be given in the beginning of the experiment. Due to the relative low age (3-5), the participants may not be able to speak out the meaning of word clearly. To solve this problem, the word assessment is formed like the Peabody Picture Vocabulary Test: for each word, we'll show four pictures to participants, and ask them to pick up a picture which is closest to the meaning of that word. After 24 hours of the test, we give each participant a word assessment as same as the one in the beginning, and compare the results between two assessments. Figure 14 shows that children learned some of the words presented during the interaction.

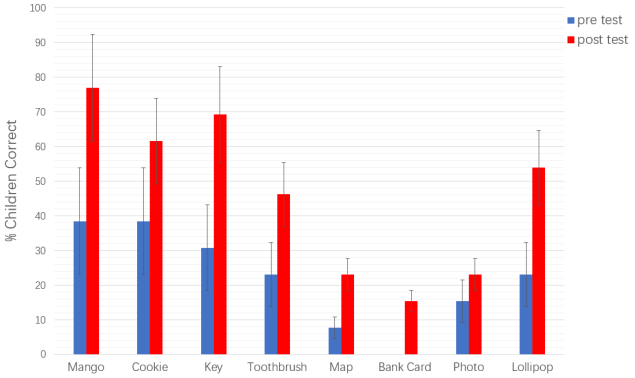


FIGURE 14. The result of the word assessment test.

As shown in Figure 14, the most learned words are "Mango", "Key" and "Lollipop", which are also repeated most frequently during the interaction. Words "Bank Card", "Map" and "Photo" are less learned among eight words. From our point of views, there are two factors caused this situation: First, the shapes of these items are similar. They are all rectangle with some colors on the surface. The most conspicuous feature of them is size, map is the biggest and bank card is the smallest. However, in word assessment, children cannot identify the size of the object in the picture, which may confuse them and lead to give a wrong answer. Second, the interactions of these words are a little bit complex than other words'. Participants need to spend more time to complete the interaction of "Bank Card", "Map" and "Photo" than other words', and thus they may not focus on the word learning, lead to the learning efficiency is lower than other words.

In the experiment, the experimenter recorded the time each participant completes word's interaction, and the time he/she stops learning a word and moves to another. Figure 15 shows the average time for each word. As can be seen in Figure 15,

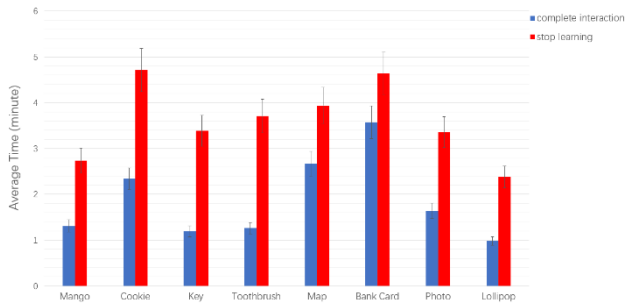


FIGURE 15. Average time of complete interaction and stop learning.

except for ''Map'' and ''Bank Card'', the average stop learning time of each word is about two to three times longer than its corresponding complete interaction time, which means these words have been learned more than once by participants. However, the words ''Map'' and ''Bank Card'' are rarely learned more than once in the experiment. We think that's because the interaction of the two words are more complex than other words. As can be seen in Figure 15, the complete interaction times of ''Map'' and ''Bank Card'' are the top two among the eight words.

During the test, the participants were observed by the experimenter on their reactions and behaviors to infer whether they were focused on word learning, such as what they were looking at, whether they followed the verbal instructions, and so on. The experimenter recorded the time that they focused on learning, as Maximum Attention Time (MAT). Figure 16 shows each word's average MAT during the test.

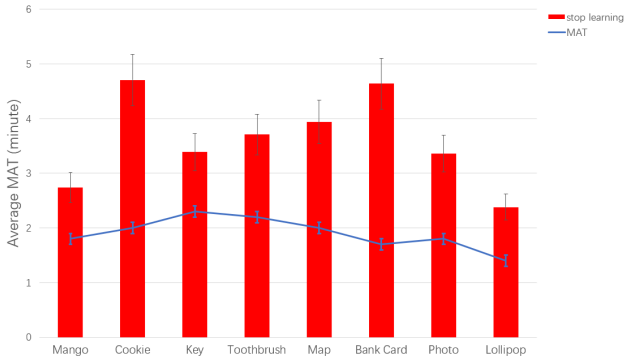


FIGURE 16. Average Maximum Attention Time and time of stop learning.

Compared to the average time of stop learning, MAT is about half of it on each word. After excluding the time taken to complete the interaction as shown in Figure 15, we can figure out that the system made participants focus on learning well after completed interaction once. While there is a decline from ''Toothbrush'' to ''Photo'' in MAT, but an increase in stop learning time. In our opinion, that may be due to two reasons. First, the interaction of these words is too complex for the young participants to concentrate on learning them. Second, these words do not show often in their daily life.

According to the results of the assessments after 1, 2 and 7 days, we draw Figure 17 to show the Saving Score of word and interaction in our system. The Saving Score is calculated using the Ebbinghaus Forgetting Curve, specifically, the formula is:

$$\text{SavingScore} = \frac{\textit{Original Learning} - \textit{Relearning}}{\textit{Oringinal Learning}} \times 100$$

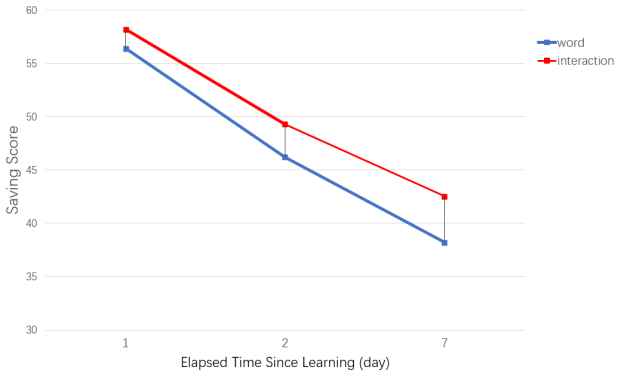


FIGURE 17. Forgetting Curve of experiment.

As can be seen above, the forgetting curve of words and interactions gradually flattens out over time. At the end of day one, the saving scores of word and interaction are close, with a gap of about 2 to 3. By day seven, the gap had increased significantly, to about 4 to 5. The results of the experiment show that learn a word with interaction is better than remember it directly.

D. EXPERIMENT CONCLUSION

Through the analysis of the test results, we come to the following conclusions.

1) For children aged 3 to 5, our system can attract their attentions better and keep their learning enthusiasm.
2) The more complex the interaction is when participants learn words, the less likely they are to repeat autonomous learning, which will lead to poorer learning outcomes.
3) After connecting English words to entities, children can understand the words more easily. Furthermore, they can immediately remember and speak them in their daily life when they see the entities in the experiment, such as ''Mango'' and ''Lollipop''.
4) For children ages 3 to 5, learning a word with interaction is better than just remembering it alone.

VII. CONCLUSION & FUTURE WORK

In this paper, a projection robot system based on object recognition technology is proposed, which can be used as a second language learning tool for preschool children in China. This study shows that object recognition is a feasible and beneficial teaching method for divergent thinking,

which can enhance and improve the language learning ability of preschool children. In addition, the experiments show that our system is attractive to children of all ages by combining physical interaction with English teaching. While by keeping children's interests in learning, our system improves the learning efficiency. The proposed system contributes to exploring new teaching methods for preschool children.

Our educational robot is still in the early stage of development. This paper only introduces the primary functions of this learning model and opens a wide field for exploration and new research. There is still much future work to be done to address all the fundamental challenges arising from technical difficulties, including signal processing, operational sensitivity and image recognition accuracy. To make our system simpler and portable, we will try to integrate the operating system into Android-based mobile devices, and design a more esthetic and humane robot appearance. In addition, more educational models for preschool children in China need to be explored. It is necessary to study the efficiency of language learning methods. This interactive form based on object recognition will have many application scenarios in the future, where it would not only be applicable in family education, but also in school education, smart appliances and the Internet of things, mobile AR and VR systems, many interactive physical environments and exciting applications of various light and shadow arts. There are many application scenarios in this interactive form combining object recognition and gesture recognition. Besides the language education involved in this paper, it can be well applied in intelligent electrical appliances, Internet of things, ''AR'' technology and virtual reality system. Through this article, we hope to inspire a wide range of researchers, educators and practitioners to explore this exciting new way of object recognition interaction.

REFERENCES

[1] C. Snow, ''Literacy and language: Relationships during the preschool years,'' *Harvard Educ. Rev.*, vol. 53, no. 2, pp. 165–189, 1983.

[2] M. Snowling, D. V. M. Bishop, and S. E. Stothard, ''Is preschool language impairment a risk factor for dyslexia in adolescence?'' *J. Child Psychol. Psychiatry Allied Disciplines*, vol. 41, no. 5, pp. 587–600, 2000.

[3] S. Yun, J. Shin, D. Kim, C. G. Kim, M. Kim, and M.-T. Choi, ''Engkey: Tele-education robot,'' in *Proc. Int. Conf. Social Robot.* Berlin, Germany: Springer, Nov. 2011, pp. 142–152.

[4] H. Sun, R. Steinkrauss, J. Tendeiro, and K. De Bot, ''Individual differences in very young children's English acquisition in China: Internal and external factors,'' *Bilingualism, Lang. Cognit.*, vol. 19, no. 3, pp. 550–566, 2016.

[5] K. A. Magnuson, M. K. Meyers, C. J. Ruhm, and J. Waldfogel, ''Inequality in preschool education and school readiness,'' *Amer. Educ. Res. J.*, vol. 41, no. 1, pp. 115–157, 2004.

[6] Z. Yu and J. Ruan, ''Early childhood English education in China,'' in *Perspectives on Teaching and Learning English Literacy in China*. Dordrecht, The Netherlands: Springer, 2012, pp. 51–65.

[7] J. M. K. Westlund, M. Martinez, M. Archie, M. Das, and C. Breazeal, ''Effects of framing a robot as a social agent or as a machine on children's social behavior,'' in *Proc. 25th IEEE Int. Symp. Robot Hum. Interact. Commun. (RO-MAN)*, Aug. 2016, pp. 688–693.

[8] J. M. K. Westlund *et al.*, ''Flat vs. Expressive storytelling: Young children's learning and retention of a social robot's narrative,'' *Frontiers Hum. Neurosci.*, vol. 11, p. 295, Jun. 2017.

[9] C. Breazeal, P. L. Harris, D. DeSteno, J. M. K. Westlund, L. Dickens, and S. Jeong, ''Young children treat robots as informants,'' *Topics Cognit. Sci.*, vol. 8, no. 2, pp. 481–491, 2016.

[10] L. Hall *et al.*, ''Map reading with an empathic robot tutor,'' in *Proc. 11th ACM/IEEE Int. Conf. Hum. Robot Interact.*, Mar. 2016, p. 567.

[11] M. Saerbeck, T. Schut, C. Bartneck, and M. D. Janse, ''Expressive robots in education: Varying the degree of social supportive behavior of a robotic tutor,'' in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 2010, pp. 1613–1622.

[12] G. Gordon *et al.*, ''Affective personalization of a social robot tutor for children's second language skills,'' in *Proc. AAAI*, Feb. 2016, pp. 3951–3957.

[13] F. Tanaka and S. Matsuzoe, ''Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning,'' *J. Hum.-Robot Interact.*, vol. 1, no. 1, pp. 78–95, 2012.

[14] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, ''Interactive robots as social partners and peer tutors for children: A field trial,'' *Hum.-Comput. Interact.*, vol. 19, nos. 1–2, pp. 61–84, 2004.

[15] T. Cremin, R. Flewitt, B. Mardell, and J. Swann, Eds., *Storytelling in Early Childhood: Enriching Language, Literacy and Classroom Culture*. London, U.K.: Taylor & Francis, 2016.

[16] K. M. Speaker, D. Taylor, and K. Ruth, ''Storytelling: Enhancing language acquisition in young children,'' *Education*, vol. 125, no. 1, p. 3, 2004.

[17] R. Isbell, J. Sobol, L. Lindauer, and A. Lowrance, ''The effects of storytelling and story reading on the oral language complexity and story comprehension of young children,'' *Early Childhood Educ. J.*, vol. 32, no. 3, pp. 157–163, 2004.

[18] R. Girshick, J. Donahue, T. Darrell, and J. Malik, ''Rich feature hierarchies for accurate object detection and semantic segmentation,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[19] C. Szegedy *et al.*, ''Going deeper with convolutions,'' in *Proc. CVPR*, Jun. 2015, pp. 1–9.

[20] X. Hu *et al.*, ''Emotion-aware cognitive system in multi-channel cognitive radio ad hoc networks,'' *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 180–187, Apr. 2018.

[21] J. Zhang *et al.*, ''Energy-latency tradeoff for energy-aware offloading in mobile edge computing networks,'' *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2633–2645, Aug. 2018.

[22] Y. Guo, X. Hu, B. Hu, J. Cheng, M. Zhou, and R. Y. K. Kwok, ''Mobile cyber physical systems: Current challenges and future networking applications,'' *IEEE Access*, vol. 6, pp. 12360–12368, 2018.

[23] X. Hu, X. Li, E. Ngai, V. Leung, and P. Kruchten, ''Multidimensional context-aware social network architecture for mobile crowdsensing,'' *IEEE Commun. Mag.*, vol. 52, no. 6, pp. 78–87, Jun. 2014.

[24] Z. Ning *et al.*, ''A cooperative quality-aware service access system for social Internet of vehicles,'' *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2506–2517, Aug. 2018.

[25] X. Hu, J. Zhao, B. C. Seet, V. C. M. Leung, T. H. S. Chu, and H. Chan, ''S-Aframe: Agent-based multilayer framework with context-aware semantic service for vehicular social networks,'' *IEEE Trans. Emerg. Topics Comput.*, vol. 3, no. 1, pp. 44–63, Mar. 2015.

[26] Q. Wu *et al.*, ''Intelligent smoke alarm system with wireless sensor network using ZigBee,'' *Wireless Commun. Mobile Comput.*, vol. 2018, Mar. 2018, Art. no. 8235127.

[27] *Kinect for Xbox One*. Accessed: May 15, 2018. [Online]. [Online]. Available: https://www.xbox.com/en-US/xbox-one/accessories/kinect

[28] *XGIMI H1 Projector*. Accessed: May 15, 2018. [Online]. Available: https://www.xgimi.com/en/h1.html

[29] M. A. Sabbagh, F. Xu, S. M. Carlson, L. J. Moses, and K. Lee, ''The development of executive functioning and theory of mind: A comparison of Chinese and U.S. preschoolers,'' *Psychol. Sci.*, vol. 17, no. 1, pp. 74–81, 2006.

[30] Z. Liu, ''Design of a cartoon game for traffic safety education of children in China,'' in *Proc. Int. Conf. Technol. E-Learn. Digit. Entertainment*. Berlin, Germany: Springer, Apr. 2006, pp. 589–592.

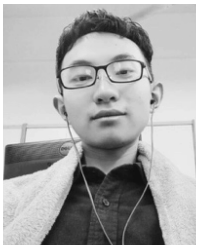
[31] P. Encarnação, L. Alvarez, A. Rios, C. Maya, K. Adams, and A. Cook, ''Using virtual robot-mediated play activities to assess cognitive skills,'' *Disability Rehabil., Assistive Technol.*, vol. 9, no. 3, pp. 231–241, 2014.

[32] M. Fridin and M. Belokopytov, ''Embodied robot versus virtual agent: Involvement of preschool children in motor task performance,'' *Int. J. Hum.-Comput. Interact.*, vol. 30, no. 6, pp. 459–469, 2014.

[33] J. S. Radesky, J. Schumacher, and B. Zuckerman, ''Mobile and interactive media use by young children: The good, the bad, and the unknown,'' *Pediatrics*, vol. 135, no. 1, pp. 1–3, 2015.

QIN WU received the B.A. degree in art design from the Chengdu University of Information Technology, China, in 2012, and the M.F.A. degree in information arts and design from Tsinghua University, Beijing, China, in 2016. She is currently a Lecturer with the Department of Computer Science, Chengdu University of Information Technology. She is also a Research Assistant with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. Her research interests include human–computer interaction, user experience, and interaction design.

SIRUI WANG received the B.S. degree in digital media technology from the Chengdu University of Information Technology, China, in 2018. His research interests include human–computer interaction, computer graphics, and digital arts.



JIASHUO CAO received the B.S. degree in computer science from the Chengdu University of Information Technology, China, in 2018. His research interests include machine learning and data visualization.

BING HE received the M.S. degree in mine spatial informatics from the China University of Mining and Technology, Beijing, China, in 2008, and the Ph.D. degree in cartography and geographical information engineering from Tongji University, Shanghai, China, in 2011. He is currently a Postdoctoral Researcher with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. His research interests include computing vision and spatio-temporal data analysis.

CHENMEI YU is currently pursuing the bachelor's degree in digital media technology with the Chengdu University of Information Technology, China. Her research interests include product design and human–computer interaction.

JIANBO ZHENG received the B.S. degree in management from the Shandong University of Science and Technology, China, in 2006, the M.S. degree in software engineering from Southeast University, China, in 2013, and the M.S. degree in computer science from the University of New Brunswick, Canada, in 2014. He is currently an Engineer with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. His research interests include cloud robotics, human–computer interaction, and data mining.