HoikuCam: Voice Recognizing Camera for In Situ Recording of Child Development

Shigeru Owada
Sony Computer Science Laboratories, Inc
Tokyo, Japan
owd@csl.sony.co.jp

Etu Terasaki† and Hiroshi Sugimura†
Kanagawa Institute of Technology
Kanagawa, Japan
terasaki.etu.1633039@gmail.com,
sugimura@he.kanagawa-it.ac.jp

Abstract—We propose the HoikuCam, a camera with voice recognition, targeted at the usage in nursery schools. It starts recording the voice when the user turns on the camera, expecting the name of child is spoken. The recognized name is used to query a database to remind the expected forthcoming behaviors of the child, which is related to the current focus of the care. The voice recognition is also triggered when a photo is taken, which allows text annotation of the photo. The image is automatically classified by the child’s name that appears in the recognized text. Our system significantly reduces the burden of caregivers when they write summaries of the care.

Index Terms—User interfaces, Human computer interaction, Child care, Camera, voice recognition

I. BACKGROUND

Early childhood care facilities usually impose heavy duty on caregivers, which causes high rate of caregiver’s job turnover [ftEoYCN04]. A survey indicates that the heaviest burden for caregivers is to write a monthly child care note [Kyo14]. The child care note describes the objective of child care and consequences, coupled with the supportive action caregivers took in monthly basis (Fig.1). This note is also used as the reference for planning the service in the next month.

Fig. 1. An example of monthly child care note (courtesy of Kuraki Nagata Nursery School)

One difficulty in writing this report is that caregivers must complete this work while they are in the school because of protecting personal information. Since they are almost always busy taking care of children, they have to write it in short spare period of time, such as while children are sleeping. Writing the note concisely is also a creative and time consuming task because they have to summarize large amount of information obtained within the month. They have to recall the past activities and pick up important moments. Currently, they take photos of children in situ and at the writing stage, use them as triggers to recall what were going on afterwards. However, if the time distance between photo taking and report writing becomes larger, they tend to forget the reason why they took the photos. It is also important to take photos that are related to the expected behaviors. However, when the number of children increases, it is difficult to recall each child’s varying objectives.

II. OUR SYSTEM

We propose a camera called HoikuCam, which supports caregivers in two ways: 1. provides reminders of expected behaviors of a child, which are related to necessary pictures to be taken, and 2. annotates the photos by spoken texts that should help recalling of the objective of photo taking. While there are plenty of commercial services already available that support caregivers [kida], [kidb], most systems provide a tablet or PC software that are independent from caregiver’s daily activities.

Fig. 2. Our camera allows text annotated photo shooting while caring the children.

Fig. 2a shows our HoikuCam, which is essentially a camera with voice recording. There is a touch sensor on top of the shutter button (Fig. 2b). When the caregiver touches the sensor, the LCD screen is turned on and starts recording (query mode). If the caregiver utters a name of a child at this mode (Fig. 2c), the camera searches the corresponding expected behaviors of the child and shows them on the display (Fig. 3).
When the caregiver presses down the shutter button, the camera takes a photo, and again starts recording the voice of the caregiver (annotation mode). This time, the caregiver can utter an arbitrary sentence to annotate the photo, while the sentence must at least contain the name of the child whom they are mentioning. The voice is transcribed to the text by Google Cloud Speech-to-Text API and saved to the cloud storage with the corresponding photo. Then the child name is picked up from the transcribed text using Google Apps Script and stored as the category tag of the photo. During this process, we access the database of child names and their variants, because the transcription API sometimes outputs a wrong name which has the similar sound. Finally, the photo and the transcribed text are forwarded to external web services using IFTTT [ift]. Currently the photos are sent to Trello and Facebook pages (Fig. 4).

Trello is designed as a task management system [Tre], but its visualization is very suitable for arranging images with categories. In Trello, one child is associated with one column, and newer images are appended on the top of the column. Facebook pages is another convenient visualization for child name-tagged images [Fac]. Facebook is also suitable to control the accessibility of the information. In Facebook, each child name is associated with an ‘album name’.

Database of expected behaviors

The expected behaviors database used in the query mode is based on ‘Kuraki-Nagata nursery school childcare process’ sheet (Fig. 5).

This sheet contains very detailed behavioral signs of typical children’s development, such as ”grab the thumb of a caregiver and get up using the abs”, ”get off the step backwards”, and so on. Since the order of signs do not change often, caregivers should observe only a few number of behaviors that will occur next.

III. Conclusion

Our prototype system was put on an informal trial at Kuraki Nagata Nursery School, Yokohama, Japan. As of other nursery schools, the school also suffers from turnover of caregivers. We got an early but very positive feedback about the annotations of photos. Especially, two dimensional photo view provided by Trello was enthusiastically accepted. On the other hand, we observed that accuracy of voice recognition significantly loses the usability of the entire system. The most critical issue is the variance of the name of children. Common names cause conflicts in identification, while less common names cause mistakes in transcription. Therefore, our future work includes adopting additional information to identify children, such as face recognition. Currently, the our system only supplies references to write monthly child care notes. We are interested in directly generating the note by photo taking. The detailed field test is another future work.

REFERENCES


[Kyo14] KYOKAI N. H.: A survey on reduction of burden of nursery school teachers, chapter 4 (japanese article, the original title is ”hoiku shi ni okeru gyoumu no futan keigen ni kan suru cyousa kenykyuu houkokusyo), 2014.