Bulletin of the Faculty of Education, Fukuyama City University, vol. 8, pp. 57-70, 2020
doi: http://doi.org/10.15096/fcu_education.08.06

Technical Report on Designing Video Documentation System on the Practice of Early Childhood Education and Care, Utilizing Automatic Face and Emotion Recognition

HIROTA Yosuke(1), ZENG Haipeng (2), UEDA Kyo (3) and YAMAMOTO Keisuke (3)

In this study, we describe the process of jointly designing the advanced digital documentation systems by university and company researchers and early childhood teachers. Our system has great novelty. It mounts videos of educational practices, which teachers can select and see, and enables them to reflect on the points that they are concerned about in their practice without much effort. In addition, the system includes a function that automatically recognizes the faces of children appearing in videos and interprets their emotions from facial expressions. This paper discusses the implications of designing the system called 'Flying baby', its technical composition and verification. At this stage, it does not have a high level of practicality because it is very difficult to increase the accuracy of recognizing children's faces and emotions from the videos, but this trial is what makes us think about the future of human-technology collaboration in the practice of ECEC.

Keywords : Digital Pedagogical Documentation System, Face Recognition, Emotion Recognition, Early Childhood Education and Care

1 Purpose and framework of this study

1.1. Purpose of this study

In recent years, pedagogical documentations have attracted the attention of early childhood teachers and related academic societies. The documentation is not a formal record such as a conventional diary, daily or weekly plan, but the panel or the portfolio of the childcare plan, the activity of children, the subjective reflection of the teachers and the comments of the observers in an easy-to-watch and -read form. In addition, the documentation is not complete and fixed, but it is flexible and thus can be used as a tool for conferences held by teachers.

Pedagogical documentations consist of photos taken with digital cameras or smartphone cameras and the comments written with personal computers. Now most of the documentations are still composed on paper and written out by teachers. In the near future, as the childcare environment becomes more digitized, documentations will be made and viewed on PCs and tablets. In recent documentation studies*1, digital documentations using 'PowerPoint' Application and websites are recommended. In that case, it will be possible to share the necessary filmed scenes with other teachers. When the documentation with videos is composed on the PCs and tablets, the comments can be written and shared on the web at any time. The main purpose is to design a digital reflection system that enables the teachers to reflect on their practices by watching videos and putting remarks on them. We are planning to develop the system that generates documentations which can be combined with video data and on which teachers can share comments.

In this study, considering this futurity of digital

(1) Department of Childhood Education, Faculty of Education, Fukuyama City University e-mail:y-hirota@fcu.ac.jp
(2) Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Graduate School
(3) viv Limited.

documentation, we designed an advanced system in collaboration with several early childhood teachers. This system also has great novelty. The system where videos of daily practices are mounted automatically recognizes the faces of children appearing in the videos and tries to interpret emotions from their facial expressions. By using PCs or tablets, teachers can watch the videos of their educational practices, focusing on the points they are especially concerned about, and share comments on the videos. We think teachers can reflect on their practices and that the system can help to follow up on newly hired teachers and to raise the level of mid-career teachers. The reflection system to recognize the child's facial expressions and their emotion is the core of this discussion. The analysis of video data by this system will include the automatic face and expression recognition system. For face matching, we use the APIs from 'Face++' which has a good reputation. The Facial Action Coding System (FACS) in the API, using facial feature points, classifies human emotions into seven categories (six emotions [surprise, fear, disgust, anger, happiness, sadness] + neutral).

The specific goal of this paper is to indicate the following three points and to describe their progress. 1) The conceptualization of the current state of pedagogical documentation and the promotion of ICT(ch.2). 2) The concept and development of our digital documentation system 'Flying baby' (ch.3) . 3) The verification of face recognition and emotion recognition from facial expression data of children(ch.4).

1.2. Framework for this study

This study was carried out by the authors of this paper sharing roles among them. Assignment of the research and writings is as follows.

Hirota: Integration of this paper, research coordination with related institutions and interview survey on them and verification of the system.
Zeng: Development and verification of the facial recognition on consignment from viv Limited.
Viv Limited.(Ueda and Yamamoto): Development and verification of the pedagogical documentation system 'Flying baby'.

We asked five Japanese kindergartens and preschools to collaborate on our research. We set up video cameras in their classrooms and received comments from the teachers and staff members, viewing the video together. We filmed the practices with fixed cameras from December,2017. Video data were automatically acquired through Arlo Pro2 video camera (resolution of video data: 1080p).

The videos data are automatically uploaded and can be viewed on Arlo's portal site. The transfer from this site to our platform 'Flying baby' enables the teachers to write comments on the places which they are concerned about and want to pay attention to. On this platform, they can review their practice from multiple perspectives through video and collaboratively share comments with colleagues. In the future, we will conduct regular research and trainings using this reflection system with the teachers including newly hired ones. We will also establish a system on the web using video and for developing and training early childhood teachers.

This study was approved by the Fukuyama City University Research Ethics Review Committee in December, 2017. In the kindergartens and preschools where the videos were filmed, we obtained consents for video shooting of children from their parents and staff members. This research is conducted by university researchers (Hirota and Zeng) in collaboration with a private company, viv Limited. (Ueda and Yamamoto), but this is not fit for profitable business at this stage and is not intended for sale now and in the future. So our research activity does not conflict with the interest as formulated in the Research Ethics Review.

2 Pedagogical documentation

2-1. The current situation of the use of ICT on the practice of ECEC in Japan and in other countries

Over the past few years, teachers' working conditions were thoroughly reviewed in the Japanese education and childcare business, and with this tendency the utilization of ICT (Information and Communication Technology) for work has been accelerated. It is said that the computerization on Japanese education so far has been lagging behind the international standard, but Japanese government has also promoted introduction of ICT technology into ECEC (Early Childhood Education and Care), and reforms have been made by allocating subsidies.

In Japanese childcare services, the system called 'Codmon' *2 has been widely introduced. It has been installed in 3600 Japanese kindergartens as of April 2019. In such a system, it is possible to create a childcare diary and plan, to make closer contact between parents and staff members, to check the child attendance and to calculate childcare expense for billing, etc. on the PCs and tablets. Other companies have created similar systems, and there are more than 10 such business management systems in Japan.

We haven't confirmed all of them, but we know that some systems have entries for pedagogical documentation. For example, a system called 'Brain' *3 is equipped with the function that allows teachers to take pictures and video of childcare practice and to attach comments to it. However, such functions are still poor, compared with the cases of other countries as described below.

Before referring to these digitized documentations, we first have to make a clear definition of the pedagogical documentation itself.

2-2. On pedagogical documentation

Pedagogical documentation is generally known to have been developed by the outstanding educational practice held in Reggio Emilia in Italy.

The documentation there is defined as follows:

'Documentation collected through photographs, recorded conversations, and visual art examples can provide an opportunity for educators to make visible the learning that is happening in the classroom not only for individual children but also for a group of children or an entire classroom. Documentation reveals not only what children are learning but how they are learning.'*4

In this study, documentation is simply described as something to make the learning visible. And it shows not only what children have learned, but also how they have learned. The visualization is also a tool to share the information and the story on the practice with colleagues and to receive direct feedback on their works. So documentation on the Reggio Emilia Approach is a bond of collaboration and an open platform for discussion.

In Japan, this view on documentation has become popular and is formulated by Japanese researchers as follows.

'Documentation is an important part for connecting teachers, children, and parents, creating mutual learning and developing the next stage of childcare and education. '*5

Documentation is not a bulletin but a tool for communication on a subject. Also, in educational studies focused on primary education, such documentation has been defined as something included in the concept of a portfolio. For example, in OECD report "Starting Strong III", the portfolio was explained as follows:

'Portfolio is a collection of documents or documentations held by an individual – e.g., child, staff member, etc. In ECEC, it can be used as an assessment, data collection or monitoring tool to examine the quality of ECEC services by observing what a child or a staff member has produced.'*6

However in German research on ECEC, portfolio is positioned in the kinds of documentation*7. So there is no clear definition globally.

In this paper, we use the term 'documentation', which is often used in early childhood education, as the comprehensive concept of 'documentation' and 'portfolio' , and the difference between them is not considered to be an issue.

2.3. On digital pedagogical documentation

In the previous section, we examined the formulation of pedagogical documentation. As previously announced, we would like to discuss what can be called digitalized documentation.

As we've already mentioned, Japanese ECEC supporting digital systems still have documentation functions and on that systems teachers can write just comments on photos, as far as we've examined them till September 2019. One of the reasons for their poor introduction could be that those digital tools are not welcomed in Japanese kindergartens and preschools. In our experience, we have never seen the Japanese kindergarten teachers using tablet devices in front of the children, except for taking their pictures. In contrast, the use of such devices in front of children is beginning to be encouraged in other countries. We see one scene in the video advertisement recommending the use of a tablet by the App 'Stepfolio' , a digital documentation system in Germany*8. The scene showing the teachers using the tablets in front of the children would be impermissible in Japan.

At present, there is demand for simplification and digitization of ECEC services, especially where Japanese teachers work excessively. App ' Codmon' and other systems meet these needs. However, we dare to suggest that it is necessary not only to shift the paper based ECEC services to the digital based, but also to build digital base platform for video sharing that includes communication among teachers. Of course, it goes without saying that direct face to face communication is necessary. However, we think that digital platforms may be more important in the future so that both teachers and children can understand the practice in reflective manners and to understand the created works objectively.

We obtained some of the ideas from instances of digital-based documentation in New Zealand and Germany. In recent years, 'Story Park' *9, childcare SNS from New Zealand, has become popular throughout the world, including Japan.

Story Park advocates 'Document and support children's learning together' (as of September 2019), and provides the format application which allows educators and parents to share photos, videos and comments on childcare facilities and at home.

Each post on 'Storypark' is tagged with the name of the child and the name of the activity. From this tag, posts are sorted according to the names of children and activities. There is a large benefit in keeping the history of learning in the way similar to SNS that teachers and parents are familiar with.

We would like to introduce another example that this study has already referred to. The online e-portfolio from Germany, 'Stepfolio' is an application aimed to allow teachers to reduce the time for paperwork and to spend more time for children. An organization called KLAX, based in Berlin, is spreading these digital tools. Using tablet devices, the teachers can record the activities of children in the form of photos, videos and comments. This 'Stepfolio' is more structured than the previous 'story park' , and the functions are divided into two parts, 'I-site' and 'Education area', which are further subdivided*10.

Both applications, 'Story Park' and 'Stepfolio' , continue to be delivered now and will change according to demand. We investigated their 2017 versions as precedent examples, and both contributed to the development of our system. Compared with the above mentioned two applications, we designed our system 'Flying baby' to allow teachers to understand their activities more reflectively.

This is because our system enables teachers to reflect on their performances using the videos and to understand them well using emotion indicators that can be recognized from the face data. We think the reflection by the videos and the visualization of emotions greatly demonstrate the novelty of our system.

Based on such previous researches, we decided to develop 'Flying baby' , and the details of the system will be explained in the following chapter.

3. 'Flying baby' system

3.1. Ground design

In this chapter, we will show you the design of the 'Flying baby' system, its data modeling and processing. In several kindergartens and preschools, videos of children were shot and uploaded to the ' Flying baby'

Bulletin of the Faculty of Education, Fukuyama City University, vol. 8, pp. 57-70, 2020
doi: http://doi.org/10.15096/fcu_education.08.06

system. This process depends on the Arlo Pro2 video cameras and the system of the Arlo portal web site. These cameras are set up in kindergartens and pre-schools and can shoot the scene of children doing activities automatically. The videos are uploaded to the Arlo portal, and we designed the system so that videos of the children's activities are transferred automatically to 'Flying baby'.

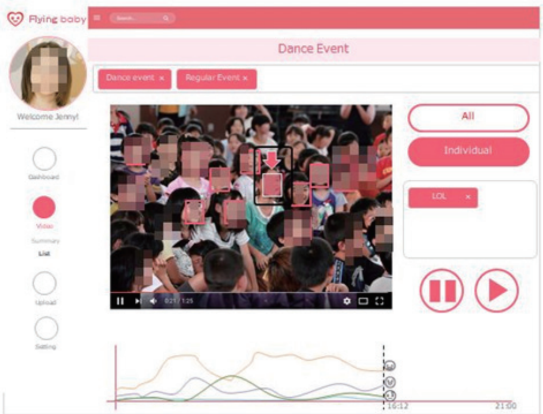
'Flying baby' provides some analytical functions, such as labeling, comments and emotion analysis, which can facilitate teachers to analyze the collected videos. The design of the 'Flying baby' system contains two parts, the frontend interface and the backend models.

As shown in [Figure 1], this screenshot is the front-end interface of the first version of the education supporting system, 'Flying baby' . The left side of [Figure 1] contains a teacher' s profile and some buttons of different functions. Once users click a button, corresponding content will be shown in the center. For example, a teacher clicks "video" button, selects a video of an event, such as Parents' Day. Children' s faces are detected and highlighted in the video. Extracted emotion information is visualized as in the graph below, from which we can see the overall emotions of the children in the video. Furthermore, teachers can add some labels and comments to this video.

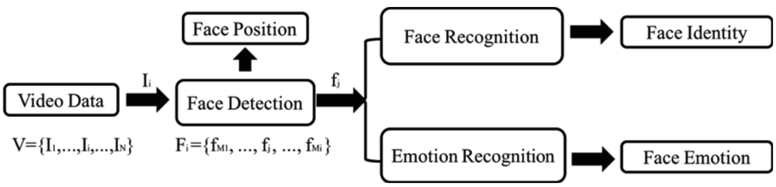
As for the backend, we first modeled the video data and then we applied some widely-used algorithms and methods to extract related information. Details are described in the next section.

3.2. Data Modeling and Processing

To analyze the videos shot in kindergartens, we use some widely-used machine learning techniques. Please note that the current methods can easily be replaced by more advanced methods. Here we mainly focus on the pipeline of data processing. A given video V is modeled as a series of images. V = {I1, I2,...,Ii,...,IN}, where Ii is the i-th frame in V and N is the number of frames in V. In this study, the rate of one frame is set to be per second. We conduct face detection first, then we conduct face recognition and emotion recognition based on the faces detected. Details for each step are described as follows [Figure 2].



[Figure 1] The first version of ' Flying baby' . Teachers can manage and analyze videos to explore children' s status in daily practices.



[Figure 2] The data processing pipeline. Given a video, we conduct face detection, face recognition and emotion recognition to acquire some related information, i.e., face position, identity, emotion.

doi: http://doi.org/10.15096/fcu_education.08.06

Face Detection

In this step, we need to detect faces in each sampling frame. As a result, we can get the position of each face fj in the frame Ii, i.e., Fi = {f1, ..., fj , ..., fMi }, where Mi is the number of faces in Ii.

Face Recognition

After locating the position of faces, we proceed to conduct face recognition, since we need to identify each face. Here we ask teachers to register children' s profile images (their names and face photos from 7 angles) and then we build a face database, which allows the system to compare the filmed images with their profile images. A general method of face recognition is to compare images with filmed image vectors. A classifier can be used to fit an input face image to its corresponding person.

Emotion Recognition

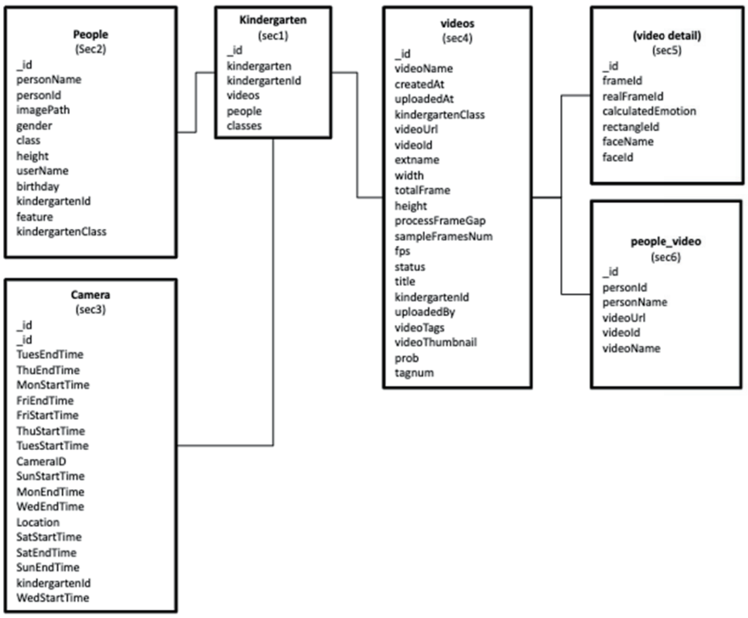
There are two widely-used categories of emotion models: categorical emotion states and dimensional emotion space. Here, as it is easier to understand, we adopt the categorical emotion models, where emotions are categorized into basic emotion types. Given a face image, we categorize the emotion to one type which is considered most likely from the output of the model.

In order to detect the face and specify each child, we use APIs from 'Face++' *11 that can extract the feature points of eyes, nose, mouth and chin and categorize the expressions into 7 kinds of emotions: surprise, fear, disgust, anger, happiness, sadness and neutral. The scores of each emotion sum to 1. And on the 'Flying baby' system, one of dominant emotion and its score are shown in the square frame. We used MongoDB as a database and developed the system tailored for this research.

3.3. Technical Process

Here, we would like to confirm the technical process of the 'Flying baby' . Cameras are connected with a Wi-Fi router that sends the videos to our AI servers. The server is a cloud server, Amazon EC2.

Data is stocked as [Figure 3] shows. (sec1) is the data related to the kindergarten or the preschool. This is entered by the researcher. (sec2) is an item related to the children, and this is also entered by the researcher. (sec3) is data related to camera recording and is automatically counted. Based on these three basic pieces of



[Figure 3] Database relationships.

information, (sec4) is generated as video information, and the emotional numbers and children' s IDs that can be read from the video analysis are in (sec5), and the information that belongs to each child is in (sec6).

After filming, the video data is analyzed. The video data is divided into 1 frame per 1 second. We can change this rate. At first, each face is detected and tried to match with the database (sec2). Each face is identified with pre-registered database data. Furthermore from the facial feature points 7 emotions are detected between the range 0-1. For instance, the score of 'happiness ' is 0.82, 'surprise ' is 0.08, the other 4 items are 0.00, and 'neutral' is 0.1 … and then the total sum becomes 1. In this survey, we added a detailed graph at the bottom of the 'Flying baby' frontend, indicating the total amount of each expression along with elapsed time[Figure 4]. If you choose one child, you can recognize the transition of his / her 7 emotions.

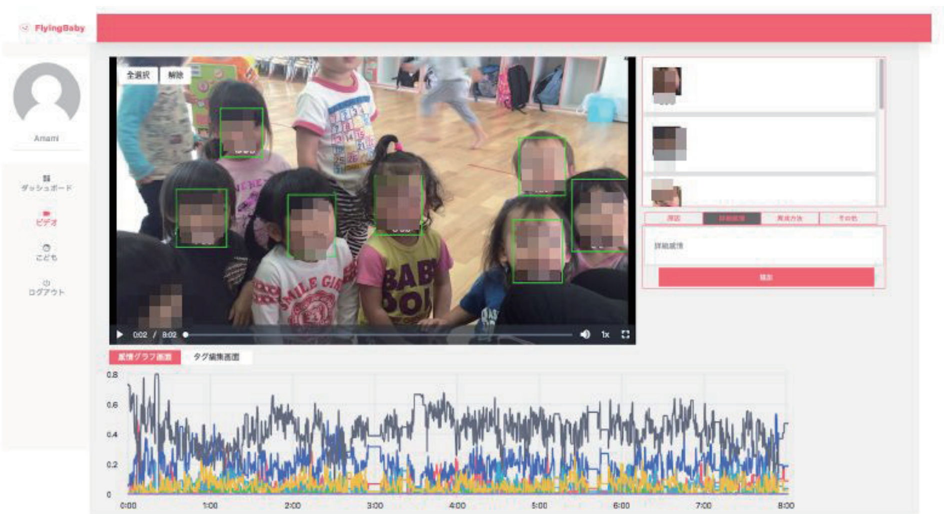
4. Validation processes and results

4. 1. Verifying the accuracy of the face recognition

As we have written, the 'Flying baby' system has been designed. This is a revolutionary system that automatically identifies a child in video data on database and visualizes their emotions. However, this system requires two-step verifications. One is to verify whether the child in the video matches a pre-registered child on database, and the other is whether the child's facial expression matches the emotion indicated on the 'Flying baby' system.

On the 'Flying baby' system, square frames and the emotion data are automatically shown on the faces of children. Behind this system the APIs from 'Face ++' are used. For example, the detect API is used to detected faces, the search API is used to obtain the person's identity and the analyze API is used to analyze emotion data. The emotion data includes the most dominant emotion and its score on top and the children's name at the bottom [Figure 5].

[Figure 5] The processing results of face detection, face recognition and emotion recognition. This photo was generated from the video recorded on Feb.8.2019.



[Figure 4] The second version of 'Flying baby' appearance. The graph at the bottom shows the total amount of each expression along with elapsed time. The amount of emotions sums to 1 and each emotion is counted between 0-1 range.

As we can see from [Figure 5], children' s faces can be detected by the API. However, the result will be affected by various elements, such as resolution, occlusion, illumination, etc. Then we check the accuracy of the face recognition. The procedure is as follows and uses the same video for the verification in the next section.

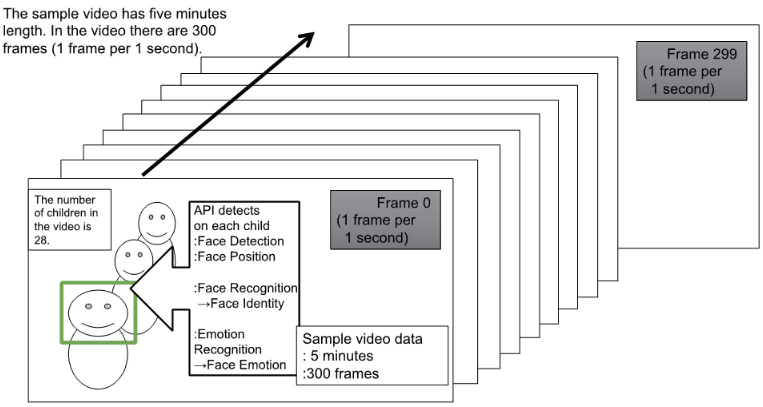
The 5 minutes sample video ,shot on February 8, 2019, kindergarten A (pseudonym), 28 children, 3 years old class, during their daily activity using cards or music, was analyzed every second with API from 'Face++' to detect each child. Face Detection, Face Position, Face Recognition, Face Identity, Emotion Recognition, Face Emotion are carried out as indicated in [Figure 2]. This procedure is modeled as follows in [Figure 6].

First, we consider as one of the big problems that not all children are face-recognized. The number of recognized children largely depends on the frames. For example, 21 out of 28 children are recognized in frame 16 (The number of total frames is 298) [Figure 7], whereas only one child is recognized in frame 143 [Figure 8]. It means that there are many frames where face recognitions were not working well. Also, the children sitting at the back and the children sitting on the left end of the photo are not recognized in most frames due to the position of the camera. For the camera was set on the upper front of the classroom in this kindergarten.

[Figure 7] Photo of Frame 16 where 21 children are recognized among 28 children.

[Figure 8] Photo of frame 143 where only one child is recognized among 28 children.

The next big problem is the difficulty of the identification between the children in the video and the children in the pre-registered database. For this verification, we chose the scenes where children were not moving. Relatively many children are recognized on two scenes ([Figure 9 and 10]). Our system detects and recognizes the child, according to the pre-registered



[Figure 6] The image of analyzing sample video data.

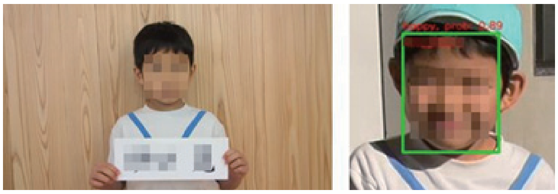
Bulletin of the Faculty of Education, Fukuyama City University, vol. 8, pp. 57-70, 2020
doi: http://doi.org/10.15096/fcu_education.08.06

photos and data. On this verification, we registered 10 children among 28 children. We asked the teacher in front of the children on photos to check the identification of the faces and the name. According to her check, our system recognized correctly 1 of 10 pre-registered photos of the children [Figure 9], and 3 of 10 [Figure 10]. As a result, we find that the accuracy of the face recognition is not so high on the verification of this video.

[Figure 9] The photo (frame 0) for verification. 1 of 10 children were identified correctly.

[Figure 10] The photo (frame 18) for verification. 3 of 10 children were identified correctly.

With these difficulties, we conducted an additional experiment. We shot additional videos (Mar.15,2019), showing large images of two children' s faces (both from 5 year old class) and compared them with the database. Unfortunately, even in this additional experiment, the correct names registered in the database were not retrieved by use of the faces of the children in the video. On the large image of face shown in [Figure 11], the API wrongly recognized the boy' s identity. Although this video was recorded with relatively high resolution (720p) and the boy' s face on the screen is large, the identification was wrong.



[Figure 11] The additional experiment of face recognition. The left photo is for profile image in the pre-registered database. The right one is cut off from the video for the re-verification.

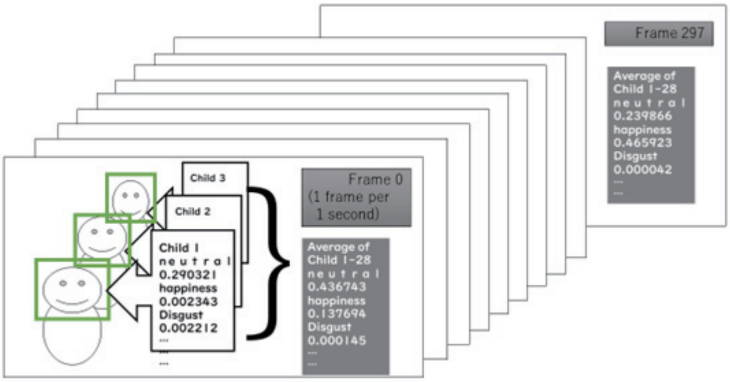
We thought the proposed workflow was smooth and could be used to analyze the collected videos. However, in reality, there are some factors that will affect the results. Therefore, in order to better use the system, there are some requirements (e.g., resolution, occlusion and illumination, etc.) for collecting good quality videos. It is worth further exploring.

4.2. Examining the emotion recognition

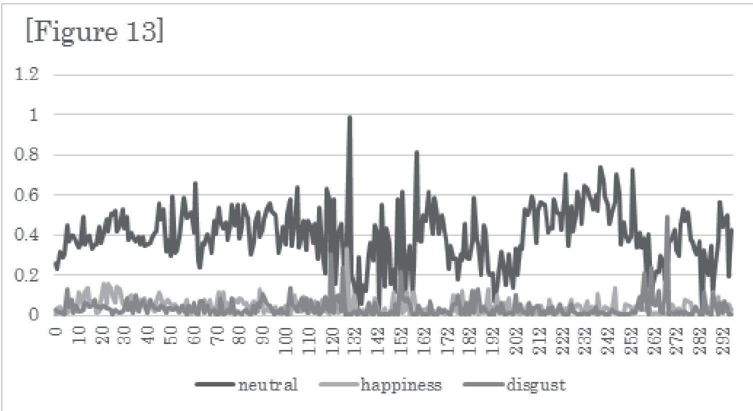
In the previous section we checked the accuracy of face recognition. Furthermore, we need to examine whether the child's facial expression matches the emotion indicated on the 'Flying baby' system. We examine how emotion data can be acquired from the same video that we used in the previous section. This video is the 5 minute sample video shot on February 8, 2019, kindergarten A (pseudonym), 28 children, 3 years old class, during their daily activity using cards or music. The API 'Face++' classifies each child's expressions into 6 emotions and 1 neutral (total 7 classifications), by scoring the characteristic points in the facial expression. The total amount of each emotion is summed up to 1 and these emotions are automatically calculated according to the feature points.

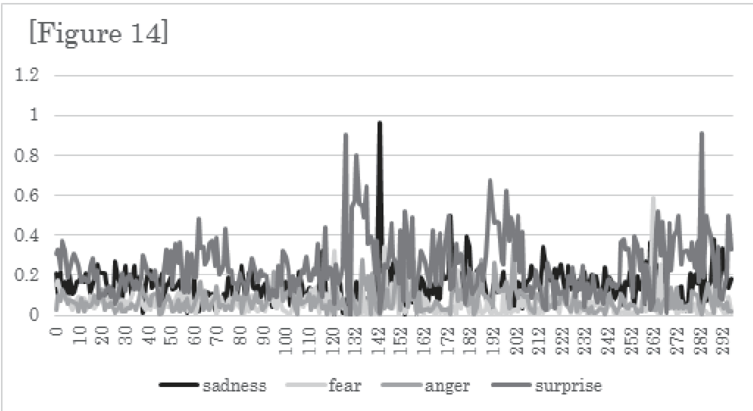
By showing the children's emotions in each frame, the overriding emotion in each frame can be inferred. In addition, by examining the transition of emotions for each frame, it is possible to know what kind of children's emotions the educational activities stimulate. In the next [Figure 12], we show the image of how the averages of the children's emotions are totaled in each frame.

The following [Figure 13 and 14] show the emotional transition of the children in each frame. These two graphs show how the seven emotions change from



[Figure 12] This is the image of how the averages of the children's emotions per each frame are totaled.





[Figure 13 and 14] These figures show the average of a total of 28 children's emotional transition. The horizontal axis represents a frame (1 frame /1 second), and the vertical axis represents the average of the transition of each emotion of all the children, which is the average calculated by dividing total amount of each emotion of each child by the number of children. Two frames where no faces were recognized were removed from the data. Also, places where only one emotion shows a prominent number close to 1 may be the places where only a few children were identified with face recognition.

frame to frame. Since it is difficult to read all seven emotions in one graph, the emotions are divided into two graphs.

We regard this emotion recognition as a prototype at this stage. There are several reasons for this. One of the main reasons is that the recent systems that interpret emotions from these features are trained to match the emotions of adult facial expressions, not of children. Considering that it is in the prototype stage, we would like to point out problems that have been discovered in this emotion recognition system.

The next problem is about emotion recognition. The average of emotions of all the frames is as follows [Table 1].

[Table 1] The average of the total of each emotion of all children in all the frames. The scores in the tables in this paper are calculated up to six decimal places. The same applies to the following tables.

neutral	0.397297
happiness	0.041458
disgust	0.03512
sadness	0.144763
fear	0.060434
anger	0.074722
surprise	0.246205

[Table 2] Emotion distribution in frame 283 where surprise was detected most highly of all the frames.

frame No.	283
neutral	0.007216
happiness	0.0134
disgust	0.005484
sadness	0.03224
fear	0.028464
anger	0.005564
surprise	0.907628

Looking at this table, the most prominent emotion is neutral, the second is surprise and the third is sadness. The rates of disgust, happiness, fear and anger are relatively low. Referring to this average value, we would like to look at the scenes where the rate of surprise is dominant. The emotional scores of the frame 283 where the rate of surprise is dominant are as follows [Table 2].

In this scene, the score of surprise is 0.907628, indicating that the most dominant emotion of many children was surprise. The picture of the scene is as follows [Figure 15].

[Figure 15] Photo of the frame where surprise is detected most highly of all the frames.

To the right of this photo, a teacher is playing the piano and children are singing with their mouths wide open. API from 'Face++' seemed to interpret these large mouths as expressions of surprise. However, they are just singing with their mouths wide open, and it surely does not mean that their facial expressions show that they are feeling surprised.

The maximum score of sadness is recorded in frame 143, which is [Figure 8] where only one child is recognized. This is not the average score of many recognized children, but it can be thought as a facial expression characteristically extracted from one child.

We examined whether the children's facial expressions in certain situations and the emotions shown there could match, taking into consideration the specific activities of the children. After some time of verification, matching proved to be very difficult. Of course, it should be taken into consideration that this experiments were based on kindergarten activities recorded with a camera at a fixed point and that the video was not accurate enough due to various elements, such as resolution, occlusion, illumination, etc. In addition, the defects of emotion recognition as shown in the examinations may be compensated just by setting the conditions and limiting the scenes. However our conclusion is that the system is not yet ready for practical use.

Though there are some problems on our system, we believe that showing the total amount of emotions of children in the activity has certain significance. We think that it is also meaningful for teachers to confirm what kind of emotions each activity brings out. However, we must conclude that more specific and step-by-step verifications are needed to identify a particular child's emotion from a certain facial expression.

Conclusion

We designed the system that utilized video and automatically recognized faces and emotions. The system will be a useful tool for reflection, with teachers reviewing the necessary parts of the videos and writing comments on them.

At our current development stage of this system, as described in the previous section, face recognition of children is incomplete, probably because the current face recognition API is not yet compatible with the use for children and videos. In addition, there were some deficiencies such as shooting conditions and video resolution. It is conceivable that the accuracy is improved to some extent by correcting the conditions and video devices. Also, the progress of technology such as API is remarkable, and the result of the experiment on face recognition may change greatly when the next experiment is performed.

However our emotion recognition also has two major challenges. One is that there is no verification of the recognition. The other is that even if the facial expression shows some kind of emotion, human emotion does not have a one-to-one correspondence with the facial expression. In other words, even if human beings have certain facial expressions, they may not equal certain emotions. Also, because children's emotions are complex and mixed, they cannot be categorized into seven emotions as stated in the division of FACS. As a result, we admit that our system might be not accurate enough to interpret emotions from children's facial expressions in current experimented videos. More explorations are needed.

While acknowledging the shortcomings, if there is an effective way to utilize this system, it may be the case where the data on facial expressions shows a large deviation. In such cases, the system can be used to show the child's mental bias in a visible way. Visualizing this bias using the actual video scenes can be one means to show the child's specificity to parents and other related members.

Now, as a conclusion of this paper, we would like to say the following. Our technology is still in the process of development. Also, no matter how much the development is advanced, the technology will never see wholly through the human mind. From now on, no matter how advanced the robots and AI technologies are, humans will continue to educate and raise children by their own abilities.

Taking this into consideration, we have the following implications for the coexistence of technology and humans: Even if the technology to interpret human facial expressions and emotions is incomplete, our reflection system can contribute to future education because this technology tells us that there are some places that humans cannot see. However, technology doesn't make clear what humans can't see. Humans must know the limits of what they can see and keep trying to see their blind spots. In this way, we think that technology and humans should do what each of them can, while knowing the limitations of each other.

That last suggestion is what a kindergarten director who cooperated with us taught us. We would like to close this article by thanking many people who cooperated with us during this project.

Acknowledgment

Our research was supported by Mayekawa foundation from August 2017 to July 2019 and JSPS KAKENdoi: http://doi.org/10.15096/fcu_education.08.06

HI Grand Numbre 19K21794, 19K02328.

Note

All internet connections in this paper were confirmed on September 14, 2019.

*1 One of the studies is following; Susan Stacey "Pedagogical documentation in early childhood" Redleaf Press, 2015.

*2 You can know the details of 'Codmon' from the website; https://www.codmon.com/

*3 See; https://brain.mj-inc.jp/

*4 G. Schroeder Yu 'Documentation: Ideas and Applications from the Reggio Emilia Approach' , in "Teaching Artist Journal" 6 (2),2008, p.37.

*5 T.Ueda 'The documentation as the method of understanding children' in F.Nakatsubo (ed.) "Methodology of understanding children" Nakanishiya Press,2012,p88.（上田敏丈「子ども理解の方法としてのドキュメンテーション」中坪史典編『子ども理解のメソドロジー』ナカニシヤ出版、2012）

*6 OECD "Starting Strong III A Quality Toolbox for Early Childhood Education and Care: A Quality Toolbox for Early Childhood Education and Care" OECD Publishing, 2011, p.323. This explanation remained on following "Starting Strong IV" and "Starting Strong V".

*7 ' In the practice of children' s ECEC centres in Germany, three forms of pedagogical documentation predominate: portfolios, wall displays and presentations of children's work' (H.Knauf ' Styles of documentation in German early childhood education' in "Early Years: An International Research Journal" Volume 35, 2015 - Issue 3, p.233.

*8 See; https://www.youtube.com/watch?v=cioHjdcawSA

*9 See; https://www.storypark.com/educators/

*10 A.Bostelmann, et al. "Das Portfolio Portfolio-Konzept digital für den Kindergarten. Mit Smartphone und Tablet zeitsparend und fundiert dokumentieren" Bananenblau, 2017.

*11 See; https://www.faceplusplus.com/

（2019 年 10 月 23 日受稿, 2019 年 11 月 26 日受理）

福山市立大学教育学部研究紀要，vol. 8, pp. 57-70, 2020
doi: http://doi.org/10.15096/fcu_education.08.06

顔・感情認識を用いた
幼児教育ビデオドキュメンテーションシステムのデザイン記録

弘田　陽介[(1)]，曾　海鵬[(2)]，上田　響[(3)]，山本　圭介[(3)]

本研究では、今日の幼児教育・保育のＩＣＴ化および保育ドキュメンテーションへの注目を鑑み、大学および企業の研究者、幼稚園・保育所の保育者が共同で先進的なデジタル・ドキュメンテーションシステムを開発するプロセスを論じている。このシステムには、大きな新規性がある。それは動画をシステムに載せ、保育者が実践の中で気になる箇所を大きな労力をかけることなく省察できることである。加えて、その動画に登場する子どもの顔を自動的に認識し、表情から感情を読み取ろうとする機能もそのシステムは包含している。本論文では、それをデザインする含意、技術的な構成、そして検証まで論じている。現段階では、そのまま使用し、子どもの顔や感情を認識できる高度な実用性をもっていないが、この試行的なシステムは、教育・保育実践における今後の人間と技術の協働の在り方まで私たちに考えさせるものであった。

キーワード：デジタル・ペダゴジカル・ドキュメンテーションシステム，顔認識，感情認識，幼児教育・保育

(1)福山市立大学教育学部児童教育学科
(2)香港科技大学大学院コンピュータ技術工学研究科
(3)viv Limited.