Liu, C.-C. et al. (Eds.) (2014). Proceedings of the 22nd International Conference on Computers in Education. Japan: Asia-Pacific Society for Computers in Education

Development of Sign Language Training Machine using Depth Sensor

Yuichiro MORI[a*], Akie FUKUHARA[a] & Shogo HAYASHIDA[b]
[a]*Department of Information Science, Kochi University, Japan*
[b]*Graduate School of Information Science, Kochi University, Japan*
*ymori@is.kochi-u.ac.jp

Abstract: In this paper, we describe a development of Japanese sign language training machine using depth sensor that targets a beginner of Japanese sign language user and supports a learner by feeding back mistake information. This system consists of two parts, one is a input device of sign language and the other is a personal computer for study. That input device consists of flex sensors and accelerometer installed on gloves, and a depth sensor. The depth sensor (Xtion) is used to acquire the position of the hand and head, because this information is very important to analyze a sign language. Correct or incorrect determinations of input sign language motions are very important to achieve the sign language study accurately. Therefore, this thesis especially describes the correction judgment of sign language. For the internal processing, we defined fundamental motions and forms of a sign language, and modeled sign language motions by combination of fundamentals. These fundamentals are defined by some information of the shape of the hand, the inclination, the position and the direction of the movement. The number of sign language words necessary for communication with a hearing impaired is about 2,000 words. As the first step, we selected 231 words that an elementary conversation is possible and constructed the system. As a result, we are able to achieve correction judgments that have high accuracy. On the other hand, the problem in a present method is clarified, too.

Keywords: Japanese Sign language, Hearing-impaired person, Xtion, Data glove

1. Introduction

The United Nations adopted 'Convention on the Rights of Persons with Disabilities' in the 61st general meeting in December 2006. In this convention, there is the following item.

> *Accepting and facilitating the use of sign languages, Braille, augmentative and alternative communication, and all other accessible means, modes and formats of communication of their choice by persons with disabilities in official interactions;*

So, it is a policy of aiming at the environment that can easily take communication even if sign language is selected and used as a communication tools. This convention is ratified by 138 countries in October 2013, keeps increasing now, and the spread activity of sign language is advanced all over the world. In Japan, 'Basic Act for Persons with Disabilities' was revised on in 2011, and sign language was taken formally as a language (Cabinet Office Japan, 2013). To make the environment where it is easy to live for a hearing-impaired or speech-disabled person, a sign language broadcasting and a sign language window, etc. are set up now. However, sign language cannot be used in other places, consideration to a handicapped person is not complete. To make that environment, it is necessary to increase a place where sign language can be used more than now. And to do that, an able-bodied person should study sign language and well understand them. There are three typical communication way between hearing impaired and able-bodied person, it is writing communication, lip speaking and sign language. However sign language requires much training, a lot of people are using it recently because the information carrying capacity is very high.

There are roughly separately two kinds of study methods of studying sign language. One is a image/video teaching material such as a book or DVD, the other is a sign language school. The book and DVD have a weak point that learner's mistake cannot be pointed out. Therefore, there is uneasiness that the learner is learned wrong motion and does not pass in practice. On the other hand, that weak point does not exist in a sign language school. However, the cost is high, and some restrictions of time and place are very large. Therefore, we developed the sign language training machine that targets beginners of sign language, and there are no restrictions of time and place, furthermore it supports the learner by feeding back mistake information.

2. Framework of Sign Language Training Machine

2.1 Outline

As shown in Figure. 1, this system consists of two parts, one is a input device of sign language and the other is a personal computer for study. This input device consists of flex sensors and accelerometer installed on gloves, and a depth sensor. The depth sensor (Xtion) is used to acquire the position of the hand and head. In our laboratory, we used a depth sensor from Kitagawa et al. (2013), and used both hands version of data glove from Matsushita (2014).

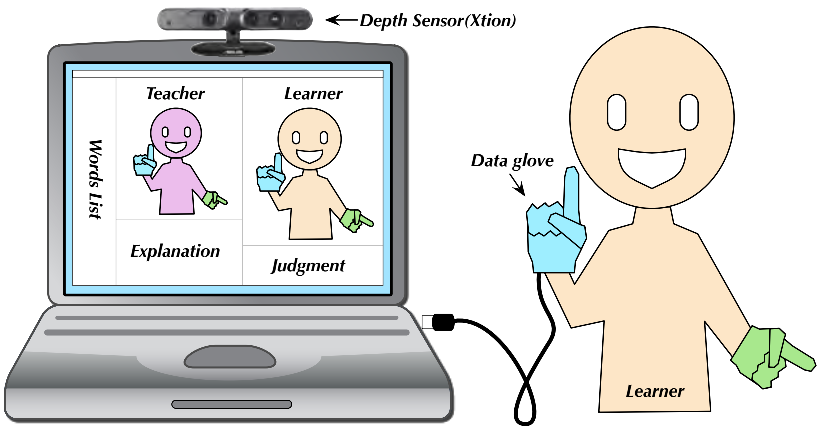

Figure 1. Configuration image of Sign Language Training Machine

Information of correct sign languages are registered in the database of this training machine as a teacher data. In the process of study, this system compares sign language information acquired from the learner with teacher's data, and judges the correction. When the learner makes a mistake, the reason is pointed out from the system, and the learner advances study again considering the mistake pointed out.

2.2 Sign language input device

That input device consists of flex sensors and accelerometer installed on gloves, and the depth sensor that is used to acquire the position of the hand and head. In present performance of a depth sensor, Xtion cannot acquire the shape and motion information of hands and fingers. Then, we use a data glove together, and acquire detailed shape, movement, and inclination of the hand.

2.2.1 Data glove

The role of the data glove in this system is acquisition of the information of hand's shape, movement, and inclination. As shown in Figure. 2, the flex sensor is installed in each finger, and we can observe the curved condition as a change in resistance. So, we convert this change into the change of voltage by principle of partial pressure ratio. We use PIC (Peripheral Interface Controller) to treat information from each sensor. Concretely, PIC digitalizes the voltage with A/D converter, and its data is sent to PC with USB communication. Similarly, the acceleration sensor also outputs the acceleration as a change in the voltage, it is converted with A/D converter in PIC, and it transmits to PC. There are a lot of signals that should convert A/D, one group is 5 signals from flex sensor of each finger, the other group is 3 signals (xyz-axes) from acceleration sensor of the hand. In total of both hands, they become 16 signals. Because the number of A/D channels of PIC is not enough, analog switches are connected outside, and data is acquired by switching the sensors. These switches are called a multiplexer (MUX), and we show MUXes in Figure. 2.

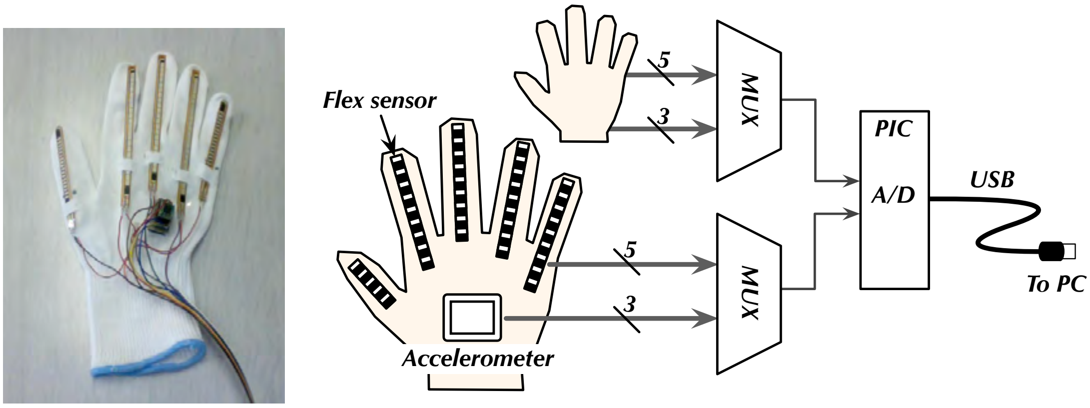

Figure 2. Data glove (left) and the peripheral circuit (right)

2.2.2 Depth sensor (Xtion)

This system uses depth sensor (Xtion) as a sensor to acquire the location information of learner's hand. By using skeleton tracking function (Figure 3) of Xtion, we acquire the three-dimension location information. As a depth sensor, we selected Xtion, which has the function equal with Kinect. Because Xtion is smaller than Kinect, and AC adaptor is not needed (USB Bus Power). Moreover, Xtion uses OpenNI, we can widely select a environment.

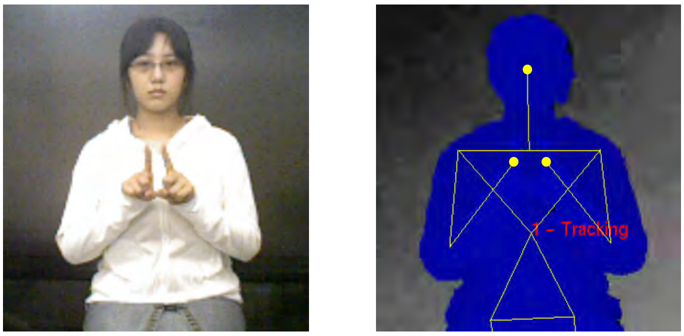
Figure 3. Real RGB image (left) and its skeleton tracking image (right)

3. Targeted words of sign language learning

The sign language word of about 2,000 words are necessary to take communication freely with a hearing impaired person, as shown in Table 1. Therefore, a practicable sign language training machine should be able to study more than 2,000 words finally. As the first step of this study, we selected 231 words that elementary conversation is possible and constructed the system. This 231 words are necessary for the level 5~7 of Sign Language Proficiency Test in Japan (Sign Language Proficiency Test Association, 2014).

Table 1. Necessary word number classified by Sign Language Proficiency Test

Level	*No. of words*	*Practicality*
1 ~ 2	about 2,000 words	A free communication is possible.
3 ~ 4	about 1,000 words	A daily conversation is possible.
5 ~ 7	about 200 words	An easy conversation is possible.

4. Correct or incorrect determination of sign language motion

In this system, sign language recognition is not a purpose. In the process of sign language study, the user perform sign language word requested by the system, and the system judges the motion which is correct or incorrect, and at the same time points out a mistake point. Seemingly, the system seems to recognizing a sign language. However, this system's processing is very simple compared with recognition. Extracting process of sign language word from time-series data can be simplified, because the user begins performance of sign language after the system directs the beginning point. Moreover, because system side knows what word motion of the input, there is no necessity for recognizing what it is, and it only has to judge the correction compared with a correct answer. If a recognition processing is requested, we need very complex processing (Toyokura et al., 2006 and Matsuo et al., 2013). However, processing can be simplified like these.

In the determination process, we should think about acceptable errors compared with the right motions. Recognition is not a purpose in this system as previously described. If this system aims at recognition, there is a necessity for considering about a tolerance of individual variation (habit) and error margin. It is a correction judgment ability that this system needs when training. Therefore, for the training of correct motion, we dared to reduce the tolerance of error margin and individual variation. So, these tolerance settings are very tight for recognition.

4.1 Determination process

In a judgment of a sign language motion, we resolve the sign language motion to fundamental motions, and are using the modeling method, which is modeling a sign language motion by combining the fundamentals. Moreover, it is possible to correspond to an increase of words in the future, because this method can express a new word by combining the fundamentals. In Table 2, we show various basic patterns of the hand used to define the fundamental motion.

Table 2. Patterns for fundamental motion definition

Basic pattern	*No. of patterns*	*Sensor*
Shape	24	Flex
Position	6	Depth
Tilt	6	Acceleration
Relocation	53	Depth & Acceleration
Shake	2	Acceleration

Even if which enumerated information in this table is missed, an accurate judgment is difficult. It became possible to acquire those information easily by not using a picture processing and using a depth sensor. Especially, the acquisition of the position of hands and head shown in Figure 4 are very difficult by using only existing sensors.

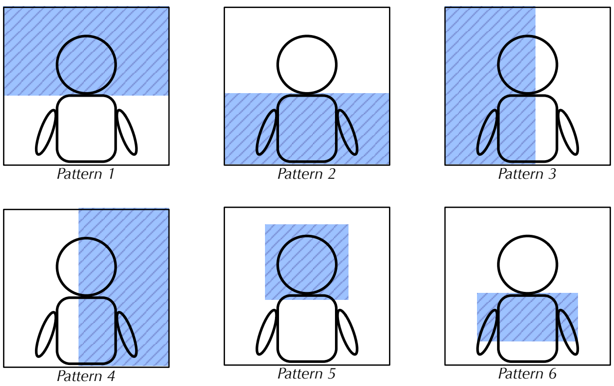

Figure 4. Position relation between hands and head

This system has teacher's database expressed by combining fundamentals, and correction of learner's motion is judged by using it. In most determination process, only initial state and end state of a sign language motion are modeled by fundamentals, and correction judgment processing is done by using it. Because the system knows the correct answer that should be judged, the correction judgment is enough by this method.

5. Learning system

At the learning part of this system, we implemented a basic function, as shown in Figure 5. The teacher's demonstration video and explanation of the sign language can be browsed. These are not different from video learning materials. Afterwards, the learner performs the sign language motion, and the evaluation is obtained. This is the feature of this system. When making a mistake, the learner confirms which part of motion is wrong by playback own video. Of course, an incorrect part is pointed out by the system. The learner can remember correct sign language motion by repeating this process. The study result is preserved by the system, and the learning scheme is made based on it. The environment of this system is shown below.

- CPU : Intel® Core™ i7-2670QM 2.2GHz
- Memory : 8 GB
- OS : Windows 7 64bit
- IDE : Microsoft Visual Studio C# 2010
- Framework : OpenNI & NITE ver.1.5.2

Figure 5. Screen shot of training machine

6. Evaluation

The determination accuracy of motion should be high to achieve correct sign language study by using this system. In other words, there is not a meaning if feedbacks of this system are not based on correct information. So that, we first executed evaluation of the determination capability.

As examinee's condition, we assumed a user of this system. It was a sign language beginner and if lecture was received, it was assumed level to be able to do correct motion. In other words, it is not a person who can do sign language with strange habits.

6.1 *Determination capability*

We experimented on all of 231 words this system targeted. Examinees were five of sign language beginners. First, the examinee selected a word with manual operation, confirmed the sign language motion, and afterwards, judged the motion by the system. Those experiments of each words were done after enough practice, and all motions were correct in check with eyes. And so, the Table 3 shows the determination rate when the examinee does correct motion. The average determination rate of the system became 88.7%, and shows a high performance. It can be said that this result has enough capability for use to learning system. On the other hand, the feature of the motion that failed in the judgment became clear, too.

Table 3. Determination rate

Examinee	***A***	***B***	***C***	***D***	***E***	***Ave.***
Determination rate(%)	90.5	90.5	90.0	86.1	86.1	88.7
No. of faults	21	23	23	32	32	26.2

This system has high determination accuracy. However, it is necessary to solve the following problems to improve accuracy from the result of experiment.

- Malfunction of skeleton tracking because of hands overlapping
- Shortage of definition of basic motion pattern "Hand's vibration (i.e. shake)"

The first is a typical problem of skeleton tracking by depth sensor. When hands overlap, it does not see from depth sensor, and malfunction occurs. It is necessary to prepare processing when losing. In addition, the middleware side should not lose even if it overlaps, too.

The next is a problem of definition of the fundamental motion based on information from acceleration sensor. In this part, we define the fundamental motion of a hand's inclination and vibration that is small shake motion. In these motions, there are a lot of individual variations more than our assumption. In this problem, it causes the policy of exclusion of individual variation and reduction of acceptable error.

7. Conclusion

In judgment of sign language motion, the position of hand and head is very important to say nothing of the shape and motion of hand. In an existing method, the location information was acquired by recognizing hand and head position by image processing. However, the processing is not easy. At the present day, that information can be easily acquired by using a depth sensor according to skeleton model with human body. And, we can acquire the motion and shape information of hand and finger accurately by using data glove. Accurate fundamental motions of sign language can be defined by combining these technologies. As a result, high determination accuracy could be achieved.

On the other hand, the problem in determination processing became clear, too. At the stage of practice of sign language, the learner should exclude an individual feature (habit), and do basic correct motion. We are aiming at a more accurate determination based on this policy now. However, when this system is enhanced to recognition processing, the individual variation problem cannot be avoided. Therefore, we should carefully consider these problems as the next one.

References

Cabinet Office Japan (2013). Annual Report on Government Measures for Persons with Disabilities 2013, Cabinet Office, Government of Japan.

Kitagawa, M. & Mori, Y. (2013). Development of the sign language training machine using the Kinect, *Technical Reports on Information and Computer Science from Kochi.* Vol. 5, No. 6.

Matsuo, T., Yamada, Y., Shirai, Y., & Shimada, N. (2013). Feature Extraction and State Decomposition for Image-based Japanese Sign Language Words Recognition Using HMM, *The Transactions of Human Interface Society 15*(1-4), 85-94.

Matsushita, T. (2014). Development of sign language training machine - About correspondence to the both hands sign language -. Bachelor thesis 2013, Dept. of Information Science, Kochi University.

Sign Language Proficiency Test Association (2014). Retrieved Aug. 7, 2014, from http://www.shuwaken.org/

Toyokura,Y., Nankaku, Y., Goto, T., Tokuda, K., & Kitamura, T. (2006). Approach to Japanese Sign Language Word Recognition using Basic Motion HMM. *Proceedings of IEICE Engineering Science Society 2006*, A-4-5, 72.