Recognition of Japanese Sign Language Words Represented by Both Arms Using Multi-Stream HMMs

Shinnpei IGARI and Naohiro FUKUMURA
Department of Computer Science and Engineering, Toyohashi University of Technology
Hibarigaoka 1-1, Tenpaku-cho, Toyohasi-shi,Aichi-ken, 441-8580 Japan

ABSTRACT

We have been studied Japanese sign Language (JSL) recognition system. In our previous research, we focused on only JSL words performed by the movement of the dominant arm. We showed that recognition rate improved when the number of states of HMM was set to the number of via-points that were extracted from the trajectory data using the minimum jerk model. In this study, we deal with recognition of JSL words performed by both arms movement. At first, we classified the JSL movements into three categories. Because the recognition results of both arm movements have to be integrated, we use multi-stream HMM which is commonly suitable for the recognition of multi-modal data. We tested the stream-weight of multi-stream HMM for each JSL word category and the number of states of HMM for left arm movement. In summary, recognition rate showed 90.2% by cross validation using 160 JSL words measured by three-dimensional electromagnetic tracking system from 15 experienced signers. From these results, it is suggested that suitable stream weights of multi-stream HMM should be set in consideration of the property of the data. Furthermore, synchronous multi-stream HMM is effective when there is strong correlation between the data to be integrated.

Keywords: Sign Language Recognition, Via-Points, multi-stream HMM, stream-weight.

1. INTRODUCTION

In a community of deaf people, sign language is one of the most common methods of exchanging information. However, deaf people experience serious problems communicating with people who hear normally and do not understand sign language. Although many deaf people lead full and fruitful lives, this communication barrier affects their lives and relationships negatively. Therefore, development of the sign language translation system for the purpose of smooth communication is expected. Automatic Sign Language Recognition (SLR) is very important in many applications, such as sign language translation, sign language tutor, and sign language learning system. The aim of SLR is to provide an efficient and accurate mechanism to translate sign language into text or speech. SLR becomes one of the important research areas of human–computer interaction (HCI), and has attracted interest more in HCI society.

There has been a resurging interest in recognizing human arm gestures. Sign language speakers also perform their signs with their heads, eyes, and facial expressions [1]. Research in sign language and gesture recognition are reviewed in [2,3]. Most of the current systems use private databases, specialized hardware [4], and are person dependent [5, 6]. As main instrumentation in this field, five kinds of devices are used for SLR. They are data glove, motion tracker, video camera, depth camera and accelerator sensor. Furthermore, most approaches focus on the recognition of isolated signs only [5, 6], or on the simpler case of gesture recognition [7] for small vocabularies. Therefore, a large vocabulary or high versatility SLR system is required. SLR research can be categorized into three major classes.

(i) computer vision based
(ii) data glove and motion sensor based
(iii) a combination of these two methods

Computer vision based SLR [8,9] relies on image processing and feature extraction techniques for capturing and classifying arm movements and handshapes. On the other hand, data-glove and motion sensor based SLR [10,11] use a sensory glove and a motion tracker for detecting handshapes and arm movements directly. The third method includes a combination of techniques from these two methods.

The major significance in a sign language is created by complex arm movements, i.e., using mainly the dominant arm or both arms. When sign languages are created using both arms, the dominant arm is more active than the non-dominant arm. We focused on the arm movement using class (ii) system and proposed a word recognition method of the sign language performed in a dominant arm movement using the motion planning model of the human arm [12]. As a motion planning model to reproduce and predict common features of reaching movements of the human arm, minimum jerk model [13] and minimum torque-change model [14] have been proposed. Wada et al. [15] assumed that more complex arm movements are reproduced by adding the via-points to such a model for reaching movements. Based on this hypothesis, we proposed to extract the via-points from arm movement data of Japanese Sign Language (JSL) word based on the minimum jerk model and used them as feature points for the JSL word recognition [12]. In addition, when Hidden Markov Model (HMM) was employed as a recognition method for JSL words represented by right arm movements, we showed that recognition rate improved by setting the number of states in HMM based on the number of via-points [16]. Because this method was very effective for recognition using HMM of the sign language word movement, we expand this method in this study.

In this study, we try to enlarge the scope of the sign language recognition system to be able to recognize the sign word performed in the both arm. We have to recognize the movement of left arm using HMM, but the number of the states of left arm HMM is also considered to influence the recognition rate. Therefore, we compare three methods to specify the number of states of the left arm HMM. In addition, because the recognition results of trajectory data of both arms have to be integrated, we employ multi-stream HMM which is commonly suitable for the recognition of multi-modal data. In our previous research [17], recognition results of both arms using DP matching were integrated equally. However, the movement information of both arms of the sign language is not equivalent, and the stream-weight for each HMM should be estimated adequately.

Moreover, we classified the JSL movements into three categories in these researches [17]. We perform a similar classification in this study, and we investigate the most effective method to set the number of the states of left arm HMM and weight of both arms HMM in each category, respectively.

2. RECOGNITION METHOD ONLY FOR RIGHT ARM

In this section, the recognition method of single stream HMM including an extraction of the via-points from the dominant arm movement data and a matching method with dictionary data in our previous research are explained [16].

2.1 Minimum jerk trajectory

Flash and Hogan [9] showed that a trajectory which minimizes sum of squares of the jerk given by (1) coincides with the trajectory of reaching movement of the human arm.

$$C = \frac{1}{2}\int_0^{t_f} \left\{ \left(\frac{d^3x}{dt^3}\right)^2 + \left(\frac{d^3y}{dt^3}\right)^2 + \left(\frac{d^3z}{dt^3}\right)^2 \right\} dt \quad \cdots \quad (1)$$

Moreover, it is possible to calculate a minimum jerk trajectory passing through some via-points. By increasing the number of via-points, the minimum jerk model is possible to reproduce more complex trajectories of human arm movements like JSL trajectories from the parameters of start, end, and via-points. The fact that it is possible to reproduce the JSL movement from a few feature points indicates that JSL trajectory data can be compressed into a small number of via-points information. Therefore, we have proposed that the via-points extracted from arm movement data were used for JSL recognition [12]. Positions and timing of via-points are specified so that the error between a measured trajectory and minimum jerk trajectory passing through the via-points became sufficiently small. After that, start, end and these via-points of JSL word trajectory were used for the recognition by DP matching. We showed the recognition rate of the proposed method were higher than the methods based on the other feature points.

2.2 SLR by HMM

In our previous research, we also used a left-to-right-model of HMM for SLR and performed the training at the word level [16]. The trajectory data of the dominant arm detected in a motion trucker were used as inputs of HMM directly. There are two kinds of methods to set the number of the states of HMM. In the first method, the constant number of the HMM state is searched so that the recognition rate becomes higher. In another method, the number of the HMM state is estimated by each trajectory to train. We showed that recognition rate improved when the number of HMM states was set according to the number of via-points that were extracted from the trajectory data using the minimum jerk model. This is related to the position of the extracted via-points. The via-points were more likely to be extracted near local minimum points of the tangential velocity of the trajectory, that is, the peak points of curvature. Because a moving direction changes at these points, such points can become the candidate of the timing of the state transition of HMM. For example, in the case of an almost linear movement like a simple reaching movement, a trajectory is reproduced only from start and end points. Thus, the number of the states becomes one. Because there were individual differences in the number of the extracted via-points in the same sign language word caused from the difference in the arm trajectories, different dictionary database of the same word depending on the number of the via-points were prepared. Thus, the plural models are sometimes learned for one word. In addition, we input test data into all HMM generated in a recognition process.

3. MEASUREMENT ENVIRONMENT

At first, we conducted a sign language measurement to perform recognition experiments of JSL words performed by both arms. We measured one data of 160 different words from fifteen JSL experienced persons (adult man and woman in their 30s and 60s) in this study. Experiments were undertaken with the understanding and written consent of each subject. Figure 1 shows the measurement environment. We used a three-dimensional position measurement device Liberty (POLHEMUS Inc.) to measure JSL movements. Sampling frequency is 240 [Hz]. The sensors were attached to the wrists of the both arms and measured their positions as hand positions of both arm. Because the movement data of different people have to be compared, we measured the positions of both shoulders and set them as the origin of each hand position. And we normalized the hand positions by the length of each arm. We defined that saggittal-horizontal axis is X-axis, frontal-horizontal axis is Y-axis, and vertical axis is Z-axis. Because every subject performed movements of the dominant arm by the right arm, we defined a right arm as a dominant arm and a left arm as a non-dominant arm in this study. The subjects put both hands on her knee as an initial posture, performed a sign language movement, and came back to the initial posture last in each task. We calculated tangential velocity from measured data and extracted sign language movement from the time of the first local minimal value to the time of the last minimal value of the velocity from the trajectory data of both arms respectively.

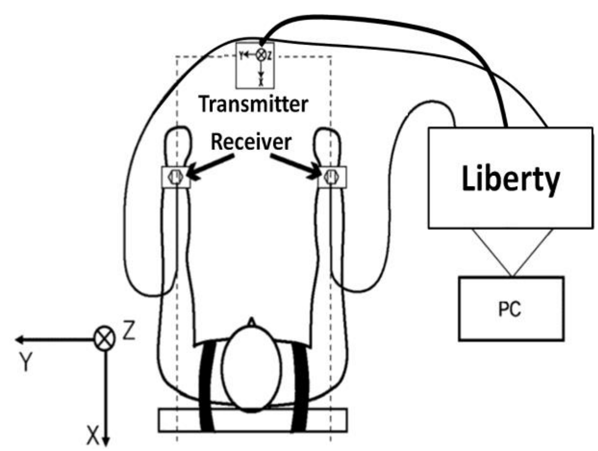


Figure 1. Measurement environment

4. METHOD

In this chapter, we explain the classification method of the sign language word and setting method of the number of the states of left arm HMM and weight setting of multi-stream HMM.

4.1 Integration method of multi-stream HMM

To recognize sign language performed by both arms, we must integrate the information of the right and left arm. SLR systems can be broadly classified according to the method in which we integrate the incoming right arm and left arm information streams. Most systems can be described [18] as performing either feature fusion or decision fusion. In feature fusion systems, the right and left arm feature vectors are integrated. By contrast, in decision fusion systems, separate classifiers are constructed for the right and left arm features, and the classifier outputs are integrated. And decision fusion integration systems can easily be made robust to known individual difference of either stream by simply weighting the right and left arm classifier decisions. In this study, we employ the simpler decision fusion systems.

4.2 Classification of one-handed signs and two-handed signs

According to a study of linguistic characteristics of the sign language of Kanda [18], the sign language is divided into "One-handed signs" that are performed by a dominant arm and "Two-handed signs" that are performed by both arms. For example, in 160 words that we measured in this study, 80 words belong to "One-handed signs" and 80 words belong to "Two-handed signs". The hand of the subject moved from the initial position on the knee to a specific start position before performing sign language movement on this measurement condition. However, when the subject performed "One-handed sign", the left hand remains on the knee during performing sign word. Therefore, when the path length of the left arm from the initial position to the final position is shorter than 5cm, the left arm was thought not to move and we classify the words into "One-handed signs".

We conducted a preliminary experiment to verify this classification. At first, the right arm HMM and the left arm HMM are learned independently using dictionary data before the recognition experiment. In the recognition method examined in this study, the input word data is classified into "One-handed sign" and "Two-handed sign", first, and matched with only the dictionary words that belong to the same class. In the case of "One-handed sign", only a likelihood of the right arm HMM is used for recognition. In the case of "Two-handed sign", we compare the recognition rates based on integration of both output likelihoods with equal weights and the rate only based on a likelihood of the right arm HMM. The recognition rates were calculated by a leave-one-out cross-validation (LOOCV). The result is shown in Figure 2.

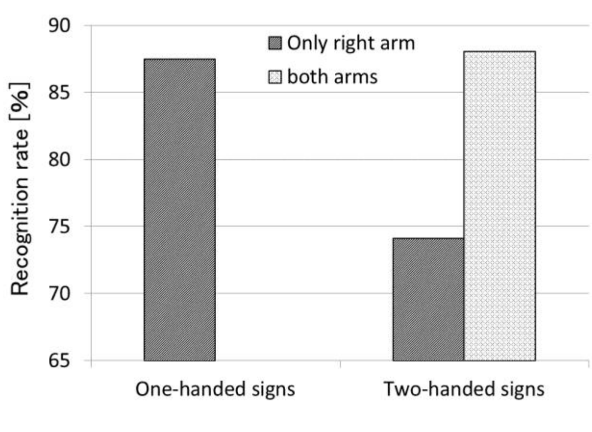
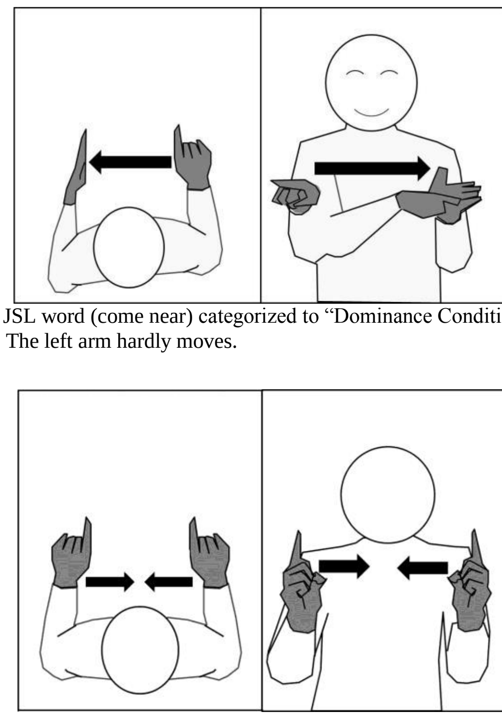


Figure 2. Recognition rate of "One-handed signs" and "Two-handed signs"

The recognition rate in "One-handed signs" became about the same as the rate based on the integration of both output likelihoods in "Two-handed sign". Moreover, there was no word that the classification was mistaken. This result suggest that integration process for "One-handed signs" became unnecessary by adequate classification of sign words.

4.3 Dominance condition and Symmetry condition

According to Kanda [18], the sign language is divided first as "One-handed signs" and "Two-handed signs", and "Two-handed signs" is divided roughly into some types by a characteristic of the movement. In this study, we classify these words into "the sign language that a non-dominant arm remains stationary" and "the sign language that both arms move". In our previous study [16], we defined the former as "Dominance Condition" and the latter as "Symmetry Condition". 80 words of "Two-handed signs" can be divided into 40 words of "Dominance Condition" and 40 words of "Symmetry Condition" according to the classification in this study. Figure 3 is an example of each classification word.



(a) JSL word (come near) categorized to "Dominance Condition". The left arm hardly moves.

(b) JSL word (meet) categorized to "Symmetry Condition". The left arm moves almost as same as the right arm.

Figure 3. Exapmples of "Dominance condition" word and "Symmetry condition" word

The path length of the left arm movement during the sign word seems well suited to classify the "Dominance Condition" and "Symmetry Condition". However, the left arm often moves unconsciously in the "Dominance Condition" word. Furthermore, the unconscious left arm movement in these words has a large individual difference. Therefore, it is difficult to classify into these categories only by the path length of the left arm.

The movements of both arms in the sign language classified into "Symmetry Condition" are mirror symmetry or the same as shown in figure 3(b). In this case, we expect that left arm trajectory can overlap with right arm trajectory by symmetric transformation, rotation, and parallel displacement. On the other hands, the left arm almost stands still in the sign language of the "Dominance Condition" as shown in figure 3(a). Even if the trajectory of the left arm in the word of the "Dominance Condition" was transformed by those transformations, may not overlap the trajectory of the right arm. Therefore, we evaluate trajectory error when we transformed the left arm trajectory so that trajectory error between the right arm trajectory and the transformed left arm trajectory becomes minimum. If it is small enough, the word is classified into "Symmetric Condition" and otherwise, the word is classified into "Dominance Condition".

In this classification algorithm, we have to calculate a transformation matrix and amount of parallel displacement for the left arm trajectory. We use the steepest descent method in this study to calculate them. To avoid falling into a localized solution, we made a symmetric transformation and shift in order to roughly coincide the start position and movement direction of both arm trajectories first and calculated the element of the rotation matrix by steepest descent method.

4.4 Number of the states of left arm HMM

When an input word is classified as "One-handed signs", we can perform HMM-matching with dictionary data of "One-handed signs" using the right arm data. On the other hand, several setting methods of the number of HMM states of the left arm exist when an input word was classified as "Two-handed signs". The first method is that the number of HMM states of the left arm is defined as a fixed number. In this case, a suitable fixed number of the states are searched such that the recognition rate becomes higher. In the second method, the number is set to the same value of the right arm HMM. This method is thought to be particularly effective for the Symmetry condition since the similarity of the trajectory data of both hands is expected to be high. The third method is that the number of via-points extracted from left arm movements in the same manner as the right arm is used. The third method is thought to be effective for Dominance condition whose trajectory data of left and right arm are different. Therefore, we compare these three methods for the number of HMM states of the left arm when the word is classified as "Two-handed signs".

4.5 Weight-setting of the both arms

The construction of the multi-stream HMM commences by first training independent right arm and left arm word-level HMMs using motion tracker data. The stream weighting parameter is related to the relative reliability of the right and left arm modalities, which is dependent on the classification of movement. It is thought that information quantity in each arm movement is equal as for the sign language of the "Symmetry Condition" whose trajectories of right and left arm resembles. Consequently, the weights of right and left arm HMM are preferred to be the same. Conversely, the weight of the right arm HMM should be bigger than that of left arm HMM because the left arm almost stands still as for the sign language of the "Dominance Condition". Accordingly, we change the weights of the right arm and the left arm in classification category respectively and examine the recognition rate.

5. RECOGNITION EXPERIMENT

At first, we test the algorithm to classify the sign words into three categories. Although Individual difference was considerably seen in the measured trajectory, the trajectory error between the right arm trajectory and transformed left arm trajectory for the word of the "Symmetry Condition" became small enough. In contrast, the trajectory error for the word of the "Dominance Condition" did not decrease. As a result, the measured sign words of "Two-handed sign" were classified into "Dominance Condition" and "Symmetry Condition" exactly.

Figure 4 and 5 show the results of the recognition experiments for each sign word category. Results of recognition rates for every combination of "stream weights" and "number of states" is shown using polygonal line graph. The numerical value in the horizontal axis of the graph expresses the stream weights of the right and left arm HMMs. We found that the fixed number of the states of the left arm HMM achieving the highest recognition rate was three.

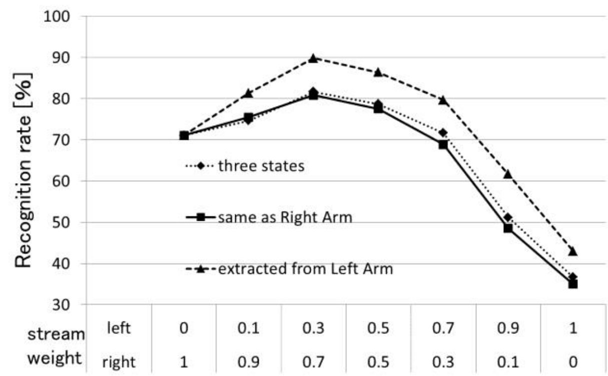


Figure 4. Recognition result of the "Dominance Condition"

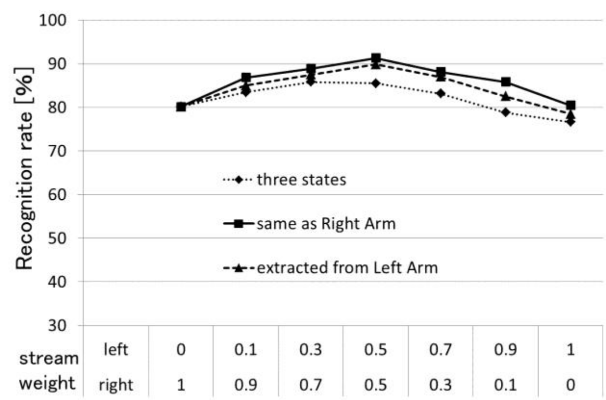


Figure 5. Recognition result of the "Symmetry Condition"

From figure 4 showing the results of "Dominance Condition", when the number of the states of left arm HMM is calculated from left arm trajectory independently, recognition rate is always higher than other HMM state number conditions. Moreover, there is little difference between condition of the fixed number and the condition of the same number as the number of the right arm HMM. We confirmed that since the left arm's movement does not resemble the right arm movement in "Dominance Condition", the suitable number of the left arm HMM states is different from the right arm HMM. Meanwhile, the recognition rate became highest when the stream weight of right arm HMM is 0.7. We think that this result reflects the relative importance of right arm movement in the words that belongs to the "Dominance Condition".

From figure 5 showing the results of "Symmetry Condition", when the number of the states of left arm HMM is same as the number of the right arm HMM, recognition rate became higher. However, difference of recognition rate between the "same as right arm" and "calculated from left arm data" is only around 2%. In the case of "Symmetry Condition" words, trajectories of both arms almost correspond and the numbers of via-points of both arms were almost equal. So, this result shows that the symmetricalness and the synchronousness of the right and left arm movements are important to improve the recognition in the sign language of the "Symmetry Condition". Moreover, the recognition rate is highest when the stream weight of the right arm HMM is 0.5 and the graph is almost symmetric to the case that the weights of right and left arm HMM are equal. This result is also thought to reflect the similarity of the right and left arm trajectories.

These results indicates that the condition like the number of HMM states or stream weight that achieved higher recognition rate is different for each category respectively. The total recognition rate using the best condition in each category reached 89.6%.

We show examples of the database based on this recognition experiment in table 1. In the case of "One-handed signs", the number of the states of left arm HMM is not necessary. In the case of “Symmetry Condition”, the numbers of the HMM states of the right and left arm are set equally. In the case of "Dominance Condition", the numbers of the HMM states of the right and left arm are set independently. Therefore the number of the states of the left arm HMM is different from that of the right arm HMM.

Table 1 : Example of a database classified in three

JSL Word	Classification	Number of the states of HMM	
		Right hand	Left hand
River	One-handed	5	---
Rein	Symmetry Condition	2	2
Tennis	One-handed	3	---
		4	---
Winter	Symmetry Condition	3	3
Soccer	Dominance Condition	2	1
		3	1

6. STATE-SYNCHRONOUS DECISION FUSION

The result of the recognition experiment in the previous section, suggests that the symmetricalness and the synchronousness of both arm movements of the words of “Symmetry Condition” are important. However, HMMs of right and left arms were learned independently because we employed a decision fusion system – also termed ‘late integration’ (LI) –. Therefore, synchronization should be considered more. On the other hand, the design of LI systems is very flexible and many variations on the theme exist. We examine a method to synchronize state transition of HMMs of left and right arms.

The LI systems can be roughly sub-classified according to the lexical level at which decision fusion occurs. At the lowest level decisions are fused on a frame by frame basis. If the streams are modeled using HMMs then this equates to combining the likelihoods of corresponding right arm and left arm HMMs model states – often termed ‘state-synchronous decision fusion’[19]. Within each state the model progresses independently through the states of the separate right and left arm HMMs under the constraint that the timing of the state transition occurs synchronously in the right and left arm domains. Therefore, we compare "state-synchronous decision fusion" in consideration of synchronicity with the normal "decision fusion" for recognition of the words in “Symmetry Condition”.

The result of two kinds of multi-stream HMM learning methods is shown in figure 6. The recognition rate of "State-synchronous decision fusion" becomes higher, especially, when the stream weights of both arm HMMs are equal. And, in the case of the right arm or the left arm HMM was used for the recognition, the difference of recognition rate almost disappears. However, recognition rate of "state-synchronous decision fusion" was slightly high. Thus, the effect of the synchronized learning was confirmed. As a result, the total recognition rate increased to 90.2% for 160 sign words.

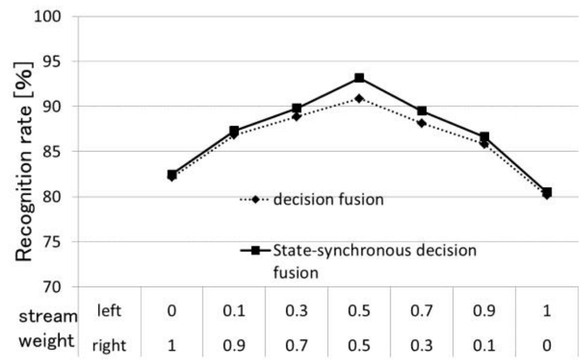


Figure 6. Recognition result of "state-synchronous decision fusion"

7. CONCLUSION

In this study, we considered a recognition method of JSL words performed by both arm movements using multi-stream HMM. As a result, we should change the matching method such as the number of the states of left arm and the stream weight depending on the category of the sign language word. In the case of the words in "One hand sign language", which are performed by only a right arm movement, a HMM for left arm movement was not necessary. Therefore, recognition rate only by a HMM for right arm movement showed the highest rate. In the case of the words in "Dominant Condition", which are performed mainly by a right arm movement and left arm is used for auxiliary, the ratio of stream weights of right and left arm HMM was 7 to 3, and the number of states of HMM for left arm movement was independently set to the number of the via-points extracted from the left arm trajectory. At last, in the case of the words in "Symmetry Condition", which are performed by symmetric movements of both arms, stream weights and the number of states were equal in the right and left arm HMM. In addition, state-synchronous decision fusion represented higher recognition rate. In summary, the recognition of the JSL of 160 words showed 90.2% by cross validation.

From these results, it is suggested that when multi-modal data is recognized using multi-stream HMM, suitable stream weights should be set in consideration of the property of the data. Furthermore, state-synchronous decision fusion is effective when there is strong correlation between the data to be integrated.

In this study, we focus on the JSL words only discernible by both arm movements. It will be necessary to include handshape data to recognize more JSL words. We will examine a weight estimation and classification technique of multi-stream HMM that uses "handshapes" or "facial expressions" in future. Some words in “Dominant Condition” express a meaning by difference of right and left handshapes without arm movements. Thus, the classification considering handshapes other than the arm movements is required.

ACKNOWLEDGMENT

This work was supported by Grant-in-Aid for Scientific Research (C) 25330335.

8. REFERENCES

[1] U.von Agris, M. Knorr, and KF. Kraiss,“The Significance of Facial Features for Automatic Sign Language Recognition,” 8th IEEE International Conference on Automatic Face and Gesture Recognition, pp.1-6, 2008.

[2] S. Ong and S. Ranganath, “Automatic sign language analysis: A survey and the future beyond lexical meaning,” IEEE Trans. PAMI, vol. 27, no. 6, pp. 873–891, 2005.

[3] T.S. Huang Y. Wu, “Vision-based gesture recognition : a review,” in Gesture Workshop, Gif-sur-Yvette, France, vol. 1739 of LNCS, pp. 103–115, 1999.

[4] G. Yao, H. Yao, X. Liu, and F. Jiang, “Real time large vocabulary continuous sign language recognition based on op/viterbi algorithm,” in Intl. Conf. Pattern Recognition, Hong Kong, vol. 3, pp. 312–315, 2006.

[5] C. Vogler and D. Metaxas, “A framework for recognizing the simultaneous aspects of american sign language,” Computer Vision & Image Understanding, vol. 81, no. 3, pp. 358–384, 2001.

[6] R. Bowden, D. Windridge, T. Kadir, A. Zisserman, and M. Brady, “A linguistic feature vector for the visual interpretation of sign language,” in European Conf. Computer Vision, vol. 1, pp. 390–401, 2004.

[7] S. B. Wang, A. Quattoni, Louis-Philippe Morency, David Demirdjian, and Trevor Darrell, “Hidden conditional random fields for gesture recognition,” in Computer Vision & Pattern Recognition, New York, USA, vol. 2, pp. 1521–1527, 2006,

[8] B. Bauer, and KF. Kraiss, “Video-Based Sign Recognition Using Self-Organizing Subunit,” In Proceedings of International Conference on Pattern Recognition, Vol.2, pp.434-437, 2002.

[9] P. Dreuw et al., “Speech Recognition Techniques for a Sign Language Recognition System,” InterSpeech-2007, pp.2513-2516, 2007.

[10] C. Oz, M.C. Leu, “Linguistic properties based on American Sign Language isolated word recognition with artificial neural networks using a sensory glove and motion tracker,” Neurocomputing, vol. 70, Issues 16–18, pp. 2891-2901, 2007.

[11] C. Oz, M.C. Leu, “American Sign Language word recognition with a sensory glove using artificial neural networks,” Engineering Applications of Artificial Intelligence, vol. 24, Issue 7, pp. 1204-1213, 2011.

[12] S. Igari, N. Fukumura, “A Study of Sign Language Recognition Using Via-Points based on a Minimum Jerk Model,” SICE Annual Conference 2013, pp.1288-1293, 2013.

[13]T. Flash, and N. Hogan, “The coordination of arm movements, an experimentally confirmed mathematical Model,” The Journal of Neuroscience, Vol.5, No7, pp.1688-1703, 1985.

[14]Y. Uno, M. Kawato, and R. Suzuki, “Formation and control of optimal trajectory in human multijoint arm movement minimum torque-change model,” Biological Cybernetics, Vol.61, pp.89-101, 1989.

[15]Y. Wada, and M. Kawato, “A theory for cursive handwriting based on the minimization principle,” Biological Cybernetics, Vol.73, pp.3-13, 1995.

[16] S. Igari, N. Fukumura, “Arm Movement Pattern Recognition of Sign Language Words Based on the Minimum Jerk Model to Develop a Sign Language Translation,” IEICE Journal, Vol.J98-D No.3 pp.437-447, 2015

[17]S. Igari, N. Fukumura, “Sign Language Word Recognition using Via-point Information and Correlation of the Bimanual Movements,” ICAICTA International Conference 2014, pp.75-80, 2014.

[18] K. Kanda, “A study of linguistic characteristics of Japanese sign language,” Fukumura publication, 2010. (in Japanese)

[19]X. Shao, J. Barker, “Stream weight estimation for multi-stream audio–visual speech recognition in a multi-speaker environment,” Speech Communication, vol. 50, Issue 4, pp. 337-353, 2008.