CONTINUOUS KOREAN SIGN LANGUAGE RECOGNITION USING GESTURE SEGMENTATION AND HIDDEN MARKOV MODEL

Jung-Bae Kim*, Kwang-Hyun Park**, Won-Chul Bang** and Z. Zenn Bien*
*Div. of EE, Dept. of EECS, KAIST, Daejeon, Republic of Korea
**HWRS-ERC, KAIST, Daejeon, Republic of Korea

ABSTRACT

In recognizing gesture words such as Korean Sign Language (KSL), it is a very difficult to segment a continuous sign into individual sign words since the patterns are very complicated and diverse. To solve this problem, we disassemble the KSL into hand motion classes according to their patterns. Observing the speed and the change of speed of hand motion and using state automata, we reject unintentional gesture motions such as preparatory motion and meaningless movement between sign words. To recognize hand motion classes we adopt Hidden Markov Model. Using these methods, we recognize 10 KSL sentences and obtain 95% recognition ratio.

INTRODUCTION

Sign language is a representative example of hand gesture with linguistic structure and is important for the hearing impaired to communicate one another. It is, however, required to develop a system capable of recognizing and/or generating sign language, in order to communicate with normal person.

Sign language recognition has been attempted since several years. Starner [1] proposed a HMM-based recognition algorithm with his vision system. It recognizes 40 words with 91.3% accuracy. Liang proposed a glove-based system, which can recognize Taiwanese Sign Language [3]. This HMM-based system recognizes 250 words with 90.5% accuracy. Kim et al. used Fuzzy Min-Max Neural Network to recognize Korean Sign Language (KSL) [2]. It recognizes 131 words with 94.3% accuracy based on 14 hand motion-direction classes, 23 hand posture classes, and 14 hand orientation classes. However, these systems concentrates on a part of sign language that has hand trajectory using glove device [1-3].

This paper presents a vision-based recognition system of continuous KSL as shown in Figure 1. For this end, it is first necessary to segment a continuous motion gestures into isolated basic forms, and then the isolated words are recognized.

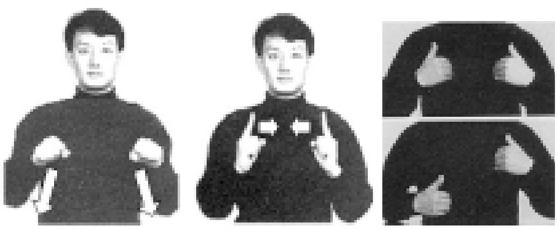


Figure 1. Continuous KSL sentence: “Hello, nice to meet you.”

GESTURE SEGMENTATION

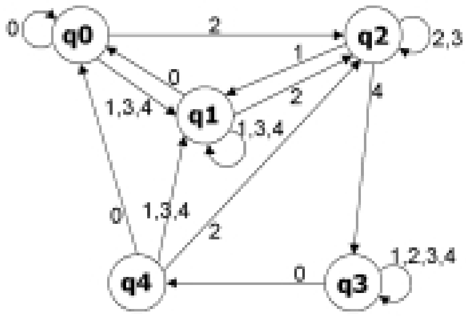
There are no explicit indications for the beginning and the end of gesture. So, the meaningless movement that changes hand position to prepare next gesture can be interpreted as a meaningful sign word. By analyzing hand motions, we can observe that the speed and the change of speed of hand motion can be used as remarkable features. For an example, meaningful sign words consist of three distinct phases: preparation, stroke and end, while meaningless gestures do not have distinct stroke.

From the feature pair, we can define 5 motion phases (stop, preparation, stroke, moving and end) and develop 6 rules based on the phase patterns. Then, the rules are adopted to discriminate meaningless gestures and segment the continuous KSL sentence into several isolated words in the framework of state automata as shown in Figure 2.

Table 1. The defined states of automata for continuous KSL

State	Description	Function
q_0	Resting state	Initialization
		Static gesture recognition
q_1	Preparation state	Initialization
q_2	Stroke/ Moving state	Feature extraction
q_3	Ending/ Repetition state	Feature extraction
q_4	End state	Dynamic gesture recognition
		Initialization

Fig. 2. State automata diagram for gesture segmentation.



In this figure, q_0, q_1, q_2, q_3 and q_4 mean 5 states, which are defined in Table 1, and arrows mean state transition functions and numbers on the arrow mean 5 motion phases such as ‘stop’, ‘preparation’, ‘stroke’, ‘moving’ and ‘end’ respectively, which are labeled as 0,1,2,3,4 respectively.

Generally meaningful sign word starts in state q0 and ends in state q4 after several state transitions. Whereas, meaningless gesture starting in state q0 cannot reach state q4. The result of the automata based gesture segmentation is the set of isolated words.

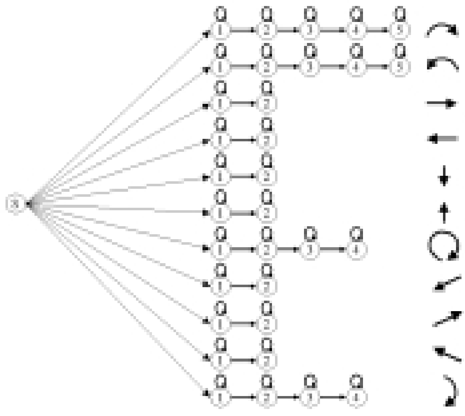


Figure 3. The HMM network for gesture recognition

GESTURE RECOGNITION

To recognize isolated words, we define 11 hand motion classes depending on motion features, and adopt Hidden Markov Model (HMM), which is effective to model spatio-temporal information. The HMM is a collection of states connected by transitions. Each transition has a pair of probabilities: a state transition probability and an output probability, which defines the conditional probability of emitting an output symbol from given state [4].

Because we use discrete HMM, the feature of hand motion has to have discrete values. The hand motion can be expressed as the set of direction feature. All motions at every sample steps are quantized into 16 directions. Here we do not consider the distance of motion at each sample step.

We have already segmented the continuous hand motion into isolated words, so we only need to recognize the hand motion of isolated word. To do so, we have constructed left-right type HMM network for gesture recognition as shown in Figure 3 because the direction sequence has clear start point and end point. In the figure, S is the dummy start state and 11 arrowed motions in the right side mean 11 hand motion classes. With the HMM network for gesture recognition, we discriminate the class of the segmented gesture. We select the likelihood model that has the biggest probability on each class.

SYSTEM CONFIGURATION AND EXPERIMENTS

To verify the proposed vision-based KSL recognition system, we have implemented the algorithm on 500 MHz Pentium PC with Matrox Genesis imaging board and PULNIX TMC-7 RGB camera. We have obtained very successful experimental results with 95% accuracy of recognition in hand motion classification.

CONCLUDING REMARKS

Continuous sign language is not simply connected form of individual sign words. They entail meaningless gestures, which make machine recognition difficult. This paper studied segmentation and recognition of continuous sign language using color vision. We have used the state automata method whose inputs are defined as the speed and change of speed of motion to segment isolated sign words from continuous sign language. And we adopted HMM method to recognize hand motion, since the method is effective on spatio-temporal data and is effective to recognize continuous Korean Sign Language.

REFERENCES

1. T. Starner and A. Pentland, “Real-time American sign language recognition from video using Hidden Markov Models,” In Proc. International Symposium on Computer Vision, pp.265-270, 1995.
2. J.-S. Kim, W. Jang and Z. Bien, “A Dynamic Gesture Recognition System for the Korean Sign Language”, IEEE Trans. on Systems, Man and Cybernetics, Vol. 26, No. 2, pp. 354-359, 1996.
3. R.-H. Liang, Continuous Gesture Recognition System for Taiwanese Sign Language, Ph.D Thesis, National Taiwan University, Taiwan, R.O.C., 1997.
4. L.R. Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proceeding IEEE, vol. 77, pp. 257-285, 1989.

ACKNOWLEDGMENTS

This research is fully supported by HWRS-ERC of KAIST.

Jung-Bae Kim,
Div. of EE, Dept. of EECS, KAIST,
373-1, Guseong-dong, Yuseong-gu, Daejeon, 305-701, Republic of Korea
+82-42-869-8418, +82-42-869-8750(fax), jbkim@ctrsys.kaist.ac.kr