2011 11th International Conference on Control, Automation and Systems
Oct. 26-29, 2011 in KINTEX, Gyeonggi-do, Korea

Motion Primitives for Designing Flexible Gesture Set in Human–Robot Interface

Suwon Shon[1], Jounghoon Beh[2], Cheoljong Yang[1], David K. Han[1], Hanseok Ko[1,2]

[1] School of Electrical Engineering, Korea University, Seoul, Korea
(Tel : +82-2-926-2909; E-mail: {swshon, cjyang}@ispl.korea.ac.kr, hsko@korea.ac.kr)
[2] University of Maryland College Park, MD, USA
(Tel : +1-301-405-2876; E-mail: jhbeh@umiacs.umd.edu)

Abstract: This paper proposes motion primitives for designing a gesture set in a gesture recognition system as Human-Robot Interface (HRI). Based on statistical analyses of angular tendency of hand movements in sign languages and hand motions in practical gestures, we construct four motion primitives as building blocks for basic hand motions. By combining these motion primitives, we design a discernable 'fundamental hand motion set' toward improving machine based hand signal recognition. Novelty of combining the proposed motion primitives is demonstrated by a 'fundamental hand motion set' recognizer based on Hidden Markov Model (HMM). The recognition system shows 99.40% recognition rate on the proposed language set. For connected recognition of the 'fundamental hand motion set', the recognition system shows 97.95% recognition rate. The results validate that using the proposed motion primitives ensures flexibility and discernability of a gesture set. It is thus promising candidate for standardization when designing gesture sets for human-robot interface.

Keywords: HMM, gesture recognition, HRI.

1. INTRODUCTION

Gesture based HRIs have been receiving wide interests in recent years due to their naturalness for human to learn/use and also the advances made in robot intelligence for recognizing human gestures. Utility of such interfaces resides on the level of difficulty for human to learn and the complexity of commands possible with the interface.

There is a variety of gestures, from static hand posture to sign language with different levels of associated complexities in hand movement. A static hand posture, such as an open palm extended toward observers to mean "stop", is obviously of low complexity. Static postures are easy for human to learn but have limitations delivering complex messages or commands. This is because the variety of visually discernable postures with a static hand is quite limited due to limited possible articulations with fingers and associated joints [1,2]. On the other hand, a sign language using two hand motions may capture a large set of complex commands, but may pose major difficulties for machine-based recognizers to adequately understand the meanings in real time [3,4]. Another problem with sign language based HRIs is that they are difficult for a human operator to learn. To develop an HRI easy to learn and yet sufficient in its variability to command a robot to perform complicated tasks, the complexity of the associated gesture set has to be at mid-level, somewhere between complexities of static postures and sign languages.

In the mid-level gesture interface research, there have been efforts in isolated gesture recognition [5,6], continuous gesture recognition [7,8] and interfaces using various input devices [9,10,11,12].

These efforts in general targeted gesture set consisting of letters, symbols and arbitrarily defined gestures. In this paper, we propose a scheme for designing flexible gesture set by using motion primitives based on American Sign Language and other simple hand gestures.

Performance is evaluated using a gesture recognizer based on HMM. HMM, well known for capturing stochastic dynamics of an information source, is applied to areas such as speech recognition, HRI, and information science. In particular, the gesture recognizer in this paper is similar as that of the phoneme based vocabulary-independent speech recognizer which can flexibly build vocabulary set by sequentially combining phonemes. Just as in spoken languages, the gesture model developed here can create a large number of vocabularies using a small number of motion primitives.

The rest of this paper is organized as follows. In Section 2, the recognizer based on HMM is described. In Section 3, the motion primitives are proposed. In Section 4, we describe how the gesture set based on the motion primitives is developed. In Section 5, experiments and the result are described. The conclusion is provided in Section 6.

2. GESTURE RECOGNITION BASED ON HMM

The gesture interface we propose targets gesturing with only one hand. This is because hand gesture is usually performed with one hand while the second hand is assumed carrying or holding an object. Although there are hand-gestures performed with two hands, it can be thought as one hand mirrored to the other hand or doing same gesturing [13]. The features we use for developing gesture recognition consist of angle and velocity of one hand trajectory. For ease of detecting hand trajectory, we use blue color glove that is used on special effect in movie. Thus the hand trajectory is obtained by continuously finding and tracking the center point of the moving glove, using webcam.

978-89-93215-03-8 98560/11/$15 ©ICROS

The recognition of hand gesture is based on HMM. The HMM of each gesture has simple left to right structure as depicted in Figure 1. The first and last state is null state that has only state transition probability. The Expectation-Maximization algorithm is used to train the HMM [14].

The HMM is represented by the parameter λ. It is

$$\boldsymbol{\lambda} = \{\mathbf{A}, \mathbf{B}, \boldsymbol{\pi}\} \qquad (1)$$

where $\mathbf{A}$ represents the state transition probability distribution, $\mathbf{B}$ is the observation symbol probability distribution and $\boldsymbol{\pi}$ is the initial state distribution. [15]. We can recognition by conditional probability such that

$$\hat{c} = \arg\max_{c=1\cdots C} P(\mathbf{O} \mid \boldsymbol{\lambda}_c) \qquad (2)$$

where $\mathbf{O}$ is observation sequence, C is total number of class.

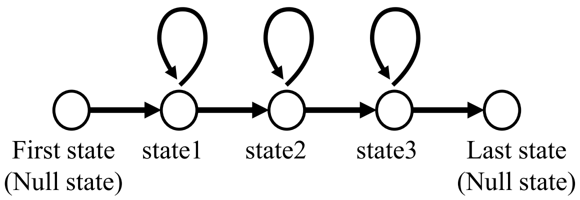

Fig. 1 Simple left to right HMM structure

3. MOTION PRIMITIVES

In vocabulary-independent speech recognition, a word is built by combining the models of phoneme. So the word model consists of phoneme models such that the smallest speech unit can be recognized.

In example, the word 'three' consists of three phonemes, /th/, /r/, /iy/ in sequence. The three phoneme HMM models are connected sequentially as shown in Figure 2.

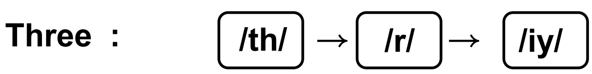

Fig. 2 Example of the word 'three' HMM

In this paper, the phoneme-word HMM relationship is adapted in our gesture recognition. In gesture task, there are motion primitives as there are phonemes in speech. In this way, we can create new gestures without training for new gestures. We can build new gesture sets by just combining the motion primitives.

We analyzed the trajectory of the hand movement in the American Sign Language and existing hand signal systems for determining the desired motion primitives.

3.1 American Sign Language

To analyze the statistics of motions in the American Sign Language, publically available database 'RWTH-BOSTON-104' is used [16]. The database consists of 201 video streams of American Sign Language sentences published by the National Center for Sign Language and Gesture Resource of Boston University. In the database, the two hands trajectory is also provided. We computed the angles in every 201 sentences trajectory and the resulting statistics of angle are shown in Figure 3. Statistically, the angles 0, 90, 180, 270 degrees exist in relatively large number.

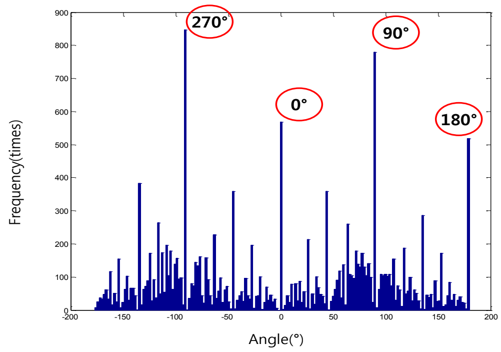

Fig. 3 Statistical result of motion trajectory angles in the 'RWTH-BOSTON-104' database

3.2 Hand signaling

Many hand signaling can be seen in daily chores and activities. A variety of hand signaling is often found in official sports refereeing. Crane operator also use hand signaling motions in accordance to the forms suggested by OSHA (Occupational Safety and Health Administration, U.S Department of Labor)[17] Usually hand signaling is performed when the it is too noisy or too far for human speech to reach. In short, hand signaling must be simple enough to capture and provide discernable messages for effective communication. Among these signals, we find that there are statistically distinguishable set of horizontal and vertical hand motions.

3.3 Four motion primitives

Based on the statistical analysis of the American Sign Language and typical hand motions in crane operation and other activities, we determine the four representative motion primitives as shown in Figure 4.

Symbol	Up	Right	Left	Down
Gesture	↑	→	←	↓

Fig. 4 Four motion primitives

4. GESTURE SET DESIGN

4.1 Fundamental hand gesture set

Combining the motion primitives, we can create various gesture set. The first step, however, is to build a "fundamental hand gesture set", similar in spirit as that of building "syllable" in speech recognition. Here, we design a fundamental hand gesture set using the motion primitives as depicted in Figure 5 and validate its performance. The proposed fundamental hand gesture set is designed by using 2 or 4 motion primitives. For example, the UP_L gesture consists of UP, LEFT, RIGHT, DOWN motion primitives sequentially.

The fundamental hand gesture set is designed wherein the end point returns back to the start point. This structure removes the non-gesture element, a critical problem in continuous gesture recognition.

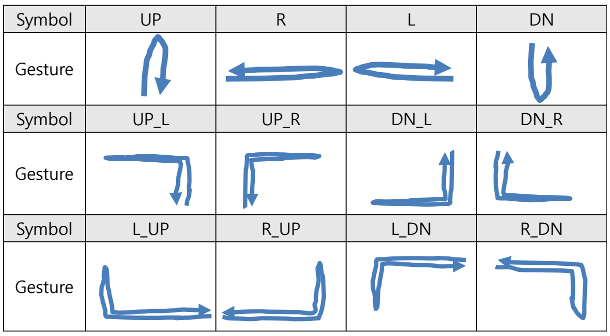

Fig. 5 Fundamental hand gesture set

5. EXPERIMENTS

The 4 motion primitives and fundamental hand gesture set database is collected and built from 28 people and each person tried 3 times. Database is recorded by webcam with blue color glove. The 4 motion primitive HMM is trained using 4 motion primitive database. Its effectiveness is evaluated by measuring the recognition performance of the fundamental hand gesture set. The optimal state and mixture for 4 motion primitives HMM were found as 5 states and 4 mixtures and shows the best recognition rate.

5.1 Fundamental hand gesture set recognition

The fundamental hand gesture set recognition rate is 99.40%. Recognition rate of each model is shown in Figure 6.

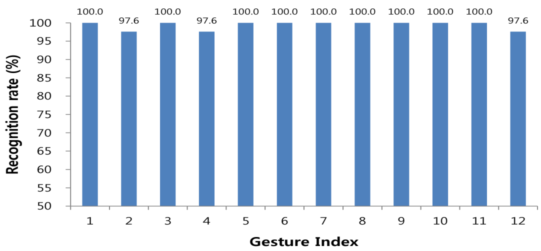

Fig. 6 Recognition result of fundamental hand gesture set

5.2 Connected fundamental hand gesture set recognition

Another experiment is connected fundamental hand gesture set recognition. The word network used is depicted in Figure 7.

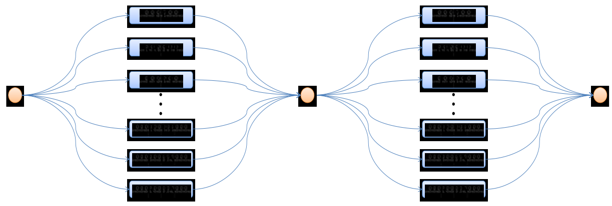

Fig. 7 Connected recognition word network

The 48 test dataset was also collected from the same 28 people and each person tried 3 times. The detailed dataset is presented in Table 1. Note that the recognition rate is 97.95% total and each model recognition rate is delineated in Table 2.

Table. 1 Database list for testing connected fundamental hand gesture set.

Index	Combination	Index	Combination	Index	Combination
1	[*U* , *R*]	17	[*UL* , *L*]	33	[*LU* , *U*]
2	[*U* , *RU*]	18	[*UL* , *LU*]	34	[*LU* , *DL*]
3	[*U* , *LU*]	19	[*UL* , *R*]	35	[*LU* , *LD*]
4	[*U* , *LD*]	20	[*UL* , *LD*]	36	[*LU* , *LU*]
5	[*R* , *D*]	21	[*UR* , *R*]	37	[*RU* , *D*]
6	[*R* , *DR*]	22	[*UR* , *UL*]	38	[*RU* , *UL*]
7	[*R* , *RD*]	23	[*UR* , *UR*]	39	[*RU* , *UR*]
8	[*R* , *UR*]	24	[*UR* , *LU*]	40	[*RU* , *RU*]
9	[*L* , *U*]	25	[*DL* , *L*]	41	[*LD* , *U*]
10	[*L* , *DL*]	26	[*DL* , *RD*]	42	[*LD* , *DL*]
11	[*L* , *LU*]	27	[*DL* , *R*]	43	[*LD* , *L*]
12	[*L* , *LD*]	28	[*DL* , *LU*]	44	[*LD* , *LD*]
13	[*D* , *L*]	29	[*DR* , *R*]	45	[*RD* , *U*]
14	[*D* , *LD*]	30	[*DR* , *LU*]	46	[*RD* , *DR*]
15	[*D* , *DR*]	31	[*DR* , *RU*]	47	[*RD* , *R*]
16	[*D* , *LU*]	32	[*DR* , *DR*]	48	[*RD* , *RD*]

Table. 2 Connected recognition result of fundamental hand gesture set.

Gesture	Number of Gestures	Results	
		correct	Detection
UP	672	646	96.13%
RIGH	840	821	97.74%
LEFT	672	662	98.51%
DOWN	504	483	95.83%
UP_L	504	493	97.82%
UP_R	588	574	97.62%
DN_L	588	570	96.94%
DN_R	672	663	98.66%
L_UP	1008	1003	99.50%
R_UP	588	578	98.30%
L_DN	840	834	99.29%
R_DN	588	572	97.28%
Total	8064	7899	97.95%

These high recognition rates show the gesture set designed using 4 motion primitive is discernable, which in turn ensures the high recognition. Although there are 4 motion primitive HMM's, a variety forms of gestures can by created and built by combining the HMM's without training again. In this way, the gesture set is flexible for building new word set and subsequent usage.

6. CONCLUSION

In this paper, we explored the American Sign Language and existing hand signal schemes to find the appropriate motion primitives. The 4 motion primitives were determined exhibiting discernability. Two relevant experiments were conducted for performance evaluation. By combining the proposed motion primitives, we verified the discernability and flexibility of designing hand gestures using the 4 motion primitives.

In the future, this work can be extended to real robot system for an operator to interact with robot via user defined gesture set.

7. ACKNOWLEDGEMENT

This research work was supported in part by the Bell lab Seoul (WR010951) project and in part by the ONR grant (N000141010068).

REFERENCES

[1] Lee, J.T. and T.L. Kunii, "Model-Based Analysis of Hand Posture.", IEEE Computer Graphics and Applications, vol. 15, no. 5, pp. 77-86, Sep 1991.

[2] James, D. and S. Mubarak, "Recognizing Hand Gestures.", European Conference on Computer Vision, pp. 331-340, Stockholm, Sweden, May 1994.

[3] Starner, T. and A. Pentland, "Real-time American Sign Language recognition from video using hidden Markov models.", International Symposium on Computer Vision, pp. 265-270, Coral Gables, FL, USA, Nov 1995.

[4] Christian, V. and M. Dimitris, "Parallel hidden markov models for american sign language recognition.", International Conference on Computer Vision, pp. 116-122, Kerkyra, Greece, Sep 1999.

[5] Just, A. and S. Marcel, "A comparative study of two state-of-the-art sequence processing techniques for hand gesture recognition.", Computer Vision and Image Understanding, vol. 113, no. 4, pp. 532-543, Apr 2009

[6] Yanghee, N. and W. KwangYun, "Recognition of Space-Time Hand-Gestures using Hidden Markov Model.", ACM symposium on Virtual reality software and Technology, pp. 51-58, Hong Kong, China, 1996.

[7] Lee, H.K. and J.H. Kim, "An HMM-based threshold model approach for gesture recognition.", Ieee Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no.10, pp. 961-973, Oct 1999.

[8] Eickeler, S., A. Kosmala, and G. Rigoll, "Hidden Markov Model based continuous online gesture recognition.", Fourteenth International Conference on Pattern Recognition, vol. 1, no. 2, pp. 1206-1208, Aug 1999

[9] Schlömer, T. et al, "Gesture Recognition with a Wii Controller.", Proceedings of the Second International Conference on Tangible and Embedded Interaction (TEI'08), pp. 11-14 Bonn, Germany. Feb 2008.

[10] Park, H.S., et al., "HMM-based gesture recognition for robot control.", Pattern Recognition and Image Analysis, Pt 1, Proceedings, vol 3522, pp. 607-614, 2005

[11] Kang, H., W.L. Chang, and K.C. Jung, "Recognition-based gesture spotting in video games.", Pattern Recognition Letters, vol. 25, no. 15, pp 1701-1714, 2004.

[12] Kortenkamp, D., E. Huber, and R.P. Bonasso, "Recognizing and interpreting gestures on a mobile robot.", Proceedings of the Thirteenth National Conference on Artificial Intelligence and the Eighth Innovative Applications of Artificial Intelligence Conference, vol. 1, no. 2, pp. 915-921, 1996

[13] McNeill, D., "Hand and mind : what gestures reveal about thought.", Chicago: University of Chicago Press, pp. 117, 1992.

[14] Baum, L.E., "An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes.", Inequalities, vol. 3, no. 9, pp. 1-8, 1972.

[15] Rabiner, L.R., "A Tutorial on Hidden Markov-Models and Selected Applications in Speech Recognition.", Proceedings of the IEEE, vol. 77, no. 2, pp. 257-286, 1989.

[16] Dreuw, P., et al., "Speech Recognition Techniques for a Sign Language Recognition System.", Interspeech, pp. 705-708, Antwerp, Belgium, Aug 2007.

[17] L.Neitzel, R., N.S. Seixas, and K.K. Ren, "A review of crane safety in the construction industry.", Applied Occupational and Enviromen-tal Hygiene, vol. 16, no. 12, pp 1106-1117, 2001.