

수신호 인식을 이용한 로봇 사용자 제어 시스템

Robot User Control System using Hand Gesture Recognizer

손수원, 배정훈, 양철중, 왕한, 고한석*

(Suwon Shon¹, Jounghoon Beh², Cheoljong Yang¹, Han Wang¹, and Hanseok Ko¹)

¹Korea University

²University of Maryland, College Park, USA

Abstract: This paper proposes a robot control human interface using Markov model (HMM) based hand signal recognizer. The command receiving humanoid robot sends webcam images to a client computer. The client computer then extracts the intended commanding human's hand motion descriptors. Upon the feature acquisition, the hand signal recognizer carries out the recognition procedure. The recognition result is then sent back to the robot for responsive actions. The system performance is evaluated by measuring the recognition of '48 hand signal set' which is created randomly using fundamental hand motion set. For isolated motion recognition, '48 hand signal set' shows 97.07% recognition rate while the 'baseline hand signal set' shows 92.4%. This result validates the proposed hand signal recognizer is indeed highly discernable. For the '48 hand signal set' connected motions, it shows 97.37% recognition rate. The relevant experiments demonstrate that the proposed system is promising for real world human-robot interface application.

Keywords: HMM, hand gesture, robot control system

I. 서론

최근 손동작을 이용한 로봇과 사람간의 인터페이스에 대한 다양한 방법이 개발되어왔다. 이에 따라 다양한 입력장치와 카메라를 이용한 연구가 진행되어 왔다[1-6]. 그 중 인식기에 가장 많이 사용되는 방법은 HMM (Hidden Markov Model)를 기반한 인식기이다. 이는 시간의 변화에 따른 신호를 통계적으로 모델링 할 수 있어 제스처 인식에서의 그 효율성이 증명되었다[7].

손동작 인식관련 연구분야에서는 고립수신호 인식[8,9]과 연속 수신호 인식[10,11]의 논문에서 제안한 인식기들은 모두 기존에 존재하는 문자나 기호, 혹은 자신들이 임의로 정의한 손동작 집합을 이용하였고, 그 종류의 가지 수도 제한적이다. 기존의 논문들에서 제안한 인식기를 쓰더라도 만약 수신호가 바뀌게 된다면 인식률을 예상하기 힘들 것이다. 또한 제스처가 늘어난다면 더욱이 인식률을 보장하지 못 할 것이다. 실제로 로봇 시스템에 적용하기 위해서는 로봇의 다양한 목적과 용도에 맞게 명령 체계 및 수신호의 구조를 변경할 수 있는 유연한 시스템이 필요하다. 하지만 수신호의 구조가 변경된다는 것은 새로운 수신호가 생기는 것이므로 그에 맞는 새로운 database를 모아 다시 훈련을 해야 한다는 번거로움이 있다. 따라서 실제 로봇을 제어하는 사용자 입장에서 사용하기 용이한 유연한 시스템을 위해 해결책이 필요하다.

본 논문에서는 영상입력과 HMM 기반의 수신호 인식기를

이용한 로봇 사용자 제어 시스템을 제안한다. 로봇과 client computer 사이의 네트워크 통신을 이용하여 인식이 이루어진다. 로봇에서는 단지 캠으로 촬영된 영상을 전송만 하면 인식은 모두 client computer에서 이루어지는 것이다. 이러한 시스템은 로봇에 부하되는 연산량을 줄이고, 또한 로봇 플랫폼에 독립적인 장점이 있다. 또한 인식기에서는 기본 손동작을 모델을 조합하여 새로운 수신호를 생성할 수 있다. 이러한 방법은 새로운 수신호를 만들 때, 훈련을 위하여 새로 database를 모을 필요가 없다는 장점이 있다. 수신호 인식기는 그 필요와 활용에 따라 수신호의 가지 수 및 동작이 다른데, 이러한 장점을 통해 database를 다시 모아 훈련을 하여 인식기를 구성할 필요 없이 기본적인 손동작들을 조합하여 이론적으로는 무한대 가지 수의 수신호를 만들 수 있는 것이다.

제안한 시스템의 성능 평가를 위해 구별성 있는 기본 손동작의 무작위 조합을 통해 50여개의 수신호 집합을 생성한다. 50여 개의 수신호 집합을 훈련하여 그에 대한 인식률을 얻고 또한, 기본 손동작 모델의 조합을 통해 50여개의 수신호 집합에 대한 인식률을 평가한다. 그리고 비교 대상을 위해 기존의 문자와 기호, 숫자를 이용한 수신호 집합을 만들어 성능을 평가하여 두 집합을 비교하여 제안한 시스템의 우수성을 판단한다.

본 논문의 나머지 부분은 다음과 같이 구성된다. II 장에는 이전까지의 연구를 소개하고 III 장에서는 본 논문에서 제안하는 영상기반의 HMM 수신호 인식기 시스템에 대한 소개가 있다. IV 장에서는 수신호 집합 생성에 대해서 설명한다. V 장에서는 인식률을 바탕으로 한 성능평가가 이루어지며 VI 장에서 결론을 맺는다.

II. Background

1. 이전까지 연구

이전까지 손동작 패턴 설계 방법에 대해 연구를 진행하였

* 책임저자(Corresponding Author)

논문접수: 2011. 1. 27., 수정: 2011. 2. 25., 채택확정: 2011. 3. 14.

손수원, 양철중, 왕한, 고한석: 고려대학교 전기전자전파공학과
(swshon@ispl.korea.ac.kr/cjyang@ispl.korea.ac.kr/hanwang@ispl.korea.ac.kr/hsko@korea.ac.kr)

배정훈: University of Maryland, Institute for Advanced Computer Studies
(jhbeh@umiacs.umd.edu)

* 본 연구는 서울시 산학협력사업(WR080951)에 의하여 연구되었음. ONR(N000141010068)의 부분지원에도 감사를 표합니다.

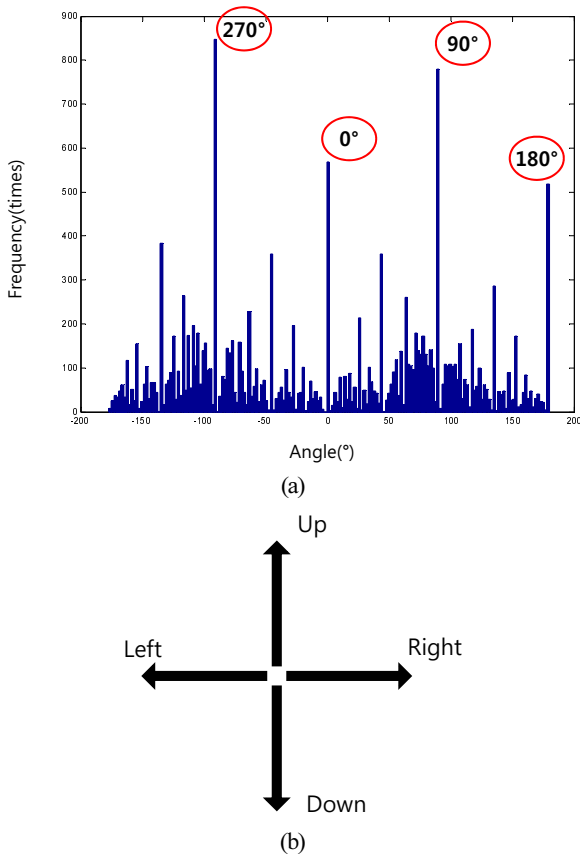


그림 1. (a) “RWTH-BOSTON-104” 데이터베이스의 손 좌표의 각도 통계 결과(b) 네 방향 모션 프리미티브.
 Fig. 1. (a) Statistical result of angle in hand position coordinates of “RWTH-BOSTON-104” database (b) Four direction motion primitive.

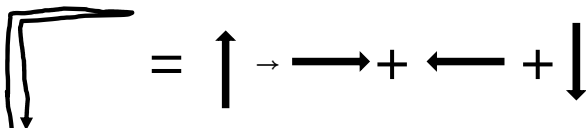


그림 2. 4개의 모션 프리미티브를 이용한 손동작 UR 조합의 예.
 Fig. 2. Example of UR hand motion combination using 4 motion primitives.

다[12]. 이는 사람의 음성이 구분 가능한 최소 단위인 음소의 조합으로 이루어지는 것을 손동작에도 적용하여, 손동작의 기본이 되는 구분 가능한 최소단위인 모션 프리미티브를 찾는 연구이다. 손동작을 분석하기 위해서 Boston university에서 공개한 미국 수화 (American Sign Language) DB를 분석하였다 [13]. DB는 총 201개의 미국 수화 문장으로 구성되어있으며, 같이 제공된 양손의 좌표를 이용하여 시간에 따른 각각의 손의 좌표의 점과 점 사이의 벡터가 이루는 각도를 구하여 그림 1(a)와 같은 통계를 얻었다. 이를 통해 0°, 90°, 180°, 270°를 이루는 벡터가 많은 것을 알 수 있고, 이 4가지 방향을 그림과 같이 모션 프리미티브로 정의하였다.

손동작의 기본이 되는 모션 프리미티브를 조합하여 기본 손동작 집합을 생성하였다. 예를 들면 그림 2의 왼쪽과 같은

표 1. 모션 프리미티브를 이용하여 만든 12가지 ‘기본 손동작 집합’.

Table 1. 12 class ‘fundamental hand motion set’ using motion primitives.

Symbol	U	R	L	D
Gesture				
Symbol	UL	UR	DL	DR
Gesture				
Symbol	LU	RU	LD	RD
Gesture				

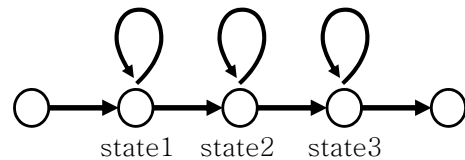


그림 3. HMM 모델.
 Fig. 3. HMM model.

손동작을 생성하고 싶다면 4개의 모션 프리미티브의 조합을 통해 만들 수 있다. 이 손동작의 이름을 UR이라 하고 이를 식으로 나타내면 식 (1)과 같다.

$$UR = [UP, RIGHT, LEFT, DOWN] \quad (1)$$

모션 프리미티브를 2개 혹은 4개를 이용하여 표 1과 같이 12개의 기본 손동작을 만들고 인식을 실험을 통해 성능을 평가한 결과 99.01%의 높은 인식률을 보였다. 이를 통해 기본 손동작이 높은 구별성을 가지는 것을 알 수 있었으며, 모션 프리미티브를 이용하여 조합하는 방법의 우수성을 알 수 있었다. 또한 모든 손동작의 시작점과 끝점을 시작점으로 돌아오게 하였기 때문에 연속 수신호 인식에 있어 중요한 문제인 비 손동작(non-gesture) [10] 구분하는 문제를 없앨 수 있을 것이다. 이러한 구별성을 보장하는 기본 손동작을 조합하여 HCI (Human Computer Interaction) 명령체계에 적합한 수신호를 만든다면 구별성이 보장되어 높은 인식률을 보일 것이다.

2. HMM

손동작을 모델링 하는 데에는 HMM이 쓰인다. HMM은 시변적인 신호를 통계적으로 모델링 할 수 있어서 음성인식을 비롯하여 HCI 및 정보과학 등 다양한 분야에 적용되어 왔다. 손동작 모델링을 위해서 다음과 같은 simple left to right 구조의 HMM을 이용하며, 위의 12가지 기본 손동작의 인식률 실험을 통해 최적의 state와 mixture가 각각 9와 4인 것을 알 수

있었다. 모델을 구성하는 각 파라미터는 기본적인 HMM의 이론과 동일하다.

III. 로봇 사용자 제어 시스템

로봇에서의 연산량을 줄이기 위해 robot-client computer 간의 네트워크 인식을 구현하였고 MFC (Microsoft Foundation Class library)를 이용하였다. 전체 시스템은 그림 4와 같다. 로봇에 달려있는 웹캠을 통해 영상이 입력되면 로봇은 이를 client computer로 바로 전송한다. 무선 랜 전송 규격은 802.11g이며 최대속도는 54Mbps이다. 처리속도를 위해 160x120 픽셀의 영상을 전송하며 웹 캠의 영상은 조명 및 환경변화에 따라 같은 색이라도 밝기차이가 있지만 최적의 환경을 제공하기 위해 입력영상 픽셀의 마우스 클릭을 통한 색상 채색정보로 조명 및 환경변화에 대응하였다. Client Computer에서 인식 과정 연산이 되어 인식결과 index만 로봇에게 다시 넘겨주고, 로봇에서는 index에 따른 동작을 실행하게 된다.

영상 입력을 통해 제스처 인식을 위해 특수효과 영상 촬영에 쓰이는 초록색 장갑을 이용하였다. 이는 연산량을 획기적으로 줄일 수 있으며, 속도를 위해 160x120의 매우 작은 크기의 영상을 사용하기 때문에 픽셀의 색상정보가 아닌 형태의 특징을 추출하여 인식하는 알고리즘에 부적합하였다. 초록색 장갑의 픽셀 영역의 중심점을 구하여 손의 x-y 좌표를 구하고 통해 특징을 추출한다. 특징 벡터는 그림 5와 같이 시간에 따른 궤적의 각도 (θ)와 프레임 당 픽셀거리 (l), 총 2

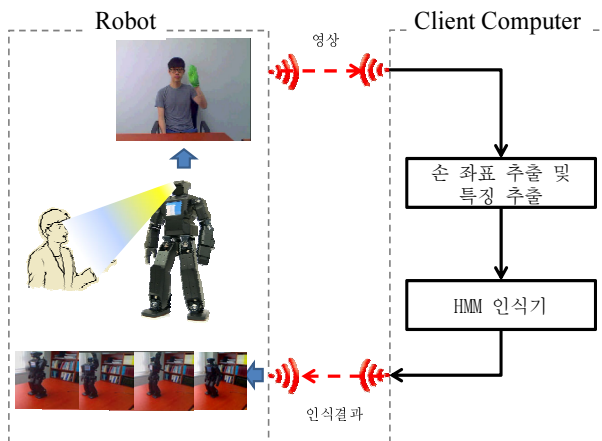


그림 4. 로봇 사용자 제어 시스템.

Fig. 4. Robot user control system.

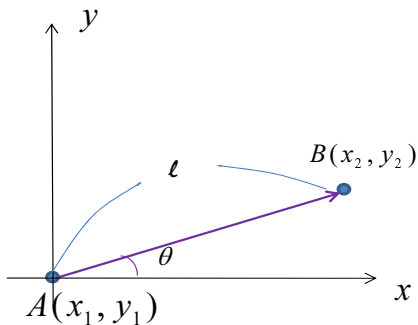


그림 5. 특징 추출.

Fig. 5. Feature extraction.

개의 특징으로 이루어지며 HMM 기반의 손동작 인식기에 적용된다[7].

인식기는 HMM을 이용하여 단일 인식기와 연속 인식기 두 가지를 설계하였다. 실험에 쓰인 휴머노이드 로봇 사양은 CPU는 임베디드 AMD 지오드 LX800 500MHz이며 메모리는 DDR 512MB 저장공간은 CF메모리카드를 이용한다. OS는 Windows XP이고, 640x480 해상도를 지원하는 130만 화소 웹캠이 달려있다.

IV. 수신호 집합 생성 및 인식

1. 48가지 수신호 집합

HCI 명령어 체계에 적합한 손동작 집합을 위해 위의 “기본 손동작 집합”의 12가지 기본 손동작들을 조합하여 최종적인 수신호 집합을 생성하였다. 12가지의 기본 손동작 중에서 2개를 선택하여 이어 붙이면 총 144개의 손동작을 생성 가능하다. 임의의 손동작 집합을 생성하기 위하여, 이어 붙이는 2개의 손동작 중 앞에 오는 손동작을 한 종류 당 4개씩 오도록 고정시킨 후 이어지는 동작을 임의로 선택하여 표 2와 같이 총 48개의 손동작을 생성하였고 ‘48가지 수신호 집합’이라 하고 이를 인식하는 인식기를 ‘48가지 수신호 집합’ 인식기라 한다.

2. Baseline 수신호 집합

비교실험으로 사용하기 위하여 정의한 수신호 집합으로, 모션 프리미티브의 조합을 이용하지 않은 손동작 집합을 정의하였다. 알파벳의 일부와 숫자, 그리고 간단한 기호를 이용해 표 3과 같이 손동작을 정의했다. 구성은 숫자(1, 2, 3, 4, 5, 6, 7, 8, 9)와 문자(A, B, C, D, X, Y, Z) 그리고 임의의 기호(↑, ↓, ←, →, @)로 구성되고 이것을 ‘baseline 수신호 집합’이라 하고 이를 인식하는 인식기를 ‘baseline 수신호 집합’ 인식기라 한다.

3. 인식

다음은 표에서 index 21번 수신호에 대한 예시를 보여준다. UR과 R을 이용하여 그림 6의 오른쪽과 같이 하나의 수신호를 생성할 수 있다. 이와 같이 하나의 수신호 자체를 훈련하

표 2. 48가지 수신호 집합.

Table 2. 48 hand signal 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]

표 3. Baseline 수신호 집합.

Table 3. Baseline hand signal set.

Symbol	1	2	3	4	5	6
Gesture	1	2	3	4	5	6
Symbol	7	8	9	A	B	C
Gesture	7	8	9	A	B	C
Symbol	D	X	Y	Z	↑	↓
gesture	D	X	Y	Z	↑	↓
Symbol	←	→	@			
gesture	←	→	@			

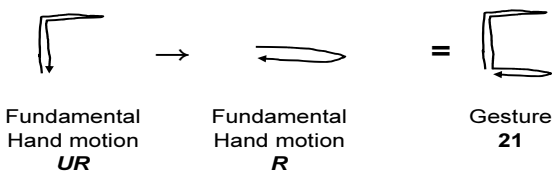


그림 6. 단일 수신호 모델의 예.

Fig. 6. Example of isolated hand signal model.

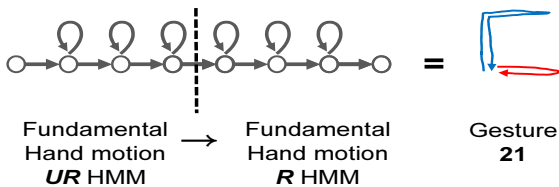


그림 7. 연결 수신호 인식의 예.

Fig. 7. Example of connected hand signal recognition.

여 인식하는 것이 일반적인 단일 수신호 인식이다.

연결 수신호 인식의 경우 이미 훈련된 기본 손동작의 조합을 이용하여 인식하는 것이다. 그림 7은 연속 수신호 인식의 경우 기본 손동작 HMM 조합의 예시이다. 그림과 같이 수신호 집합을 새로 생성하더라도 기존에 있는 기본 손동작 모델의 조합을 이용하여 인식할 수 있다.

V. 실험

실험에 쓰인 기본 손동작 집합에 대한 데이터베이스는 28명이 12개의 모델을 3번씩 촬영하여 모았다. ‘48가지 수신호 집합’에 대한 데이터베이스는 28명이 48개의 모델을 3번씩 촬영하여 모았다. ‘baseline 수신호 집합’에 대한 데이터베이스는 22명의 사람들에게 대해 ‘A, B, C, D, X, Y, 1, 2, 3, 4, 5, 6, 7, 8, 9 →, ←, ↑, ↓, @’의 글자를 3번씩 촬영하여 얻었다. 기본 손동작 집합의 인식률을 최대화 하는 state와 mixture는 이

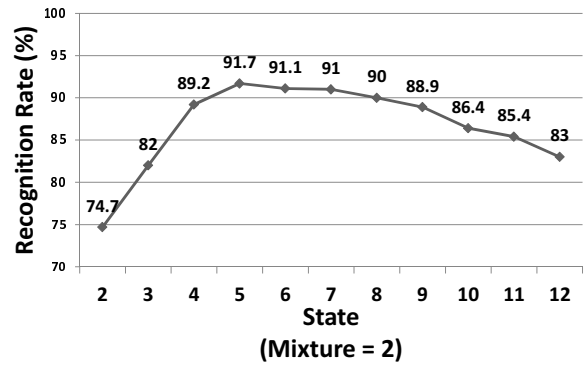


그림 8. State 수에 따른 인식률 변화.

Fig. 8. Recognition rate corresponding state number.

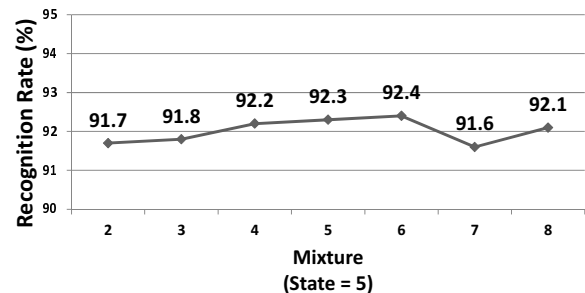


그림 9. Mixture 수에 따른 인식률 변화.

Fig. 9. Recognition rate corresponding mixture number.

전의 연구를 통해 각각 9와 4인 것을 알았다. 따라서 기본 손동작의 2개를 조합한 ‘48’가지 수신호 집합의 경우 state와 mixture를 각각 기본 손동작 집합의 두 배인 18과 8로 놓았다. 실험은 ‘baseline 수신호 집합’의 인식률을 최대화 하는 최적의 state와 mixture를 찾고 두 집합의 인식률을 비교하며 추가로 연결 수신호 인식을 통해 단일 수신호 인식 성능과 비교한다. 실험에 쓰인 인식률을 얻는 방법은 ‘Leave-one-out cross-validation’ 방법을 적용하였다. 즉, 기본 손동작 집합과 ‘48가지 수신호 집합’에 대해서 27명의 데이터베이스를 훈련용으로 이용하고 남은 1명의 데이터베이스를 인식용으로 쓰는 방법으로 총 28명을 돌아가며 인식률을 내어 평균을 얻었다. ‘baseline 수신호 집합’에 대해서 21명의 데이터베이스를 훈련용으로, 1명의 데이터베이스를 인식용으로 쓰는 방법으로 총 22명을 돌아가며 인식률을 내어 평균을 얻었다.

1. 수신호 인식 task 위한 최적의 mixture, state 평가

Baseline 손동작 집합의 최적 mixture와 state를 찾는 실험이다. 첫 번째로 mixture를 2로 고정하고 state 수의 변화 따른 인식률의 변화를 보았다. 그림 8과 같이 ‘baseline 수신호 집합’의 경우 state가 5일 때 91.7%로 최대의 인식률을 보였다.

각각의 인식률이 최대로 나온 state로 고정시킨 후 mixture를 변화시켜가며 실험해 보았다. 그 결과 그림 9과 같이 ‘baseline 수신호 집합’은 mixture가 6일 때 92.4%로 최대의 인식률을 보였다.

2. 단일 수신호 인식 task에 대한 각 인식기 성능 평가

두 번째 실험은 ‘48가지 수신호 집합’의 인식률을 알아보고, 기존의 문자와 숫자를 이용한 ‘baseline 수신호 집합’의 인식률을 비교하기 위한 실험이다. 실험환경은 위에서 얻은

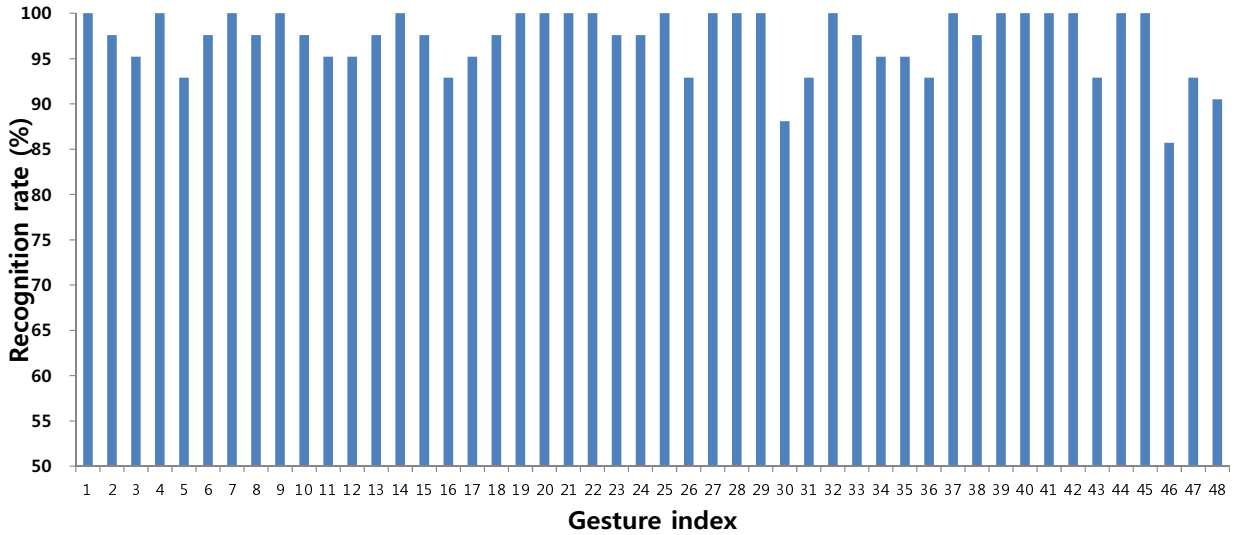


그림 10. ‘48가지 수신호 집합’ 인식기의 인식률.
Fig. 10. Recognition rate of ‘48 hand signal set’ recognizer.

각 손동작 집합 별 최적의 state와 mixture에서의 인식률을 구하였다.

2.1 ‘48가지 수신호 집합’ 인식기

기본 손동작 집합의 12가지 모델의 조합을 통해 생성한 수신호 집합의 DB를 이용해 총 48개의 모델을 생성하여 인식률을 구하였다. 그 결과 인식률을 97.07%가 나왔고 인식률을 그림 10과 같다.

2.2 ‘Baseline 수신호 집합’ 인식기

‘Baseline 수신호 집합’의 21가지 모델에 대한 총 인식률은 92.4%로 각 모델에 대한 인식률은 그림 11과 같다.

‘baseline 손동작 집합’의 각 모델 별 인식률을 살펴보면 인식률이 비교적 높은 모델들은 90% 대에 있지만 특정 모델들에 대해서, 특히 ‘1’과 ‘9’, ‘C’ 등의 모델에서 현저하게 낮은 인식률을 보이는 것을 확인할 수 있다.

3. 연결수신호 인식 task에 대한 인식기 성능 평가

수신호 집합은 기본 손동작 집합의 조합을 통하여 생성되었으므로 그림 12와 같은 기본 손동작 집합의 HMM의 네트워크를 통하여 수신호 집합(인식기 48)을 인식할 수 있다. 이

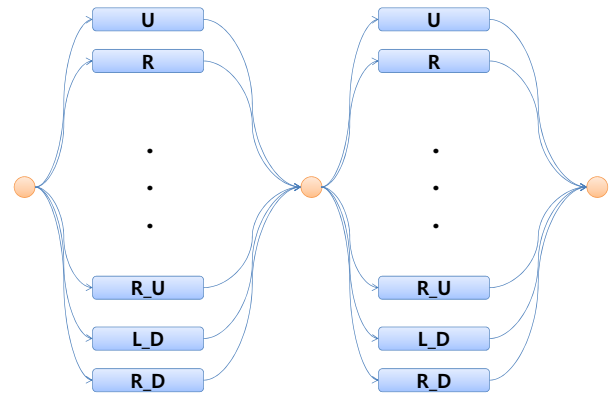


그림 12. 연결 수신호 인식의 워드 네트워크.
Fig. 12. Word network of connected hand signal recognition.

표 4. ‘48 가지 수신호 집합’의 연결 수신호 인식 결과.

Table 4. Result of ‘48 hand signal set’ connected hand signal recognition.

Gesture	Number of Gestures	Results	
		correct	detection
UP	672	671	99.85%
RIGH	840	823	97.98%
LEFT	672	666	99.11%
DOWN	504	498	98.81%
UP_L	504	490	97.22%
UP_R	588	588	100.00%
DN_L	588	588	100.00%
DN_R	672	632	94.05%
L_UP	1008	942	93.45%
R_UP	588	588	100.00%
L_DN	840	840	100.00%
R_DN	588	526	89.46%
total	8064	7852	97.37%

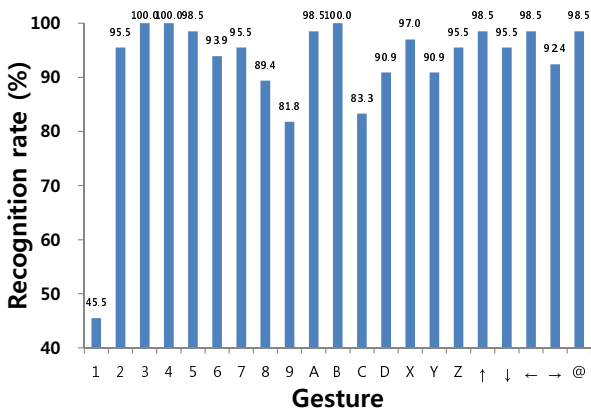


그림 11. ‘Baseline 수신호 집합’ 인식기의 인식률.
Fig. 11. Recognition rate of ‘baseline hand signal set’ recognizer.

는 기본 손동작의 끝점을 시작점으로 다시 돌아오게 설계하였기 때문에 특별한 연결 손동작 인식 알고리즘이 없더라도 연결 손동작 인식에 문제가 없을 것이다. 실험결과 총 인식률은 97.37%이고 각각의 모델 별 총 손동작 개수와 인식률은 앞의 표 4와 같다.

4. 토의

수신호 인식 task를 위한 최적의 mixture, state 평가 실험을 통하여 'baseline 수신호 집합'의 최적의 state는 5개, mixture는 6개인 것을 확인하였다. 단일 수신호 인식 task의 각 인식기 성능 평가 실험의 결과를 보면 기존 문자와 숫자, 기호를 이용한 'baseline 수신호 집합'을 이용한 실험에서의 인식률은 92.4%로 비교적 낮은 결과를 보였다. 특히 특정 모델에서 현저하게 낮은 인식률을 보였는데 이는 기존 문자를 이용한 손동작 집합이 구별성을 가지 못하고 하나의 모델이 다른 모델의 부분집합이 되어버리는 subset problem으로 설명할 수 있다. 그림 13의 예와 같이 숫자 1은 B, D, ↓의 부분집합이 되고 문자 C는 9의 부분집합이 되어서 오인식이 일어나게 된다.

이를 모션 프리미티브를 이용하여 만든 손동작 집합과 비교한다면 '기본 손동작 집합'과는 7%, '48가지 수신호 집합'과는 5%의 차이로 'baseline 수신호 집합'이 더 낮은 인식률을 보인다. 특히 그림 10의 각 모델 별 인식률을 보면 subset problem이 일어나지 않고 대부분의 인식률이 90% 후반인 것을 알 수 있다. 또한 '48가지 수신호 집합'이 'baseline 수신호 집합'에 비해 2배 정도 많은 모델을 가지고 있음에도 불구하고 인식률이 더욱 높은 것을 알 수 있다. 따라서 모션 프리미티브를 이용하여 손동작을 만드는 것이 손동작간에 큰 구별성을 갖는다는 것을 알 수 있다.

또한 이 실험에서 모션 프리미티브를 기반으로 만든 두 개의 손동작 집합을 비교할 수 있다. 12개의 모델로 구성되어 있는 '기본 손동작 집합'의 인식률은 99.01%이고 48개의 모델로 구성되어있는 '48가지 수신호 집합'의 인식률은 97.07%이다. 조합을 통해 12개의 모델에서 48개의 모델로 모델의 개수가 4배로 많아졌지만 인식률은 1.94%의 비교적 적은 하락을 보인다. 이로써 모션 프리미티브의 조합으로 구별성이 보장된 손동작 집합을 이용하여 한번 더 조합한 손동작 집합도 역시 구별성을 보장하는 것을 보여주고 있다.

연결 수신호 인식 task의 각 인식기 성능평가 실험에서는 '기본 손동작 집합' HMM의 단순한 네트워크를 통하여 '48가

지 수신호 집합'의 손동작을 연결 인식 하였다. 손동작의 구조가 끝점이 모두 시작점으로 돌아오므로 다음 손동작이 이어지더라도 비손동작 부분이 생기지 않기 때문에 실험결과와 같이 97.37%의 높은 인식률을 보였다. 이는 '48가지 수신호 집합'의 단일 수신호 인식 결과인 97.07% 보다 더 높은 수치임을 알 수 있다.

VI. 결론

본 논문에서는 실제 로봇 시스템에 적용하기 위해 수신호를 변경 및 확장의 중요성을 인지하여, HMM 기반의 수신호 조합을 이용한 로봇 사용자 제어 시스템을 제안하였다. 임의의 수신호 집합을 생성하여 인식률 실험을 통해 시스템의 성능을 검증하였다.

그 결과 모션 프리미티브를 바탕으로 생성된 기본 손동작 집합을 이용하여 임의로 생성한 48가지 손동작 집합의 인식률은 97.07%로 높게 나타났다. 기존 문자를 이용한 'baseline 수신호 집합'의 인식률인 92.4%에 비해 7% 더 높은 인식률을 보였다. 끝점이 시작점으로 돌아오는 구조를 통한 손동작은 두 단어의 연결 손동작 인식 실험을 통해 97.37%의 인식 결과를 보여 로봇에 연속적으로 명령을 내릴 수 있는 것을 확인하였다. 이러한 실험 결과들은 로봇의 사용자 제어를 위해 새로운 손동작을 만들거나 그 손동작이 많아지더라도 충분히 구별성이 보장되는 인식률이 높은 로봇 사용자 제어 시스템을 구성할 수 있는 근거가 된다.

향 후, 연속 인식을 통하여 명령어의 체계를 성립하여 주어, 동사, 목적어의 형태를 갖는 명령 체계를 만들고, 언어 모델을 적용시킨다면 더욱 정교한 제어를 하며 높은 인식률을 갖는 로봇 사용자 제어 시스템을 구성할 수 있을 것이다.

참고문헌

- [1] T. Schlömer, et al. "Gesture recognition with a wii controller," *Proc. of the Second International Conference on Tangible and Embedded Interaction (TEI'08)*, Bonn, Germany. pp. 11-14, Feb. 2008.
- [2] H. S. Park, et al., "HMM-based gesture recognition for robot contro," *Pattern Recognition and Image Analysis, Pt 1*, vol. 3522, pp. 607-614, Jun. 2005.
- [3] H. Kang, W. L. Chang, and K. C. Jung, "Recognition-based gesture spotting in video games," *Pattern Recognition Letters*, vol. 25, no. 15, pp. 1701-1714, Nov. 2004.
- [4] D. Kortenkamp, E. Huber, and R. P. Bonasso, "Recognizing and interpreting gestures on a mobile robot," *Proc. of the Thirteenth National Conference on Artificial Intelligence and the Eighth Innovative Applications of Artificial Intelligence Conference*, vol. 2, pp. 915-921, Aug. 1996.
- [5] J. Y. Oh and C. W. Lee, "Survey: Gesture recognition techniques for intelligent robot," *Journal of Control, Automation and System Engineering(in Korean)*, vol. 10, no. 9, pp. 771-778, Sep. 2004.
- [6] I. M. Kim, W. C. Kim, K. S. Yun, and J. M. Lee, "Navigation of a mobile robot using hand gesture recognition," *Journal of Control, Automation and Systems engineering(in Korean)*, vol. 8, no. 7, Jul. 2002.
- [7] Y. Wu and T. S. Huang, "Vision-based gesture recognition: A

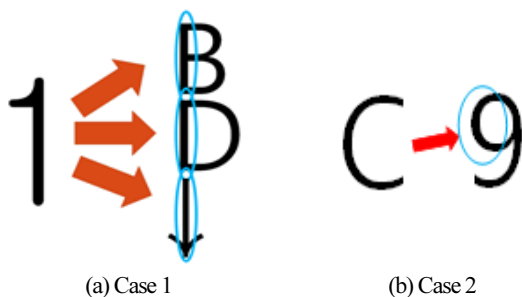


그림 13. Baseline 손동작 집합에서 오인식이 일어나는 예.
(a) 경우 1 (b) 경우 2.

Fig. 13. Recognition error example in baseline hand signal set (a) case 1 (b) case 2.

- review,” *Gesture-Based Communication in Human-Computer Interaction*, vol. 1739, pp. 103-115, 1999.
- [8] A. Just 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.
- [9] N. Yanghee and W. KwangYun, “Recognition of space-time hand-gestures using hidden markov model,” *VRST: ACM symposium on Virtual reality software and Technology*, Hong Kong, China, pp. 51-58, Jul. 1996.
- [10] H. K. Lee 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.
- [11] S. Eickeler, 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.
- [12] S. W. Shon, J. Beh, C. J. Yang, H. Wang, and H. S. Ko, “Hand motion design for performance enhancement of vision based hand signal recognizer,” *Journal of IEK, SP*, vol. 48, no. 4, Jul. 2011.
- [13] P. Dreu, et al., “Speech recognition techniques for a sign language recognition system,” *Interspeech, Antwerp, Belgium*, pp. 705-708, Aug. 2007.



손수원

2010년 고려대학교 전기전자전파공학부 학사 졸업. 2010년~현재 고려대학교 전기전자전파공학과 석·박사통합과정 재학중. 관심분야는 음성처리, 패턴인식.



배정훈

2001년 고려대학교 전기전자전파공학부 학사 졸업. 2003년 고려대학교 전자컴퓨터공학과 석사 졸업. 2008년 고려대학교 전자컴퓨터공학과 박사 졸업. 2010년~현재 University of Maryland에서 research associate로 재직중. 관심분야는 음향신호처리, 영상신호처리, human-robot interaction.



양철중

2010년 고려대학교 전기전자전파공학부 학사 졸업. 2010년~현재 고려대학교 영상정보처리협동과정 석·박사통합과정 재학중. 관심분야는 영상처리, 모션 추적, 제스처 인식.



왕한

2003년 ShenYang Ligong University 전자공학과 학사 졸업. 2007년 Northeastern University 전자공학과 석사 졸업. 2009년~현재 고려대학교 전자전기전파공학과 박사과정 재학중. 관심분야는 영상처리, 신호처리, HCI.



고한석

1982년 Carnegie Mellon Univ., 전기공학과 학사 졸업. 1986년 Maryland Colleague Park 시스템 공학과 석사 졸업. 1988년 Johns Hopkins Univ. 전기공학과 석사 졸업. 1992년 Catholic Univ. of America, 전기공학과 박사 졸업. 1995년~현재 고려대학교, 전기전자전파공학부 교수 재직중. 관심분야는 음향신호처리, 영상신호처리, 멀티모달리티를 이용한 human-robot interface.