

스테레오 영상 추적 자동초기화와 HMM을 이용한 원격 작업용 제스처 인식

정 지원[†] · 이 용 범^{††} · 진 성 일^{†††}

요 약

본 논문에서는 원격작업용 제스처 명령의 인식을 위해 컴퓨터 시각 센서와 HMM을 이용한 알고리즘을 사용하였다. 스테레오 정합을 이용한 자동 손영역의 추출은 원격작업용 제스처 추적의 자동 초기 영역 설정을 위해 제안하였다. 이는 좌, 우 두 카메라의 시각차를 이용하였으며, 손의 상대적인 거리정보를 이용하게 된다. 손영역의 추적은 일반적인 흑백 영상에서도 배경이나 잡음에 구애받지 않도록 예지값의 양의 시간차(PDOE : positive difference of edges) 영상이 사용되었으며, 제스처 명령 모델로는 실제 원자력 발전소 내 원격작업용으로 정의된 K/K(KNU/KAERI) 제스처 명령을 이용하였다. 이산 HMM을 이용한 복합 제스처 인식모델은 전명령어(pre-orders)와 기본명령어(basic-orders), 후명령어(post-orders)의 세 부분에 대한 각각의 명령 모델로 구성되며, 각 명령 모델의 조합과 연동에 의해서 복합적인 명령의 인식을 수행하도록 하였다. 최종 제스처 인식 결과는 2개 이상의 명령어 복합된 연속명령에 대해 97% 이상의 안정된 인식율을 나타내었다.

Gesture Recognition Using Stereo Tracking Initiator and HMM for Tele-Operation

Jee-Won Jeong[†] · Yong-Bum Lee^{††} · Sung-Il Chien^{†††}

ABSTRACT

In this paper, we describe gesture recognition algorithm using computer vision sensor and HMM. The automatic hand region extraction has been proposed for initializing the tracking of the tele-operation gestures. For this, distance informations(disparity map) as results of stereo matching of initial left and right images are employed to isolate the hand region from a scene. PDOE(positive difference of edges) feature images adapted here have been found to be robust against noise and background brightness. The KNU/KAERI(K/K) gesture instruction set is defined for tele-operation in atomic electric power stations. The composite recognition model constructed by concatenating three gesture instruction models including pre orders, basic orders, and post-orders has been proposed and identified by discrete HMM. Our experimental results showed that consecutive orders composed of more than two ones are correctly recognized at the rate of above 97%.

1. 서 론

제스처는 인간의 의도를 손이나 얼굴, 몸 등을 이용

하여 의사를 표현하는 대화 방법으로 오랫동안 의사전달의 보조수단이 되어왔다. 특히, 청각 장애자를 위한 수화의 경우는 제스처가 유일한 의사소통의 수단으로 사용되어진다. 최근에 사람과 컴퓨터와의 접속 기술(HCI : human-computer interaction)은 인공 지능 기술을 주축으로 하여 여러 관련 기술이 접목되어 사람들

† 정 회 원 : 델타정보통신(주) 기술연구소 연구원
†† 정 회 원 : 한국원자력기술연구소 로봇핵심기술개발실 실장
††† 정 회 원 : 경북대학교 전기전자공학부 교수
논문접수 : 1997년 11월 24일, 심사완료 : 1999년 6월 21일

이 대화하는 것과 유사한 기술의 개발에 중점을 두고 있다[1]. 이러한 제스처 입력 방식은 폭 넓고 자연스러운 컴퓨터와의 의사소통 방식의 한 방안이라 할 수 있다. 특히, 손 제스처 영상을 컴퓨터 상에서 해석하기 위해 어떻게 모델화 할 것인가에 관한 연구가 계속 있어왔다. 이러한 모델화를 위해 Kendon은 하나의 제스처를 준비(preparation), 스트로크(stroke), 철회(re-traction)로 구분하였으며, Queck[2]는 손이 움직이는 동안 손의 이동 궤적이 의미를 가지며, 손의 모양이나 손가락의 움직임 등은 거의 의미를 가지지 않는다는 연구결과를 얻었다. 손 제스처를 인식하는데는 크게 글로브를 낀 상태의 제스처 인식(glove-based), 컴퓨터 시각에 기반한 제스처의 인식(vision-based), 필기 상태의 제스처(drawing gesture) 인식으로 분류할 수 있다[1].

컴퓨터 시각에 기반한 제스처 인식에 관한 연구는 3차원 손모델에 기초한 해석[3], 마커(marker)나 색상정보를 이용한 해석[4][5], 영상 특성을 이용한 해석[6]으로 구분되어질 수 있다. 이러한 해석 능을 통해 획득된 손의 영상정보를 사용하여 제스처를 특정작업의 명령으로 모델링 하려면 이들 동작들이 시간에 따라 가변적일 수 있으며 환경이나 사람에 따라 달라지는 다양성을 고려하여야 한다. 이러한 경우에 적절히 대처할 수 있는 은닉 마르코프 모델(Hidden Markov Model)[7]을 이용한 제스처 인식[4][8][9]이 많이 연구되어져 오고 있다. 이들 제스처 인식모델에서는 한 제스처 명령이 한 명령어를 나타내는 단어 기반(word-based) 인식으로서 필요로 하는 명령어 늘어날 때마다 HMM 인식 모델의 개수가 계속 늘어나야 하고 제스처가 더욱 복잡해질 수 있다.

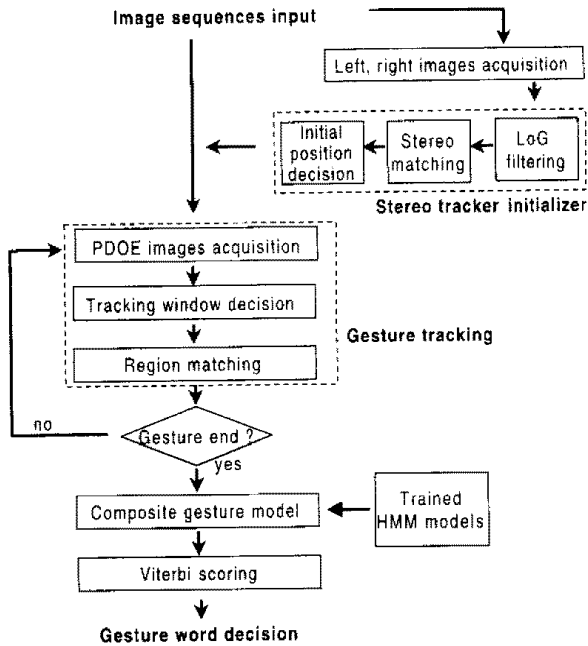
본 논문에서는 스테레오 정합을 이용하여 손영역을 자동으로 추출하고 제스처의 확장이 용이한 복합 제스처 인식모델을 제안하였다. 제스처 인식을 위해 손모양이나 이동궤적을 분석할 때 초기 손 위치의 설정이 필요하다. 따라서, 초기에 지정된 위치에 손을 위치시키거나 인위적으로 원격 작업자가 손 위치를 직접 설정 해 주어야 한다. 그러나 본 논문에서 사용된 손영역 추출은 이런 작업을 없애고 자동으로 손영역을 추출할 수 있도록 좌, 우 2대의 카메라를 사용한 스테레오 정합을 수행하였다. 사용된 정합법은 잡음에 둔감하고 빠른 처리시간을 가지는 특징점 기반 방식 알고리즘을 적용하였으며 손과 배경사이의 상대적인 거리

정보를 이용하기 때문에 사용자가 인위적으로 초기 위치설정을 할 필요 없이 좌, 우 두 영상을 이용함으로써 자동으로 초기위치를 설정하게 된다. 이때 구해지는 거리 정보인 이차원 변이영상을 가로축과 세로축의 일차원상으로 투영시켜 이들을 분석함으로써 최종적인 초기영역인 손영역을 자동으로 추출하게 된다. 그리고, 일반적인 흑백 카메라에서도 손영역을 안정되게 추적하기 위해서 물체 추적알고리즘을 적용하였다. 추적을 통해 획득된 제스처 정보는 음성이나 온라인 문자와 같은 가변적인 크기의 데이터 인식에 주로 사용되어지는 HMM을 이용하여 제스처 명령으로 인식하게 된다. 단어 기반 인식에서는 명령어가 늘어날 경우 그에 따른 HMM 제스처 모델 또한 명령어 개수만큼 계속 늘어나야 한다. 따라서 본 논문에 사용된 제스처 명령 형태는 명령을 전달할 경우 자세한 동작속도에 대한 설명이 표현된 전명령어(pre-orders)와 기본 동작에 대한 명령인 기본명령어(basic-orders), 원격 작업에서 쌍방향 정보 전송을 위해 의문문인지를 결정하게 되는 후명령어(post-orders)의 세 부분으로 나누었다. 그리고 이들을 조합함으로써 작은 수의 제스처 명령으로도 많은 의미의 제스처 명령으로 확장할 수 있는 K/K(KNU/KAERI) 제스처 명령을 구성하였다. 제안된 제스처 인식 모델은 각 세 부분에 대해 학습된 HMM 모델을 구성하여 이들의 복합적인 결합을 통해 전체적인 제스처 명령어를 인식하도록 하였다. 구성된 인식 모델은 기본 명령어를 중심으로 작은 수의 HMM 모델을 조합시켜 다양한 의미의 제스처 명령어에 대한 인식으로 확장되었다. 실제 실험에서는 연속된 2개 이상의 제스처 명령으로 이루어진 복합 명령에 대한 인식결과는 97%이상의 인식율을 나타내었다.

2. 시스템의 구성

본 논문에서 구현한 원격작업용 제스처 인식 시스템은 (그림 1)과 같은 구조를 가진다. 구현된 시스템은 손영역의 자동 추출, 손영역의 추적, 최종 제스처 인식의 세 부분으로 구성된다. 연속적으로 입력되는 제스처 영상에 대해 이들 세 영역에 대한 수행함으로써 원격 작업용 제스처를 인식하게 된다. (그림 1)에 나타난 스테레오 추적 초기화(stereo tracker initializer)에서는 두 대의 카메라로부터 입력되는 좌, 우 두 영상에 대해 스테레오 정합을 수행하여 초기 손영역을 추출한

다. 이렇게 추출된 초기 손영역 정보로써 PDOE 영상 추적 알고리즘을 수행하여 전체적인 손의 이동 궤적을 구한다. (그림 1)에 나타난 인식기에서는 HMM 인식 모델을 사용하여 입력된 8방향성 코드로써 연속적인 복합 제스처 명령어를 인식하게 된다.



(그림 1) 제안한 제스처 인식기의 구성도

3. 스테레오정합을 이용한 추적 초기화

3.1. 스테레오 정합을 이용한 제스처 초기 영역 설정

평행한 두 대의 카메라로부터 입력된 좌, 우 두 영상의 스테레오 정합은 잡음에 강하면서 빠른 연산을 위해 계산량이 적은 에지 정합 방식을 수행하였다.

LoG 필터에 의한 에지 추출은 영교차점(zero-crossing point)을 사용하지만 에지를 특징영상으로 할 경우 구해지는 최종 변이영상이 충분한 정보를 갖지 못할 수 있다. 따라서, 손영역을 찾을 만큼 거리 정보가 충분히 나타나지 않을 수 있다. 이러한 단점의 보완을 위해서 LoG 필터를 거친 영상의 영교차점 부분에서 음의 값을 가지는 영역도 정합을 위한 특징영상에 포함시켜 부분적으로 영역 정합 특성을 나타내게 하였다[10]. (그림 2)의 (a)에서는 초기영역 설정을 위한 좌, 우 두 영상을 나타내며, (b)에서는 LoG 필터를 거친 영교차점 영역과 음의 값 영역을 모두 나타낸 영상이다.



(a) 입력된 좌, 우영상



(b) 영역 특징 영상

(그림 2) 입력된 좌·우 두 영상과 영역 특징 영상

좌, 우 두 영상의 변이값을 구하기 위한 스테레오 정합에 사용된 SAD(sum of absolute differences)는

$$SAD(dx, dy) = \sum_{i,j} \sum_{i \in MW} |I_1(i, j) - I_2(i + \Delta x, j + \Delta y)| \quad (1)$$

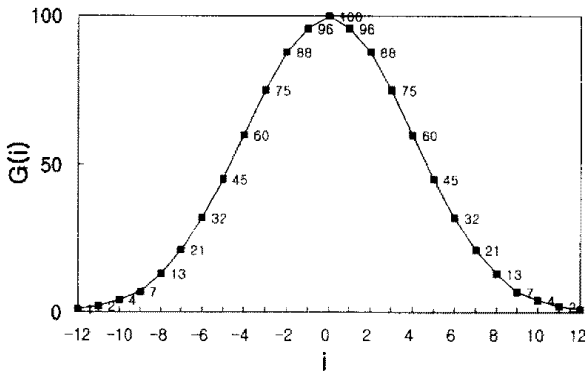
이다. 좌영상의 특징영상과 우영상의 특징영상 사이의 크기차의 비교로서 가장 작은 SAD값을 가지는 x, y축으로의 변이를 구한다. 이때 I_1 과 I_2 는 각각 우영상과 좌영상의 영역특징영상을 나타내고, MW는 좌, 우영상이 정합되는 창영역을 나타낸다.

구해진 x, y축으로의 변이값으로서 전체 변이영상을 구하게 되며, 그 값이 클수록 카메라와의 거리가 가깝고 거리가 멀수록 그 값이 작다. 구해진 최종 변이영상은 (그림 4)의 (a)에서 나타나듯이 손을 중심으로 높은 변이값이 몰려 있음을 나타내게 된다. 이 부분을 추출함으로써 우리가 원하는 초기 손영역을 구할 수 있다. 이와 같이 높은 변이값을 가지는 손영역의 추출을 위해 실험적 문턱치를 중심으로 이진화 시킨 영상을 (그림 4)의 (b)에서 볼 수 있다. 그리고, 이진화된 영상을 x, y축으로 투영시켜서 누적 분포를 구하게 된다. x, y축으로 투영된 누적치는 급격한 변화를 보여주므로 후보영역을 설정하기 위해서는 넓은 범위의 평탄화(smoothing)를 하여 영역을 구간화하는 과정이 필요

이다. 평탄화 방법에는 평균값이나 중심값을 이용하거나 가우시안 필터(Gaussian filter)를 이용한 방법 [11][12]등이 있다. 본 논문에서는 잡음 제거 등의 기본적인 평탄화 효과와 함께 중심값을 강조하는 효과를 같이 가지는 가우시안 평탄화를 사용하였다. 사용되어진 가우시안 필터는 화소의 중심에서 12화소를 반경으로 하여 평균이 0이고 σ 가 1인 정규분포성격을 가지는 함수에서 얻어진 값을 샘플링해서 사용하였다. 사용된 가우시안 필터는

$$G(i) = ke^{-\frac{i^2}{2\sigma^2}} \quad (2)$$

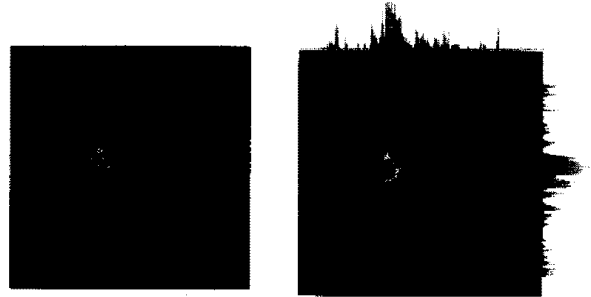
으로 구할 수 있다. 여기서 k는 임의의 상수이고 σ^2 은 분산값을 나타낸다. 사용되어진 필터는 (그림 3)과 같다.



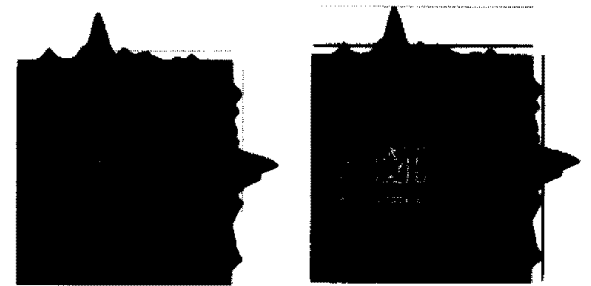
(그림 3) 1차원 투영영상을 평탄화 시키기 위한 가우시안 필터

평탄화 시킨 결과는 최댓값을 기준으로 정규화 시키며, 투영영상을 정규화 시킨 영상은 (그림 4) (c) 영상의 가로축과 세로축에 나타난 영상과 같다. 정규화된 데이터는 중에서 평균값 이상의 데이터 영역을 중심으로 다시 평균값을 취하여 이 값이상인 후보 영역을 탐색한다. (그림 4)의 (c), (d)는 x, y축으로의 후보 영역들 중 변이값의 합이 가장 높은 영역을 손영역으로 추출한 결과이다. 평탄화는 가우시안 필터를 사용하였다. 영상에서 후보영역을 정규화된 데이터는 값을 가지지 않은 영역이 많으므로 우선 평균값 이상을 가지는 데이터 값만을 고려한다. 고려된 데이터 값들을 평균하여 이를 기준으로 전 영역에 걸쳐 후보영역을 탐색한다. (그림 4)의 (d)에서는 탐색된 후보영역을 나타내고 있다. 이들 후보 영역 중 영역내에서 초기에 구해진 변이값들의 합이 가장 큰 영역을 선택하게 된다. (그림

4) (c)는 최종 선택된 손영역을 보여 주고 있다.



(a) 최종 변이영상 (b) 이진화와 1차원투영영상



(c) 투영영상의 smoothing영상 (d) 탐색된 후보 영역들



(e) 최종 설정된 초기영역

(그림 4) 변이영상을 이용한 제스처 자동 초기화 과정

3. PDOE 영상을 이용한 손영역 추적

추적을 수행하기 위한 특징영상은 PDOE 특징 영상을 이용한다. 이는 에지 검출과 차영상 획득으로 구성 되어 시공간적 이동의 개념이 있는 '에지값의 양의 시간차(PDOE)' 영상이 된다. 구해진 에지 영상의 차영상(DE)은

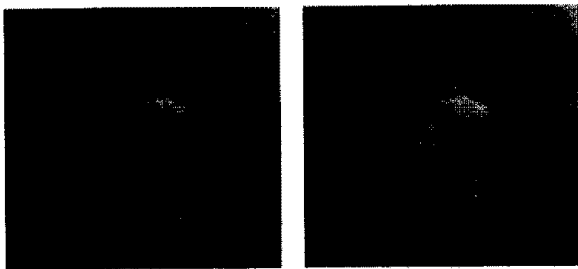
$$DE = I_c(n) - I_c(n-1) \quad (3)$$

로 획득할 수 있다. 여기서 $I_c(n)$ 은 n번째 프레임에서 8방향 Prewitt 연산자를 취한 에지 영상이 된다.

$$PDOE = U(DE)$$

$$U(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (4)$$

그러나, PDOE 영상은 식 (4)와 같이 DE값 중 양의 값만 취하게 된다. DE 크기영상은 (그림 5) (b)에서 나타나듯이 움직이는 물체의 현재와 이전 프레임의 에지성분을 모두 보여주고 있다. (c)는 PDOE 영상으로 (b)와는 달리 움직이는 손영역의 현재프레임의 에지성분만을 보여주고 있음을 알 수 있다.



(a) 실제 연속된 두 프레임 영상



(b) DE 크기영상 (c) PDOE 영상

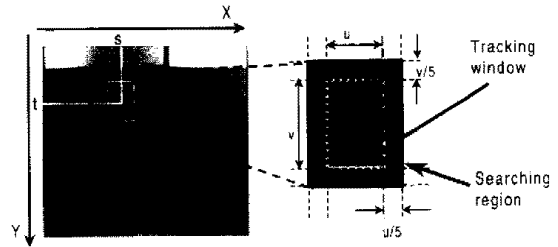
(그림 5) 탐색영역내의 DE 크기영상과 PDOE 영상의 비교

추적을 위한 정합법은 cross-correlation(CC)을 사용하였는데 이는 계산이 간단하여 수행속도가 빠르지만, 전처리를 하지 않은 영상을 정합할 경우, 영상의 갑작스런 밝기값 변화에 적절히 대응하지 못한다는 단점이 있다. 하지만 입력영상의 밝기변화는 크더라도 PDOE 영상은 밝기 변화가 크지 않기 때문에 영향을 거의 받지 않는다[13]. 추적을 행할 때 이전 프레임에서 결정된 손영역은 표적추적창이 되며, 현재 프레임에서 손영역을 찾기위한 영역은 탐색영역이 된다. 그림 6에 대해서 $M \times N$ 의 탐색 영역(searching region)내에서 추적을 위한 CC의 기본 정합식은

$$c(s, t) = \sum_y \sum_x R(x-s, y-t)S(x, y) \quad (5)$$

으로 표현되며, $0 \leq s \leq M-1$, $0 \leq t \leq N-1$ 가 된다. $R(x-s, y-t)$ 는 이전 프레임에서 획득된 추적창내 특정

영상이고, $S(x, y)$ 는 현재 프레임의 탐색영역 내의 특정영상이다.



(그림 6) 추적창과 조사 영역

4. HMM을 이용한 제스처 분할 인식

4.1 입력 특성 추출

본 논문에서 제스처는 카메라 영상으로부터 입력받은 손동작의 이동 궤적을 제스처 동작으로 정의하였다. 추적창의 이동 궤적을 입력 데이터로 바꾸기 위하여 추적창의 시간에 따른 중심 좌표 변화를 벡터로 나타내었다. 이들 벡터는 8방향을 갖는 코드로 양자화한다[5][14]. 그리고, 이 8방향 벡터 성분을 인식기의 기본 입력 데이터로 설정하였다. 그러나, 단순한 이전 프레임과의 방향성 각도만으로 8방향 코드를 구한다면 손의 떨림이나 손의 느린 움직임 등에 의해 잘못된 코드값이 들어갈 수 있다. 이러한 잡음성분들을 최소화하기 위해 5프레임 단위로 DA를 평균한 후 8방향 코드로 정규화 시킨다. n 번째 방향성 각도(DA_n)는

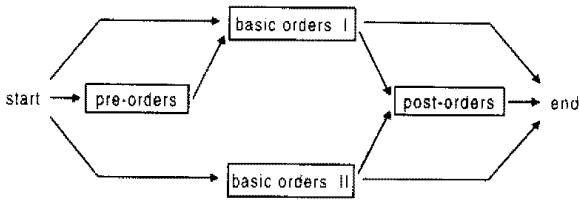
$$DA_n = \frac{1}{5} \sum_{i=n-4}^n \tan^{-1} \frac{y_{i-1} - y_i}{x_{i-1} - x_i} \quad (6)$$

이다. DA_n 는 추적중인 창에서 $n-1$ 번째 프레임에서 표적추적창의 중심점(center point)의 중심 좌표와 n 번째 프레임에서 표적 추적창의 중심점사이의 각도를 이전 5프레임 단위로 평균한 값을 나타낸다. 위 식에서 (x_i, y_i) 는 i 번째 프레임의 추적창의 중심점을 나타낸다. 방향성 코드는 방향성 각도(direction angle)를 8등분하여 0~7사이 값으로 양자화 시킨 것이다.

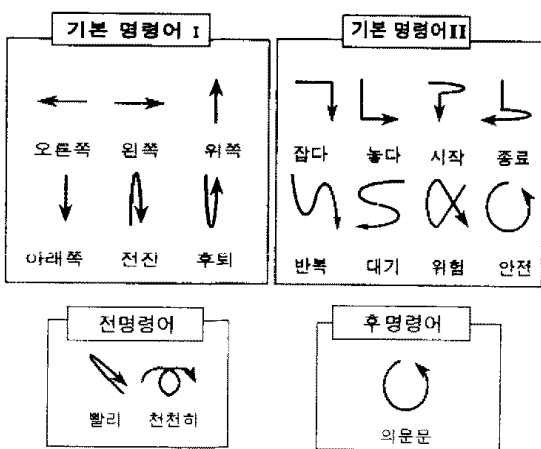
4.2 복합 제스처 인식 모델의 구성

사용된 K/K 원격 작업용 제스처의 구성은 원격 작업을 위한 기본 명령 단어와 동작설명에 대한 전명령어, 명령의 최종 형태인 후명령어로 나눌 수 있다. 기본명령어 I은 주로 방향에 대한 명령을 나타내고, 기본

명령어 II는 상황을 나타내는 명령어로 구성하였다. 전 명령어는 기본 명령 단어의 제스처 동작을 행하기 전에 앞서 표현을 하며, 후명령어는 기본 명령어에 대한 동작이 끝난 후 행한다. '빨리', '천천히'라는 의미의 전명령어는 그 자체로서는 명령어로서의 의미를 가지지 않으므로 기본명령어 I과 연결되어 의미를 가지게 된다. 이때 전명령어는 동작에 대한 설명이므로 상황을 나타내는 기본명령어 II와는 연결되지 않는다. 원격 작업에서 다음 명령을 수행하려면 현재 상태를 점검하기 위해서 쌍방향 정보전송이 필요할 경우가 있다. 이러한 경우에 후명령어가 필요하며, 이러한 후명령어가 기본 명령어에 연결되어 있으면 전체 명령어의 형태가 의문문이 된다. 전체적인 명령어의 구조는 기본명령어가 항상 존재하며 전명령어는 기본 명령어 I에만 연결되고 후명령어는 기본명령어에 연결되어 형태를 결정하게 된다. (그림 7)은 K/K 원격 작업용 제스처 명령어가 구성되는 형태를 나타내며, (그림 8)은 손의 이동 궤적에 따른 K/K 원격 작업용 제스처 명령어를 나타낸다.



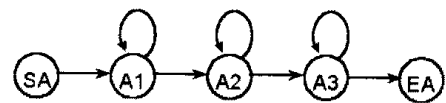
(그림 7) 제스처 명령의 기본 구분



(그림 8) K/K 원격작업용 제스처 명령어

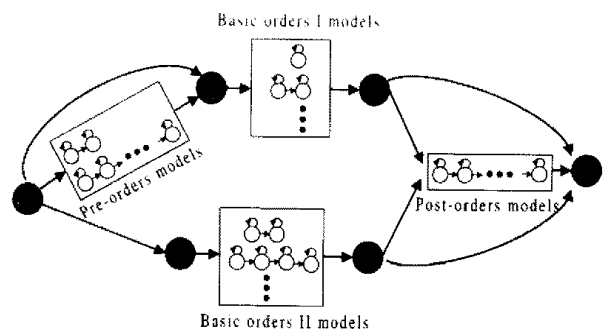
K/K 원격 작업용 제스처의 기본 명령어 I, II, 전명령어, 후명령어에 대한 각각의 명령 모델을 구성시킨

다. 각 세 부분 대한 명령 모델들은 (그림 9)와 같이 left right model을 구성하여 각 단어마다 통제된 학습을 실시한다. 이들 명령 모델들은 그 복잡성에 따라 상태(state)의 수가 1~5개로 구성되며 starting state와 end state를 가지게 된다. 이 모델의 파라미터는 Baum-Welch 재추정 알고리즘을 적용하여 구해졌다. 제스처 인식은 조합된 인식모델에서 최종적인 인식을 수행하게 된다.



(그림 9) 제스처 HMM 명령모델

HMM으로 훈련된 모델의 각 상태(state)는 훈련된 패턴의 일부를 내재적으로 표현하고 있으므로, 훈련된 상태를 이용하여 각 제스처 동작의 인식을 수행하게 된다. 본 논문에 사용된 제스처에 대한 모델은 3부분에 대한 HMM 모델을 구성하여 각각의 조합으로 최종적인 인식을 수행하게 되며 제스처 모델의 인식은 각 모델에 대한 Viterbi scoring을 구하게 된다. 이는 음성 인식에서 음소단위로 인식할 때 각각의 음소단위의 모델을 구성하여 이들을 복합적으로 연결하여 음성을 인식[15]하는 것과 같다. (그림 10)은 최종 제스처 인식 모델을 나타내며, pre-orders models는 전명령어를 인식하고 basic orders I models는 기본명령어 I을, basic orders II models는 기본 명령어 II를 인식한다. post-orders models는 후명령어를 인식하게 된다. 실제 인식되는 명령어는 52개가 되며, 이들을 전명령어 2개, 후명령어 1개, 기본명령어 14개의 조합에 의해 인식이 이루어진다. 따라서 17개의 인식 모델로써 52개의 명령어를 인식하게 된다.

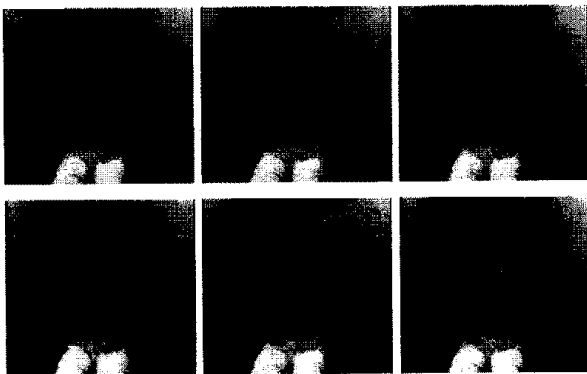


(그림 10) 복합 제스처 인식 모델

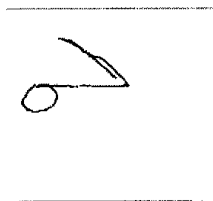
5. 실험 및 결과

본 논문에서 제안한 PDOE 추적(Tracking) 알고리즘과 HMM을 이용하여 원자력 발전소 내 원격 작업 환경용 제스처 인식기를 설계하였다. 이 시스템은 Pentium-150MHz PC상에서 구현되었으며, 영상 입력을 위한 하드웨어환경으로는 CCD 카메라와 DT 3851보드를 사용하였다. 제스처 영상의 크기는 256x256 화소이며, 각 화소는 256개의 명암도를 가진다. CCD 카메라로부터 입력되는 모션 영상을 물체 추적알고리즘을 적용하였을 때 초당 5 프레임의 처리속도를 나타내었다.

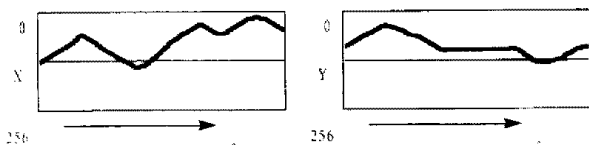
(그림 11)은 실제 제스처 입력 영상에 대해 PDOE 영상을 이용해서 추적 알고리즘을 수행한 결과이다. (그림 11) (a)에서는 실제 제스처 입력 영상에 대해 PDOE 영상을 이용해서 추적 알고리즘을 수행한 것을 나타내며, (그림 11)의 (b), (c)는 2차원 영상에서 추적상의 이동궤적과 x, y축에서의 이동궤적을 나타낸 것이다.



(a) 제스처('빨리 오른쪽으로 갈까요?')에 대한 추적 영상



(b) 추적상의 이동궤적



(c) x, y축으로의 이동궤적

(그림 11) 실제 입력영상에 대한 추적과정과 이동궤적

<표 1>은 각 명령 모델 별로 학습시킨 데이터개수를 나타내고, 기본 명령어 I, II에 대한 인식결과는 <표 2>와 같다. 2가지 명령어 이상이 결합된 복합 제스처 명령의 인식 결과는 <표 3>과 같다. K/K 제스처 데이터 베이스중 2개 동작이상이 연속된 복합명령 즉, 전명령어+기본명령어 I, 전명령어+기본명령어 I+후명령어, 기본명령어 I+후명령어, 기본명령어 II+후명령어에 해당하는 36개 제스처 단어에 대해 실험을 수행하였다. <표 2>에서 '?'는 후명령어를 나타낸다.

<표 1> 실험에 사용한 데이터

종 류		학 습 용
전명령어	천천히(slowly)	93
	빨리(fast)	111
후명령어(question : '?')		348
기본명령어 I		20
기본명령어 II		30

<표 2> 기본명령어 I 과 기본명령어 II로 구성되는 제스처 명령에 대한 인식결과

제스처종류		정인식갯수/ 테스트갯수	제스처종류	정인식갯수/ 테스트갯수	
가 본 명 령 어 I	오른쪽	13/13	기 본 명 령 어 II	잡다	19/20
	왼쪽	13/13		위험	20/20
	위쪽	13/13		종료	19/20
	아래쪽	13/13		놓다	19/20
	전진	13/13		반복	17/20
	후퇴	13/13		안전	19/20
전체	229/238 = 96.22%		시작	19/20	
			대기	19/20	

<표 3> 2개이상의 명령어가 복합된 제스처 명령어의 인식결과

제스처종류	정인식개수 /테스트개수	제스처종류	정인식개수 /테스트개수
빨리+오른쪽	19 / 19	천천히+왼쪽+'?'	16 / 16
빨리+왼쪽	19 / 19	천천히+아래쪽+'?'	9 / 11
빨리+위쪽	17 / 19	천천히+전진+'?'	12 / 12
빨리+아래쪽	19 / 19	천천히+후퇴+'?'	10 / 10
빨리+전진	19 / 19	오른쪽+'?'	20 / 20
빨리+후퇴	19 / 19	왼쪽+'?'	23 / 27
천천히+오른쪽	15 / 15	위쪽+'?'	33 / 33
천천히+왼쪽	18 / 18	아래쪽+'?'	23 / 23

〈표 3〉 계속

제스처종류	정인식개수 / 테스트개수	제스처종류	정인식개수 / 테스트개수
천천히-아래쪽	15 / 15	전진+?’	30 / 30
천천히+전진	16 / 16	후퇴+?’	28 / 28
천천히-후퇴	14 / 14	잡다+?’	19 / 20
빨리+오른쪽+?’	15 / 15	위험+?’	20 / 20
빨리+왼쪽+?’	15 / 15	종료+?’	20 / 20
빨리+위쪽+?’	13 / 15	놓다+?’	19 / 20
빨리+아래쪽+?’	15 / 15	반복+?’	20 / 20
빨리+전진+?’	15 / 15	안전+?’	20 / 20
빨리+후퇴+?’	15 / 15	시작+?’	20 / 20
천천히+오른쪽+?’	14 / 14	대기+?’	20 / 20
전	체	648/666 = 97.3 %	

6. 결 론

본 논문에서는 컴퓨터 시각에 기반해서 연속적으로 입력되는 영상을 통해 손영역의 추적을 행하여 이동 궤적을 분석하였다. 손영역을 추적함에 있어서 PDOE 영상을 이용한 추적 알고리즘을 수행함으로써 일반적인 흑백영상에서도 복잡한 배경이나 잡음 성분, 조명의 변화에도 안정된 추적을 나타내었다. 추적 알고리즘은 실험에 사용된 모든 제스처에 대해 올바른 손영역의 추적을 행하였다. 그리고 이동궤적을 8방향 성분으로 정규화 하여 인식기의 입력으로서 사용하게 된다. 손동작의 이동궤적을 효율적으로 찾기 위하여 추적 범위를 설정하고 이동 궤적을 분석함으로써 제스처를 인식한다. 그리고, 인식에서는 이산 HMM을 이용한 각 부분에 대한 명령모델을 연동시킴으로써 제스처 명령의 확장을 꾀하였으며, 2개 명령 이상으로 구성된 복합 제스처 명령에서 97% 이상의 안정된 인식율을 얻었다. 앞으로 자연스러운 명령을 인식하기 위해 더 많은 제스처 데이터 베이스를 구축하고 다양한 제스처의 분할과 복합적인 제스처 인식을 위한 다양한 알고리즘의 개발 및 통합에 관한 연구가 필요하다.

참 고 문 헌

[1] T. S. Huang and V. I. Pavlovic, "Hand Gesture Modeling, Analysis, and Synthesis," *IWAFGR*, pp.73-79, Zurich, June, 1995.
 [2] F. Quek, "Toward a vision based hand gesture

interface," *in Virtual Reality Software and Technology Conference*, Vol.2, 1990.

[3] J. M. Rehg and T. Kanade, "Digiteyes: Vision-based human hand tracking," *Tech. Rep. CMU-CS-93-220*, School of Computer Science, CMU, 1993.
 [4] J. Davis and M. Shah, "Gesture Recognition," Department of Computer Science Technical Report, CS-TR-93 11, University of Central Florida, 1993.
 [5] 이현규, 김호연, 김진형, "은닉 마르코프 모델을 이용한 연속 제스처 Spotting", *제9회 영상처리 및 이해에 관한 워크샵*, pp.114-119, 1997.
 [6] W. T. Freeman and M. Roth, "Orientation Histograms for Hand Gesture Recognition," *IWAFGR*, pp.296-301, Zurich, June, 1995.
 [7] L. R. Rabiner and B. H. Juang, "An Introduction to Hidden Markov Models," *IEEE ASSP Magazine*, pp.4 16, Jan., 1986.
 [8] J. Yang, Y. Xu, and C. S. Chen, "Human Action Learning via Hidden Markov Model," *IEEE Trans. on SMC*, Vol.27, No.1, pp.34 44, Jan., 1997.
 [9] T. Starner and A. Pentland, "Visual Recognition of American Sign Language Using Hidden Markov Models," *IWAFGR*, pp.189-194, Zurich, June, 1995.
 [10] 구분기, 전성일, 박순용, 이용범, "다중 특징 정보 병합에 기반한 스테레오 정합", *제9회 신호처리 합동학술대회 논문집*, Vol.9, pp.277-280, 1996.
 [11] R. N. Bracewell, *Two Dimensional Imaging*, Prentice-Hall, Englewood Cliffs, pp.267 272, N. J., 1995.
 [12] R. Crane, *A Simplified Approach to Image Processing Classical and Mode Techniques in C*, Prentice Hall, Inc. pp.67-76, N. J. 1997.
 [13] 성시훈, 전성일, 김문갑, 김준년, "다단계 크기 조절이 가능한 적응창과 Sequential Joint 방법에 의한 영상추적", *제9회 신호처리 합동학술대회 논문집*, Vol.9, pp.517 520, 1996.
 [14] 배창석, 전병태, 윤호섭, 민병우, "손의 이동 궤적 분석에 의한 제스처 인식", *제8회 영상처리 및 이*

해에 관한 워크샵, pp.144-148, 1996.

[15] J. R. Deller, Jr., J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*, pp.677-740, Memillan Publishing Company, 1993.

[16] U. R. Dhond, "Structure from Stereo: A Review." *IEEE Tran. on PAMI*, Vol.19, No.6, pp.1489-1989, Nov., 1989.



정지원

e-mail : jeewon@www.delta.co.kr

1996년 2월 경북대학교 전자공학과 졸업(공학사)

1998년 2월 경북대학교 대학원 전자공학과 졸업(공학석사)

1998년 3월~현재 델타정보통신주 기술연구소 연구원

관심분야 : 컴퓨터 비전, AI, Network Management System 등



이용범

e-mail : yblee2@nanum.kaeri.re.kr

1985년 2월 경북대학교 전자공학과 졸업(공학사)

1987년 2월 경북대학교 대학원 전자공학과 졸업(공학석사)

1999년 2월 경북대학교 대학원 전자공학과 졸업(공학박사)

1987년 3월~현재 한국원자력연구소 로봇핵심기술개발실 실장

관심분야 : 입체영상시스템, 가상현실, 로봇비전 등



진성일

e-mail : sichien@ee.kyungpook.ac.kr

1977년 서울대학교 전자공학과 졸업(공학사)

1981년 한국과학기술원 전기 및 전자공학과 졸업(공학석사)

1988년 미국 Carnegie Mellon University Electrical and Computer Engineering 졸업(공학박사)

1981년~현재 경북대학교 전기전자공학부 교수

관심분야 : 신경망, 컴퓨터 비전, 패턴인식, 문서인식 등