

3차원 손 모델링 기반의 실시간 손 포즈 추적 및 손가락 동작 인식

(Real-Time Hand Pose Tracking and Finger Action
Recognition Based on 3D Hand Modeling)

석 흥 일 † 이 지 홍 †† 이 성 환 †††
(Heung-Il Suk) (Ji-Hong Lee) (Seong-Whan Lee)

요 약 손 포즈 모델링 및 추적은 컴퓨터 시각 분야에서 어려운 문제로 알려져 있다. 손 포즈 3차원 복원을 위한 방법에는 사용되는 카메라의 수에 따라 다중 카메라 또는 스테레오 카메라 기반 방식과 단일 카메라 기반 방식이 있다. 다중 카메라의 경우 여러 대의 카메라를 설치하거나 동기화를 시키는 등에 대한 제약사항이 따른다. 본 논문에서는 확률 그래프 모델에서 신뢰 전파 (Belief Propagation) 알고리즘을 이용하여 단안 카메라에서 획득된 2차원 입력 영상으로부터 3차원 손 포즈를 추정하는 방법을 제안한다. 또한, 은닉 마르코프 모델(Hidden Markov Model)을 인식기로 하여 손가락 클릭 동작을 인식한다.

은닉 노드로 손가락의 관절 정보를 표현하고, 2차원 입력 영상에서 추출된 특징을 관측 노드로 표현한 확률 그래프 모델을 정의한다. 3차원 손 포즈 추적을 위해 그래프 모델에서의 신뢰 전파 알고리즘을 이용한다. 신뢰 전파 알고리즘을 통해 3차원 손 포즈를 추정 및 복원하고, 복원된 포즈로부터 손가락의 움직임에 대한 특징을 추출한다. 추출된 정보는 은닉 마르코프 모델의 입력값이 된다. 손가락의 자연스러운 동작을 위해 본 논문에서는 한 손가락의 클릭 동작 인식에 여러 손가락의 움직임을 함께 고려한다. 제안한 방법을 가상 키패드 시스템에 적용한 결과 300개의 동영상 테스트 데이터에 대해 94.66%의 높은 인식률을 보였다.

키워드 : 3차원 손 포즈 추적, 신뢰 전파(Belief Propagation), 확률 그래프 모델, 은닉 마르코프 모델, 가상 키패드, 인간-컴퓨터 상호작용

Abstract Modeling hand poses and tracking its movement are one of the challenging problems in computer vision. There are two typical approaches for the reconstruction of hand poses in 3D, depending on the number of cameras from which images are captured. One is to capture images from multiple cameras or a stereo camera. The other is to capture images from a single camera. The former approach is relatively limited, because of the environmental constraints for setting up multiple cameras. In this paper we propose a method of reconstructing 3D hand poses from a 2D input image sequence captured from a single camera by means of Belief Propagation in a graphical model and recognizing a finger clicking motion using a hidden Markov model.

We define a graphical model with hidden nodes representing joints of a hand, and observable nodes with the features extracted from a 2D input image sequence. To track hand poses in 3D, we use a Belief Propagation algorithm, which provides a robust and unified framework for inference in a graphical model. From the estimated 3D hand pose we extract the information for each finger's motion, which is then fed into a hidden Markov model. To recognize natural finger actions, we consider the movements of all the fingers to recognize a single finger's action. We applied the proposed method

· 이 연구에 참여한 연구자는 '2단계 BK21 사업'의 지원비를 받았다

† 학생회원 : 고려대학교 컴퓨터학과
hisuk@image.korea.ac.kr

†† 비회원 : 고려대학교 컴퓨터학과
jhlee@image.korea.ac.kr

††† 종신회원 : 고려대학교 컴퓨터·통신공학부 교수
swlee@image.korea.ac.kr

논문접수 : 2008년 10월 6일
심사완료 : 2008년 11월 6일

Copyright©2008 한국정보과학회: 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 소프트웨어 및 응용 제35권 제12호(2008.12)

to a virtual keypad system and the result showed a high recognition rate of 94.66% with 300 test data.

Key words : 3D Hand Pose Tracking, Belief Propagation, Probabilistic Graphical Model, Hidden Markov Model, Virtual Keypad, Human-Computer Interaction

1. 서론

손 포즈를 모델링하거나 손 또는 손가락의 움직임을 분석 및 이해하는 것은 컴퓨터 시각에서 커다란 관심을 받고 있으며, 매우 어려운 문제로 알려져 있다. 손 포즈 또는 손 움직임 모델링 방법은 HCI, 수화 인식, 제스처 인식과 같은 여러 응용에 사용될 수 있다.

손은 여러 개의 관절을 포함하고 있어 높은 자유도를 가지기 때문에, 손가락 관절의 위치를 추정하기 위해서는 고차원의 공간상에서 방대한 양의 검색 시간을 수반한다. 손 추적 및 동작 분석을 위해 다양한 방법들이 제안되어 왔으며, 이는 크게 세 가지로 구분해 볼 수 있다. 첫 번째는 뉴턴의 운동법칙과 주성분 분석(Principal Component Analysis)을 기반으로 한 변형이 가능한 손 모양 모델의 사용[1]이다. 그러나 이 방법은 손의 정확한 포즈를 추정할 수 없다는 단점이 있다. 두 번째는 입력 영상의 특징 공간과 손의 움직임에 대한 두 공간 사이의 매핑을 이용하는 외형 기반 방법[2,3]이다. 이는 두 공간 사이의 매핑 함수를 학습하기 힘들고, 1:1 매핑이 이뤄지지 않을 수도 있으며, 방대한 양의 훈련 데이터를 수집하기가 어렵다는 문제점을 가지고 있다. 세 번째는 입력 영상과 3차원 모델 투영 영상과의 매칭을 추정하는 3차원 모델 기반 방법이다. 3차원 모델과 입력 영상과의 대응점을 찾기 위해 영상에서 손가락 끝[4-6], 선[7,8], 윤곽[9], 실루엣[4,10,11] 등과 같은 다양한 특징들을 사용한 방법들이 연구되어 왔다.

손 동작 인식을 위해 추적은 중요한 문제이다. 최근 포즈 추정을 위해 신뢰 전파(Belief Propagation) 알고리즘[7,12]이 많이 사용되고, 좋은 성능을 보이고 있다. T. Han 등은 다양한 시점에서의 영상 시퀀스를 사용하여 3차원 휴먼 포즈를 추적하는 방법을 제안했다[7]. O. Bernier 등은 파티클 필터와 신뢰 전파 알고리즘을 결합하여 실시간으로 3차원상에서 휴먼의 상체를 추적하는 방법을 제안했다[13].

본 논문에서는 단안 컬러 카메라를 통해 획득되는 영상에서 확률 그래프 모델에서의 신뢰 전파 알고리즘을 이용하여 손의 포즈를 3차원으로 복원하는 방법을 제안한다. 복원된 3차원 모델로부터 손가락 움직임에 대한 특징 벡터를 추출하여 은닉 마르코프 모델을 이용한 손가락 클릭 동작을 인식하는 방법을 제안한다. 전체적인 시스템의 구성은 그림 1과 같다.

논문의 구성은 다음과 같다. 2절에서는 2차원 입력 영상에서 3차원 손 포즈를 추정하는 방법과 추정된 3차원 모델에서 손가락 움직임의 특징 벡터를 추출하고, 은닉 마르코프 모델을 인식기로 하여 클릭 동작을 인식하는 방법에 대해 설명한다. 3절에서는 제안한 방법을 가상 키패드 시스템에 적용한 결과 및 기존 연구와의 성능 평가를 보이고, 4절에서는 결론 및 향후 연구에 대해 설명한다.

2. 3차원 손 모델링 및 손가락 동작 인식

이절에서는 3차원 손 포즈 추정을 위한 확률 그래프 모델을 정의한다. 신뢰 전파 알고리즘을 이용하여 단안 컬러 카메라를 통해 획득된 영상으로부터 3차원 손 포즈를 추정하고, 은닉 마르코프 모델을 이용하여 손가락의 클릭 동작을 인식하는 방법에 대해 설명한다. 확률 그래프 모델에서 은닉 노드들은 고유의 자유도를 가지는 각 손가락의 관절을 표현하고, 관측 노드들은 2차원 입력 영상에서 획득된 특징값들을 표현한다. 실시간 응용을 위해 손가락 관절 간의 특성을 이용하여 자유도를 낮추어 확률 그래프 모델의 계산 시간 복잡도를 줄인다.

2.1 3차원 손 모델의 구성

2.1.1 관절(Articulated) 손 모델

본 논문에서는 24 자유도를 가지는 손 모델을 사용한다. 손가락의 관절들에 대해 굴곡 (Flexion), 외전 (Abduction), 회전(Rotation)를 각각 하나의 자유도로 고려했을 때, 엄지 손가락은 5, 나머지 네 손가락들은 4,

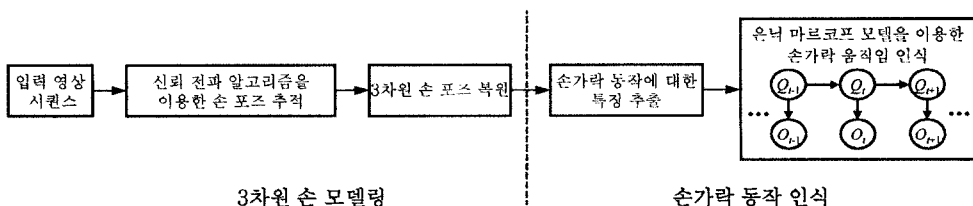


그림 1 제안하는 시스템의 구성도

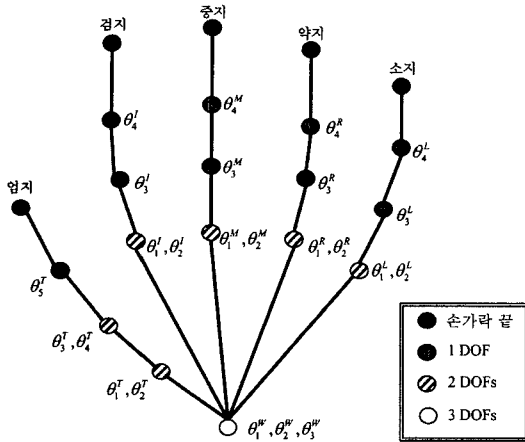


그림 2 24 자유도(DOF)를 가지는 손 모델

손목은 3이다. 각 손가락에 대한 자유도는 그림 2에 나타나 있다. 그림에서 θ_i^H 는 각 손가락 관절에 대한 정보를 표현한다. 여기서 $i \in \{1,2,3,4,5\}$ 는 각 손가락 관절에서의 자유도를 번호로 나타낸 것이고, $H \in \{T, I, M, R, L, W\}$ 는 각각의 손가락과 손목을 의미한다.

2.1.2 손 형태 구성의 특징

손을 표현하는 기존 모델들은 손의 특징을 고려하지 않아 컴퓨터 시각이나 애니메이션과 같은 응용에 적용하기가 매우 어렵다. 본 논문에서는 H. Rijpkema와 M. Girard가 제안한 다음의 특성[14]들을 이용한다.

- 엄지를 제외한 나머지 네 손가락에서 θ_3^F 와 θ_4^F 는 상호 의존관계를 가진다. 여기서 $F \in \{Index(\text{검지}), Middle(\text{중지}), Ring(\text{약지}), Little(\text{소지})\}$ 를 의미한다.

$$\theta_4^F = \frac{2}{3}\theta_3^F \tag{1}$$

- 엄지 손가락의 경우 손가락을 움직이는데 사용되는 관절이 손바닥에 포함되어 있어 동작 구성이 더 복잡하다. 그러나 엄지 손가락의 운동학적 특징은 실험적 관측값에 따라 다음과 같은 수식에 따라 계산될 수 있다.

$$\theta_1^T = 2 \times \left(\theta_3^T - \frac{\pi}{6} \right) \tag{2}$$

$$\theta_2^T = \theta_4^T \times \frac{7}{5} \tag{3}$$

- 관절의 각도 θ_1^F 와 θ_3^F 는 다음의 상관관계를 가진다.

$$\theta_1^F = k\theta_3^F, \quad 0 \leq k \leq \frac{1}{2} \tag{4}$$

- 각 손가락에서 θ_2^F 의 외전은 거의 존재하지 않는다.

$$\theta_2^F = 0 \tag{5}$$

- 엄지 손가락에서 두 관절 θ_4^T 와 θ_5^T 는 다음의 상관관계를 갖는다.

$$\theta_5^T = \alpha\theta_4^T, \quad \alpha \geq 0 \tag{6}$$

식 (1), 식 (4), 식 (5)로부터 각 손가락에서 자유도를 3씩 줄일 수 있고, 식 (2), 식 (3), 식 (6)로부터 엄지 손가락의 자유도 3을 각각 줄일 수 있다. 따라서, 손가락 관절들의 상관관계를 이용하여 손 모델에서의 자유도를 24에서 9로 줄일 수 있다.

2.2 제안하는 손 포즈 추적 방법

본 논문에서는 3차원 손 포즈를 확률 그래프 모델로 표현하고, 그래프 모델에서 확률값을 추정하여 손 포즈를 추적하는 방법을 제안한다. 손 포즈와 손가락 움직임 추적은 파티클 필터의 변형과 신뢰 전파 알고리즘을 적용하여 추정한다. 3차원 손 모델과 확률 그래프 모델을 입력 영상에서 추출된 특징을 관측값으로 사용한다.

2.2.1 관절(Articulated) 포즈 추적을 위한 신뢰 전파 알고리즘

손 포즈의 추정을 위해 그림 3의 확률 그래프 모델을 정의한다. 그림 3에서 원 노드는 은닉 노드이고, 사각형 노드는 관측 노드를 표현한다. 확률 그래프 모델은 손목과 각 손가락의 상태를 모델링하는 6개의 은닉 노드 X^n 로 구성되어 있고, 각 은닉 노드는 영상에서의 특징값 Y^n 을 관측값으로 가진다. 여기서 $n \in \{Wrist, Thumb, Index, Middle, Ring, Little\}$ 이다. 방향성이 있는 화살표는 노드간의 시간적 의존 관계를 표현하고, 방향성이 없는 선은 동일 시간에서 노드간의 상관관계를 표현한다. 시간 t 에서 손가락의 상태를 모델링하는 X_t^n 은 시간 $t-1$ 에서의 상태 X_{t-1}^n 에 의존하며 이는 $P(X_t^n | X_{t-1}^n)$ 로 표현되고, 시간 t 에서의 관측값 Y_t^n 은 은닉 노드 X_t^n 에 의존하여 $P(Y_t^n | X_t^n)$ 로 표현될 수 있다. 그리고 시간 t 에서 각 손가락 관절 상태는 상호작용을 모델링하는 전위(Potential) 함수 $\Phi(X_t^n, X_t^i)$ 로 표현한다. 확률 그래프 모델에서의 세 가지 관계를 모두 고려하면 시간 t 에서 손 포즈에 대한 확률은 다음과 같이 계산된다.

$$P(M_t) \propto \prod_{n=1}^N P(Y_t^n | X_t^n) \int P(X_t^n | X_{t-1}^n) dX_{t-1}^n \prod_{i \in L} \Phi(X_t^n, X_t^i)$$

확률 그래프 모델에서 추론은 신뢰 전파 알고리즘[15]을 사용한다. 현재 프레임에서 계산된 주변 확률 분포는 다음 프레임에서의 사전 확률 분포가 된다. 그러나 정확한 확률 분포를 알 수 없으므로 샘플들을 이용하여 확률 분포를 근사적으로 표현하는 Monte-Carlo 방법을 이용한다. 이는 가중치를 가지는 샘플 집합으로 확률 분포를 표현하는 방법으로 본 논문에서는 파티클 필터 알고리즘[16]을 이용한다. 그리고 각 샘플들의 가중치는

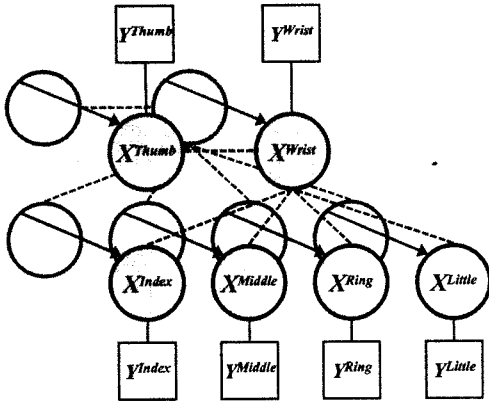


그림 3 손 모델 표현을 위한 확률 그래프 모델

노드간의 연결성을 고려한 신뢰 전파 알고리즘을 통해 계산된다. 이는 각 샘플들에 대해서 우도값을 계산하고, 노드간의 관계를 고려한 전위 함수의 경우에도 각 샘플들의 관계만을 고려하여 계산하므로 속도가 빠르다는 특징을 가진다.

2.2.2 관측 확률

관절의 상태 n 에 대해 관측된 특징값의 우도를 S 로 정의하자. 관측값은 피부색 검출, 에지 필터를 적용한 결과 영상이다. 그러나 여러 필터를 적용한 결과들 간의 조건부 상관관계를 식 (7)의 형태를 갖는 Gibbs 모델 [12]을 사용하여 결합 확률 밀도를 혼련함으로써 명시적인 모델을 생성한다. 식 (7)에서 N 은 검출된 손에 대한 에지에서의 픽셀수, P_n 은 픽셀의 좌표, h_n 은 얼굴 영역에서 생성한 확률 히스토그램, c_n 은 신뢰 요소값(상수)을 의미한다. 얼굴 영역의 검출은 Haar-like 얼굴 검출기[17]를 사용하였다.

$$P(Y|X) \propto \exp \left[\sum_{n=1}^N S(P_n, h_n) \cdot c_n \right] \quad (7)$$

관측값들은 각 은닉 상태 노드와 픽셀들에 대해 독립적이라고 가정한다. 손바닥과 각 손가락 마디에 대한 S 는 각각을 평평한 실린더로 표현했을 때 실린더와 손의

윤곽 에지의 가우시안(Gaussian) 거리와 피부색 확률의 곱으로 계산된다.

2.2.3 상호 작용 전위(Potential) 함수

손가락 마디를 표현하는 은닉 노드들 간에 전달되는 메시지는 가중치를 가지는 샘플들의 집합으로 표현한다. 신뢰 전파 알고리즘에서 각 노드는 전달되는 모든 메시지를 곱하여 각 노드의 확률 분포를 갱신한다. 알고리즘에서의 모든 연산은 가우시안 분포로 계산된다. 손가락 마디들 간 상호작용에서의 전위 함수는 두 손가락 마디사이의 거리에 대한 가우시안 분포로 표현된다. 손바닥과 중지의 손가락 관절간의 참조거리를 가우시안 분포의 평균으로 한다. 관절 손 포즈에 대한 추적 결과가 그림 4에 나타나 있다.

2.3 제안하는 손가락 움직임 인식 방법

2.3.1 특징

본 논문에서는 검지, 중지, 약지 세 손가락에 대한 움직임을 인식하기 위해 각 손가락의 길이 변화를 특징값으로 사용한다. 그림 5는 클릭 동작에 대한 손가락의 길이 변화를 보여준다.

특징값들은 식 (8)과 같이 계산한다. 여기서 l_t 는 t 시간에 손목 중심으로부터 손가락 끝까지의 길이를 의미한다.

$$\Delta l = l_t - l_{t-1} \quad (8)$$

자연스러운 손가락 움직임을 인식하기 위해 다른 손가락의 길이 변화도 함께 고려한다. 따라서 모든 손가락의 길이 변화를 특징값으로 하는 5차원 특징 벡터를 사용한다.

$$L = (\Delta l^{Thumb}, \Delta l^{Index}, \Delta l^{Middle}, \Delta l^{Ring}, \Delta l^{Little})$$

5차원 특징 벡터에 대해 가우시안 혼합 모델(Gaussian Mixture Model)을 사용하여 특징 벡터에 대한 클러스터링을 수행한다. 특징 공간에서의 밀도 추정은 가우시안 함수의 파라미터들로 계산된다. 클러스터를 다음과 같이 표현하자.

$$C = \{C_1, \dots, C_K\}$$



그림 4 관절(Articulated) 손 포즈 추적 결과 예

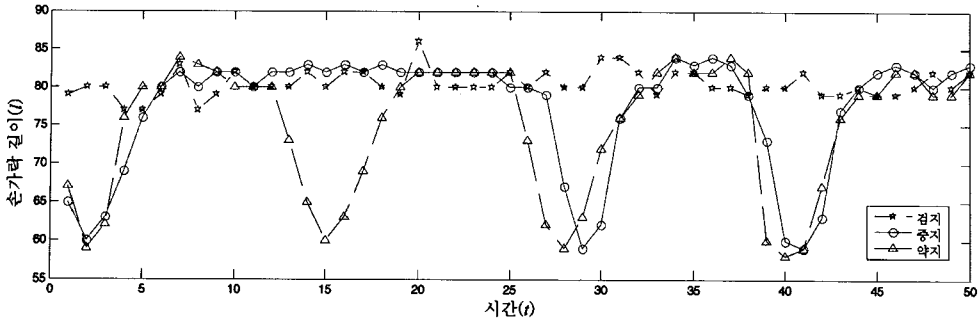


그림 5 클릭 동작에 따른 손가락 길이 변화

여기서 K 는 클러스터의 수이고 각 클러스터는 특징 공간에서의 특정 영역을 표현한다. 특징 벡터 L 은 GMM 기반의 확률 분포 함수에 의해 다음과 같이 모델링된다.

$$P(C_k | L) = \frac{P(L | C_k)P(C_k)}{P(L)}$$

위의 식에서 우도(likelihood) $P(L | C_k)$ 는 식 (9)의 다변량 가우시안으로 계산한다. 식에서 μ_k 와 V_k 는 k 번째 클러스터에 대한 평균과 공분산을 의미한다.

$$P(L | C_k) = \frac{1}{(2\pi)^{n/2} |V_k|^{n/2}} \exp \left[-\frac{1}{2} (L - \mu_k)^T V_k^{-1} (L - \mu_k) \right] \quad (9)$$

2.3.2 은닉 마르코프 모델

손가락 움직임 인식을 위해 본 논문에서는 은닉 마르코프 모델[18]을 인식기로 사용한다. 은닉 마르코프 모델은 시계열 데이터를 모델링하는데 적합한 모델로 알려져 있으며, 다음 세 개의 파라미터로 표현된다.

$$\lambda = \{A, B, \pi\}$$

A 는 은닉 노드간의 상태 전이 확률, B 는 은닉 노드에서 입력의 특징벡터가 관측될 확률 분포, π 는 시작 상태 확률 분포를 각각 나타낸다. 은닉 마르코프 모델의 파라미터는 식 (10)에 표현된 것과 같이 주어진 훈련 데이터에 대한 우도값이 최대가 되도록 Baum-Welsh 알고리즘[18]을 사용하여 훈련한다.

$$\lambda^* = \arg \max_{\lambda} P(O | \lambda) \quad (10)$$

새로운 입력에 대해서는 전진(forward) 알고리즘[18]을 사용하여 입력 손가락의 움직임을 인식한다. 전진 알고리즘은 시간 t 까지의 관측 입력 시퀀스 $O_1 O_2 \dots O_t$ 와 시간 t 에서 상태 $q_t = j$ 에 머무를 결합 확률을 다음과 같이 재귀적 방법으로 계산한다.

$$\alpha_t(j) = P(O_1 O_2 \dots O_t, q_t = j | \lambda) = \left\{ \sum_{i=1}^N \alpha_{t-1}(i) A_{ij} \right\} B_j(O_t)$$

본 논문에서는 손가락의 길이를 특징 벡터로 사용하기 때문에 이를 이산값으로 양자화하게 되면 정보의 손실을 가져오게 된다. 따라서 연속 은닉 마르코프 모델을 사용하고, 각 상태에서의 관측 확률 분포는 가우시안 혼합 모델을 사용한다.

$$B_j(O_t) = \sum_{k=1}^M c_{jk} N(O_t; \mu_{jk}, \Sigma_{jk}), \quad 1 \leq j \leq N$$

식에서 M 은 혼합된 가우시안 모델의 개수, c_{jk} 는 k 번째 가우시안 N 에 대한 가중치, μ_{jk} 와 Σ_{jk} 는 각각 k 번째 가우시안에 대한 평균과 공분산을 나타낸다.

은닉 마르코프 모델에서 상태 전이에 대한 위상(topology)은 그림 6의 left-right 모델을 사용한다. 입력 시퀀스에 대한 특정 손가락 움직임에 대한 인식은 관측된 특징 벡터에 대해 그림 7과 같이 여러 개의 은닉 마르코프 모델에서 병렬적으로 계산을 하여 우도값이 가장 높은 것으로 결정한다.

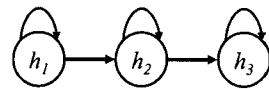


그림 6 Left-right 모델

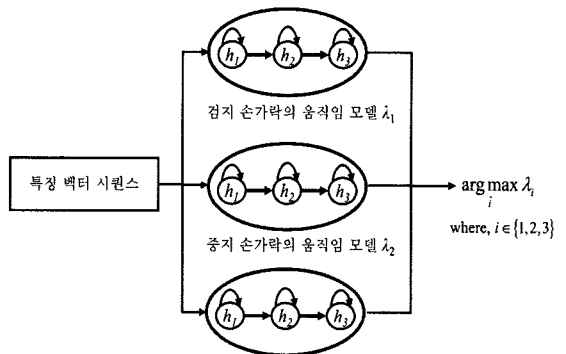


그림 7 손가락 움직임 인식을 위한 은닉 마르코프 모델 구조

3. 실험 결과 및 분석

제안하는 방법의 효율성을 보이기 위해 그림 8과 같은 환경에서 가상 키패드 시스템을 구축하였다. 시스템은 그림 8(c)와 같이 화면에 나타난 가상의 숫자들을 보고 손가락 클릭 동작을 취함으로써 이동 전화 서비스와 같은 환경에서 인터페이스로 활용될 수 있다. 본 논문에서는 세 손가락을 이용한 클릭 동작을 은닉 마르코프 모델을 이용하여 인식할 수 있도록 하였다. 그리고 2차원 상에서의 컨투어(Contour)와 손가락 길이 변화에 대

해 임계치를 적용한 기존 방법[19]과의 성능 비교도 함께 실험하였다.

3.1.1 실험 환경

소형 CMOS 카메라를 이용하여 그림 9와 같이 10명의 사용자로부터 가상 키패드를 클릭하는 동작에 대한 데이터를 획득하였다. 테스트 데이터는 카메라와 손이 평행을 이루는 경우와 평행하지 않은 경우로 나누었고, 두 경우 모두 검지, 중지, 약지 손가락 각각에 대해 100번씩 촬영하였다. 데이터는 초당 30 프레임으로 캡처하였으며, 각 프레임의 크기는 320×240이고, 24비트 컬러 색상이다.

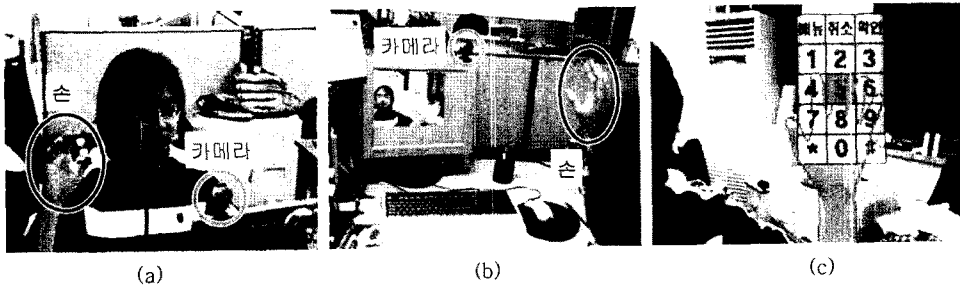


그림 8 가상 키패드 시스템: (a), (b) 시스템 환경, (c) 프로그램 실행 화면

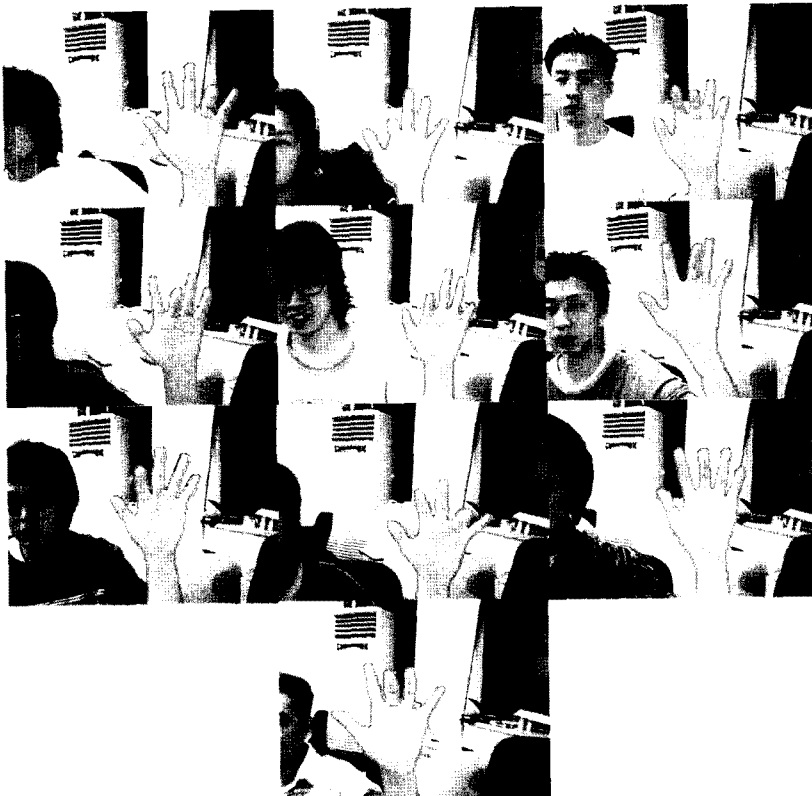


그림 9 10명의 사용자로부터 실험 데이터 획득

3.1.2 실험 결과

입력 2차원 영상에 대한 3차원 손 포즈 복원 결과가 그림 10에 나타나 있다. 그림 10에서 위쪽 열에 있는 그림들은 각 손가락 관절의 추정 위치와 손가락의 끝을 검은색 점으로 표시하여 보여주고 있고, 아래쪽 열에 있는 그림들은 3차원으로 복원했을 때의 결과를 다양한 각도에서 보여주고 있다.

3차원으로 복원한 결과로부터 손바닥의 중심과 손가락 끝의 길이 변화를 관측함으로써 손가락 클릭 동작에 대한 인식을 수행하였다. 실시간 인식을 위해 본 논문에서는 10 프레임 크기의 윈도우를 사용하였다. 이전 윈도우와 5 프레임씩 겹치면서 이동하였고, 윈도우 내에서의 관측값에 대한 우도값을 HMM을 이용하여 계산하였을 때, 미리 정한 임계값보다 우도값이 큰 경우 손가락 동작이 발생한 것으로 인식하였다. 제안하는 방법의 성능 평가를 위해 인식 결과를 정인식과 오인식으로 나누었고, 오인식의 경우 False Positive, False Negative

로 에러의 종류를 분류하여 표 1에 나타냈다. False Positive는 클릭 동작이 없으나 있다고 인식한 경우이고, False Negative는 클릭 동작이 발생하였으나 인식하지 못한 경우를 의미한다. 표 1에서 M. Tosas와 B. Li의 방법[19]은 검출된 손을 키투어로 표현하고, 키투어로부터 손가락 길이의 변화를 관측하여 임계치보다 작은 경우 클릭 동작이 발생하였다고 인식하는 것으로 제안하는 방법보다 인식 성능이 크게 떨어짐을 알 수 있다.

표 2는 카메라의 수직 방향 벡터를 기준으로 사용자가 손을 회전(Yaw)하여 카메라와 평행하지 않은 경우에 대한 실험 결과를 보여주고 있다. 사용자의 손이 카메라와 평행한 경우보다 인식 성능이 떨어지긴 하였지만 여전히 높은 인식률을 보이고 있으며, M. Tosas와 B. Li의 방법[19]보다 높은 인식 성능을 보이고 있다.

4. 결론 및 향후 연구

본 논문에서는 2차원 입력 영상으로부터 손 포즈를



그림 10 3차원 손 포즈 추적 결과

표 1 제안하는 방법과 기존 방법과의 인식 성능 비교

		손가락 명령어 인식 및 오인식 수			Hits
		검지	중지	약지	
M. Tosas와 B. Li의 방법 [19]	True Positive	92	72	70	234 (78%)
	False Positive	3	18	18	39 (13%)
	False Negative	5	10	12	27 (9%)
제안하는 방법	True Positive	96	95	93	284 (94.66%)
	False Positive	1	2	3	6 (2%)
	False Negative	3	3	4	10 (3.33%)
계		100	100	100	300

표 2 손을 회전했을 경우(Yaw: ±10도)에 대한 제안하는 방법과 기존 방법과의 인식 성능 비교

		손가락 명령어 인식 및 오인식 수			Hits
		검지	중지	약지	
M. Tosas와 B. Li의 방법 [19]	True Positive	63	57	58	178 (59.33%)
	False Positive	19	23	21	63 (21%)
	False Negative	18	20	21	59 (19.66%)
제안하는 방법	True Positive	89	85	80	254 (84.66%)
	False Positive	3	6	9	18 (6%)
	False Negative	8	9	11	28 (9.33%)
계		100	100	100	300

추정하기 위해 확률 그래프 모델을 정의하고, 신뢰 전파 알고리즘을 이용하여 3차원으로 복원하는 방법을 제안하였다. 확률 그래프 모델에서 관절에 대한 정보는 은닉 노드, 2차원 입력 영상에서 추출된 특징은 관측 노드로 표현하였다. 관절에 대한 정보를 표현하는 은닉 노드는 각 손가락의 관절이 가지는 자유도에 따라 다른 차원으로 표현하였다. 확률 그래프 모델의 추론에서 계산량을 줄이기 위해 H. Rijkema와 M. Girard[14]가 제안한 방법에 따라 자유도를 24에서 9로 줄여 3차원으로 복원하였다. 복원된 3차원 손 모델에 기반하여 손가락의 움직임에 대한 특징을 추출하여 은닉 마르코프 모델을 인식기로 한 손가락 동작 인식 방법을 제안하였다.

제안한 방법의 효율성을 확인하기 위해 가상 키보드 시스템에 적용하였다. 적용 결과 제안한 방법은 M. Tosas와 B. Li의 방법[19]보다 높은 인식률을 보였으며, 불안정한 손의 움직임 또는 회전에 대해도 강한 성능을 보였다. 본 논문에서의 제안 방법은 두 손을 이용한 가상 키보드나 복잡한 손동작 인식에 대한 향후 연구의 이정표가 될 수 있다.

참 고 문 헌

[1] A. Heap and D. Hogg, "Improving Specificity in PDMS using a Hierarchical Approach," Proc. British Machine Vision Conference, Essex, UK, Vol. 1, pp. 80-89, Sept. 1997.
 [2] R. Rosales, S. Sclaroff, and V. Athitsos, "3D Hand Pose Reconstruction using Specialized Mappings,"

Proc. 8th IEEE International Conference on Computer Vision, Vancouver, Canada, Vol. 1, pp. 378-385, July 2001.
 [3] Y. Wu and T. Huang, "View-Independent Recognition of Hand Postures," Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, South California, USA, Vol. 2, pp. 88-94, June 2000.
 [4] J. Deutscher, A. Blake, and I. Reid, "Articulated Body Motion Capture by Annealed Particle Filtering," Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, South California, USA, Vol. 2, pp. 126-133, June 2000.
 [5] N. Shimada, Y. Shirai, Y. Kuno, and J. Miura, "Hand Gesture Estimation and Model Refinement using Monocular Camera Ambiguity Limitation by Inequality Constraints," Proc. 3rd IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, pp. 268-273, 1998.
 [6] Y. Wu and T. Huang, "Capturing Articulated Human Hand Motion: A Divide-and-Conquer Approach," Proc. 7th IEEE International Conference on Computer Vision, Kerkyra, Greece, Vol. 1, pp. 606-611, 1999.
 [7] T. Han, H. Ning, and T. Huang, "Efficient Nonparametric Belief Propagation with Application to Articulated Body Tracking," Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, USA, Vol. 1, pp. 214-221, June 2006.
 [8] J. Rehg and T. Kanade, "Model-based Tracking

of Self-Occluding Articulated Object," Proc. 5th International Conference on Computer Vision, Cambridge, USA, pp. 612-617, June. 1995.

- [9] B. Stenger, A. Thayananthan, P. Torr, and R. Cipolla, "Hand Pose Estimation Using Hierarchical Detection," Proc. European Conference on Computer Vision, Lecture Notes in Computer Science, Prague, Czech Republic, Vol. 3058, pp. 105-116, May 2004.
- [10] J. Kuch and T. Huang, "Vision based Hand Modeling and Tracking for Virtual Teleconferencing and Telecollaboration," Proc. 5th International Conference on Computer Vision, Cambridge, USA, pp. 666-671, June 1995.
- [11] J. Lee and T. Knuii, "Model-based Analysis of Hand Posture," Proc. IEEE Computer Graphics and Application, New York, USA, Vol. 15, No. 5, pp. 77-86, 1995.
- [12] M. Vittrup, M. Srensen, and B. McCane, "Pose Estimation by Applied Numerical Techniques," Proc. Image and Vision Computing New Zealand, Auckland, New Zealand, Vol. 2, pp. 35-38, Nov. 2002.
- [13] O. Bernier and P. Cheung-Mon-Chan, "Real-Time 3D Articulated Pose Tracking using Particle Filters Interacting through Belief Propagation," Proc. 18th IAPR/IEEE International Conference on Pattern Recognition, Hong Kong, China, Vol. 1, pp. 90-93, Aug. 2006.
- [14] H. Rijkema and M. Girard, "Computer Animation of Knowledge-based Human Grasping," Proc. International Conference on Computer Graphics and Interactive Techniques, New York, USA, Vol. 25, No. 4, pp. 339-348, Aug. 1991.
- [15] C. Bishop, *Pattern Recognition and Machine Learning*, Chapter 8, Springer, 2007.
- [16] M. Isard and A. Blake, "CONDENSATION - Conditional Density Propagation for Visual Tracking," International Journal of Computer Vision, Vol. 29, No. 1, pp. 5-28, Aug. 1998.
- [17] P. Viola and M. Jones, "Robust Real-Time Face Detection," International Journal of Computer Vision, Vol. 57, No. 2, pp. 137-154, 2004.
- [18] L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proceedings of the IEEE, Vol. 77, No. 2, pp. 257-285, Feb. 1989.
- [19] M. Tosas and B. Li, "Virtual Touch Screen for Mixed Reality," Proc. European Conference on Computer Vision, Lecture Notes in Computer Science, Prague, Czech Republic, Vol. 3058, pp. 48-59, May 2004.



석 홍 일

2004년 부경대학교 멀티미디어공학과(학사). 2007년 부경대학교 컴퓨터공학과(석사). 2008년~현재 고려대학교 컴퓨터학과 박사과정. 관심분야는 패턴인식, 컴퓨터 시각, 뇌공학, 기계학습 등



이 지 홍

2006년 한성대학교 컴퓨터시스템학부(학사). 2008년 고려대학교 컴퓨터학과(석사). 관심분야는 패턴인식, 컴퓨터 시각, 영상처리 등



이 성 환

1984년 서울대학교 계산통계학과(학사) 1986년 KAIST 전산학과(석사). 1989년 KAIST 전산학과(박사). 1989년~1995년 충북대학교 컴퓨터학과 전임강사, 조교수. 2001년~2002년 MIT Brain and Cognitive Sciences 학과 방문교수. 1995년~현재 고려대학교 정보통신대학 정교수. 1997년~현재 고려대학교 인공지능연구센터 소장. 관심분야는 패턴인식, 컴퓨터 시각, 뇌공학 등