
CONDENSATION 알고리즘을 이용한 입자필터 기반 동작 인식 연구

이 양 원*

A Study on the Gesture Recognition Based on the Particle Filter Using
CONDENSATION Algorithm

Yang-Weon Lee*

이 논문은 2006년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아
수행된 연구임(KRF-2006- D00356(100068))

요 약

연속되는 이미지 중에서 인간의 동작을 인식하는 것은 인간과 컴퓨터의 상호 작용에서 매우 중요하고 도전할 분야이다. 본 논문에서는 CONDENSATION 알고리즘을 이용하여 입자 필터(particle filter)에 기반 한 동작 인식 알고리즘을 제안한다. 입자 필터는 조건 확률 전파 모델(Conditional Density Propagation)인 베이시안(Bayesian) 추정 규칙을 적용하는 추적구조를 갖고 있기 때문에 일반적으로 기존 추적 알고리즘보다 뛰어난 성능을 갖는 경향이 있다. 본 논문에서는 알고리즘의 성능 평가를 위해서 두 개의 동작 모델을 가정하였고, 영상에 대한 전처리를 위해서는 MATLAB를 이용하였으며 입자필터는 고속 처리를 위하여 C++로 구현하였다. 두 개의 동작 실험 결과를 통해, 동작 인식 입자 필터가 근접한 동작 환경 속에서 강인한 추적 성능을 나타냄을 확인하였다.

ABSTRACT

The recognition of human gestures in image sequences is an important and challenging problem that enables a host of human-computer interaction applications. This paper describes a gesture recognition algorithm based on the particle filters, namely CONDENSATION. The particle filter is more efficient than any other tracking algorithm because the tracking mechanism follows Bayesian estimation rule of conditional probability propagation. We used two models for the evaluation of particle filter and apply the MATLAB for the preprocessing of the image sequence. But we implement the particle filter using the C++ to get the high speed processing. In the experimental results, it is demonstrated that the proposed algorithm prove to be robust in the cluttered environment.

키워드

Gesture recognition, particle filter, CONDENSATION

I. 서론

인간의 동작은 행위와 밀접한 관계가 있다. 즉 우리가 제스처를 이용하여 정보를 전달하고자 할 때 몸을 움직이게 된다. 이런 종류의 인간 동작을 이해하고 추정하는 것은 인간과 컴퓨터의 상호작용, 의약, 로봇, 애니메이션, 비디오 데이터베이스, 이름 검색 등의 다양한 응용분야에 적용이 가능하므로 매우 관심이 큰 분야이다. 본 논문에서는 입자필터를 이용하여 인간의 동작을 인식하는 알고리즘에 대한 연구에 중점을 둔다.

입자필터[1]은 사전분포(prior distribution)와 사후분포(posterior distribution)와 같은 베이시안 조건 확률(Bayesian conditional probability)에 기반을 두고 있다.

본 논문에서는 먼저 CONDENSATION 알고리즘에 기반을 둔 입자필터를 동작 인식을 위한 모델로 유도하기 위해서 Black and Jepson[2]의 결과를 확장해서 유도하였다. 그리고 실험을 위한 영상으로서 2개의 수화 모델(그대로 둠, 노젓기)을 실험 동작으로 하여 알고리즘 평가를 수행하였다. 이미지에 대한 기본 처리는 MATLAB를 이용하여 수행하였고, 추적은 속도를 고려하여 C++ 언어를 사용하여 구현하였다. 논문의 전체적인 수행내용은 그림 1과 같다.

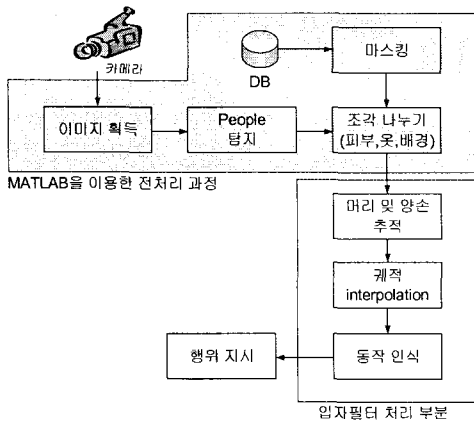


그림 1. 전체 시스템 동작 과정
Fig.1 Overall operation block diagram of recognition system

본 논문의 구성은 다음과 같다. 서론에 이어 II장에서는 제안하고자하는 CONDENSATION 알고리즘의 개요

와 유도과정 및 관련 모델을 기술하고, III장에서는 본 논문에서 제안된 알고리즘을 실험하기 위한 동작 모델 추출 과정을 설명한다. IV장에서는 제안한 기법에 대한 실험 결과를 기술하고, 마지막으로 결론을 맺는다.

II. CONDENSATION 알고리즘

2.1. CONDENSATION 알고리즘 개요

CONDENSATION 알고리즘은 CONDitional dENSity propaGATION의 글자를 조합해서 만든 것으로서 시간에 따른 조건 밀도 전파(probability density propagation)를 이용하고 있다. 즉 측정 데이터를 어떤 특정식에 적용하기 보다는 데이터가 설명하는 궤적을 근사화하기 위해서 N 개의 가중치 표본을 사용한다. 각 표본은 상태(state)와 가중치(weight)로 구성되어 있으며, 가중치는 상태(state)가 입력 데이터에 의해서 예측되는 확률에 비례해서 결정된다. 표본 수가 증가함에 따라서 표본들이 가지는 관측된 pdf의 정확도는 증가하게 된다.

일단 시간 t_1, \dots, t_n 동안 측정값들이 만들어졌다고 가정하자. 그러면 시간 $t+1$ 에서 새로운 표본 집합을 생성하기 위해서 상태(states)들을 시간 t 에 있는 표본 집합으로부터 각각의 가중치를 근거로 하여 랜덤하게 선택한다. 즉 각 표본의 가중치는 상태가 선택(selection)되어질 확률을 결정하게 된다. 랜덤하게 표본화된 상태 s_t 가 주어지게 되면 시간 $t+1$ 에서 새로운 상태 s_{t+1} 의 예측은 예측 모델에 기반을 두어 만들어진다. 이것은 프로세스 밀도 $p(x_{t+1}|x_t = s_t)$ 로부터 표본화(sampling)를 수행하는 것과 같게 된다. 여기서 X_t 는 대상 물체의 상태를 설명하는 파라미터들의 벡터다.

마지막으로 s_{t+1} 은 확률 $p(z_{t+1}|s_{t+1})$ 에 비례하는 가중치로 할당된다. 여기서 z_{t+1} 은 시간 $t+1$ 에서 대상 물체의 관측 상태(observed state)를 나타내는 파라미터들의 집합이다. 그리고 나서 프로세스는 다음 관측을 위하여 반복하게 된다. 이 같은 방법으로 데이터에 가장 근접한 예측 상태들이 더 큰 가중치를 갖게 된다. 임의의 복소 pdf들이 모델 될 수 있기 때문에 임의의 경쟁 가설들(충분히 큰 N 이 있다고 가정함)이 하나의 주도적인 가설이 남을 때까지 존재하게 된다. 이 같은 과정을 그림 2에 보였다.

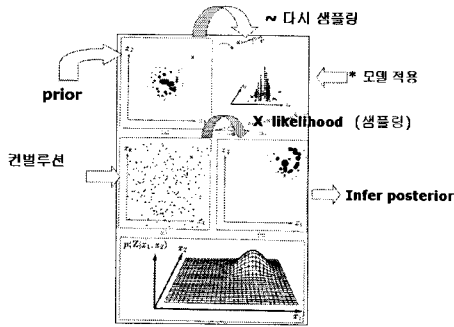


그림 2. Particle Filter 계산 과정
Fig. 2 The calculation process of the particle filter

2.2. 동작 인식을 위한 CONDENSATION 적용

신호 동작 인식을 위하여 CONDENSATION 알고리즘을 적용하기 위해서 본 논문에서는 Black and Jepson[2]이 제시한 방법을 확장하였다. 특히 시간 t 에서 상태(state)는 파라미터 벡터로서 $s_t = (\mu, \phi^i, \alpha^i, \rho^i)$ 와 같이 표현하였다. 여기서 각 파라미터들은 다음과 같은 의미를 갖고 있다.

- μ : 모델이 매치되는 것을 나타내는 정수 지수
- ϕ^i : 모델에서 현재 위치
- α^i : 진폭 크기 조정 파라미터
- ρ^i : 시간 축에서 압축-신장 파라미터

여기서 i 는 손의 움직임 궤적을 나타내는 것으로서 $i \in \{l, r\}$ 이다. 본 논문에서는 왼손과 오른손의 궤적에 대해서 각각 천이와 스케일을 조정할 수 있도록 하였으므로 독립적인 두 개의 파라미터가 필요하다. 즉 ϕ^l 은 왼손의 궤적에 대한 모델의 현재 위치를 나타내고, ϕ^r 은 오른손의 궤적에 대한 모델에서 위치를 나타낸다.

요약하면 각 상태를 나타내기 위해서 본 논문에서는 7가지의 파라미터를 사용하게 된다.

가. 초기화 단계(initialization)

표본 집합은 처음단계에서 초기화할 때 N 개의 표본들이 균등한 가중치인 $\frac{1}{N}$ 로 시작 가능한 상태들에 대해서 균등하게 분포하고 있는 것으로 한다. 구체적으로 하면 초기 파라미터들은 다음식과 같이 정해진다.

$$\begin{aligned} \mu &\in [1, \mu_{\max}] \\ \phi^i &= \frac{1 - \sqrt{y}}{\sqrt{y}} \quad \text{단, } y \in [0, 1] \\ \alpha^i &\in [\alpha_{\min}, \alpha_{\max}] \\ \rho^i &\in [\rho_{\min}, \rho_{\max}] \end{aligned} \tag{1}$$

나. 예측 단계(prediction)

예측단계에서는 랜덤하게 표본화된 s_t 의 각 파라미터들의 다음 단계를 결정하는 것이다. 즉 각각의 특수한 s_t 의 파라미터들에 기반을 두어 s_{t+1} 를 결정하는데 사용된다. 각각의 과거 상태 s_t 는 표본 집합으로부터 각 표본의 가중치에 기반을 두어 랜덤하게 선택된다. 즉 각 표본의 가중치는 선택된 표본의 확률을 결정한다. 이것은 축적 확률 테이블(cumulative probability table)을 생성하여서 $[0,1]$ 사이의 균등 랜덤 번호(uniform random number)를 선택함으로써 효율적으로 선택할 수 있다. 그리고 표본을 추출하기 위해서는 이진 탐색(binary search)[3] 방법을 사용한다.

새로운 상태를 예측하기 위해서는 다음 식을 사용한다.

$$\begin{aligned} \mu_{t+1} &= \mu_t \\ \phi_{t+1}^i &= \phi_t^i + \rho_t^i + N(\sigma_\phi) \\ \alpha_{t+1}^i &= \alpha_t^i + N(\sigma_\alpha) \\ \rho_{t+1}^i &= \rho_t^i + N(\sigma_\rho) \end{aligned} \tag{2}$$

여기서 $N(\sigma^*)$ 는 표준 편차 σ^* 를 갖는 정규 분포에 따라서 랜덤하게 선택한 값이다. 이것은 각 예측치에 불확실 요소를 더해줌으로서 표본 집합이 잡음 데이터를 충분히 포함할 수 있는 확률을 유지하도록 하고 있다.

주어진 추출 샘플에 대해서 예측은 모든 파라미터들이 허용 범위 내에 있을 때까지 발생된다. 만일 시도를 하였으나 유효한 예측을 발생하는 것이 불가능하다면 새로운 표본이 위의 초기화 절차에 따라서 발생된다. 추가로 새로운 표본 집합에서 모든 표본들의 10%는 앞에서 기술한 초기화 단계에서와 같이 랜덤하게 초기화된다. 이렇게 함으로서 국부 최대값이 과정을 완전히 장악하지 못하게 할 수 있다. 즉 새로운 가설들이 항상 주도할 수 있는 기회가 주어지게 된다.

다. 갱신 단계(updating)

위의 예측단계 다음으로는 새로운 N 개의 예측 표본들의 집합이 있는데 이것들에 대해서 가중치를 할당해 주어야 한다. 각 표본의 가중치는 관측된 데이터 $Z_t = (z_t, z_{t-1}, \dots)$ 가 주어졌을 때 그것에 대한 우도(likelihood)의 측정(measure)가 된다.

시간에 대하여 i 번째 계수에 대한 측정값들의 열은 다음과 같이 정의한다.

$$Z_{t,i} = (z_{t,i}, z_{(t-1),i}, \dots) \quad (3)$$

본 논문에서 사용하는 측정값들은 4개의 계수를 가지고 있다. 즉 왼손의 수평속도, 왼손의 수직속도와 오른손의 수평속도, 수직속도이다. 따라서 측정값 열은 각각 $Z_{t,1}, Z_{t,2}, Z_{t,3}, Z_{t,4}$ 와 같이 표시한다.

참고문헌[2]의 Black and Jepson의 식을 확장하여 본 논문에서는 다음과 같은 가중치를 계산하는 식을 사용하였다.

$$p(z_t | s_t) = \prod_{i=1}^4 p(Z_{t,i} | s_t) \quad (4)$$

단,

$$P(z_{t,i} | s_t) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{\sum_{j=0}^{w-1} (z_{(t-j),i} - \alpha^* m_{(\phi^* - \rho^* j), i}^{(\mu)})^2}{2(w-1)} \right\} \text{이고, } w$$

는 시간 축에서 뒤쪽으로 후진하는 임시 창의 크기이다. 위식에서 α^*, ϕ^* 와 ρ^* 는 이미지의 덩어리(Blob)에 대한 모델의 적절한 파라미터를 나타내고 있으며, $\alpha^* m_{(\phi^* - \rho^* j), i}^{(\mu)}$ 는 모델 μ 가 시간 $\phi^* - \rho^* j$ 에서 interpolate 됨과 동시에 α^* 에 의해서 스케일된 모델 μ 의 i 번째 계수에 주어진 값을 나타낸다.

라. 판별(classification)

지금까지 과정이 정확하게 정립되었다면 이제 남은 것은 입력되는 비디오 시퀀스에서 두 개의 동작 중에서 어느 것인가를 판별해내는 것만 남게 된다. CONDENSATION의 전체 아이디어는 가장 그럴듯한 가설이 중국에는 지배하도록 하는 것이기 때문에 본 논문에서는 전체 비디오 시퀀스의 분류를 결정하기 위해서 비디오 시퀀스의 마지막에 있는 것을 가장 그럴듯한 것

으로 간주하는 것을 모델의 조건으로 선택하였다. 각 모델에 할당된 확률을 결정하는 것은 간단하게 주어진 순간에 표본 집합에서 각 표본의 가중치들을 더하여서 구한다. 여기서 상태는 찾고자하는 모델을 의미한다.

III. 동작 모델 및 사전 이미지 처리

제안된 입자필터의 성능을 시험하기 위하여 본 논문에서는 두 개의 동작 모델을 선택하였다. 그림 3에 보인 것과 같이 동작 1은 잔류(그대로 둬)를 나타내는 동작이고 동작 2는 노젓기 동작을 나타낸 것이다.

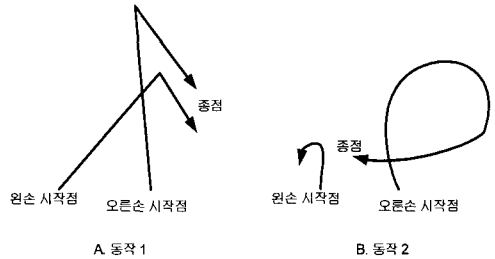


그림 3. 2개의 동작 모델
Fig. 3 Two gesture model

가. 실험 영상 준비

실험 영상은 디지털 캠코더(SONY DCR)를 이용하여 촬영하였고, 촬영된 영상은 프리미어 패키지를 이용하여 프레임으로 분해한 후에 320x240 픽셀 단위로 JPG 형식으로 저장하였다. 저장된 프레임 중에서 동작 과정을 나타내는 핵심적인 내용을 포함하고 있는 7장을 2개의 동작 모델별로 선택하여 실험 영상으로 채택하였다(그림 4 참조).



그림 4. 2개의 모델 동작 과정 영상
Fig. 4 gesture images of the two models

나. 피부색 픽셀 도출

피부색 픽셀 부분을 조각(segment)으로 도려내기 위하여 MATLAB 툴을 이용한 프로그램을 작성하였다. 동

작 이미지의 각각의 장에 대하여 피부(skin), 배경(background), 의복(clothes), 외곽(outlier) 등의 4가지 요소로 구분하여 분리작업을 수행하였다.

피부색 도출 과정은 먼저 피부, 배경, 의복에 대한 마스크를 설정하고 마스크를 이용하여 각각에 대한 평균값(mean)과 공분산값(covariance)을 구한 후에 이것을 가우스 분포로 잡아서 피부색 조각을 도출하였다. 결과는 그림 5와 같다.

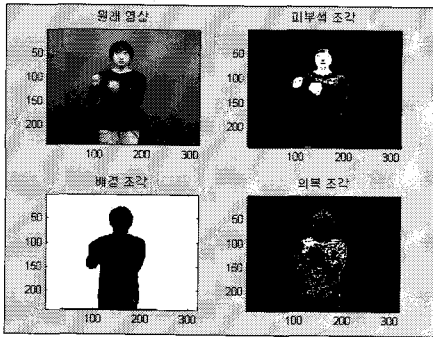


그림 5. 조각 구하기 결과
Fig. 5 Output of segmentation

다. 피부색 덩어리 중심 구하기

이제는 각 이미지에서 세 개의 가장 큰 피부색 덩어리(왼손, 오른손, 얼굴)의 중심을 구한다. 중심을 구하기 위해서 본 논문에서는 피부 픽셀 마스크를 왼쪽에서부터 오른쪽으로 위에서 아래로 스캔하면서 반복적으로 조사하였다. 만일 피부 픽셀이 기존의 덩어리에 오게 되면 주변을 조사해서 주변이 덩어리에 같이 있는가를 조사해서 만일 같이 있지 않으면 새로운 덩어리로 인식하는 방법을 사용하였다. 이렇게 해서 구한 덩어리는 세 개 되었으며 각각의 중심을 구하였다.

라. 덩어리의 이동 궤적 계산

이제 덩어리의 궤적을 추적하는 것은 간단한 문제다. 먼저 각 프레임에서 세 개의 덩어리의 중심의 목록을 만든다. 그리고 나서 이미지 열(sequence)중에서 첫 번째 프레임을 조사해서 어떤 것이 왼손과 오른손하고 가장 먼가를 결정한다. 왜냐면 하나만 결정하면 나머지 하나는 자동으로 결정되기 때문이다. 계속되는 열(sequence)에서는 어떤 것의 중심이 전번 것의 왼쪽 중심과 가장 가까운가를 결정한다. 그리고 이것을 새로운 왼쪽 중심으

로 갱신한다. 이것은 오른쪽에 대해서도 같은 과정을 반복한다. 일단 두 개의 덩어리가 결정되면 수평 및 수직 속도를 계산한다. 그리고 이 값을 나중에 모델 인식을 위해서 저장해 둔다.

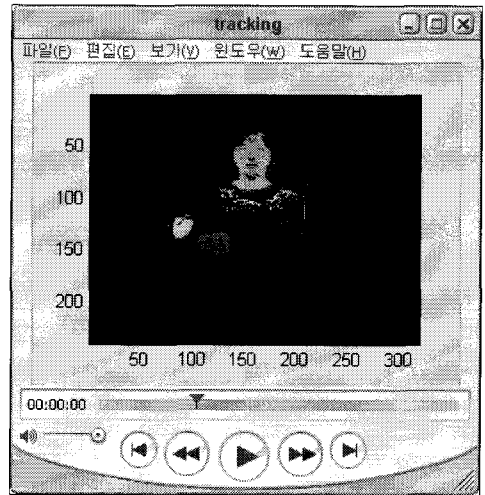


그림 6. 중심을 이용한 추적 결과
Fig. 6 Tracking result using centroid calculation

마. 운동 모델의 생성

각 신호에 포함된 손 운동의 모델을 생성하기 위해서 각 프레임에서 3 개의 훈련 예제를 이용하여 양손의 수직 및 수평 속도를 계산하였다. 다음 그림 7은 도출된 모델의 수평 및 수직 움직임 속도를 보인 것이다.

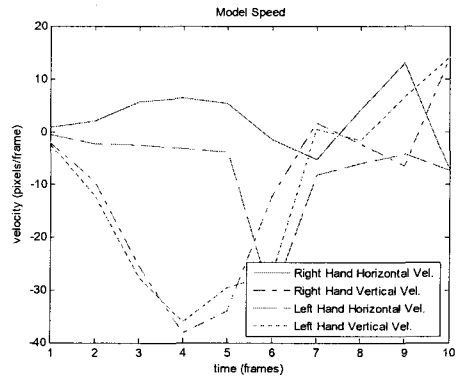


그림 7. 모델 1의 속도
Fig. 7 Velocity of Model 1

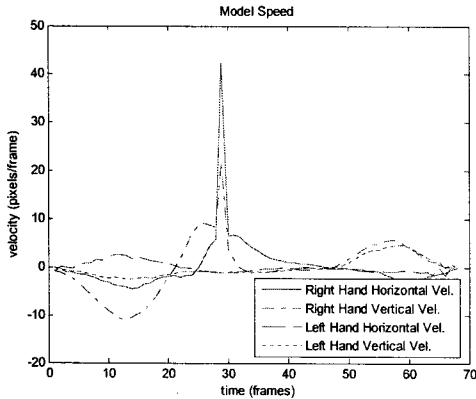


그림 8. 모델2의 속도
Fig. 8 Velocity of Model2

IV. 실험 결과 및 고찰

제안된 입자 필터의 성능을 시험하기 위하여 본 논문에서는 두 개의 동작 모델을 선택하였다. 즉 그림 2에 보인 것과 같은 모델을 이용하였다. 입자필터의 계수는 50%의 스케일을 하기 위하여 $\mu_{max} = 2, \alpha_{min} = 0.5, \alpha_{max} = 1.5, \rho_{min} = 0.5, \rho_{max} = 1.5$ 와 같이 설정해 주었다. 한편 $\sigma_\phi = \sigma_\alpha = \sigma_\rho = 0.1$ 로, 식(4)의 $w = 10$ 으로 설정해 주었다.

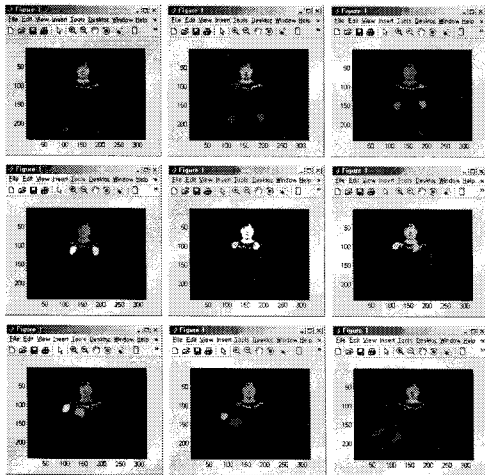


그림 9. 입자 필터를 이용한 모델 1추적 과정(왼쪽에서 오른쪽으로, 위에서 아래로)
Fig. 9. The tracking process of particle filter for the Model 1(From left to right, top to down)

입자 필터를 이용하여 모델 1과 2에 대한 추적 및 동작 인식과정을 그림 9, 10에 나타냈다. 그림 11,12는 CONDENSATION 필터의 추적 결과를 보였다. 그림 11에서는 이미지 픽셀은 왼쪽 상단이 원점이므로 머리 부분이 아래쪽에 오고 손이 위에 보인다. Model1의 추적은 상승부분에서 분포도가 커지는 점은 있으나 추적을 놓치지 않은 것을 확인할 수 있다. 그림 12는 Model2의 추적 결과를 보인 것이다. 그림 11과 마찬가지로 머리 부분이 아래쪽에 위치하고 있으며, 또한 추적 결과는 매우 양호한 것을 알 수 있다.

추적결과를 바탕으로 모델 결정은 이루어진다. 모델 인식 결정에 대한 확률은 그림 13,14에 보였다. 그림 13은 한 개의 제스처에 대한 인식 확률을 보였고, 그림 14는 두 개의 동작을 연속해서 수행할 때 인식확률을 보였다.

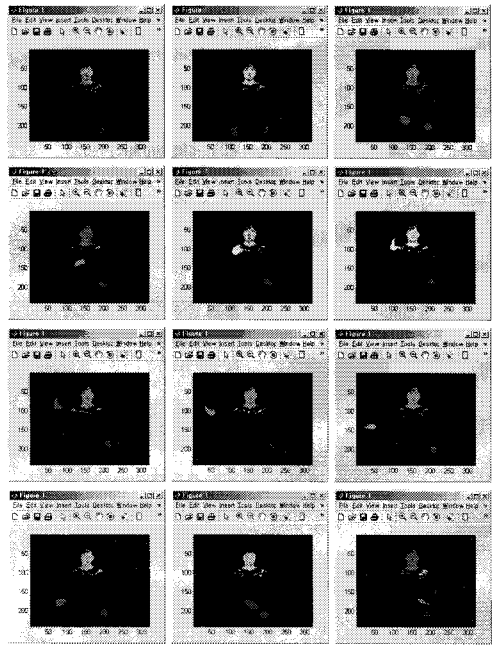


그림 10. 입자필터를 사용한 모델 2 추적 과정
Fig. 10 The tracking process of particle filter for the Model 2

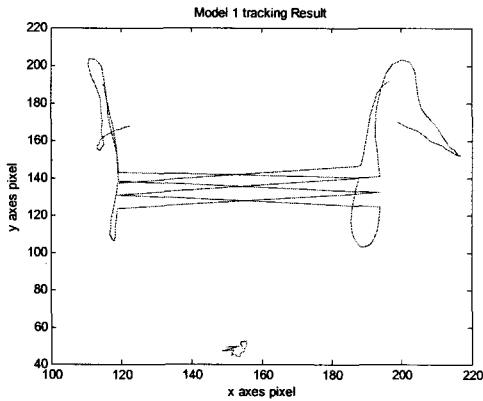


그림 11. Model1의 추적 결과
Fig11. Tracking result of Model1

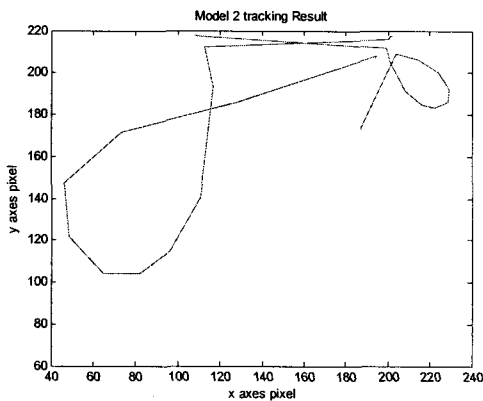


그림 12. Model2의 추적결과
Fig 12. Tracking Result of Model2

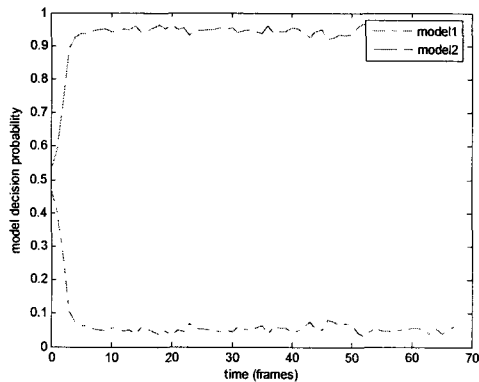


그림 13 1개의 동작에 대한 모델 판정 확률
Fig.13 Model Decision Probability of Model1

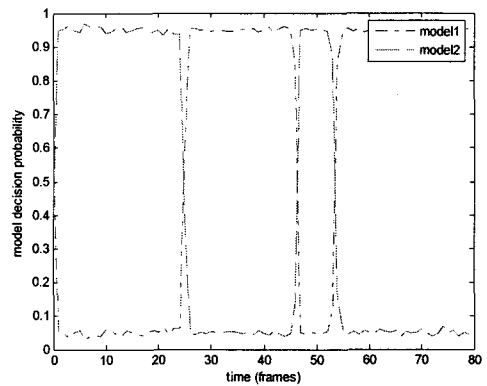


그림 14 두 개의 동작이 연속으로 이루어질 때 모델 판정 확률
Fig. 14 Model decision probability for the continuous gestures

지금까지의 실험결과 CONDENSATION에 기반을 둔 입자필터는 두 개의 동작인식 및 추적 성능에서 매우 강한 특성을 갖고 있음을 확인하였다.

V. 결 론

본 논문에서는 피부색, 의복, 배경과 같은 다중 관측 모델로 확장된 CONDENSATION 알고리즘을 이용하여 인간이 동작할 때 사용되는 양손을 추적하였다. 사전 밀도(prior density)와 사후 확률(posterior density)로 측정된 다중 물체의 특징은 추적 대상 물체의 위치, 척도 요소, 기준 좌표에 대한 정보의 상태를 평가하기 위해서 전파된다. 실험 결과를 통해, 본 논문에서 제안하고 있는 필터는 동작 인식을 위해서 사용 될 수 있음을 증명하였다.

참고문헌

- [1] M. Isard and A. Blake, "CONDENSATION - conditional density propagation for visual tracking", Int. J. Computer Vision, 1998
- [2] Michael J. Black and Allan D. Jepson, "A probabilistic framework for matching temporal trajectories: Condensation-based recognition of gestures and expressions", In Proceedings 5th European Conf. Computer Vision, Vol. 1, pp. 909-924, 1998
- [3] Michael Isard and Andrew Blake, "A mixed-state condensation tracker with automatic model-switching," In Proceedings 6th Internal Conf. computer Vision, pp. 107-112, 1998

저자소개



이 양 원(Yang-Weon Lee)

1992년 서울대학교제어계측공학과 졸업(공학석사)

1996년 포항공과대학교전자전기공학과졸업(공학박사)

1996-현재:호남대학교 정보통신공학과 교수

1982-1996년 국방과학연구소 선임연구원

※ 관심분야 : 신호처리, 인터넷 응용기술, stochastic filtering