

동적 베이스망(Dynamic Bayesian Network)을 이용한

양손 제스처 인식

석홍일 신봉기^o

부경대학교 컴퓨터공학과

daedalos@pknu.ac, bkshin@pknu.ac.kr

Two-Hand Gesture Recognition with Dynamic Bayesian Network

Heung-II, Suk Bong-Kee, Sin^o

Computer Engineering, Pukyong National University

사람은 사람 또는 사물과의 상호 작용에 있어 끊임없이 손을 사용한다. 예를 들면, 특정 사물 또는 위치를 가리키거나 다른 사람과의 대화 속에서 자신의 감정이나 생각을 표현하기 위해 손을 이용한다. 그리고, 손은 신체의 다른 어떤 부분보다도 많은 것을 표현할 수 있고, 움직임 또한 자유롭다. 따라서, 손동작은 컴퓨터와 자연스러운 상호작용을 하는 방법 또는 인터페이스가 될 수 있다. 본 논문에서는 미디어 플레이어를 제어하기 위한 10가지 동작들을 정의하고 인식한다. 비디오 영상에서 손과 손, 손과 얼굴 사이에서의 겹침 현상에서도 정확한 추적이 가능한 방법을 소개하며, 인식 모델로는 동적 베이스망(DBN) 프레임워크[1]를 이용한다. 한 손 제스처 인식 중심의 기존 연구[2, 3, 4]와 달리, 한 손을 이용한 것(5가지)과 두 손을 모두 이용하는 것(5가지)을 포함한다.

입력 비디오 영상에서 피부 영역을 검출하기 위해서는 두 가지의 방법을 결합하여 사용한다. 첫 번째는 가장 간단하고 널리 이용되는 것 중 하나로 RGB 색상 모델을 YIQ 색상 모델로 변환하여 I의 값이 미리 정해진 임계치의 범위를 만족하는 경우 피부색으로 선택한다. 두 번째는 Haar-like 얼굴 검출기[5]로 검출된 얼굴 픽셀들을 이용하여, 현재의 빛 조건 및 사용자의 피부색 특성을 반영한 피부색 모델을 생성한다. 이는 Haar-like 얼굴 검출기가 빛 조건의 변화에서 정면을 향하고 있는 얼굴을 잘 찾는다는 것을 이용한 것이다. 영상 내에 있는 사용자의 얼굴을 검출하고, 검출된 얼굴 부분에서 너무 어둡거나, 너무 밝지 않은 픽셀들을 이용하여 그 사용자에게 맞는 피부 색상 모델을 생성한다. Haar-like 얼굴 검출기를 이용한 피부 색상 모델은 HSV 색상 모델에서의 색상 hue값의 히스토그램이다. 임의의 픽셀에 대하여 그 픽셀의 피부색으로서의 확률값을 히스토그램에서 계산하여 임계치보다 큰 경우 피부색으로 결정한다. 이 방법은 조명의 변화에도 빠르게 적응하며, 임의의 사용자에게 대해서도 강인한 검출을 할 수 있다는 장점이 있다. 손과 얼굴 영역의 추적을 위해 개개의 손 및 얼굴에 해당하는 각 영역을 하나의 가우스(Gaussian) 모델로 표현하여 추적하는 Argyros 등[6]의 방법을 이용하되, 현재 프레임에서의 위치를 예측하기 위해 이전 두 프레임에서의 변화율(velocity)을 사용한 기존의 방법과는 달리 광류(optical flow)를 사용한다. 이전 두 프레임에서의 변화율을 이용하는 경우, 손의 움직임이 일정한 속도를 가질 때 비교적 정확한 추적이 가능하지만, 그렇지 못한 경우에는 오차를 범하게 되고, 추적에 실패하게 된다. 그러나, 광류는 이전 프레임에서 현재 프레임으로의 움직임에 대한 정보를 직접적으로 제공해주므로 추적하고자 하는 영역이 일정한 속도를 가지거나 선형적으로 움직여야 한다는 제약이 없어도 추적이 가능하다.

손의 움직임 정보를 표현하기 위해서는 손 영역을 모델링하는 가우스 분포에서의 평균의 변화를 이용한다. 그러나 손의 움직임에 대한 방향 정보만으로는 서로 다른 동작 간의 명확한 구분이 어려우므로, 오른손과 왼손의 상대적 위치에 대한 정보와 손과 얼굴과의 상대적 위치 정보를 또 다른 특징으로 이용한다. 그러면 방향 코드만을 사용했을 때 발생하는 동작들 간의 모호성을 해결할 수 있다.

본 논문에 정의된 동작들은 한 손 동작뿐 아니라 양손을 모두 사용하는 동작을 포함하며, 양손을

이용한 동작의 경우 각 손의 동작들은 서로 연관성을 가진다. 이런 동작들은 coupled HMM으로 모델링 될 수도 있겠으나 한 손 동작을 고려하게 되면, 방향 코드만을 관측 데이터로 사용하는 경우에는 다른 동작으로 잘못 인식될 수도 있다. 이를 해결하기 위해 coupled HMM에 변형을 가하여 두 손의 상대적 위치를 관측 데이터로 가지는 은닉 상태 노드 X^3 를 추가하여 새로운 동적 베이스망을 정의한다. 이때, 각 손을 모델링하는 기존의 두 노드(X^1, X^2)가 새로운 노드 X^3 에 대한 원인이 되도록 화살표를 연결하고, 연속한 두 시간에서의 베이스망의 관계는 1차 마르코프(Markov) 과정을 가정한다. 이웃한 두 시간에서의 X^3 노드간의 연결($X_{t-1}^3 \rightarrow X_t^3$)은 각 시간에서의 다른 두 노드 (X_{t-1}^1, X_{t-1}^2), (X_t^1, X_t^2)에 대한 정보를 X_{t-1}^3 와 X_t^3 가 각각 가지고 있으므로 이는 이웃한 두 시간에서의 X^1, X^2 를 간접적으로 coupling 시키는 결과를 가진다.

모델의 훈련 및 입력 시퀀스에 대한 출력값의 계산을 위해서 interface 알고리즘[7]을 이용하였다. 실험 데이터는 PC 카메라를 이용하여 촬영하였고, 7명이 10개의 제스처에 대해서로 다른 날 각각 7번씩 촬영하여 490개의 비디오 데이터로 구성하였다. 초당 30프레임으로 촬영하였으며, 각 프레임의 크기는 320×240이고, 24비트 컬러 색상이다. 실험에 사용된 프로그램 코드는 Visual C++ 6.0과 Matlab 7.0으로 작성하였으며, Intel OpenCV 라이브러리와 Bayes Net Toolbox(BNT)를 이용하였다. 모델의 성능은 교차 검증(leave-one-out) 방법을 이용하여 각 제스처에 대해 42개의 비디오 데이터로 모델을 훈련을 하고, 나머지 7개의 비디오 데이터로 인식률을 평가하였다. 총 7번씩을 실험하여 평균 확률을 인식률로 결정하였다. 각 제스처 모델의 은닉 노드에 대한 값의 범위는 동작의 복잡성을 고려하여 2부터 4사이로 변경하며 실험한 결과 최대 99.57%의 인식률을 얻었다.

참고 문헌

- [1] K. P. Murphy, "Dynamic Bayesian Network: Representation, Inference and Learning," PhD Dissertation, University of California, Berkeley, 2002.
- [2] H. K. Lee, and J. H. Kim, "An HMM-Based Threshold Model Approach for Gesture Recognition," *IEEE Transaction on PAMI*, vol.21, no.10, pp. 961-973, 1999.
- [3] H. H. Avilés-Arriaga, L. E. Sucar, and C. E. Mendoza, "Visual Recognition of Similar Gestures," In *Proceedings of ICPR*, pp.1100-1103, 2006.
- [4] S. F. Wong, and R. Cipolla, "Continuous Gesture Recognition using a Sparse Bayesian Classifier," In *Proceedings of ICPR*, pp.1084-1087, 2006.
- [5] P. Viola, and M. J. Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision*, vol.57, no.2, pp.137-154, 2004.
- [6] A. A. Argyros, and M. I.A. Lourakis, "Real-time tracking of multiple skin-colored objects with a possibly moving camera," In *Proceedings of ECCV*, vol. 3, pp.368-379, 2004.
- [7] A. Darwiche, "Constant Space Reasoning in Dynamic Bayesian Networks," *International Journal of Approximate Reasoning*, vol. 26, no. 3, pp. 161-178, 2001.