

실시간 제스처 인식을 위한 클러스터링 기반 다수 비제스처 거부 모델

윤종원[○] 민준기 조성배

연세대학교 컴퓨터과학과

jwyoons@scslab.yonsei.ac.kr, loomlike@scslab.yonsei.ac.kr, sbcho@cs.yonsei.ac.kr

Clustering-based Multiple Non-gesture Rejection Model for On-line Gesture Recognition

Jong-Won Yoon[○] Jun-Ki Min Sung-Bae Cho

Dept. of Computer Science, Yonsei University

최근 사람이 특정 동작을 취함으로써 컴퓨터와 상호작용을 수행할 수 있는 제스처 기반 인터페이스가 주목받고 있다. 그 중 3축 가속도 센서 기반 제스처는 입체적인 움직임을 인지하는 데에 유리하다는 장점이 있어 널리 사용되고 있다[1].

제스처 기반 인터페이스를 사용할 때에는 정의된 제스처가 아닌 의미없는 비제스처도 상당히 많이 포함된다. 특히 제스처와 유사한 비제스처가 존재하면 제스처 인식기만으로는 정확한 인식이 힘들기 때문에 보다 정확한 인식을 위해서는 비제스처를 인식 대상에서 거부시키는 것이 중요하다[2]. 그러나 정의되어있는 제스처와는 달리 비제스처는 다양한 상황에서 발생할 수 있으며 분포가 매우 광범위하고 복잡하기 때문에, 하나의 인식 모델을 이용한 거부[3]는 모든 비제스처에 대한 정확한 학습 및 거부가 어렵다는 한계가 존재한다. 따라서 본 논문에서는 은닉 마르코프 모델(Hidden Markov Model, HMM) 기반 제스처 인식에 있어서 보다 정확한 인식을 위해, 클러스터링 기법을 이용하여 비제스처를 세부 종류들로 구분한 뒤, 각 세부 종류를 담당하는 국소 비제스처 거부기들을 생성하여 비제스처를 거부하는 모델을 제안한다. 그림 1은 제안하는 비제스처 거부 모델을 보여준다.

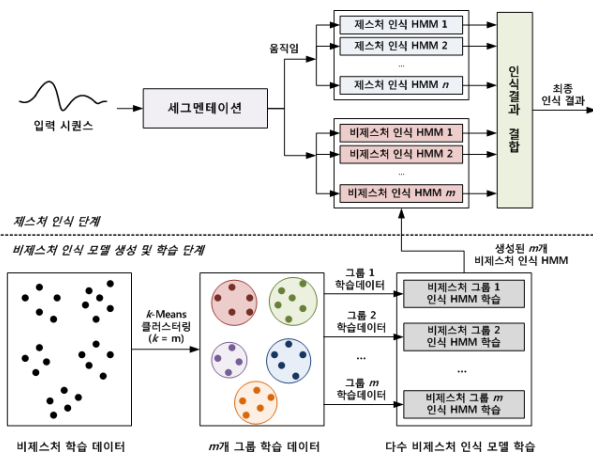


그림 1. 제안하는 비제스처 거부 모델

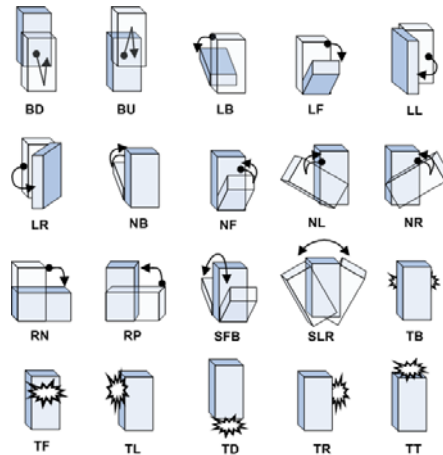


그림 2. 실험에 사용된 모바일 UI 제스처 셋

비제스처를 보다 정확하게 인식하기 위해서는 앞서 거론된 비제스처 분포의 특성을 고려할 때, 비제스처를 몇 개의 그룹으로 나눈 뒤 특정 비제스처 유형에 최적화된 국소 비제스처 모델 다수를 사용하는 것이 유리하다. 이를 위해 제안하는 방법은 전체 비제스처 학습 데이터를 k -Means 클러스터링 기법을 이용하여 그룹화 한 뒤, 각각의 그룹에 특화된 인식 모델을 생성한다. N 개의 미리 정의된 제스처에 대한 학습 데이터와 더불어 전체 비제스처 학습 데이터로부터 클러스터링을 수행하여 M 개의 비제스처 그룹 데이터를 얻을 수 있으며, 이후 Baum-Welch 알고리즘을 이용한 학습을 통해 최종적으로 N 개의 제스처 인식 모델 $\lambda_g (g=1,2,\dots,N)$ 과 M 개의 비제스처 인식 모델 $\lambda'_h (h=1,2,\dots,M)$ 을 얻을 수 있다.

인식 단계에서 비제스처의 검출 및 거부를 위해 본 논문에서는 제스처 인식 결과와 비제스처 인식 결과의 차이를 이용하는 방법을 제안한다. 입력 구간 x 에 대한 인식 결과의 차이 $\delta(x)$ 는 수식 (1)과 같

이 정의된다. w 는 비제스처 거부를 수행하는 정도를 의미하는 가중치이다. 이후 $\delta(X)$ 를 이용한 X 에 대한 최종 인식 결과 $D(X)$ 는 수식 (2)와 같이 정의된다.

$$\delta(X) = \max_g P(x|\lambda_g) - (w \times \max_h P(X|\lambda'_h)) \quad (1)$$

$$D(x) = \begin{cases} g = \operatorname{argmax}_g P(X|\lambda_g), & \text{if } \delta(X) \geq 0 \\ \text{non-gesture}, & \text{if } \delta(X) < 0 \end{cases} \quad (2)$$

최종 인식 결과 $D(X)$ 는 $\delta(X)$ 가 0 이상일 경우, 입력 구간 X 를 제스처로 판단하여 가장 높은 결과 값을 가지는 제스처를 인식 결과로 사용하며, 0보다 작을 경우 비제스처로 판단하여 인식을 거부한다.

제안하는 방법의 인식 성능 평가를 위해 그림 2와 같이 모바일 UI에서 사용될 수 있는 20가지의 제스처를 정의하였다. 제스처는 사용자가 모바일 기기를 한 손에 쥐고 손쉽게 수행할 수 있는 동작들로 정의하였으며, 3축 가속도 센서를 내장한 삼성 T-Omnia 스마트폰을 이용하여 5명의 20~30대 남성을 대상으로 한 제스처당 10개씩 총 2000개가 수집되었다. 비제스처 데이터는 수집된 데이터 중 제스처가 아닌 움직임을 추출하여 총 350개가 수집되었다.

인식에서 사용되는 HMM은 실험을 통해 최적의 결과를 보여주는 10개의 상태수와 20개의 심볼수를 갖도록 설정하였다. 실험에서는 비제스처를 거부하지 않는 인식 방법과, 제스처 인식 모델의 결과와 임계치를 이용하여 비제스처를 거부하는 모델과의 성능 비교를 수행하였다. 모든 실험 결과는 Five-folded cross validation을 통해 얻었으며, 실험 과정을 10회씩 수행 후 평균값을 사용하였다.

그림 3은 비교 실험 결과를 보여준다. PM은 제안하는 방법이며, No NGD는 비제스처 거부를 수행하지 않는 방법, 그리고 THR -10 ~ -50은 각각 임계치를 지정된 값으로 사용하여 비제스처를 거부하는 방법들이다. 실험 결과 제안하는 방법은 인식 정확도에 있어서 다른 모델들에 비해 최소 18.8%, 최대 28.6%의 인식 정확도가 향상된 인식 성능을 보임을 확인할 수 있다.

또한 제안하는 방법과 임계치 기반 비제스처 거부 방법 간의 성능 차이를 보다 자세히 분석하기 위해 실험 결과를 이용하여 그림 4와 같이 ROC 곡선을 얻었다. 제안하는 방법은 ROC 곡선 상에서 동일한 False positive rate(FPR)에서 임계치 기반 방법에 비해 높은 True positive rate(TPR)을 가지며 보다 우수한 비제스처 거부 성능을 보임을 확인할 수 있다.

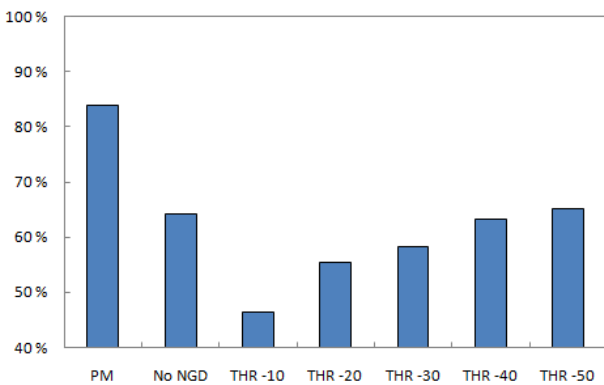


그림 3. 인식 정확도 비교 실험 결과

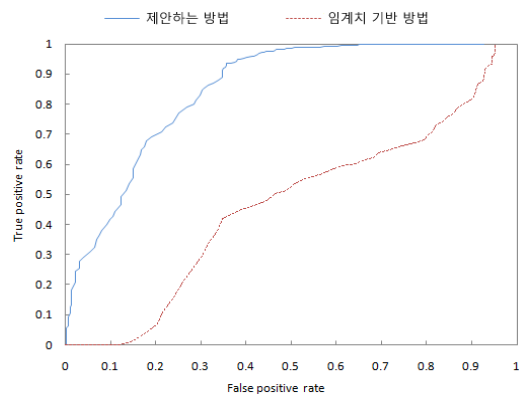


그림 4. 제안하는 방법 및 임계치 기반 방법의 ROC 곡선

감사의 글

본 논문은 지식경제부 산업원천기술개발사업(10035348, 모바일 플랫폼 기반 계획 및 학습 인지 모델 프레임워크 기술 개발)의 지원으로 수행되었음.

참고 문헌

- [1] J. Liu, L. Zhong, J. Wickramasuriya, and V. Vasudevan, "uWave: Accelerometer-based personalized gesture recognition and its applications," *Pervasive and Mobile Computing*, vol. 5, no. 6, pp. 657-675, 2009.
- [2] X.-H. Liu and C.-S. Chua, "Rejection of non-meaningful activities for HMM-based activity recognition system," *Image and Vision Computing*, vol. 28, no. 6, pp. 865-871, 2010.
- [3] D.-H. Kim, J.-Y. Song, and D.-J. Kim, "Simultaneous gesture segmentation and recognition based on forward spotting accumulative HMMs," *Pattern Recognition*, vol. 40, no. 11, pp. 3012-3026, 2007.