

관성 센서를 이용한 공간상의 제스처 입력 시스템

조성정^o 방원철, 장욱, 최은석, 양징, 오종구, 강경호, 조준기, 김동원
삼성종합기술원

{sung-jung.cho, wc.bang, wook.chang, eunseok.choi, jing.yang, jong.oh, kyoungho.kang, handle.cho, kdy2891}@samsung.com

Gesture Input System in 3-D Space by Using Inertial Sensors

Sung-Jung Cho^o, Won-Chul Bang, Wook Chang, Eunseok Choi, Yang Jing, Jonggu Oh, Kyungho Kang, Joonkee Cho, Dong Yoon Kim, Samsung Advanced Institute of Technology

요약

본 논문은 3차원 상에서 사용자의 동작을 관성센서로 입력받아 제스처를 인식하는 시스템을 소개한다. 사용자가 취한 제스처 동작은 관성 센서를 통하여 각속도 및 가속도 신호열로 변환된다. 궤적 추정 알고리즘은 이를 2차원 상의 동작 궤적으로 변환한다. 인식 알고리즘은 이 동작 궤적을 입력받아 베이지안 네트워크에 기반한 제스처 모델들로부터의 likelihood를 계산한 후, 최대 likelihood를 갖는 모델을 선택하여 인식을 수행한다. 16 명의 필자로부터 13개의 제스처 동작을 각 24회씩 수집하여 실험한 결과 평균 99.4%의 인식률을 얻었다.

1. 서론

컴퓨터가 우리 생활 곳곳에 널리 퍼져 활용되는 Ubiquitous computing 환경의 확산에 따라 사람과 컴퓨터 사이의 자연스러운 interaction에 대한 필요성이 증대되고 있다. 이에 따라 기존의 키보드 및 마우스 기반 입력에서 확장되어, 사용자의 음성, 필기, 영상 등을 인식하여 입력하는 방식이 많이 활용되고 있다. 최근에는 센서가 급속히 소형화, 저가화되고, 메모리와 CPU 등의 컴퓨팅 기술, 패턴 인식 등의 인공지능 기술이 발전함에 따라 새로운 입력 장치들이 대두되고 있다.

3 차원 공간상의 사용자의 동작을 인식하는 제스처 입력 방식은 곳곳에 편재한 컴퓨터가 내재된, 지능적 객체와의 자연스럽고 직관적인 interaction으로 주목을 받고 있다. 대표적인 연구에는 비전에 기반한 제스처로 가전 제품을 조정하고 가상공간상의 객체를 조작하는 시스템 [1], TV, VTR, DVD 등의 가전제품을 제어하는 리모콘을 들 수 있다 [2].

제스처를 입력받는 방식은 센서의 설치 방식에 따라 크게 두 가지로 나눌 수 있다. 첫째는 카메라, 초음파 거리 센서 등을 사용자 외부에 설치하는 방식이다. 둘째는 관성센서 등을 사용자에게 부착하는 내부설치(self-contained) 방식이다 [3-5]. 첫째 방식의 경우에는 비교적 정확하게 사용자의 동작을 입력받을 수 있지만, 설치 장소에서만 제스처를 입력하는 단점이 있다. 둘째 방식은 관성센서의 특성상 동작 궤적 추정의 정밀도가 제한되지만, 장소에 구애받지 않고 어디에서나 제스처를 입력하는 장점이 있다.

본 논문에서는 관성센서를 사용하여 사용자의 제스처를 인식하는 시스템을 제시한다. 사용자의 동작은 가속도계(accelerometer) 및 각속도계(gyroscope)를 통하여 가속도 및 각속도 신호로 입력된다. 이는 궤적

추정 알고리즘을 통하여 2 차원상의 동작 궤적으로 변환된다. 제스처 인식 알고리즘은 이 궤적들을 베이지안 네트워크로 학습하여 인식한다.

2. 제스처 입력 장치의 소개

2.1. 입력 장치 하드웨어

그림 1 은 제안하는 입력장치의 하드웨어 구성이다[2, 5,6]. 기관위에 X, Y, Z 축상의 각속도를 측정하는 3 개의 1 축 자이로스코프, X-Z 축을 측정하는 2 축 가속도계, Y 축의 1 축 가속도계가 실장되어 있다. 그리고, 아날로그 센서 신호를 디지털값으로 변환하는 A/D 변환기, 궤적 추정과 인식 SW 실행용 DSP (digital signal processor), 프로그램 적재용 플래시 메모리가 존재한다. 외부 통신용으로 가전제품 제어용 적외선 통신 포트, PC 상으로 데이터 전송을 위한 시리얼 포트가 있다. 또한 사용자가 누르는 동안 동작을 입력받는 버튼이 존재한다.

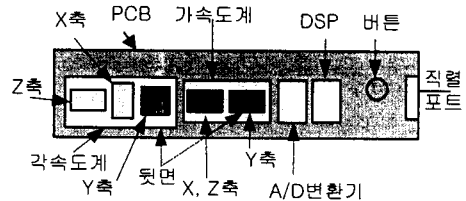


그림 1 입력장치의 하드웨어 구성도

2.2 제스처 입력

사용자는 그림 1 의 입력장치를 손에 쥐고 공간상에서 그림 2 의 13 개 제스처 동작을 취한다. 이 입력장치로 숫자열을 입력하는 task를 실험하기 위하여 0-9 까지의 숫자와 3 개의 제어 동작(delete, cancel, enter), 도합 13 개의 제스처를 정의하였다. 제스처 모양을 디자인 할 때는 모양간 유사도가 작아서 인식이 높고, 동작을 입력시 사용 편의성이 증대되도록 고려하였다 [6].

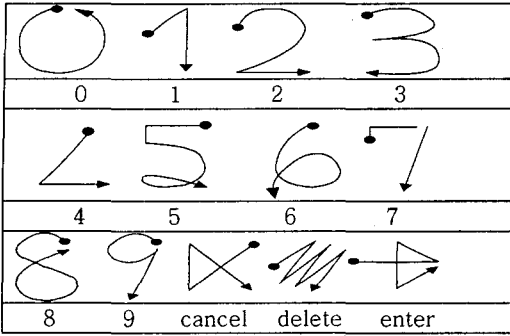


그림 2. 13개 제스처 동작 [6]

3. 제스처 인식 시스템

사용자의 손동작으로부터 제스처를 인식하는 전체 시스템의 흐름도는 그림 3 과 같다 [6,7]. 관성 센서는 사용자 동작에 따른 가속도 및 각속도 신호를 생성한다. 이 신호는 궤적 추정 알고리즘을 통하여 2 차원 평면상의 필기체적으로 변환된다. 최종적으로 필기체적은 제스처 인식 알고리즘을 통하여 제스처로 변환된다.

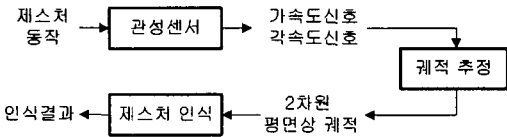


그림 3. 제스처 인식 흐름도

3.1 궤적 추정 알고리즘

그림 4 는 궤적 추정 알고리즘의 개요이다 [5-7]. 센서로부터 입력된 가속도 및 각속도 신호를 통하여 입력 장치의 현재 공간상의 자세가 계산된다. 이로부터 자세 변화에 무관한 외부 좌표계를 기준으로 계산된 가속도 신호로 변환한다. 이 가속도 신호를 두번 적분하여 3 차원 상의 위치를 추정하고, 일련의 궤적을 생성한다. 마지막으로 이 궤적을 2 차원 필기 평면상으로 투영하여 2 차원 좌표를 추출한다. 오류:

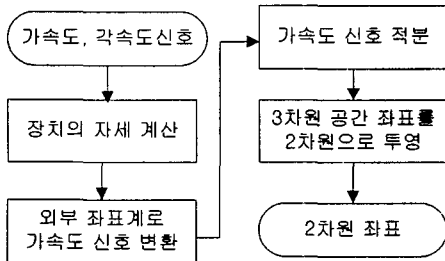


그림 4. 궤적 추정 알고리즘 개요

3.2 인식 알고리즘

추정된 궤적으로부터 제스처를 판별하기 위하여, 베이지안 네트워크 [8]에 기반한 온라인 필기 인식 시스템을 사용하였다 [9,10]. 이 시스템에서는 각 제스처 (m)마다 하나의 제스처 모델 (λ_m)을 갖는다.

T 개의 점 O_1, \dots, O_T 순서대로 입력되면, 인식은 아래와 같이 가장 높은 likelihood 를 출력하는 모델 λ^* 를 탐색하여 수행된다.

$$\lambda^* = \arg \max_m P(\lambda_m | O_1, \dots, O_T) = \arg \max_m P(\lambda_m)P(O_1, \dots, O_T | \lambda_m)$$

베이지안 네트워크 기반 인식기는 직선에 준하는 기본획과 이들 사이의 관계를 명시적이고 확률적으로 모델링하기 때문에, 기존의 전통적인 은닉마르코프 모델이나 원형정합 방법론에 비하여 강한 인식성능을 보인다 [9,10]. 각 제스처는 구성요소와 이들의 의존관계로 계층적으로 표현된다. 1 단계에서는 제스처가 기본획의 조합과 이들간 관계로 표현되고, 2 단계로 기본획은 점들과 이들간 관계로 표현된다. 점의 X-Y 좌표는 2 차원 가우시안 분포로 모델링된다. 그리고 구성요소간 의존관계는 이들의 확률적 의존관계로 표현한다.

그림 5 (a)는 두개의 기본획을 갖는 제스처 모델과 이에 상응하는 베이지안 네트워크 모델을 보인다. 이 모델에서 노드들은 점모델에 해당하고, 아크는 이들간 의존관계에 해당한다. 이 제스처 모델은 3 개의 기본획 끝점 모델 EP_0, EP_1, EP_2 와 6 개의 기본획 내부점 모델 $IP_{1,1}, \dots, IP_{2,3}$ 을 갖는다. 그림 5 (b)는 이 점 모델들에 매칭되는 점들을 보여준다($ep_0, ep_1, ep_2, ip_{1,1}, \dots, ip_{2,3}$). 제스처 모델의 매칭 확률은 각 점모델과 이에 상응하는 X-Y 상의 점의 좌표를 매칭한 확률을 결합하여 계산된다.

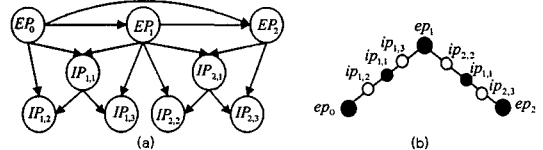


그림 5.(a) 두 기본획을 갖는 제스처 모델 (b) 이 모델에 상응하는 입력 제스처에

4. 실험 및 결과

4.1 데이터 집합

본 제스처 입력 시스템의 인식 성능을 평가하기 위하여 16 명의 사용자에게서 데이터를 수집하였다. 이중 8 명은 본 입력 장치에 익숙하며, 나머지 8 명은 사용 경험이 없다. 사용자들은 데이터 수집 화면상에 나타난 제스처 레이블을 보고, 동작 입력 장치로 해당 제스처의 동작을 취하였다. 13 개 제스처에 대하여 각 사용자별로 24 회 동작을 취하였다. 사용자들에게는 제스처 동작의 시작 순간과 종료 순간에는 입력장치를 잡고 잠시 정지하도록 주지하였다. 그림 6 은 본 입력 장치로 수집한 데이터 중 일부의 그래프를 보여준다.

인식 성능은 학습에 참여하지 않은 데이터로 인식률을 측정하여 평가하였다. 16 명의 데이터중 사용자 1 부터 8 까지의 데이터를 학습집합으로 사용하였고, 사용자 9 에서 16 까지의 데이터를 테스트 집합으로 사용하였다. 두 집합에는 숙련도가 낮은 사람과 높은 사람이 골고루

분포하였다.

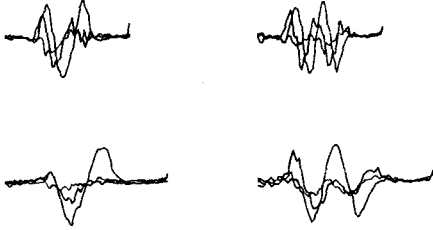


그림 6.관성 센서 신호 입력

4.2 실험 결과

궤적 추정 알고리즘에 의하여 그림 6의 관성 신호로부터 추정된 궤적이 그림 7에 나타나 있다. 각 궤적의 왼쪽 위에 놓여진 문자는 데이터의 레이블을 표시한다. 부분적인 궤적 모양은 자연스럽게 보이고, 제스처들간의 모양에 차이가 뚜렷하다. 그러나 전체적인 궤적 모양은 획의 시작과 끝부분에 noise가 존재하고, 실제 궤적 모양에서 다소 왜곡되었음을 볼 수 있다.

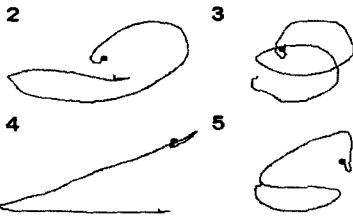


그림 7. 그림 6의 가속도 신호에서 추정된 궤적 예

각 사용자별 인식률이 그림 8에 나타나 있다. 필자 9-16에 대하여 모두 97% 이상의 인식률을 보여준다. 모든 사용자에게 대한 평균 인식률은 99.4%이다. 본 실험에 사용된 데이터의 1/2 정도가 장치를 처음 사용한 사용자에게서 얻었음을 감안할 때, 본 장치의 입력 성능이 상당히 신뢰성있음을 볼 수 있다.

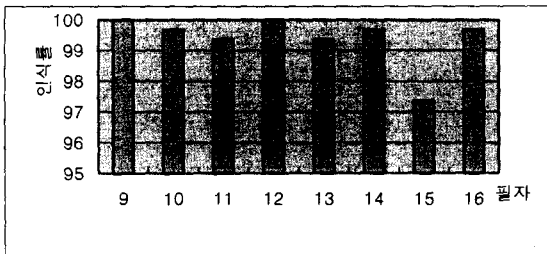


그림 8. 사용자별 인식률

5. 정리 및 결론

본 논문에서는 3 차원 상의 제스처 동작을 관성 센서로 입력받아 인식하는 시스템을 소개하였다. 관성 센서를

장착시에는 외부에 별도의 시스템이 설치될 필요가 없기 때문에, 임의의 장소에서 제스처 입력이 가능한 장점이 있다. 반면에 궤적을 관성 신호를 처리하여 추정하기 때문에 다소 궤적 추정의 정밀도가 떨어진다. 추정된 궤적은 베이지안 네트워크를 이용하여 인식하였다.

16 명의 초보자 및 숙련자에 대하여 13 개 제스처 동작을 입력하는 실험을 수행한 결과, 평균 99.4%의 인식률을 획득하였다. 이는 본 인식 시스템의 제스처 인식 성능이 상당히 신뢰할 수준임을 보여준다.

향후 연구로는 관성 센서의 특성을 심도있게 분석하여 인식 시스템의 전처리를 특화하는 것과 사용자에게 편리한 제스처 동작을 정의하는 것을 들 수 있다.

참고 문헌

- [1] V. Pavlovic, R. Sharma, T.S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review", IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, 1997, pp 677-695
- [2] 방원철, 장욱, 강경호, 최은석, 양정, 조성정, 외 3인, "관성 센서를 이용한 동작기반 리모컨", 13회 HCI-CG-VR-DESIGN 학술대회, 2004, pp 298-302
- [3] 조성정, 이택헌, 최현일, 김진형, "3 차원 가속도 신호의 문자 인식 방법에 관한 연구", 제 14 회 영상처리 및 이해에 관한 워크샵 발표 논문집, 제주, Jan. 2002
- [4] 장욱, 강경호, 최은석, Alexy Potanin, 방원철, 김동윤, "소형 관성 센서를 이용한 팬형 입력 장치의 개발", 12회 HCI-CG-VR-DESIGN 학술대회 2003, pp.318-322
- [5] W.-C. Bang, W. Chang, K.-H. Kang, E.-S. Choi, A. Potanin, and D.-Y. Kim, "Self-contained spatial input device for wearable computers," in Proc., 7th IEEE Int. Symp. on Wearable Computers, 2003, pp. 26- 34.
- [6] 조성정, 권순주, 강경호, 방원철, 장욱 외 5, "적합도 개선을 통한 동작 기반 입력 장치의 제스처 디자인 방법", 13회 HCI-CG-VR-DESIGN 학술대회 2004, pp 363-368
- [7] W. Chang, J. Yang, E.-S. Choi, W.-C. Bang, K.-H. Kang, S.-J. Cho, and D.-Y. Kim "A miniaturized attitude estimation system for a gesture-based input device with fuzzy logic approach," in Proc, 4th Int. Symp. on Advanced Intelligent Systems, Sep., 2003, pp. 616-619
- [8] F. Jensen, An Introduction to Bayesian Networks, Springer, New York, 1996
- [9] S.-J. Cho, J.H. Kim, "Bayesian network modeling of strokes and their relationships for on-line handwriting recognition," Pattern Recognition, Vol. 37, Issue 2, Feb., 2004, pp 253-264
- [10] S.-J. Cho, J.H. Kim, "Bayesian Network Modeling of Hangeul Characters for On-line Handwritten Recognition," Seventh Int. Conf. on Document Analysis and Recognition, Scotland, Aug., 2003, pp 207-211.