

---

## 모바일 사용자를 위한 3 차원 가속도기반 제스처 인터페이스

Gesture interface with 3D accelerometer for mobile users<sup>↓</sup>

최봉환, BongWhan Choe<sup>\*</sup>, 홍진혁, Jin-Hyuk Hong<sup>\*\*</sup>, 조성배, Sung-Bae Cho<sup>\*\*\*</sup>

---

**요약** ~ 최근 많은 시스템이 사용자에게 착용되어 사용자의 의도를 추론하고, 그에 맞는 서비스를 제공한다. 항상 사용자가 지니게 되는 모바일 기기에 장착될 센서에 의해 진행되는 이러한 흐름에서 가속도 센서는 이미 선두적인 역할을 하고 있다. 가속도 센서는 각종 움직임 정보를 수집하며, 제스처 기반의 사용자 인터페이스의 개발에 매우 유용하다. 보통 제스처 등의 시계열 패턴을 인식하기 위해서 많은 연산이 필요하며 연산능력이 상대적으로 부족한 모바일 환경에서는 보다 효율적인 기법이 요구된다. 본 논문은 저수준과 고수준으로 이루어진 모션 라이브러리 기반 2 단계 인식을 제안한다. DTW 를 기반으로 동작하는 3 차원 가속도 기반 저수준 인식기와 언어적으로 기술된 정보를 기반으로 하여 복합적인 동작에 대한 인식기로 구성된 모션 라이브러리를 구성하여 모바일 환경에 적합하도록 하였다.

**Abstract** ~ In these days, many systems are equipped with people to infer their intention and provide the corresponding service. People always carry their own mobile device with various sensors, and the accelerator takes a role in this environment. The accelerator collects motion information, which is useful for the development of gesture-based user interfaces. Generally, it needs to develop an effective method for the mobile environment that supports relatively less computational capability since huge computation is required to recognize time-series patterns such as gestures. In this paper, we propose a 2-stage motion recognizer composed of low-level and high-level motions based on the motion library. The low-level motion recognizer uses the dynamic time warping with 3D acceleration data, and the high-level motion is defined linguistically with the low-level motions.

**핵심어:** 동적 시간왜곡, 제스처 인식, 언어적 처리, 제스처 기반 인터페이스, *Gesture recognition, Dynamic time warping, Linguistic processing, Gesture based interface,*

---

<sup>\*</sup>주저자 : 연세대학교 컴퓨터과학과 석사과정; e-mail: [bitbyte@sclab.yonsei.ac.kr](mailto:bitbyte@sclab.yonsei.ac.kr)

<sup>\*\*</sup>공동저자 : 연세대학교 컴퓨터과학과 박사과정; e-mail: [hjinh@sclab.yonsei.ac.kr](mailto:hjinh@sclab.yonsei.ac.kr)

<sup>\*\*\*</sup>교신저자 : 연세대학교 컴퓨터과학과 교수; e-mail: [sbcho@yonsei.ac.kr](mailto:sbcho@yonsei.ac.kr)

## 1. 서론

최근 컴퓨터 시스템은 사용자와 일체화 되어 사용자의 의도를 추론하고, 그에 맞는 서비스를 제공하는 방향으로 발전하고 있다 [1]. 이러한 흐름에서 중요한 요소는 사용자의 현재 행위를 인식하여 사용자의 의도를 분석하는 기술이다. 사용자 행위 인식은 소리, 영상, 가속도 센서 등 가용한 모든 센서들을 사용해 시도되고 있다. 그 중 가속도 센서는 현재 하는 사용자의 동작을 추론하기에 적합하면서도, 장비의 이용에 사용자의 부담이 적다는 점에서 사용이 용이하다. 실제로 최근 많은 휴대폰 및 게임용 기기에서 3 차원 가속도 센서를 이용한 인터페이스가 개발되고 있다. 대표적인 예로 아이팟 터치나 PS3 의 6axis 같은 경우가 있다.

가속도를 사용한 사용자 행위(행동) 인식을 위한 연구는 지속적으로 있어 왔으나 대부분의 경우 행위 별로 별도의 인식기를 개발하여 사용하고 있다[13]. 모바일 기기의 처리 능력이 향상되고 있으나 그 이상으로 정교하게 행위를 인식하도록 요구되고 있다. 일반적으로 인식기의 개수 증가는 속도의 저하와 동일한 의미를 가진다. 따라서 인식기를 행위 별로 제작하는 것은 모바일에서 실시간 인식을 위한 용도로 사용하기 적합하지 않다.

이러한 와중에 사용자의 행위를 기본 단위로 구분하고 단위의 언어 조합으로 정리하고자 하는 시도가 있다[2]. 행위를 언어적으로 이해하여 처리하는 방식은 실시간에 인식기의 개수를 줄여 시스템의 부하를 적게 하고, 표현의 확장을 통한 인식 가능 행위가 증가하는 확장 가능한 방식이라는 점에서 의미가 있다.

가속도 센서를 사용한 순차 데이터로부터 인식을 하기 위한 알고리즘으로는 DTW(Dynamic Time Warping)를 많이 사용한다[3]. DTW 는 패턴 비교 알고리즘으로 입력과 기존의 패턴을 비교하여 그 차이를 구할 때 사용할 수 있다. 따라서 사용자의 예시를 통해 단일 패턴을 구하고 이를 학습 데이터로 사용하는 것도 가능하다[3-5].

이 논문에서는 DTW 기반으로 사용자 별로 학습이 가능한 모션 라이브러리를 개발하고자 한다. 본 논문에서 제안하는 라이브러리는 크게 1) 단위 자세 또는 동작 인식을 위한 저 수준 모션 라이브러리와 2) 복합적으로 구성된 동작 인식을 위한 언어적 모션 라이브러리의 2 단계로 나누어져 있다. 이렇듯 2 단계로 나누는 것으로 모션 라이브러리의 동작 성능을 올리고 더 적은 자원으로 많은 동작을 인식할 수 있다.

## 2. 관련 연구

최근 모바일 단말 형태의 기기를 이용한 사용자 인터페이스가 많은 관심을 받고 있으며, 가속도 등을 이용하여 움직임 정보를 수집하고 이를 해석하는 기술이 활발히 연구되고 있다[6-9]. Junker 등은 사용자의 팔에 가속도 센서를 부착하여 악수, 문열기 등의 일상 생활과 관련된 제스처를 인식하였다. Ferscha 등은 가속도 센서가 탑재된 모바일 단말 형태의 작은 기기로 표현 가능한 각종 제스처를 구성하여 제스처 라이브러리를 구축하였고 다양한 응용 어플리케이션과의 연동을 통해 그 유용성을 확인하였다[10]. 최근에는 가속도 정보를 단순한 수치 값이 아닌 언어적 해석을 통해 사용자의 행동이나 제스처를 인식하는 연구가 이루어지기도 하였는데, Guerra-Filho 와 Aloimonos 는 문맥자유문법을 이용하여 사용자의 움직임을 언어적으로 해석하는 기술을 제안하였다[2].

## 3. 제스처 인터페이스

### 3.1 개요

제스처 인식은 일반적으로 패턴 인식기술의 한 가지로 이해할 수 있다. 센서를 통해 입력되는 시계열 데이터를 그림 1 에서와 같은 과정을 통해 처리하고 그 결과로 인식된 패턴을 응용 프로그램에서 입력으로 사용한다.

본 논문에서는 XSens 社의 MTi 가속도 센서를 사용한다. MTi 센서에서 X, Y, Z 축으로 표현되는 가속도 값을 사용한다. 가속도 값은 센서를 기준으로

하는 좌표를 사용하며,  $m/s^2$  을 단위로 하는 값을 가진다[11].

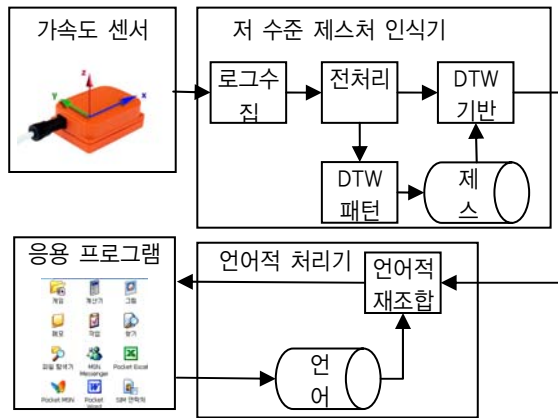


그림 1 가속도 센서 기반 제스처 인식 시스템

수집된 로그는 전처리를 통과하여 노이즈를 제거하고, 사용이 용이한 벡터 형태로 변환된다. 사용시에는 전처리에서 나온 출력을 DTW 기반 패턴 비교를 통해 한가지를 선택하고, 선택된 결과는 언어적 재조합기로 전달된다.

언어적 재조합기는 응용프로그램에서 요구되는 언어적 정보로부터 DTW 기반 패턴 비교에 의해 결정된 동작을 재구성하며 그 결과를 응용 프로그램에 전달한다. 이러한 과정을 통해 응용 프로그램은 한 개의 기본 동작을 한가지 의미로 사용할 수 있을 뿐만 아니라 다수의 동작을 조합하여 사용하는 방식으로 사용이 가능하다.

### 3.2 제스처 분석

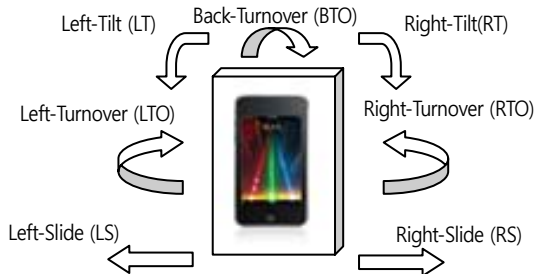


그림 2 인터페이스에서 사용할 제스처의 종류

제스처 인터페이스를 구성하기 위한 명령 동작은 입력을 명확하게 하기 위한 구분이 가능한 동작이어야 한다. 따라서 본 논문에서는 그림 2에서 제시된 형태의 동작을 수행한 후 원래의 위치로 돌아오는 것을 합쳐

명령 동작으로 간주했다. 해당 동작들은 오인식의 가능성이 적고, 일상 생활에서 의도하지 않고 나타나기 힘든 동작이기 때문에 인터페이스로 사용하기 용이하다. 또한 3 차원 가속도로 인식을 하고자 할 때 짧은 수행 시간에 비하여 특징이 명확한 동작이기도 하다.

동일한 특성을 가진 동작이라면 동적 시간 왜곡을 사용한 인식기의 특성상 사용자로부터 입력을 받은 패턴을 저장하는 것만으로도 인식하고자 하는 동작을 추가할 수 있는 장점을 가진다.

### 3.3 로그 데이터 전처리

본 논문에서는 최소한의 손떨림과 같은 진동을 배제하기 위해 단순 이동 평균 평활법(SMA Smoothing: Simple Moving Average Smoothing)을 사용하였다. 단순 이동 평균 평활법은 식 (1)과 같이 시간  $t$  에서  $t-n$  시간 동안의 평균을 가중치 없이 구해  $t$  의 값으로 사용하는 알고리즘이다.

$$SMA = (S_{t-n} + S_{t-n+1} + \dots + S_{t-1} + S_{t-n}) / n \quad (1)$$

동적 시간 왜곡 알고리즘은 실수 데이터를 입력 받아 두 패턴을 비교하는 알고리즘이기 때문에 기존에 입력된 실수 데이터와 비교하는 경우 성능이 좋다. 따라서 별도의 이산화 과정은 필요하지 않다.

입력된 X, Y, Z 축의 가속도 값은 X, Y, Z 로 된 벡터의 크기로 나누어 단위 벡터로 바꾸고, 가속도의 크기는 중력 값을 1 로 하는 최대-최소 정규화 기법을 사용해 정규화 했다. 이러한 과정을 통해 가속도의 변화와 가속 방향 변화를 구하기 쉬운 형태로 변경이 가능하다.

### 3.4 저수준 제스처 인식

인식을 위해서는 동적 시간왜곡 기법을 사용해 주어진 패턴과 입력 패턴의 거리를 측정하고, 제안된 방법에 의해 계산된 임계치 보다 작은 값이 나온다면 해당 패턴과 같은 동작으로 결정하는 방식을 사용한다. 동적 시간 왜곡 기법의 거리 함수는 다음 식 (2)와 같이 구했다.

다만, 동적 시간 왜곡 알고리즘의 시간 복잡도는  $O(w^2 n/s)$ 으로 매우 크다. 특히 윈도우 크기  $w$  에

기하급수적으로 비례하기 때문에 매우 느린 알고리즘이다. 따라서 본 논문에서는 Sakoe-Chuba Band 방식으로 변형된 동적 시간왜곡 알고리즘을 스트림 형태에서 사용할 수 있도록 변형하여 사용했다[12].

$$\sqrt{(X_L - X_R)^2 + (Y_L - Y_R)^2 + (Z_L - Z_R)^2} + |D_L - D_R| \quad (2)$$

인식기는 다중 슬라이딩 윈도우 기법을 적용해 동작하도록 했다. 즉 윈도우 크기( $w$ )에 대해 슬라이딩 크기( $s$ )가 주어 져 있다면,  $w/s$  개의 window 가 동시에 동작하여 위의 인식알고리즘을 수행하는 형태로 진행된다. 이러한 조치는 변형 알고리즘에서 DTW 의 왜곡 범위를 전체 윈도우 크기가 아닌  $D$  로 제한하는 것으로 속도와 정확도를 동시에 잡기 위한 방법으로 고안하였다. 다만, 이때 왜곡 범위  $D$  는 슬라이딩 크기와 동일하거나 크게 잡아야 한다.

인식기에서 특정 패턴이 채용될 지 여부를 결정하는 임계 치 값은 입력되어 있는 패턴 사이의 거리의 최대 값으로 결정한다. 즉, 입력 패턴이  $n$  개 있다고 할 때,  $n$  개의 패턴을 상호 비교한 거리의 최대값이 해당 패턴의 임계 치로 사용된다.

### 3.5 언어적 재조합

저 수준 제스처 인식기에서 인식된 결과는 응용프로그램에서 원하는 형태의 언어적 정보를 사용해 재 조합된다. 단일 동작만으로 구성된 경우보다 언어적으로 조합된 경우가 더 다양한 입력이 가능하다. 언어적 재조합을 위해서는 우선 인식할 동작에 대한 언어적 기술이 필요하다. 그림 3 은 이러한 언어로서 행동을 기술하는 매우 간단한 문법을 나타낸다.

```

GESTURE := BTO | LT | LTO | LS | RT | RTO | RS,
ACTION := GESTURE |
          ( GESTURE then ACTION ) |
          ( GESTURE or GESTURE )
COMMAND := [name] :: ACTION
    
```

그림 3 제스처의 언어적 기술 문법

CLICK :: BTO	PLAY/PAUSE :: BTO
DBLCLICK :: (BTO then BTO)	FF :: ( ( RS then BTO ) or ( RS then RS ) )
HOME :: LTO	FB :: ( ( LS then BTO ) or ( LS then LS ) )
ADDRESS :: RTO	STOP :: (RT then LT) or (LT then RT)
NEXT :: (RT then RT)	NEXT :: RTO
PREV :: (LT then LT)	PREV : LTO
LEFT :: LS	
RIGHT :: RS	
STOP :: (LT then RT)	

(a) 웹브라우저와 연동되는 (a) 음악 재생기와

그림 4 언어적으로 표현된 제스처 정보

그림 4 는 이러한 조합기능을 사용해 각각 웹 브라우저와 음악 재생기와 연동할 때의 언어적으로 기술된 제스처 조합의 예이다.

## 4. 실험 및 분석

### 4.1 수집 데이터

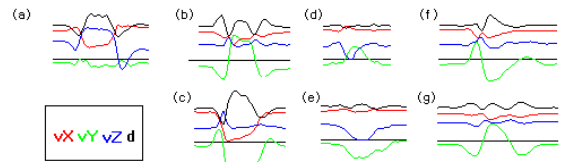


그림 5 전처리 된 데이터 예

(a) Back-Turnover, (b) Right-Tilt, (c) Left-Tilt, (d) Right-Turnover, (e) Left-Turnover, (f) Right-Slide, (g) Left-Slide

7 종류의 입력 제스처를 각 10 회씩 반복하여 입력한 데이터는 그림 5 와 같은 형태를 띠는 것을 확인 할 수 있다. 서로 다른 동작이 명확하게 다른 형태의 입력으로 나타남을 직접 확인 가능하다. 각 수집 횟수 별로 미묘한 차이는 있으나 전반적으로 동일한 형태의 값을 가지는 것을 알 수 있었다. 한가지 특징적인 사실은 값의 굴곡은 중력으로부터 영향을 가장 많이 받기 때문에 가속도 센서만으로도 장치의 회전 상태를 일부 알 수 있다는 사실을 확인하였다.

### 4.2 결과 및 분석

이렇게 수집된 데이터를 기반으로 종류당 5 개씩 임의로 묶은 뒤 나머지 종류당 5 개의 로그 정보를 임의로 재 배치하여 인식률을 실험하였다. 실험은 동일 실험을 5 회 반복 하였고, 반복한 결과 인식률은 표 1 과 같다.

표 1 5 회 실험 시 인식률

회차	1	2	3	4	5
정확도	0.67	0.64	0.49	0.64	0.62

대략 60%내외의 인식 성능을 보이는 것을 확인 할 수 있다. 이중에서 결과가 잘 나온 1 회 차의 혼동행렬은 표 2 와 같은데 이를 보면 전반 적으로 다른 동작이 BT 와 LS 동작으로 오인식 하는 경우가 많고, 나머지 동작에서는 비슷한 정도의 오 인식률을 가지는 것으로 확인 된다. 오인식의 원인은 1) DTW 의 거리 함수의 정확성, 2) 동작의 크기/세기에 따른 영향을 고려해야 할 것 같다. 또한 인식 구간을 설정 하는 문제에 있어서도 더 정확한 인식 구간을 사용할 수 있다면 높은 성능을 기대할 수 있을 것으로 보인다.

표 2 1 회 차 실험 결과의 혼동행렬

	맞 음	B T	L S	L T	LT O	R S	R T	RT O
BT	26 7	2 67	31	0	3	17	5	42
LS	36 8	19	36 8	16	0	2 0	0	2
LT	28 3	5	57 5	28 3	0	0	41	9
LT O	20 4	7 0	10 4	18	20 4	4	25	11
RS	31 2	1	75	22	0	31 2	3 9	0
RT	28 7	2 6	32	0	0	31	2 87	0
RT O	19 2	6 5	44	22	0	0	25	19 2

## 5. 결론 및 향후 연구

본 논문에서는 모바일 환경을 위한 인식기 기반 인터페이스를 위한 저 수준 인식기 모델에 대한 구현 및 성능, 사용성을 점검했다. 본 논문에서 한 이러한 저수준 인식기를 바탕으로 확장된 인식기를 개발 함으로써 사용자는 단순하고 이해 하기 쉬운 인터페이스를 제공 받을 수 있으며, 개발자는 인식 결과를 사용한 3.5 에서 제시된 시나리오와 같은 고수준 응용 프로그램을 개발하기 용이 할 것이다.

DTW 의 특성상 한 개의 입력에 대한 반복 인식 이나, 오인식의 가능성이 높으므로 오동작을 걸러 내는 문제를 해결 하기 위한 연구가 필요할 것이다.

## 참고문헌

- [1] Y.-P. Chen, J.-Y. Yang, S.-N. Liou, G.-Y. Lee and J.-S. Wang, "Online classifier construction algorithm for human activity detection using a tri-axial accelerometer," *Applied Mathematics and Computation*, vol. 205, no. 2, p.p. 849-860, 2008.
- [2] Y. Aloimonos, "HAL: Human activity language." *Journal of Vision*, vol. 8, no. 6, pp.1050, 1050a, 2008.
- [3] P. Capitani and P. Ciaccia, "Warping the time on data streams", *Data & Knowledge Engineering*, vol. 62, no. 3, pp. 438-458, 2007.
- [4] J. S. Yi , Y. S. Choi, J. A. Jacko and A. Sears, "Context awareness via a single device-attached accelerometer during mobile computing," *MobileHCI* vol. 111, pp. 303-306, 2005.
- [5] B. Hartmann, L. Abdulla, M. Mittal and S.R. Klemmer, "Authoring sensor-based interactions by demonstration with direct manipulation and pattern recognition," *Proc. of Human Factors in Computing System*, pp.145-154, 2007.
- [6] B. Coley, B. M. Jolles, A. Farron and K. Aminian, "Arm position during daily activity," *Gait & Posture*. vol. 28, no. 4, pp. 581-587, 2008.
- [7] R. Watson and T. College , "A survey of gesture recognition techniques," *computer science technical report TCD-CS-1993-11*, 1993.
- [8] S. Kallio, J. Kela, J. Mantyjarvi , "Online gesture recognition system for mobile interaction," *IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, pp. 2070-2076, 2003.
- [9] V.-M. Mantyla, J. Mantyjarvi, T. Seppanen and E. Tuulari, "Hand gesture recognition of a mobile device user," *Multimedia and Expo. ICME.*, vol. 1, pp. 281-284, 2000.
- [10] A. Ferscha, S. Resmerita, C. Holzmann and M. Reicho, "Orientation sensing for gesture-based interaction with smart artifacts," *Computer Communications*, vol. 28, no. 13, pp. 1552-1563, 2005.
- [11] MTi and MTx User Manual and Technical

Documentation, XSens Technologies,  
www.xsens.com.

- [12] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.
- [13] S. Mitra, T. Acharya, "Gesture recognition: a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 37, no. 3, pp. 311–324, 2007.