

# 연속시간 베이스 분류기를 이용한 손동작 인식

이인희<sup>1</sup>, 권우영<sup>1</sup>, 서일홍<sup>2</sup>

<sup>1</sup>한양대학교 전자컴퓨터통신공학부

<sup>2</sup>한양대학교 컴퓨터·소프트웨어학과

e-mail : *inhee.lee19@gmail.com, wykwon@hanyang.ac.kr, ihsuh@hanyang.ac.kr*

## Continuous Time Bayes' Classifier-Based Hand Activity Recognition

Inhee Lee<sup>1</sup>, Woo Young Kwon<sup>1</sup>, Il Hong Suh<sup>2</sup>

<sup>1</sup>Department of Electronics and Computer Engineering  
Hanyang University

<sup>2</sup>Department of Computer and Software  
Hanyang University

### Abstract

In this paper, we propose continuous time Bayes' classifier to recognize hand activities which is robust solution to both spatial and temporal uncertainty. In our proposed method, conditional probabilities of temporal events can be explicitly represented by defining temporal nodes in a Bayes' classifier. Utilizing these nodes, both temporal and spatial pattern can be simultaneously inferred in a unified framework. To validate our proposed method, we show the experimental results on hand gesture recognition.

### I. 서론

사람의 동작을 인식하고 이를 이용해 사람의 의도 및 상태를 예측하는 것은 인간-로봇 상호작용에서 중요하게 다루는 문제이다. 그 중에서도 사람의 손동작은 사람의 의도를 파악하는데 중요한 정보를 담고 있다.

손동작을 인식하기 위해서 다양한 특징 추출 방법이 사용되고 있다. 영상이나, 관성센서, 모션 캡처 장비로부터 손의 위치 정보획득한 후 시계열 패턴으로 변환하여 사용하는 방법 등이 가장 많이 사용되고 있으며, 손의 이동 방향을 이산시간 시계열 패턴(Discrete

time time series)으로 변환하여 사용하는 방법 또한 널리 사용되고 있다[1]. 시계열 패턴을 이용하여 손동작을 인식하는 방법으로는 은닉 마르코프 모델(Hidden Markov Model)[2], 동적 시간 정합 방법(Dynamic Time Warping)[3], 조건 랜덤 필드(Conditional Random Field) 등이 널리 이용된다[4].

본 논문에서는 이산시간 시계열 데이터를 사용하는 대신, time-tagged 이벤트 데이터를 사용하여 손동작을 인식하고자 한다. time-tagged 이벤트 데이터는 신호의 변화가 일어나는 시점에서의 신호 변화의 종류( $d$ )와 그 시간( $t$ )들의 집합으로 구성된다. 총  $N$ 번의 신호변화가 관측되었다면 time-tagged event는  $\{\langle d_1, t_1 \rangle, \dots, \langle d_N, t_N \rangle\}$ 와 같이 표현된다. 일반적으로 time-tagged 이벤트 데이터는 이산시간 시계열 패턴에 비해 적은 개수의 특징점만으로 신호를 표현할 수 있기 때문에 적은 계산량으로 높은 정확도의 분류가 가능하다. 하지만 time-tagged 이벤트 데이터는 노이즈에 민감하다는 문제점이 있다. 본 논문에서는 time-tagged 이벤트 데이터를 노이즈에 강인하게 분류할 수 있는 연속시간 베이스 분류 모델을 제안한다. 제안하는 방법은 이벤트의 종류와 시간을 각각 이산 확률변수와 연속 확률변수로 모델링하는 방법에 의해 공간적인 특징과 시간적인 특징을 동시에 고려하며, 이를 통해 손동작의 공간적 변화와 시간적 변화 모두에 강인한 특성을 보인다.

## II. 연속시간 베이스 분류기

### (Continuous Time Bayes' Classifier)

제안하는 연속시간 베이스 분류기에서는 인식하고자 하는 동작을 특정 시간에 발생하는 time-tagged 이벤트들의 집합으로 표현한다. 이는 시계열 패턴에 비해 보다 적은 데이터를 가지고도 인식하고자 하는 대상을 표현할 수 있다는 특징을 갖는다.

하나의 이벤트에 대한 종류와 시간은 결합확률로 표현한다. 예를 들어 오전 11시에서 오후 1시 사이에 점심식사를 시작할 시간적 확률은 매일 식사를 하는 확률과 식사를 할 경우에 그 시간이 오전 11시에서 오후 1시 사이가 될 확률의 곱으로 주어질 수 있다. 이를 수학적으로 표현하면, 일정 시간간격  $(t_1, t_2)$  사이에 어떤 이벤트  $X$ 가 관측될 확률은

$$\begin{aligned} P(X = true, t_1 < T_X < t_2) \\ &= P(X = true)P(t_1 < T_X < t_2 | X = true) \\ &= P(X = true) \int_{t_1}^{t_2} f_{T_X}(t_X) dt_X \end{aligned} \quad (1)$$

와 같이 표현될 수 있다.

여기서  $f_{T_A}(t_A)$ 는 이벤트의 발생 시간에 대한 확률변수인  $T_A$ 의 확률분포함수이며,  $t_A$ 는  $T_A$ 의 특정 시간에서의 값을 나타낸다. 본 논문에서는 확률 분포함수를 가우시안 분포함수로 나타내었다. 또한, 두 이벤트간의 조건부 확률은 다음과 같이 정의된다.

$$\begin{aligned} P(x_i, t_1 < T_X < t_2 | y_i, t_Y) \\ &= P(x_i | y_i) P(t_1 < T_X < t_2 | x_i, y_i, t_Y) \\ &= P(x_i | y_i) \int_{t_1}^{t_2} f_{T_X | T_Y}(t_X | t_Y) dt_{X|Y} \\ &= P(x_i | y_i) \int_{t_1}^{t_2} f_{T_Z}(t_X - t_Y) dt_Z \end{aligned} \quad (2)$$

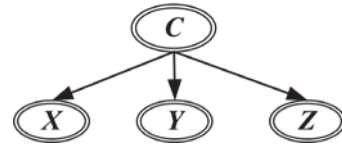
여기서  $Z$ 는  $T_X$ 와  $T_Y$ 의 시간차이에 대한 연속시간 확률 변수이며,  $x_i$ 는  $X = x_i$ 에 대한 단축 표현이다. 식 (2)의 연속시간 확률 분포함수는 가우시안 분포를 이용하면

$$\begin{aligned} P(x_i, t_1 < T_X < t_2 | y_i, t_Y) \\ &= P(x_i | y_i) \int_{t_1}^{t_2} N(t_X - t_Y; \mu, \sigma^2) dt_X \end{aligned} \quad (3)$$

으로 표현된다. 연속시간 베이스 분류기는 기존의 베이스 분류기를 시공간 패턴의 분류가 가능하도록 식(1)과 식(2)의 복합 확률표현을 사용하여 확장한 모델이다.

본 논문에서는 이벤트의 종류를 나타내는 이산확률변수  $X$ 와 그 시간에 대한 연속확률변수  $T_X$ 를 하나의 복합 노드  $\mathbf{X} = \{X, T_X\}$  로 표현한다. 이 복합노드를 이용한 시간적 베이지안 네트워크가 기존 연구로서 제

안되었다[5]. 시간적 베이지안 네트워크 구조를 베이스 분류기와 유사하게 만든 것이 본 논문에서 제안하는 연속시간 베이스 분류기이다. 그림 1은 연속시간 베이스 분류기의 예시이다. 여기서 두 줄로 나타낸 타원 표현은 이벤트의 종류와 그 시간 분포가 동시에 표현되는 복합노드를 의미한다.



[그림 1. 연속시간 베이스 분류기의 예]

어떤 이벤트들의 종류와 시간으로 구성된 데이터 집합  $\{\mathbf{x}_i, \mathbf{y}_j, \mathbf{z}_k\} = \{x_i, t_X, y_j, t_Y, z_k, t_Z\}$ 이 관측되었을 때, 이 데이터가 어떤 특정 클래스로부터 나왔을 확률은  $P(\text{class} | \mathbf{x}_i, \mathbf{y}_j, \mathbf{z}_k)$ 와 같이 표현된다. 베이스 규칙과 조건부 독립에 대한 정의를 활용하면 다음과 같은 식을 얻을 수 있다.

$$P(\mathbf{c} | \mathbf{x}_i, \mathbf{y}_j, \mathbf{z}_k) \approx P(\mathbf{c})P(\mathbf{x}_i | \mathbf{c})P(\mathbf{y}_j | \mathbf{c})P(\mathbf{z}_k | \mathbf{c}) \quad (4)$$

식 (2)와 (3)에 나타난 조건부 확률에 대한 정의를 이용하고, 복합 노드를 이산 확률 변수와 연속시간 확률 변수로 구분하여 표현하면 식 (4)는 다음과 같이 표현될 수 있다.

$$\begin{aligned} P(\mathbf{c} | \mathbf{x}_i, \mathbf{y}_j, \mathbf{z}_k) \approx P(\mathbf{c})P(x | \mathbf{c})P(y | \mathbf{c})P(z | \mathbf{c}) \\ N(t_C; \mu_C, \sigma_C^2)N(t_X - t_C | \mu_{XC}, \sigma_{XC}^2) \\ N(t_Y - t_C | \mu_{YC}, \sigma_{YC}^2)N(t_Z - t_C | \mu_{ZC}, \sigma_{ZC}^2) \end{aligned} \quad (5)$$

여기서,  $t_X, t_Y, t_Z$ 가 관측된 값이기 때문에 식(5)는 결과적으로  $t_C$ 에 대한 가우시안 분포로 나타나며, 이를 적분하면 관측된 데이터가 어떤 클래스에 속하는지에 대한 확률값을 얻을 수 있다. 식(5)로 표현되는 연속시간 베이스 분류기는 이벤트의 종류가 구성하는 패턴에 의한 확률값과 이벤트의 시간이 구성하는 패턴에 따른 확률값에 동시에 영향을 받기 때문에 시간적 변화와 공간적 변화에 강인한 분류가 가능하다. 또한 식(5)에 나타난 연속시간 확률 분포는 모두  $t_C$ 에 대한 상대시간 분포로 표현되기 때문에, 데이터가 관측되는 시간이 변하더라도 상대적인 특성이 유사하면 성공적으로 분류가 가능하다는 장점이 있다.

## III. 실험 및 결과

### 3.1 연속시간에서의 손동작 특징 추출

본 논문에서는 손동작에 대한 시계열 패턴 대신,

손동작에 따른 각 좌표축(x,y,z)의 속도변화가 최대 또는 최소가 되는 속도의 크기 및 그 시간을 특징으로 이용하였다. 그림 2는 학생 실험자가 책을 보고 공부하는 과정에서 발생하는 손동작의 예를 보이고 있다.

각 동작마다 손의 위치변화가 다른 패턴을 보이게 되며 그림3과 그림4는 그 예시이다. 그림 3(a)는 책에 밀줄 치는 동작에 대한 위치 궤적을 x,y 축으로 나누어 표현하고 있다. 이 위치궤적으로부터 속도를 추출한 결과가 그림3(b)이다.

각 좌표축의 속도 중에서 지역적인 최대점과 최소점을 peak와 valley 로 표시할 수 있다. 지역적인 최대점과 최소점을 추출하기 위해서 [6]에서 제시한 방법을 이용하였다. 실제 실험에서는 x,y,z 축의 3차원 정보를 이용하였으며, 손동작에 대한 좌표정보를 획득하기 위해 모션캡처 장비를 사용하였다.

### 3.2 손동작 인식을 위한 연속시간 베이스 분류기 설계

손동작에 따라서 속도 궤적의 극대점과 극소점은 복수의 데이터로 취득될 수 있다. 예를 들어, 그림 4(b)의 펼치하는 동작의 경우에는 5개 이상의 극대점과 극소점으로 데이터가 구성된다. 손동작의 종류에 따라서 극대점의 종류, 시간, 그리고 그 크기가 모두 다른 패턴을 보인다. 연속시간 베이스 분류기에서 관측되는 확률변수는 각 축의 속도 궤적에서 나타나는 극대·극소점의 크기 및 시간으로, 다음과 같이 정의하였다: x 축 극대점( $x_p$ ), x축 극소점( $x_v$ ), y축 극대점( $y_p$ ), y축 극소점( $y_v$ ), z축 극대점( $z_p$ ), z축 극소점( $z_v$ ). 극대점 및 극소점은 복수의 데이터셋으로 구성될 수 있으므로,  $x_{1p}$ ,  $x_{2p}$ 와 같이 관측된 순서에 따라 인덱싱 된다.

x,y,z 각 좌표축에 대한 속도의 극대점과 극소점은 6개로 구성되며, 5회 정도 반복되는 데이터라면 30개의 입력 변수를 갖는 분류기가 생성된다. 그림 5는 손동작 인식을 위한 분류기를 나타낸다. 또한 이산 변수로 구성되는 이벤트의 종류는 극대점과 극소점에서의 속도 크기로 정의하였다. 실험에서는 속도를 high, low 두 값 중 하나를 갖도록 정의하였다.

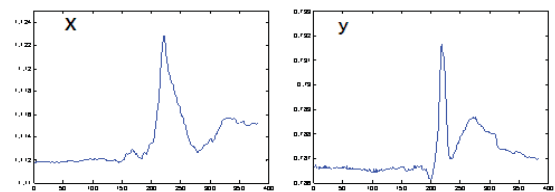
시공간 패턴을 이용한 손동작 분류 실험을 위하여 분류할 손동작을 밀줄치기, 펼치기, 지우기, 책장넘기기의 네 가지로 나누었다. 모션 캡처 장비를 이용하여 각 동작을 60번 씩 수행하여 얻어진 데이터를 바탕으로, 총 240개의 데이터 중에서 200개는 연속시간 베이스



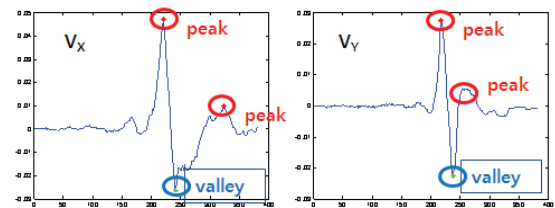
(a) 책에 밀줄치기 (b) 지우개 사용하기



(c) 책장 넘기기 (d) 책에 펼치기  
[그림2. 책 위에서의 손동작의 예]

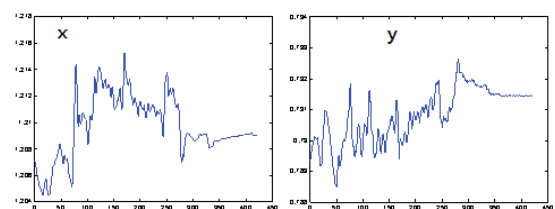


(a) 위치 궤적

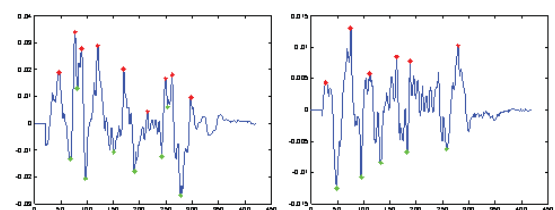


(b) 속도 궤적

[그림 3. 밀줄치는 동작의 3차원 공간 위치궤적 및 속도 궤적]



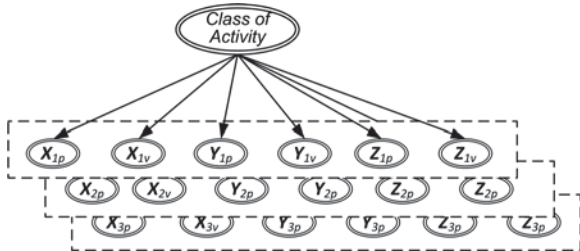
(a) 위치 궤적



(b) 속도 궤적

[그림4. 펼치하는 동작의 3차원 공간 위치궤적 및 속도 궤적]

스 분류기를 학습시키는데 사용되었으며, 40개는 성능을 검증하기 위한 테스트 데이터로 사용되었다. 연속시간 베이스 분류기의 성능 검증을 위하여 학습하는데 사용되었던 데이터 중 각 동작에서 10개씩 총 40개의 데이터와 테스트 데이터 40개의 인식 결과를 비교하였다. 그 결과는 표 1과 같다.



[그림5. 손동작 인식을 위한 연속시간 베이스 분류기]

[표 1 인식 결과]

동작 종류	트레이닝 데이터를 이용한 인식정확도	테스트 데이터를 이용한 인식정확도
밀줄긋기	100%	80%
책장넘기기	100%	100%
지우기	100%	100%
펼기하기	100%	100%

테스트 데이터를 이용한 결과를 보면 밀줄긋기의 정확도가 다른 동작에 비해 떨어지는 것을 볼 수 있다. 이는 각 동작의 길이 및 추출되는 특징의 개수와 관련이 있다. 밀줄긋기는 매우 짧은 시간동안 지속되는 동작이고, 다른 세 가지는 밀줄긋기에 비해 상대적으로 긴 시간동안 지속되는 동작이다. 즉, 밀줄긋기 동작은 다른 동작의 일부분과 유사할 가능성이 높기 때문에 다른 동작으로 오인식 될 가능성이 높다.

#### IV. 결론 및 향후 연구 방향

사람의 손동작은 대화를 할 때이나 매니플레이션 임무를 수행할 때에 사람의 의도를 내제하고 있어, 자연스러운 상호작용을 위해 중요한 정보로 인식되어 왔다. 본 논문에서는 손동작을 인식하기 위해 연속시간에서의 시공간 특징 패턴을 이용하여 분류하는 방법을 제안하였다. 그 결과, 공간적인 특징과 시간적인 특징을 동시에 고려한 특징을 추출하고, 이를 이용하여 만든 분류 모델을 사용함으로써 공간적 변화와 시간적

변화에 모두 강인한 성능을 보일 수 있다.

향후 연구 방향으로, 우선 제안한 방법을 기존 방법과 상세하게 비교하여 우수성을 검증할 필요가 있다. 또한, 시간정보를 포함한 동작 패턴을 이용하여 분류된 손동작 정보를 공부하는 동안 학습자의 상태를 추론해 내는 일에 유용한 관측 정보로 활용할 수 있으며, 사람과 로봇이 협업하는 임무를 수행할 때에 로봇이 사람의 행동을 인식하고 다음 행동을 예측하여 적절한 행동을 하는 일에도 적용 될 수 있다.

#### 감사의 글

이 논문(저서)은 2012년 교육과학기술부의 재원으로 한국연구재단의 지원을 받아 <실감교류 인체감응솔루션> 글로벌프론티어사업으로 수행된 연구임 (NRFM1AXA003-2011-0028353)

본 논문은 지식경제부 산업융합원천기술개발사업으로 지원된 연구결과입니다[10044009]

#### 참고문헌

- [1] 조선영, 변해란, 이희경, 차지훈, 키넥트 센서 데이터를 이용한 손 제스처 인식, 방송공학회논문지, 17권 3호, 2012년
- [2] Moni, M.I., HMM based hand gesture recognition: A review on techniques and approaches, Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on.
- [3] S. B. Wang,, Hidden Conditional Random Fields for Gesture Recognition, Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on.
- [4] Tan Wenjun , Dynamic hand gesture recognition using motion trajectories and key frames, Advanced Computer Control (ICACC), 2010 2nd International Conference on
- [5] W.Y Kwon, Il Hong Suh, A temporal Bayesian network with application to design of a proactive robotic assistant. IEEE International Conference on Robotics and Automation, 2013
- [6] G. Palshikar, Simple algorithms for peak detection in time-series, Proc. 1st Int. Conf. Advanced Data Analysis, Business Analytics and Intelligence, 2009