

# 키넥트센서 데이터를 이용한 동작 인식 및 분류

조병준<sup>1</sup> 장하영<sup>2</sup> 장병탁<sup>2</sup>

한국과학영재학교<sup>1</sup> 서울대학교 전기컴퓨터공학부<sup>2</sup>

josdee7@gmail.com, hyjang@bi.snu.ac.kr, btzhang@bi.snu.ac.kr

## Motion Recognition and Classification using Kinect Sensor Data

Byung-jun Cho<sup>01</sup> Ha-Young Jang<sup>2</sup> Byoung-Tak Zhang<sup>2</sup>

Korea Science Academy<sup>1</sup>

School of Computer Science and Engineering, Seoul National University<sup>2</sup>

### 요 약

사람들간의 의사소통에서 언어만큼 중요한 역할을 수행하는 것이 제스처이다. 사람은 제스처를 쉽게 이해하고 제스처를 이용한 의사소통 또한 자연스럽게 이루어지지만, 사람의 동작이 어떠한 의미를 가지고 있는지 파악하는 것은 컴퓨터에게 상당히 어려운 부분이다. 본 논문에서는 컴퓨터가 사람의 동작이 어떠한 의미를 가지고 있는지 파악하기 위한 전 단계로서 사람의 제스처를 인식하고 인식된 동작들을 다양한 기계학습 알고리즘을 이용하여 분류를 해보았다. MultiLayer Perceptron, Naïve Bayes, Decision Tree의 세가지 알고리즘의 실행결과 MultiLayer Perceptron이 가장 높은 정확도를 보여주었다.

### 1. 서 론

센서 기술의 발달로 인해서 키보드나 마우스를 대신하는 새로운 인터페이스들이 등장하고 있고, 최근에는 물리적이고 직관적인 인터페이스를 제공하는 NUI (Natural User Interface)가 많은 사람의 관심을 끌고 있다[1]. 센서를 이용하여 관측된 사용자의 자유로운 움직임에 기반한 인터페이스를 제공하는 NUI 기술은 다양한 센서들을 이용해서 구현되고 있는데 그 중 대표적인 것으로 Microsoft Kinect[2], Nintendo Wii[3] 등이 있다. 이러한 센서들을 이용해 컴퓨터는 사람의 동작을 인식 할 수 있게 되는데, 사람의 동작을 각 관절의 좌표값의 변화로 받아들이는 컴퓨터가 그러한 좌표값의 변화로 사람의 동작의 의미를 파악하는 것은 결코 쉬운 일이 아니다.

예를 들어 로봇이 사람이 움직이는 동작의 관절의 좌표값의 변화값을 입력 받아 그대로 움직임을 따라하는 것은 계산만 해주면 되는 것이기 때문에 크게 어렵지 않은 반면, 사람의 동작에 따라서 그 동작에 대응하는 적절한 동작을 로봇이 수행하는 것은 어려운 일이다. 즉, 사람의 행동에 로봇이 능동적으로 대처하는 것은 위의 좌표값 계산만으로는 힘든 일이다. 사람이 할 일을 대신 해주는 로봇은 관찮을지 모르나 사람을 위해서 어떠한 일을 수행하는 로봇에게는 사람의 동작의 의미를 알아내는 것이 중요하다.

본 연구에서는 센서를 통해서 감지된 동작의 의미 이해에 대한 전단계로서 다양한 기계학습 알고리즘을 이용한 동작의 인식 및 분류에 대해서 연구를 진행하였다.

### 2. 동작의 정의 및 분류 알고리즘

Microsoft 사에서 개발한 Kinect 라는 컨트롤러를 사용하여 사람의 동작을 인식하였고 Kinect에서 뽑아낸 관절의 좌표값들을 모션 센서 데이터 라고 정의하였다. Kinect는 적외선 센서와 카메라로 사람에서 20개의 관절을 정의하여 각 관절의 좌표값을 알아낼 수 있다.

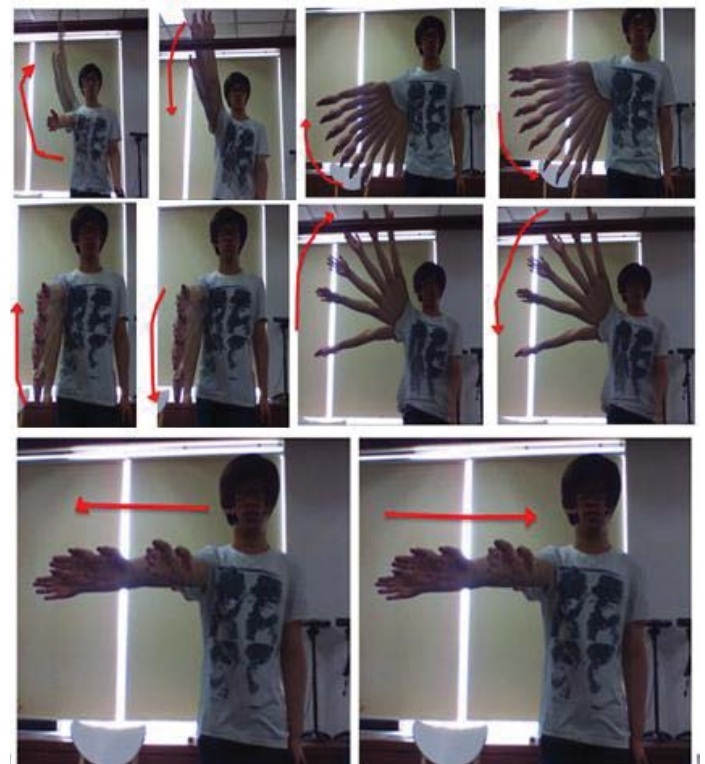


그림. 1 왼쪽 맨 위부터 차례로 동작 1번~10번

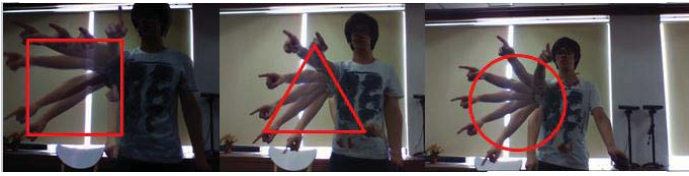


그림. 2 왼쪽부터 사각형, 삼각형, 원을 그리는 동작

동작은 그림. 1에 있는 10가지의 간단한 팔을 흔드는 동작들로 한 번, 그림. 2의 원, 사각형, 삼각형의 도형을 그리는 복잡한 동작으로 한 번 나누어서 분류를 해 보았고 13가지의 동작들을 한꺼번에 넣어서 분류를 진행해 보지는 않았다. 먼저 알고리즘이 동작 데이터의 분류를 얼마나 잘 하는지 알아보기 위한 테스트로 간단한 10가지의 동작을 정의해보았고, 10가지의 동작의 분류를 잘하는 것을 확인하고 실제 사람의 제스처와 비교적 비슷한 원, 사각형, 삼각형 그리기를 실험해 보았다.

기계학습 알고리즘 중에서는 MultiLayer Perceptron(MLP), Naïve Bayes, Decision Tree의 세 가지 알고리즘을 사용해 보았다.

MLP는 입력값을 받는 Input Layer, 출력값을 내놓는 Output Layer, 그 사이에 위치해 있는 Hidden Layer로 나누어진다. MLP는 Backpropagation 학습 알고리즘에 의해 각 Layer들 사이에 있는 weight값을 데이터에 가장 잘 맞도록 변화되고 이러한 과정을 학습이라 한다.[4]

Naïve Bayes 알고리즘은 주어진 데이터들을 통해서 새롭게 입력 받은 데이터가 각각의 클래스일 확률을 비교하여 가장 높은 것으로 분류한다. 그러한 분류과정은 다음과 같은 수식으로 나타내어 진다

$$\text{classify}(f_1, f_2, \dots, f_n) = \underset{c}{\text{argmax}} p(C = c) \prod_{i=1}^n p(F_i = f_i | C = c)$$

[5]

Decision Tree 알고리즘은 데이터를 가장 잘 나누는 feature를 골라서 그것을 기준으로 데이터들을 나누고 나누어진 데이터들 속에서 각각 같은 알고리즘을 수행한다. 데이터를 가장 잘 나누는 feature는 Information Gain이 가장 큰 feature로 결정하고 Information Gain은 다음과 같이 나타난다.

$$\text{Gain}(S, A) \equiv \text{Entropy}(S) - \sum_{\text{values}} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$H(x) = - \int p_x(x) \log_2 p_x(x) dx$$

[6]

### 3. 실험 결과 및 고찰

Weka[7]를 이용하여 학습을 시켰으며 MultiLayer Perceptron(MLP), Decision Tree, Naïve Bayes의 세 가지 알고리즘을 사용해 보았다.

[표. 1]에 있는 동작들의 경우에는 동작이 진행되는 시간 중 다섯 프레임을 뽑아내어 각 프레임별로 20개의 관절값들을 feature로 정해 총 300개의 feature로 Weka에 값을 주어 학습시켰다. [표. 2]의 제스처와 같은 경우에는 손 끝의 좌표값만을 따로 떼어내어 총 동작이 진행되는 시간 중 24프레임을 골라내어 총 72개의 feature들로 학습을 진행시켰다.

또한, 학습 데이터는 각 동작마다 400개의 데이터를 주어 10-fold cross-validation 으로 테스트를 진행하여 보았다. Decision Tree는 J48알고리즘으로 학습을 시켰고 MLP의 조건은 hidden layer=2, Learning Rate=0.3, Momentum=0.2 이었다.

번호	동작
1	90도튼상태에서90도들기
2	180도튼상태에서90도내리기
3	90도벌리기
4	90도오므리기
5	90도들기
6	90도내리기
7	90도벌린상태에서90도벌리기
8	180도벌린상태에서90도오므리기
9	90도튼상태에서90도벌리기
10	90도들고벌린상태에서90도오므리기

[표. 1] [그림. 1], [그림. 2], [그림. 3]의 동작과 동일

번호	동작
1	동그라미 그리기
2	삼각형 그리기
3	사각형 그리기

[표. 2] [그림. 4]의 동작과 동일

[표. 1]의 동작들에 대한 결과는 [표. 3]에, [표. 2]의 동작들에 대한 결과는 [표. 4]에 정리되어 있다.

알고리즘	정확도
MLP	97.20%
Naïve Bayes	100%
Decision Tree	100%

[표. 3] [표. 1]의 동작들에 대한 정확도

알고리즘	정확도
MLP	96.27%
Naïve Bayes	90.46%
Decision Tree	79.67%

[표. 4] [표. 2]의 동작들에 대한 정확도

[표. 1]의 동작들의 경우에는 [표. 3]에서 볼 수 있듯이 상당히 높은 정확도를 보여주는데, 이는 데이터들이 feature에서 하나의 값의 증가만 주시하면 되기 때문에 분류하기가 상대적으로 쉬워서 이렇게 나타난 것으로 보인다. [표. 2]의 동작들은 [표. 1]의 동작들에 비해서 복잡하기 때문에 정확도가 낮게 나왔고, MLP가 다른 알고리즘들에 비해서 분류를 모션 데이터 분류를 더 잘하고 상대적으로 Decision Tree가 잘못한다는 것도 알 수 있었다. 이는 원, 사각형, 삼각형을 그리는 동작이 [표. 1]의 동작과 비교하여 더 많은 feature가 변화하여 feature들의 변화값에 일정한 기준을 두어 판단하는 Decision Tree가 판단하기 힘들기 때문이라고 생각된다.

#### 4. 결론 및 향후 연구 계획

표에서 볼 수 있는 정확도들은 전부 90퍼센트대를 상회하는 아주 높은 값이었다. 이것을 통해 키넥트를 이용해서 얻은 동작 데이터를 기계학습 알고리즘이 굉장히 잘 분류해낸다는 것을 알 수 있다. 그러나, 각 동작들이 주어진 feature들 중의 보통 한 개 또는 두 세 개가 복잡한 변화 없이 증가 또는 감소하는 것을 파악하면 되는 것이었기 때문에 각 알고리즘들이 분류하기가 쉬웠던 것으로 생각된다. 그래도 기초적인 동작들만 정의를 했으므로 조금 더 복잡한 동작들, 사람들이 실제로 신호를 보낼 때 사용하는 동작들을 정의하여 실험을 더 해볼 필요는 있을 것 같다. 예를 들어 재난 시에 사용한다는 수신호나 두 팔이 동시에 움직이는 복잡한 동작들을 정의하여 분류를 실시해 보면 실제로 이용할 수 있을 확률도 높아 보인다. 만약 다른 동작들에서 눈에 띄는 정확도의 감소가 보인다면 그 원인을 분석하고 정확도를 더 올릴 수 있는 여러 가지 방안을 고안해 보고 분류된 결과를 토대로 로봇에게 적절한 행동을 하도록 명령을 전달하는 단계로 나아가면 될 것 같다.

#### 참고문헌

- [1] 이민규, 전재봉, “키넥트를 이용한 개인용 컴퓨터 제어”, *한국정보과학회*, 2012한국컴퓨터종합학술대회 논문집 제19권 1호, pp. 343-345
- [2] 조선영, 변혜란, 이희경, 차지훈, “키넥트 센서 데이터를 이용한 손 제스처 인식”, *방송공학회논문지* 제17권 제3호, pp. 447-458
- [3] Thomas Schlomer, Benjamin Poppinga, Niels Henze, Susanne Boll, “Gesture recognition with a Wii controller”, *Proceedings of the 2<sup>nd</sup> international conference on Tangible and embedded interaction*, pp. 11-14
- [4] H. Muhlenbein, “Limitations of multi-layer perceptron networks – steps towards genetic neural networks”, *Parallel Computing*, Vol. 14, Issue 3, August 1990, pp. 249~260
- [5] Caruana. R, Niculescu-Mizil. A, “An empirical comparison

- of supervised learning algorithms”, *Proceedings of the 23<sup>rd</sup> international conference on Machine learning*, pp. 161~168
- [6] Sagavian S. R., “A survey of decision tree classifier methodology”, *Systems, Man and Cybernetics*, Vol. 21, pp. 660~674
- [7] The University of Waikato, Weka, [www.cs.waikato.ac.nz/ml/weka/](http://www.cs.waikato.ac.nz/ml/weka/)