실시간 사용자 관절과 YOLOv3를 이용한 사용자 행동 검출

*오예준 **김상준 ***최희조 ****박구만
*동아방송예술대학교 콘텐츠제작학과
**서울과학기술대학교 정보통신미디어공학전공
***서울과학기술대학교 미디어IT공학과
****서울과학기술대학교 전자IT미디어공학과
yejun2108@naver.com

Detection of User Behavior Using Real-Time User Joints and YOLOv3

*Oh, Ye-Jun **Kim, Sang-Joon ***Choi, Hee-Jo ****Park, Goo-Man
*Dept. of Content Product, Dong-Ah institute of media and arts
**Dept. of Information Technology and Media Engineering
***Dept. of Media IT Engineering
****Dept. of Electronics and IT Media Engineering
Seoul National University of Science and Technology

요약

인물의 행동 및 이동을 인식하는 것은 다양한 분야에서 활용될 수 있다. 사람의 행동을 파악하여 니즈를 예상하고 맞춤형 콘텐츠를 제공하거나 행동을 예측하여 범죄나 폭력을 예방하는 등 여러 방면으로 활용 가능하다. 그러나 이동과 현재 위치 정보만으로 인물의 행동을 예측하기에는 한계가 있다. 본 논문에서는 실시간으로 사람의 이동과 행동을 인식하기 위해 Kinect v2가 제공하는 관절 정보와 YOLOv3를 이용하여 실시간으로 사람의 행동을 인식하는 시스템을 제작하였다.

1. 서론

사람의 행동이나 이동을 인식하는 것은 다양한 분야에서 연구되고 있으며, 실생활에서 활용될 수 있다. 예를 들면, 무인편의점인 'amazon go(아마존 고)'에서 사람의 행동을 파악하고 어떠한 물건을 구매하려고 하는지 인식하여 맞춤형 서비스를 제공하거나 폭력을 행사하기 전에 미리 행동을 인지하여 범죄를 예방할 수 있다[1]. 기존의 행동 인식 연구는 사람의 위치 정보를 사용하거나 관절 정보를 활용하는 방식을 사용했다. S. Kyo, K가 제안한 방법은 사용자의 위치 정보를 이용한 방식으로 센서로 부터 얻어진 위치 데이터를 머신러닝 모델로 학습하여 행동을 인식하는 방법이다[2]. W, Hee, Y가 제안한 방법은 사람의 관절 정보를 이용하는 방식으로 연속된 두 프레임 내에서 얻어진 시공간상 관절 데이터의 상관관계를 파악한 후 딥러닝 모델로 학습하여 행동을 인식하는 방법이다[3]. 기존의 연구 방식은 가만히 서 있기, 걷기와 같은 간단한 행동은 인식하지만 쇼핑하기, 버스 탑승하기와 같은 행동들은 센서에서 얻어진 데이터로만 행동을 구분하기에 한계가 있다는 것을 보여준다. 또한, 같은 행동에서 사람마다 자세가 다르거나, 한 사람이 같은 행동을 해도 자세의 변형이 있을 수 있어 실시간 관절 데이터로 행동을 인식하는 것은 낮은 정확도를 보인다.

이에 본 논문에서는 Kinect v2를 통해 얻은 사람의 관절 정보를 이용하여 이동과 방향을 실시간으로 인식하고 YOLOv3를 이용하여 행동을 분석한 정보와 Kinect v2관절 정보를 결합하여 좀 더 높은 정확도로 실시간 행동을 검출하는 시스템을 만들었다.

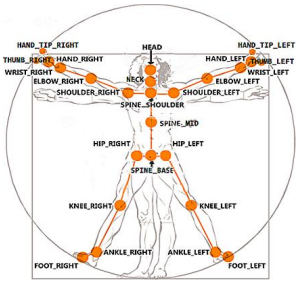
2. 관련연구

2.1 openFrameworks

openFrameworks는 C++을 기반으로 직관적이며 간단한 프레임워크로 창의적인 프로세스를 지원하도록 설계된 오픈 소스 라이브러리이다. 또한, Windows, OSX, Linux, iOS, Android와 같이 다양한 OS에서 작동하는 크로스 플랫폼이다. openFrameworks의 장점은 OpenCV나 OpenGL과 같은 라이브러리를 쉽게 추가할 수 있다는 것이다[4].

2.2 Kinect v2

Kinect는 사용자의 신체를 이용하여 다양한 콘텐츠를 경험할 수 있는 동작인식 기기이다. Kinect는 다양한 센서들로 RGB 영상, 깊이 정보뿐만 아니라 사용자의 관절 추적 정보를 제공한다. 사람의 이동에 대한 판단과 검출된 행동의 검출을 위해 Kinect에서 제공하는 25개의 관절 추적 정보, RGB 정보, 깊이 정보를 이용한다[5].



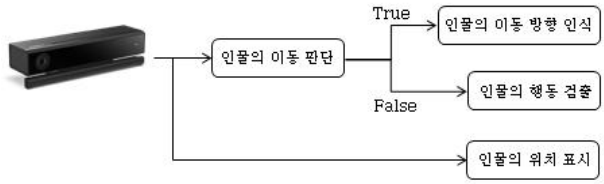
[그림1] Kinect v2에서 제공하는 관절 추적 정보

2.3 YOLO v3

기존의 R-CNN과 같은 object detection 모델은 복잡한 처리 과정으로 인해 실시간으로 물체를 검출하는데 있어 어려움이 있다. YOLOv3는 간단한 처리과정으로 빠른 연산 속도와 전체 이미지를 넣고 물체의 영역과 클래스를 분류하는 과정을 한 단계로 진행해 실시간으로 객체를 검출할 수 있는 모델이다. 본 연구에서는 실시간으로 인물을 검출하기 위해 YOLOv3를 사용하였다[6].

3. 실시간 행동 검출 시스템의 설계

본 연구의 시스템은 [그림2]와 같이 인물의 이동 판단, 인물의 이동 방향 인식, 인물의 행동 검출, 인물의 위치 표시로 구성하였다. 인물의 이동 판단은 Kinect를 통해 관절 정보를 얻어 움직임의 유, 무를 판단하는 부분이다. 인물의 이동 방향 인식은 이동이 있다고 판단되었을 때 어느 방향으로 움직였는지 인식하는 부분이다. 인물의 행동 검출은 이동이 없다고 판단되었을 때 인물의 행동을 검출하는 부분으로 YOLOv3를 사용하여 인물의 행동을 분석하고 실시간 관절 정보와 결합하여 행동을 검출한다. 인물의 위치 표시는 이동의 유, 무와 관계없이 실험 공간 안에서 인물이 어디에 위치하고 있는지 표시하는 부분이다.



[그림2] 시스템 설계도

본 연구에서는 인물의 행동을 판단하기 위해 기존의 연구에서 사용한 인물의 이동과 관절로 행동을 파악한 후 딥러닝으로 학습하는 방식과 달리 YOLOv3를 사용하여 인물의 행동을 이미지로 학습하고 실시간 Kinect에서 제공하는 실시간 관절 정보를 결합하여 행동을 인식한다. 관절 정보로만 학습하여 행동을 인식하는 것은 정확도가 낮아 데이터를 얻는 것에 있어 어려움이 있다. YOLOv3를 이용하여 이미지로 학습하는 방식은 행동의 특징을 파악하여 인물의 팔다리 같은 신체 위치가 조금씩 달라져도 행동을 인식할 수 있으며, 관절 정보와 결합했을 때 정확도와 인식의 속도를 높일 수 있다.

3.1 인물의 이동 판단

인물의 이동을 판단은 Kinect 센서의 관절 정보를 이용한다. 인물의 위치 오차를 줄이기 위해 척추에 있는 3개의(Spine Shoulder, Spine Mid, Spine Base)를 평균을 내어 인물의 위치 좌표로 사용한다. 인물의 현재 위치 좌표(x_1, z_1) 값과 일정 시간 이후에 인물의 좌표(x_2, z_2) 값을 좌표 평면에서 두 점사이의 거리 공식을 통해 움직임의 크기를 계산한다. 움직임의 크기가 사용자 지정 값 이상이면 움직임이 있는 것으로 설정한다.

3.2 인물의 방향 인식

인물의 방향 인식은 이동이 있다고 판단되었을 때 실행되는 부분이다. 인물의 정면으로 이동한 것을 N(북쪽)으로 정했을 때, N에서 시작하여 시계방향으로 45도씩 회전한 범위를 각각 8방위(N, NE, E, SE, S, SE, W, NW)로 나누어 이동의 방향으로 설정한다. 앞 부분에서 얻은 두 개의 인물 위치 좌표(x_1, z_1) 와 (x_2, z_2) 사이의 각도 값을 계산한다. 얻어진 각도 값이 8개의 방향과 비교했을 때 범위 안에 있는 경우 움직임의 방향으로 인식한다.

3.3 인물의 행동 검출

인물의 행동 검출은 이동이 없다고 판단되었을 때 실행되는 부분이다. Kinect의 RGB이미지를 YOLOv3에 입력해 '팔짱을 끼는 행동', '주머니에 손을 넣는 행동', '스마트폰을 사용하는 행동'과 같이 3가지의 행동에 대해 인식한다. 인식된 행동의 예측률이 60% 이상일 때, Kinect에서 얻은 관절 좌표와 3가지 행동에 대해 일반적인 모습의 관절 좌표를 비교하는 조건에 맞는 행동만 검출한다. 행동 인식을 위한 관절 조건은 [표1]과 같다.

Elbow는 사용자의 팔꿈치 좌표, Hand는 사용자의 손 좌표, Hip은 사용자의 골반 좌표, Shoulder은 사용자의 어깨 좌표를 의미한다. L, R은 왼쪽과 오른쪽을 의미하고, Spine_B는 척추 좌표 중 골반에 붙어있는 좌표를 의미한다. X, Y, Z는 3차원 좌표계의 각 축에 맞는 좌표를 말한다.

[표1] 행동 인식을 위한 관절 조건

행동	관절 조건
Folded Arms	$Elbow\_R(X,Y) - 10 < Hand\_L(X,Y) < Elbow\_R(X,Y) + 10$
	$Elbow\_L(X,Y) - 10 < Hand\_R(X,Y) < Elbow\_L(X,Y) + 10$
Pocket Hand	$HandState = Unknown$
	$Hand\_L(Z), Hand\_R(Z) < Spine\_B(Z)$
	$Hip\_L(X,Y) - 10 < Hand\_L(X,Y) < Hip\_L(X,Y) + 10$
	$Hip\_R(X,Y) - 10 < Hand\_R(X,Y) < Hip\_R(X,Y) + 10$
Smart Phone	$Hand\_L(Y) > Elbow\_L(Y) \&\& Hand\_R(Y) > Elbow\_R(Y)$
	$Shoulder\_L(X) < Hand\_R(X) < Shoulder\_R(X)$

3가지 행동을 인식하기 위해 각각의 행동에 맞춰 조건을 주었다. 첫 번째로 팔짱을 끼는 행동은 인물의 손이 반대 팔의 팔꿈치와 맞닿아 있어야 하므로, 왼손의 좌표(X, Y)가 오른쪽 팔꿈치의 $(X, Y) \pm 10$ 범위 안에 있으며, 오른손의 좌표(X, Y)가 왼쪽 팔꿈치의 $(X, Y) \pm 10$ 범위 안에 있어야 인식한다. 두 번째로 손을 주머니에 넣는 행동은 골반 위에 손이 위치해 있으며, 보이지 않아야 하고, 뒷짐을 지는 행동과 비슷하기 때문에 손의 관절 좌표(X, Y)는 골반 좌표 $(X, Y) \pm 10$ 범위 안에 있고, 상태가 인식되지 않는 Unknown상태이며, 양 손의 좌표(Z)가

골반의 좌표(Z)보다 앞에 있어야 인식된다. 마지막으로, 스마트폰을 사용하는 행동은 팔이 위로 접혀있어야하고, 손이 양쪽 어깨의 범위 안에서 스마트폰을 들고 있어야 하므로 손의 좌표(Y)가 팔꿈치의 좌표(Y)보다 위에 있어야 하고, 양 손의 좌표(X)가 양 어깨의(X) 범위 안에 있어야 인식된다.

3.4 인물의 위치 표시

인물의 이동의 유, 무와 관계없이 실행되는 부분이다. 실험 공간에서 Kinect의 정면에서 1.8m 떨어진 곳을 중심으로 설정한 후, 인물의 위치를 지름이 60px인 원 안에서 화면에 표시한다.

4. 실험

4.1 실험 방법

이동과 행동 인식에 대해 Kinect만 사용했을 때와 Kinect와 YOLOv3를 사용했을 때 정확성을 알기 위해 5가지 실험을 진행한다. 첫 번째는 Kinect를 사용하여 이동을 인식하는 실험, 두 번째는 Kinect를 사용하여 행동을 인식하는 실험, 세 번째는 YOLOv3와 Kinect를 사용하여 행동을 인식하는 실험, 네 번째는 Kinect를 사용하여 이동과 행동을 모두 인식하는 실험, 마지막으로 YOLO와 Kinect를 사용하여 이동과 행동을 모두 인식하는 실험으로 나누어 진행했다.

4.2 이동 인식 실험

1) 실험 공간에 Kinect를 고정한다.
2) 고정된 Kinect로부터 1.8m 떨어진 곳에 중심을 표시한다.
3) 정해진 8개의 방향 중 하나의 방향을 화면에 표시하여 문제를 제출한다.
4) 실험 대상이 화면에 표시된 문제의 방향으로 약 30cm(한걸음) 움직인다.
5) 실험 대상의 이동과 인식한 방향이 일치하면 성공, 두 방향이 다르면 실패로 정의한다.

4.3 행동 인식 실험

1) 실험 공간에 Kinect를 고정한다.
2) 실험 공간 어디에서나 행동이 검출되어야 하므로 다른 설정을 하지 않는다.
3) 정해진 3가지 행동 중 하나를 화면에 표시하여 문제를 제출한다.
4) 실험 대상이 화면의 문제에 맞는 행동을 취한다.
5) 실험 대상의 행동과 인식한 행동이 일치하면 성공, 두 행동이 다르거나 2초가 지나도 행동을 인식하지 못하면 실패로 정의한다.

[그림3] 이동 방향 인식 실험 모습

[그림4] 행동 인식 실험 모습

4.4 실험 결과

[표2]는 Kinect를 이용한 이동 인식 실험으로 관절 정보로 이동 방향을 인식하는데 있어 94%의 인식률로 좋은 성능을 보여주며, 다른 센서나 프로그램이 추가적으로 필요 없다는 것을 알 수 있다. 반면, [표3]과 [표4]를 비교해 보았을 때 Kinect 만 이용한 행동 인식 실험 은 27%의 인식률로 낮은 정확도를 보이며, Kinect와 YOLOv3를 모두 이용한 행동 인식 실험은 95.5%의 인식률로 두 실험의 정확도 차이가 큰 것을 보여준다. [표5]와 [표6], [그림5]를 보았을 때도 마찬가지로 Kinect 만 이용한 실험의 인식률은 26%인 반면, Kinect와 YOLOv3 모두 이용한 실험의 결과가 94.5%로 Kinect와 YOLOv3 모두 사용했을 경우에 정확도가 더 높은 것을 확인할 수 있다.

[표2] 이동 방향 인식 실험 결과(Kinect)

실험 회차	실험 횟수	성공 횟수	인식률(%)
1	40	36	90
2	40	39	97.5
3	40	39	97.5
4	40	36	90
5	40	38	95
평균	40	37.6	94

[표3] 행동 인식 실험 결과(Kinect)

실험 회차	실험 횟수	성공 횟수	인식률(%)
1	40	5	12.5
2	40	17	42.5
3	40	13	32.5
4	40	8	20
5	40	11	27.5
평균	40	10.8	27

[표4] 행동 인식 실험 결과(Kinect+YOLOv3)

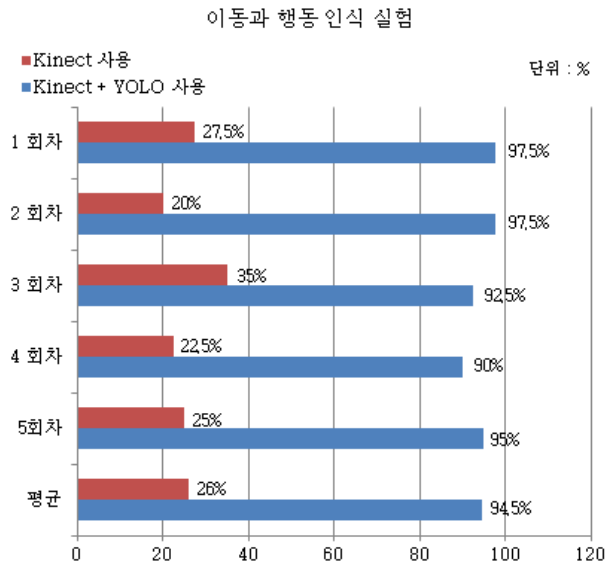
실험 회차	실험 횟수	성공 횟수	인식률(%)
1	40	38	95
2	40	37	92.5
3	40	38	95
4	40	38	95
5	40	40	100
평균	40	38.2	95.5

[표5] 이동과 행동 인식 실험 결과(Kinect)

실험 회차	실험 횟수	성공 횟수	인식률(%)
1	40	11	27.5
2	40	8	20
3	40	14	35
4	40	9	22.5
5	40	10	25
평균	40	10.4	26

[표6] 이동과 행동 인식 실험 결과(Kinect+YOLOv3)

실험 회차	실험 횟수	성공 횟수	인식률(%)
1	40	39	97.5
2	40	39	97.5
3	40	37	92.5
4	40	36	90
5	40	38	95
평균	40	37.8	94.5



[그림5] 이동과 행동 인식 실험 비교 그래프

5. 결론

본 논문에서는 인물의 이동과 방향을 파악하고 실시간으로 인물의 행동을 인식하는 방법에 대해 제안하였다. Kinect를 통해 얻은 관절 정보를 이용하여 인물의 위치를 알 수 있었고 이전 위치와 현재 위치를 수식에 대입하여 인물의 이동 유, 무와 방향을 파악할 수 있었다. 또한, 영상에서 실시간으로 인물의 행동을 인식하기 위해 Kinect와 YOLOv3를 결합하여 사용하였고, 실시간으로 인물의 행동을 인식할 수 있음을 실험을 통해 확인하였다.

향후 연구에서는 다양한 데이터를 축적하고 RNN과 CNN을 결합하여 좀 더 많은 행동을 인식하는 것을 목표로 한다.

참고문헌

[1] amazon go [Internet]
https://www.amazon.com/b?node=16008589011

[2] 김선교, 이영설, 손동운, 조성배 (2010). "계층적 은닉 마르코프 모델을 이용한 이동 센서 기반 행동 인식". 한국정보과학회 학술발표 논문집, 37(2A), pp. 137-138

[3] 양우희, 권기룡, 신봉기 (2020). "3차원 골격정보 기반 관절 간 상관관계를 이용한 행동인식". 한국정보과학회 학술발표 논문집, pp. 809-811.

[4] openFrameworks [Internet]
https://openframeworks.cc/about/

[5] 이새봄, 정일홍 (2014). "키넥트를 사용한 NUI 설계 및 구현". 한국디지털콘텐츠학회 논문지, 15(4), pp. 473-480

[6] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016), “You only look once: Unified, real-time object detection.” Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788.