KIICE 한국정보통신학회지 14권 1호
THE MAGAZINE OF KIICE
http://www.kiice.org

게임 콘텐츠를 위한 3차원 동작인식 기술

김희권* · 이재호*

*한국전자통신연구원 차세대콘텐츠연구소

목 차

Ⅰ. 서 론

IT 기술과 콘텐츠 기술이 융합된 3D동작인식 기반 기술은 최근 산업, 의료, 엔터테이먼트, 교육, 문화 체험, 국방, 건축 등 적용이 널리 증가 되었다. 기존의 UI의 개념에서 UX/NUI 개념으로의 진화를 통해서 사용자가 콘텐츠나 기기동작을 위해서 특정한 입력도구에 대한 교육 없이 자신의 동작이나 직관적인 인터페이스를 통하여 쉽고 빠르게 적응이 가능하다.

최근 들어 다양한 센서들과 터치스크린이 상용화됨에 따라 인터랙션을 가미한 콘텐츠가 증가하고 있으며 특히 게임분야에서는 다양한 형태의 센서 등을 활용하여 기존의 패드 기반의 입력 장치의 한계를 뛰어넘어 사용자가 보다 게임에 몰입할 수 있는 형태의 입력장치들이 사용화 되어 출시되고 있다. 현재 가장 많이 사용되고 있는 게임관련 인터랙티브 콘텐츠의 인터페이스는 자이로센서, Kinect, 터치스크린 등이 있다.

기존의 특정키를 입력하는 방식에서 벗어나 사용자의 특정 동작을 입력받아 그에 따른 콘텐츠의 반응은 인터랙티브 콘텐츠의 기본 구성이다. 동작인식 분야는 현재 다양한 형태의 사용자 입력을 보다 효과적이고 빠르게 분석하여 콘텐츠에 적용하는 부분에 대한 연구가 널리 진행되고 있다. 그림1에서와 같이 기존의 특정키를 입력하거나 터치하는 방식에서 벗어나 사용자의 특정 동작을 입력받는 형태로 변화되고 있다. 그에 따른 콘텐츠의 반응은 인터랙티브 콘텐츠의 기본이다. 동작인식 분야는 현재 다양한 형태의 사용자 입력을 보다 효과적이고 빠르게 분석하여 콘텐츠에 적용하는 부분에 대한 연구가 널리 진행되고 있다.

그림 1. 입력 장치 및 인터페이스의 변화

Ⅱ. 3D 동작인식

동작인식이란 공간상의 사람의 연속적 동작의 의미를 파악하거나 특정 자세에 대한 의미를 파악하고 분석하는 것이다. 동작인식분야는 80년대 중반부터 꾸준히 연구되어 온 분야이다. 동작인식이나 사용자의 동작을 모방하여 3D 오브젝트에 연동하는 방법에 대해서 그동안 많은 연구와 기기들이 개발되었다. 그림2.에서와 같이 광학을 이용하여 실험자 얼굴의 움직임을 모방하고자 할 때 얼굴의 표정에 관련되는 근육 위에 특별한 소자를 설치하여 근육의 움직임을 체크하여 연동하는 방법, 전신의 움직임을 정확하게 측정하기 위하여 기기를 몸에 부착한 후 움직임을 측정하는 방법,

각 동작에 대한 이미지를 저장하여 인식에 활용하는 방법, 거리 센서를 이용한 사용자 스켈레톤을 이용한 동작인식 방법 등이 존재한다.

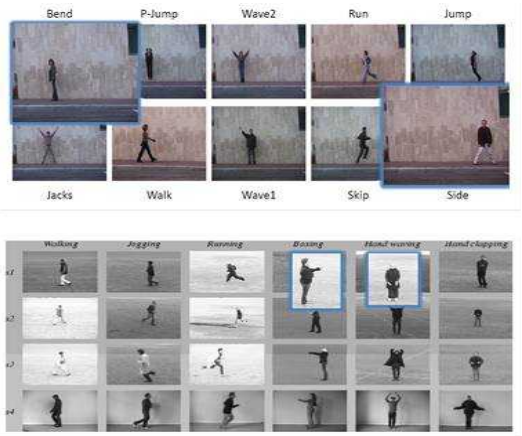

그림 2. 이미지 기반 동작인식 셋 대표적인 예
(상) Weizmann Data Set (하) KTH Data Set

그림 3은 성능 평가를 위해 Weizmann 및 KTH 데이터 셋이 주로 이용되고 있는데 이에 가장 대표적인 데이터 셋의 예를 보여준다. 비록 두 데이터 셋이 동작인식 분야에서 공정성과 범용성을 인정받아 널리 사용되고 있지만 3차원 동작 인식 알고리즘의 성능 평가를 위해서는 적용하기 어려운 문제가 있다. 이는 개발되는 3차원 동작 인식 알고리즘을 두 데이터 셋에 적용하기 불가능한 문제와 HCI에 적합한 동작이 부재하기 때문이다.

2010년 11월 저가의 거리센서의 등장으로 동작인식 분야의 획기적인 진화를 가지고 오게 되었다. 동작인식을 위한 전처리 과정으로 많은 문제를 발생시켰던 사용자의 영역 분리가 거리센서를 통해서 보다 쉽고 빠르게 획득이 가능하게 된 것이다. 특히 사람의 관절에 대한 정보를 획득하여 그에 따른 분석과 연구를 통해서 사람의 동작인식이 가능하게 되면서 그 동안 동작인식에 활용되던 알고리즘에서 다양한 알고리즘의 적용이 가능하게 되었다. 거리센서는 사람의 전신을 이용한 형태와 근거리의 손동작에 초점을 맞춘 형태로 각 센서의 특징을 구분하여 연구 개발이 되고 있다. 대표적으로는 OpenNI SDK[10] , MS Kinect SDK[11]의 사람의 전신에 대한 관절 정보를 이용한 개발툴이 존재하며 Intel perceptual computing SDK[12], Leap Motion SDK[13]의 경우 손동작이나 손에 대해서 특화된 SDK를 선보이고 있다. 또한 사람의 생체 신호를 이용한 MYO[14]가 최근 등장하면서 동작인식 분야에 새로운 패러다임을 제시하고 있다.

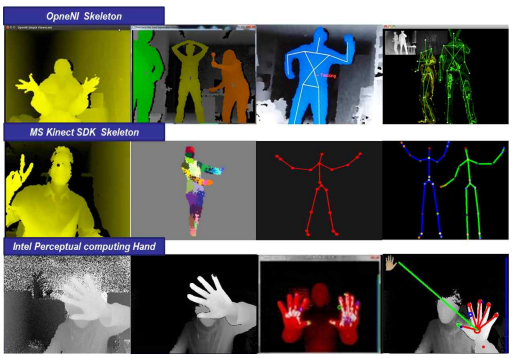

그림 3. 거리센서를 이용한 동작인식을 위한 전처리 과정의 예

Ⅲ. 3D 동작인식 기술

3.1. 동작인식의 구분 방법

3D 동작인식은 그 동안 꾸준히 연구되어 온 분야중의 하나로 다양한 방법과 기술들이 존재한다. 특히 거리센서의 등장으로 인해서 기존의 방법에서 탈피하여 다양한 방법의 적용 사례가 나타나고 있는 추세이다. 동작 인식의 기술은 인식의 결과를 도출해 내는 방법에 따라서 구분할 수 있는데 크게 싱글 레이어드(Single‐layered)기반과 계층(Hierarchical)기반으로 구분 할 수 있다[1]. 싱글 레이어드기반의 임의의 기준 영역을 동작인식의 특징으로 한 방법으로는 Interest point를 활용하는 방법[4], 시공 특징값(Space-time neighborhood features)들을 학습해 인식하는 방법[5], Dense Trajectories를 이용한 인식 방법[6]등이 있다. 다음 시간적 템플릿을 이용한 방법으로 MEI(Motion Energy Image), MHI(Motion History Image)를 생성한 후 Hu의 움직임의 특징을 이용한 동작인식 방법[7]과 이에 3차원을 적용하여 다양한 시점에 강인한 동작인식이 가능하도록 하는 방법[8]으로 MHV(Motion History Volumes)을 이용하기도 한다.

동작인식의 결과에 대한 분류 방법은 기본적으로 크게 4가지로 나눌 수 있으며 그 분류는 Gesture, Action, Interaction, Group Activity 이다. Gesture는 사용자의 가장 기초가 되는 움직임으로 “팔을 들어 올리다.”, “다리를 들다” 등으로 들 수 있다. Action은 Gesture가 여러 개가 모여 하나의 행위를 하는 것을 말하는 것으로 “걷다”, “뛰다”, “앉다”처럼 여러 개의 동작이 연속적으로 모여 하나의 정의된 행위를 하는 것을 말한다. Interaction은 두 명 이상의 사용자가 서로 Action이나 Gesture등으로 인해 발생되는 것으로 “싸우다”, “공을 패스를 하다”등으로 표현할 수 있다. Group Activity는 여러 명의 사용자가 같은 행위를 하여 발생하는 것으로 “행진하다.”, “두개 그룹이 싸운다.” 등으로 나타낼 수 있는 행위이다. 이렇게 동작인식의 범위에 대한 분류를 통해서 보다 정확하고 동작인식의 허용 범위에 대한 정의를 가능하게 한다.

본 절에서는 동작인식을 위한 접근 방법에 따른 대표적인 인식 기술에 대한 설명과 그 중 가장 많이 연구되고 있는 분야에 대해서 설명하겠다.

3.2 DTW(Dynamic Time Wrapping)

사람의 관절에 대한 연속적 흐름과 미리 훈련된 결과를 비교하여 유사도가 높은 결과를 도출하는 알고리즘으로 음성인식 등에 많이 활용된 알고리즘[3]이다. 이 알고리즘을 사람의 관절 또한 연속된 동작과 그 동작에 대한 시간적 흐름에 따른 구분을 통해서 비교하기 적합한 형태를 가지고 있다. 특징 벡터를 일정 시간 동안 누적하여 동작 시퀀스로 구성하며 각 시퀀스와 훈련된 결과를 비교하여 유사도가 높은 결과를 도출한다.

$$A = < a_1, ..., a_k >, B = < b_1, ..., b_T >$$

δ 는 시퀀스 요소들 간의 거리를 나타냄
DTW는 다음 식을 재귀적으로 계산하여 유사도를 측정함

$$D(A_i, B_j) = \delta(a_i, b_j) + \min \begin{Bmatrix} D(A_{i-1}, B_{j-1}) \\ D(A_i, B_{j-1}) \\ D(A_{i-1}, B_j) \end{Bmatrix}$$

이때 A_i는 시퀀스 A의 부분 시퀀스 $< a_1, ..., a_i >$를 나타냄

위의 유사도 측정의 특징 값으로 신체 중요 관절을 이용하게 되면 인식되어야 하는 동작 별 관절의 가중치가 서로 다를 수 있다는 가정을 바탕으로 두 시퀀스 c_i와 c_j사이의 유사도 계산은 아래 식과 같이 수행한다.

$$d(c_i, c_j) = \sqrt{\sum_{p=1}^{|c_i|} ((c_i^p - c_j^p) v^p)^2}$$

$|c_i|$ 와 v^p는 각각 c_i의 길이, p번째 관절의 중요도를 의미한다.

DTW를 사용하게 되면 트레이닝에 대한 시간이 필요하지 않으므로 실제 동작인식을 이용한 콘텐츠 제작이나 실험을 위한 환경에 효과적으로 사용하기에 적합하며 각 동작에 대한 시간적인 차이에 대해서도 강인하다. 하지만 동작에 대한 시작점 및 끝점을 찾기가 어려운 경우에 동작인식 성공률은 많이 낮아지는 단점을 가지고 있다. 또한 최적화 방법을 잘못 구성하게 되면 오류 동작의 구분이 쉽지 않게 되는 경우가 있다.

3.3 HMM(Hidden markov model)

관측된 값에서 결과를 도출하는 통계모델로 음성인식, 광학문자인식, 자연어 처리 등에 널리 활용되는 알고리즘[9]이다. 구현하기 쉽고 유연한 모델링 능력과 높은 성능을 가지고 있어 연속적인 흐름을 이용한 인식 분야에 많이 활용되고 있다. 동작인식 또한 연속적인 흐름을 기반으로 하여 만들어지므로 HMM의 이용 시 높은 성능과 다양한 적용이 가능한 장점을 가지고 있다. 하지만 연속적인 흐름에 강인한 모델이지만 시작점과 끝점의 구분에 대한 부분이 명확하지 않게 되면 문제가 생기며 또한 유연한 모델링을 통한 다양한 동작인식을 적용시킬 수 있으나 동작인식에 활용되는 트레이닝 기술의 성숙도에 따라 많은 차이가 보이게 된다.

Ⅳ. 콘텐츠 적용 사례

본 절에서는 다양한 동작인식 알고리즘 중에서 비교적 구현이 용이하고 선행과정의 시간이 적은 알고리즘을 택하여 실제 콘텐츠에 적용한 사례에 대해서 설명하겠다. 테이블 탑과 프론트 디스플레이를 병합한 새로운 형태의 인터페이스를 가진 인터랙티브 시스템을 사용자의 손 동작 과 손의 움직임을 이용하여 제어할 수 있는 동작인식에 적용하였다.

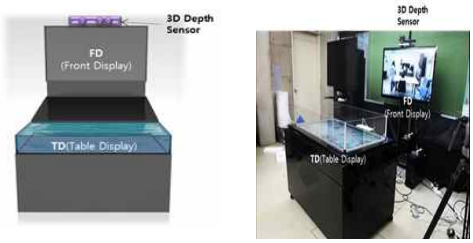

그림 4. 인터랙티브 시스템 (IVA)

동작인식을 적용한 인터랙티브 시스템은 테이블탑, 프론트 디스플레이와 3D 센서로 구성되어 있다. 테이블 탑은 순수 디스플레이로 구성되어 있으며 동작인식 및 공간인지를 통하여 가상 멀티터치, 손의 프로젝터 위치를 보여준다. 프론트 디스플레이는 사용자의 실제 모습과 3D오브젝트의 합성(Augmented Reality)을 통하여 사용자의 몰입감을 높여준다. 3D센서는 사용자의 동작인식을 위한 사용자 관절 정보를 넘겨주는 부분으로 실제 모든 연산이나 동작인식의 관련된 일련의 기능들이 수행된다.

인터랙티브 시스템에 적용한 동작은 크게 두 가지로 나눌 수 있다. Gesture 3종류, Action 4종류이며 각 종류에 대한 것은 아래 표와 같다.

표 1. 동작인식 구성

분 류	명 칭
Gesture1	손을 앞으로 하고 움직이다(G1)
Gesture2	손을 바닥에서 움직이다(G2)
Gesture3	왼손을 공중에서 움직이다.(G3)
Action1	고기를 낚다(A1)
Action2	먹이를 주다(A2)
Action3	거북이를 낚다(A3)
Action4	낚시 하다(A4)

본 논문에서는 DTW를 이용하여 동작인식에 활용하였다. 콘테츠를 구성하는데 있어서 동작인식을 통한 제어를 위해서는 다양한 형태의 실험과 선행 작업이 필요하다. 하지만 콘텐츠를 기획할 때 와 실제 구성이 되는 콘텐츠간의 차이가 존재하게 된다. 이러한 차이를 쉽게 극복하고 빠르게 적용하기 위해서 트레이닝 시간이 필요없는 알고리즘에 대해서 생각하게 되었고 가장 적합한 DTW를 이용하여 동작인식을 구현하게 되었다. 이때 각 동작에 대한 데이터 셋은 실험자 5명에게서 10 개의 동작을 받아서 데이터 셋을 구성하였다. 또한 Gesture의 경우 동작에 연속성에 대한 결과값을 기준으로 하기 때문에 정확도 측정에서는 실제 손의 위치와 비교를 통해서 확인하였다.

표 2. 10-묶음 교차 검증 방법의 결과

	G1	G2	G3	A1	A2	A3	A4
G1	0.95	0	0	0	0	0	0
G2	0	0.92	0	0	0	0	0
G3	0	0	0.93	0	0	0	0
A1	0	0	0	1	0	0	0
A2	0	0	0	0	0.97	0	0
A3	0	0	0	0	0	0.90	0
A4	0	0	0	0	0	0	1

표 3. 훈련 샘플 하나만 사용한 실험 결과

	G1	G2	G3	A1	A2	A3	A4
G1	0.92	0	0	0	0	0	0
G2	0	0.95	0	0	0	0	0
G3	0	0	0.89	0	0	0	0
A1	0	0	0	1	0	0	0
A2	0.01	0	0	0	0.91	0	0
A3	0	0	0	0	0	0.89	0
A4	0	0	0	0.02	0	0	1

표2는 각 10묶음 교차 검증을 이용한 실험 결과값에 대한 정확도이다. 각각의 Gesture, Action을 각 표와 같이 구분하여 동작의 정확도에 대해서 비교 실험을 하였다. 10묶음 교차 검증시 실제 동작에 대한 에러 확률이 낮게 나옴을 볼 수 있었다. 표3은 훈련샘플 하나

를 이용하여 동작인식에 활용할 경우 DTW의 경우 데이터 셋으로 등록 된 데이터가 실제 사용자가 행동하는 데이터 셋과의 유사도를 비교하게 되므로 많은 데이터 셋을 가지고 비교 분석하는 것 보다 정확한 동작에 대해서 몇 개의 데이터 셋을 구성하고 사용자로 하여금 정확한 동작을 취하게 학습하는 것이 보다 동작인식을 이용한 콘텐츠에 적합하다.

V. 결 론

최근 다양한 입력장치를 이용한 콘텐츠가 개발이 되면서 동작인식에 대한 관심이 높아지고 있다. 하지만 동작인식을 실제 콘텐츠에 적용하여 사용하기에는 많은 제약 조건과 문제점들이 존재한다. 이러한 제약 조건들을 극복하는 방법으로는 다양한 형태의 동작인식 기술들에 대한 이해가 필요하며 이를 바탕으로 콘텐츠를 구성하는데 있어서 적합한 동작인식 방법을 찾아 적용하는 것이 가장 바람직할 것이다. 본 논문에서는 깊이 영상에서 검출된 신체 기관의 정보를 이용한 실시간 동작 인식 에 적합한 알고리즘에 대해 논하였다. 기존의 다양한 동작인식 방법에 대해서 분석하고 깊이정보 기반 동작인식에 적용 시 문제점들과 사용자가 직접 콘텐츠에 직접 적용 가능한 동작인식 알고리즘에 대한 방법을 연구하였다. 그 중에 하나의 알고리즘을 깊이정보 기반의 데이터 셋을 이용하여 성능 및 동작인식 테스트에 대해서 분석 하였다. 분석 결과 DTW를 이용한 방법은 실제 콘텐츠 적용에 있어서 많은 장점을 가지고 있었다. 동작인식의 동작의 추가 및 삭제 등이 자유로웠으며 또한 성능에서도 뛰어난 효과를 보였다.

참고문헌

[1] M.S Ryoo, J.K.AGGRAWAL “Human Activity Analysis : A Review”, ACM Computing Surveys, Vol.43, No. 3, Article 16, April 2011

[2] http://www.softkinectis.com

[3] Al-Naymat, G., Chawla, S., & Taheri, J. (2012). SparseDTW: A Novel Approach to Speed up Dynamic Time Warping

[4] A. Gilbert, J. Illingworth, and R. Bowden, “Fast Realistic Multi-Action Recognition using Mined Dense Spatio-temporal Features” IEEE Int’l conf. Computer Vision, 2009

[5] A. Kovashka and K. Grauman, “Learning a Hierarchy of Discriminative Space-Time Neighborhood Features for Human Action Recognition,” Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition, 2010

[6] H. Wang, A. Kläser, C. Schmid, and C.-L. Liu, “Action Recognition by Dense Trajectories,” Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition, 2011.

[7] J.W. Davis and A.F. Bobick, “The Representation and Recognition of Action Using Temporal Templates,” Proc. IEEE Int’l Conf.

[8] D. Weinland, R. Ronfard, and E. Boyer, “Free Viewpoint Action Recognition Using Motion History Volumes,” Computer Vision and Image Understanding, vol. 103, nos. 2-3, pp. 249-257, 2006.

[9] Hyeon-Kyu Lee and Jin H. Kim. “An HMM-Based Threshold Model Apporoach for Gesture Recognition”, IEEE Transaction on pattern analysis and machine intelligence vol.21. no10 October 1999

[10] www.OpenNI.org

[11] http://www.microsoft.com/

[12] http://software.intel.com/en-us/vcsource /tools/perceptual-computing-sdk

[13] https://www.leapmotion.com/

[14] https://www.thalmic.com/myo/

[그림1] MS DOS, MS Window XP, ISO, MS Window8의 바탕화면 및 UI, 영화 “마이너 리포트”의 한 장면

[그림2] www.forum.libcinder.org
www.blog.msdn.com
www.software.intel.com

저자소개

김희권(Kim, Hee-Kwon)

2012 ~ 한국 전자 통신 연구원 재직
2012 충남대 컴퓨터공학 석사 졸

※관심분야 : 패턴 얼굴 인식, 동작 인식

이재호(Lee, Jae-Ho)

2005 ~ 한국전자통신연구원 선임
연구원 재직
2005 한양대학교 컴퓨터공학과 박사

※관심분야 : 패턴 얼굴 인식, 동작 인식