

# 모션 인식을 통한 춤 동작 비교 분석 시스템

황치현<sup>1</sup>, 한민재<sup>1</sup>, 김의찬<sup>1</sup>, 황광일<sup>1</sup>

<sup>1</sup>인천대학교 임베디드시스템공학과

clgusdl234@naver.com, gksalswo723@naver.com, dmlcks1998@naver.com, hkwangil@inu.ac.kr

## Dance Comparing and Analyzing System Using Pose Estimation

Chi-Hyun Hwang<sup>1</sup>, Min-Jae Han<sup>1</sup>, Eui-Chan Kim<sup>1</sup>, Kwang-il Hwang<sup>1</sup>

<sup>1</sup>Dept. of Embedded Systems Engineering, Incheon National University

### 요 약

영상처리 기술의 발달로 영상처리 기술을 이용한 다양한 어플리케이션이 출시되고 있다. 영상처리 기술로 영상의 정보를 디지털화 할 수 있는 점에 착안해 춤 실력을 평가하는 시스템을 고안했다. 본 작품에서는 Human Pose Estimation 기술로 사람의 관절 위치 정보를 파악하고, 춤 전문가의 관절 위치와 사용자의 관절 위치를 동작 비교 알고리즘을 통해 비교해 사용자가 춤을 얼마나 정확하게 추는지 수치적으로 점수화해 제공한다.

### 1. 서론

SNS의 보급이 증가하면서 사람들이 자신의 모습을 영상으로 촬영해 서로 공유하는 것이 활발해졌다. 공유되는 영상 중 자신이 춤추는 것을 촬영하는 사람들이 늘어나게 됨에 따라 춤을 배우고 싶어 하는 사람들도 늘어날 것이다. 하지만 춤을 배움에 있어 학원을 다니고자 하여도 시간적, 금전적으로 어려움이 있다. 또한 Covid-19 바이러스 확산으로 인해 대면이 어려워진 현재 춤을 배우기 위해 학원에 가는 것조차 어렵게 되었다.

본 작품에서는 PC와 웹캠을 활용하여 사용자가 장소, 시간, 금전에 제약 없이 춤을 연습할 수 있는 시스템을 설계했다. Human Pose Estimation 기술로 사람의 주요 관절 위치를 파악하고, 관절 위치를 비교해 동작 유사도를 분석하는 시스템을 구현했다. 춤 전문가의 주요 관절 위치와 사용자의 주요 관절 위치를 알고리즘을 통해 비교해 사용자가 얼마나 전문가가 추는 춤과 비슷하게 춤을 추는지 분석한다.

### 2. 관련 연구

춤을 장소, 시간, 금전에 제약 없이 연습하기 위해서는 온라인 어플리케이션을 이용하는 방법이 제격이다. 거기에 더해 사용자가 원하는 춤 영상을 보고 연습할 수 있고, 최소한의 장비로 그것이 가능하다면 더 좋을 것이다. 춤 동작을 따라하는 어플리케이션 중 가장 유명한 어플리케이션에는 유튜브의 저스

트 댄스(Just Dance) [1]가 있다. 사용 방법은 곡을 선택한 후 댄서 픽토그램의 동작을 보며 따라하는 방식이다. 저스트 댄스의 경우 어플리케이션 내에 있는 곡 중에서만 선택해서 이용할 수 있다. 어플리케이션을 실행할 수 있는 전용 기기(PlayStation, Xbox 등)가 필요하다는 단점도 있다. 전용 기기 뿐만 아니라 어플리케이션의 춤 따라하기 기능을 이용하기 위해 컨트롤러를 손에 쥐고 춤을 춰야 한다. 접근성을 늘리기 위해 PC 환경에서도 사용할 수 있는 저스트 댄스 나우(Just Dance Now) [2] 어플리케이션을 출시했지만 컨트롤러의 제약에서는 벗어나지 못했다. 스마트폰을 컨트롤러로 사용해 여전히 손에 쥐고 춤을 춰야 한다.

저스트 댄스의 춤 동작 비교 기술은 인공시각과 패턴인식을 이용한 기술이다. [3] 인체의 동작을 촬영하는 카메라 앞에서 기본 춤으로 설정된 동영상 캐릭터의 춤을 흉내내면 그 결과를 정지된 상태의 비교대상자 영상과 비교하여 점수를 채점한다. 사용자의 모든 동작을 비교하지 않기 때문에 직전 동작과 직후 동작 사이에 어떤 연결 동작을 하든 상관없이 특정 타이밍에 같은 동작을 따라 하면 같은 동작으로 판단한다. 또한 표준영상과 입력영상이 겹쳐지는 면적을 단순하게 비교함으로써 일치도를 평가하는 방법이기 때문에 어느 부위의 동작이 틀렸는지를 정확하게 지적해 낼 수가 없다는 문제점이 있다.

많은 사용자층을 보유한 PC 어플리케이션을 만들기 위해 했다. 원하는 춤 영상 선택, 춤을 출 때 자유로운

신체, 신체 부위별로 유사도 점수를 매길 수 있는 점들을 추가로 고려하기로 했다.

3. 제안 시스템

춤의 연속적인 동작 비교를 위해 동영상의 프레임마다 사람을 검출하고 33 개의 주요 관절 위치를 파악하는 Google MediaPipe Pose solution [4]을 사용했다. MediaPipe Pose solution 은 2 단계 검출기-추적기 ML 파이프라인을 활용한다. 먼저 프레임 내에서 사람의 영역(region-of-interest, ROI)을 찾는다. ROI 내에서 33 개의 주요 관절 포인트들을 예측한다. 학습된 추적 모델로 예측한 주요 관절 포인트들을 확정한다. 배우고 싶은 춤 영상 속의 사람과 사용자의 주요 관절 포인트들을 검출한다.

전문가와 사용자의 춤 동작의 유사도를 분석하기 위해 코사인 유사도 알고리즘을 사용했다. 코사인 유사도(Cosine Similarity)는 두 벡터 간의 코사인 각도로 구할 수 있는 두 벡터의 유사도를 의미한다. [5] 코사인 유사도의 공식은 아래와 같다.

$$similarity = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

코사인 유사도는 두 벡터의 방향이 완전히 동일하여 각도가 0° 인 경우에는 1 의 값을 가지고, 90° 의 각을 이루면 0 의 값을 가진다. 이러한 특성을 이용하여 사용자와 전문가의 영상을 프레임 단위로 코사인 유사도를 비교한다.

DTW 는 속도가 다른 두 데이터 집합의 유사성을 측정하는 알고리즘이다. [6] DTW 알고리즘의 수식을 표현하면 아래와 같다.

$$DTW(X, Y) = c_p^*(X, Y) = \min\{c_p(X, Y), p \in P^{N \times M}\} \dots (1)$$

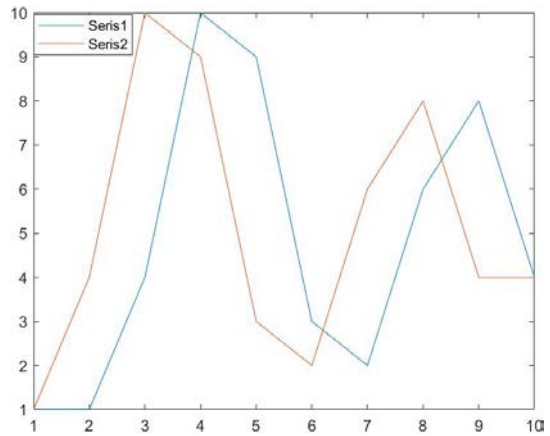
$$D(1, j) = \sum_{k=1}^j c(x_1, y_k), j \in [1, M] \dots (2)$$

$$D(i, 1) = \sum_{k=1}^i c(x_k, y_1), i \in [1, N] \dots (3)$$

$$D(i, j) = \min\{D(i-1, j-1), D(i-1, j), D(i, j-1)\} + c(x_i, y_j), i \in [1, N], j \in [1, M] \dots (4)$$

<표 1> DTW Table

Y \ X	1	2	3	...	j-1	j
1	D(1, 1)	D(1, 2)	D(1, 3)	...	D(1, j-1)	D(1, j)
2	D(2, 1)	D(2, 2)	D(2, 3)	...	D(2, j-1)	D(2, j)
3	D(3, 1)	D(3, 2)	D(3, 3)	...	D(3, j-1)	D(3, j)
...	...	...	...	...	...	...
i-1	D(i-1, 1)	D(i-1, 2)	D(i-1, 3)	...	D(i-1, j-1)	D(i-1, j)
i	D(i, 1)	D(i, 2)	D(i, 3)	...	D(i, j-1)	D(i, j)



(그림 1) Series Data

<표 2> 완성된 DTW Table

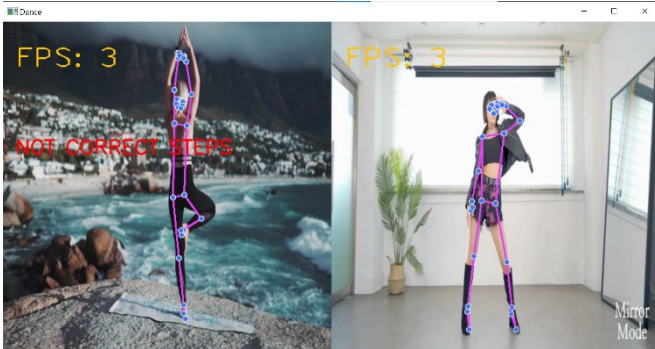
4	41	41	20	26	27	11	12	6	8	0
4	38	38	20	24	22	10	11	4	4	0
8	35	35	20	18	17	9	10	2	0	4
6	28	28	16	18	16	4	4	0	2	4
2	23	23	14	16	13	1	0	4	10	11
3	22	22	12	8	6	0	1	4	9	10
9	20	20	11	1	0	6	13	16	17	22
10	12	12	6	0	1	8	16	18	18	24
4	3	3	0	6	11	12	14	16	20	20
1	0	0	3	12	20	22	23	28	35	38
	1	1	4	10	9	3	2	6	8	4

<표 1>은 두 시계열 데이터 X,Y 의 유사도를 측정하기 위한 DTW 테이블을 나타낸 예시이다. 수식 (2)와 (3)은 각각 DTW 테이블의 첫 행과 첫 열의 원소를 구하고 수식 (4)는 이를 이용하여 테이블의 나머지 원소를 구한다. 완성된 DTW 테이블과 수식 (1)을 이용하여 테이블의 나머지 원소를 구한다.

(그림 1), <표 2>의 DTW 거리는 두 데이터가 유사할수록 0 에 가까워지고 유사하지 않을수록 값이 커

진다는 것을 보여주는 예시이다.

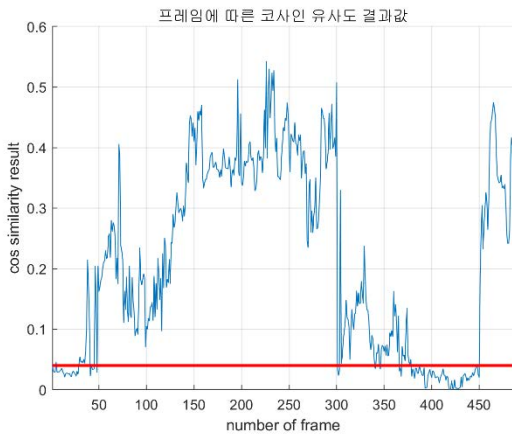
#### 4. 예상 구현 결과, 실험 결과



(그림 2) 일치하지 않은 동작 코사인 유사도 비교 화면



(그림 3) 일치하는 동작 코사인 유사도 비교 화면



(그림 4) 프레임에 따른 코사인 유사도 결과값

(그림 4)은 사용자와 전문가의 영상을 프레임 별 코사인 유사도 값을 나타낸 것이다. Y 축 값이 클수록 두 동작이 다름을 나타낸다. 최대로 동작이 다른 코사인 유사도 한계값을 0.5로 설정하고, 동작을 비교한 코사인 유사도 값이 10분의 1인 0.05 보다 작을 경우 같은 동작으로 판단하도록 했다. 즉, 동작 유사도가 90프로 이상인 경우 잘 따라한 동작이라고 판단한 그래프이다.

#### 5. 결론

SNS 의 보급이 증가하며 사람들이 자신의 모습을

영상으로 촬영해 서로 공유하는 것이 활발해졌다. 자신이 춤추는 것을 촬영하는 사람들이 늘어나게 됨에 따라 춤을 배우고 싶어하는 사람들도 늘어날 것이다. 하지만 춤을 배움에 있어 학원을 다니고자 하여도 시간적, 금전적으로 어려움이 있다. 혼자 유튜브, 틱톡 등 콘텐츠의 영상을 보고 따라하더라도 자신의 동작이 맞는 동작인지 아닌지 판단하기 쉽지 않을 것이다.

본 작품은 사용자가 원하는 영상을 시스템에 입력하면 Human Pose Estimation 기술로 사람의 주요 관절 위치를 파악하고, 관절 위치를 비교해 동작 유사도를 분석하는 시스템을 구현했다. 사용자가 원하는 춤 영상을 보고 어디서든 춤 연습이 가능하다. 실시간으로 사용자와 춤 영상을 비교하고 이를 점수화하여 제공하기 때문에 춤을 잘 따라하는지 수치적으로 판단할 수 있다. 춤 뿐만이 아닌 원하는 영상을 입력하고 모방하기에 다양한 분야에서도 활용할 수 있을 것으로 기대된다.

#### Acknowledgement

본 논문은 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행된 ICT 멘토링 프로젝트 결과물입니다.

#### 참고문헌

- [1] <https://www.ubisoft.com/ko-kr/game/just-dance/2022>
- [2] <https://justdancenow.com/?lang=ko>
- [3] 정영균, 인공지능과 패턴인식을 이용한 체감형 게임 장치(대한민국등록실용신안 등록번호 : 20-0239844 호)
- [4] <https://google.github.io/mediapipe/solutions/pose>
- [5] 길선웅, 이기영 코사인 유사도 측정을 통한 행위 기반 인증 The Journal of The Institute of Internet, Broadcasting and Communication (IIBC) Vol. 20, No. 4, pp.17-22, Aug. 31, 2020. DOI : <https://doi.org/10.7236/IIBC.2020.20.4.17>
- [6] Senin and Pavel, "Dynamic time warping algorithm review," Information and Computer Science Department University of Hawaii at Manoa Honolulu, Vol.40, No.855, pp.1-23, 2008.