저가형 3D 카메라를 이용한 K-POP 댄스 안무 검색

김도형, 장민수, 윤영우, 김재홍
한국전자통신연구원 융합기술연구소 지능형인지기술연구부
e-mail:dhkim008@etri.re.kr

K-POP Dance Choreography retrieval with low-cost depth cameras

Dohyung Kim, Minsu Jang, Youngwoo Yoon, Jaehong Kim
Electronics and Telecommunications Research Institute

요 약

본 논문에서는 대용량의 K-POP 모션캡처 데이터베이스에서 특정 안무구간을 검색하는 방법을 제안한다. 제안 기술은 저가형 3D 카메라를 이용하여 사용자가 직접 검색하고자 하는 동작을 생성하고 이를 질의동작으로 입력하여 원하는 안무동작을 검색하는 직관적인 검색 기술로서 구간 동작의 명칭이 존재하지 않는 K-POP 댄스를 검색하기 위한 핵심기술이다. 역동적인 댄스 자세를 표현하고 매칭하는 방법으로 관절 및 바디파트 간의 상대적인 각도 정보를 추출하고 비교하는 방법을 설명한다. 대용량의 모션캡처 데이터베이스를 고속으로 검색하기 위해서 안무동작의 핵심 자세를 분석하여 후보구간 집합을 빠르게 생성하고, 이들 집합에서 Dynamic Time Warping(DTW) 알고리즘으로 안무동작 간의 매칭거리를 보다 정밀하게 산출한다. 약 358분의 K-POP 댄스 곡 100곡에 대한 성능평가에서 92%의 검색정확도를 보였으며, 이는 K-POP 댄스 동작의 복잡성을 고려할 때 경쟁력 있는 성능치이다.

1. 서론

모션캡처(motion capture)란 애니메이션 등에서 캐릭터의 자연스러운 움직임을 표현하기 위해 사람의 몸에 센서를 부착하거나, 적외선을 이용하는 등의 방법으로 인체의 움직임을 디지털 형태로 변환하는 작업을 의미한다. 인간의 실제적인 동작을 모사하고자 하는 영화, 게임, 애니메이션 등의 분야와 인간의 동작을 정밀하게 분석하고자 하는 의료, 재활, 체육 등 실로 다양한 분야에서 많은 양의 모션캡처 데이터를 구축하고 있다.

모션캡처 데이터가 상당량 확보되어짐에 따라 비용 절감을 위해 모션캡처 데이터를 새로 획득하지 않고 기존의 모션캡처 데이터를 수정해서 새로운 분야에 다시 적용하려는 시도가 증가하고 있다. 이러한 재사용 목적을 위해서는 대용량 데이터베이스에서 원하는 모션 구간을 찾아 낼 수 있는 검색 방법이 필수적으로 요구되며, 빠르고 정확한 검색 방법의 중요성이 점차 강조되고 있다 [1].

K-POP 댄스, 발레, 무용 등의 안무 동작에 대해서도 앞서 언급한 다양한 목적으로 방대한 양의 모션캡처 데이터가 구축되고 있으며, 대용량의 안무 데이터베이스에서 사용자가 원하는 안무 구간을 검색하기 위한 방법에 관한 연구도 계속 수행되고 있다 [2].

모션캡처 데이터베이스에서 특정 안무 구간을 검색하는 방법은 크게 3가지로 분류된다.

첫 번째로 곡명, 안무가명, 단위동작명 등의 텍스트 질의어로 모션캡처 데이터베이스 내에서 원하는 안무 구간을 검색하는 방법이다. 이러한 질의어 기반 검색을 위해서는 데이터베이스 내에 있는 모든 단위 동작에 대해서 이름을 부여하는 색인(indexing) 과정이 선행되어야 하는 어려움이 있다. 또한 발레나 무용 등에서는 단위 동작의 전문 명칭이 존재하지만 K-POP 댄스의 경우에는 단위 동작에 대한 명칭조차 없어 질의어로 안무 구간을 검색하는 것이 원천적으로 불가능하다.

두 번째로는 사용자가 스케치 또는 간단한 캐릭터 등의 방법을 통해 특정 자세를 설정하고 설정된 자세와 유사한 자세를 시스템이 제시하는 방법이다. 사용자는 시스템과 상호작용하면서 특정 자세를 점차 구체화 할 수 있으며, 비교적 간편하고 직관적으로 모션캡처 데이터를 검색할 수 있는 장점이 있다. 하지만 이러한 방법은 특정 자세(pose)를 생성하기는 용이하나 일련의 연속된 자세로 이루어진 동작(motion)을 생성하기가 쉽지 않다는 문제점이 있어, 정확하게 안무를 검색하기에는 한계가 있다 [3].

마지막으로, 3차원 카메라로 캡처된 안무 동작을 질의동작으로 생성하고, 이를 입력으로 안무 구간을 검색하는 방법이다. 마이크로소프트 사의 키넥트(Kinect) 등의 저가형 3차원 카메라를 이용하여 사용자가 직접 검색하고자 하는 동작을 촬영할 수 있기 때문에 검색 동작의 생성이 용이하고 매우 직관적인 장점이 있다 [4]. 하지만 검색 대상 안무 데이터는 모션캡처 기술에 의해 정밀하게 캡처된 3차원 관절 위치 정보인데 반하여, 저가형 3D 카메라로 획득한 영상에서 추정된 관절의 3차원 정보는 그 정밀도

가 상대적으로 낮아, 이러한 차이점을 극복하고 검색 정확도에 있어 신뢰성을 확보하는 것이 기술의 관건이다.

본 논문은 위 3가지 방법 중 검색하고자 하는 안무 동작을 저가형 3D 카메라로 직접 생성하여 안무 구간을 검색하는 방법에 관한 것으로, 특히 대용량의 K-POP 댄스 모션캡쳐 데이터베이스를 그 대상으로 한다. 본 기술은 사용자가 직접 추는 안무 동작 자체를 질의동작으로 입력하여 안무를 검색하는 직관적인 검색 인터페이스이다. 특히 구간 동작의 명칭이 존재하지 않는 K-POP 댄스를 검색하기 위해서 반드시 필요한 기술이다.

K-POP 안무 동작을 대상으로 하는 동작인식 및 검색 기술은 다음과 같은 문제점을 해결하여야 한다.

첫째, K-POP 안무 동작은 일상 행동이나 일반 게임 등에서의 동작에 비해 매우 역동적이고 복잡하다. 바디파트의 회전 및 겹침이 빈번하므로 관절 위치의 추적 오류가 크다. 따라서 관절 오류에 강인한 특징추출이 요구된다.

둘째, 사용자의 춤 실력에 따라 퍼포먼스의 변화 폭이 매우 크다. 같은 안무라고 하더라도 사람마다 체형, 동작 크기, 안무속도 등의 차이가 발생한다. 이러한 차이에도 불구하고 같은 동작을 일관된 방법으로 표현할 수 있는 동작 기술자의 개발이 필요하다.

셋째, 안무의 특성상 유사 동작이 상당히 많이 발생할 수가 있어 이를 구별할 수 있는 정교한 동작인식 기술이 필요하다.

본 논문에서는 역동적인 K-POP 댄스 안무 동작으로 구성된 대용량 데이터베이스에서 검색 대상 안무 구간을 빠르고 신뢰성 있게 검색할 수 있는 방법을 제안한다.

2. K-POP 댄스 안무 검색 시스템

본 논문에서 제안하는 K-POP댄스 안무 검색 시스템의 흐름은 그림 1과 같다.

먼저 사용자가 입력하는 3차원 질의동작(3D 깊이 영상)은 골격정보추적 방법 [5]을 적용하여 매 프레임 마다 관절들의 위치 정보로 구성된 골격 정보로 변환될 수 있다. 자세 기술모듈은 골격정보를 입력받아 한 프레임에서의 동작의 자세를 기술하는 자세 기술자를 출력한다. 골격정보추적 방법으로 추출된 골격정보는 그림 2와 같으며 본

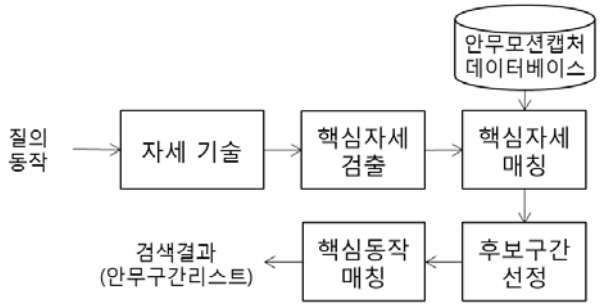

그림 1. K-POP 댄스 안무 검색 시스템 흐름도

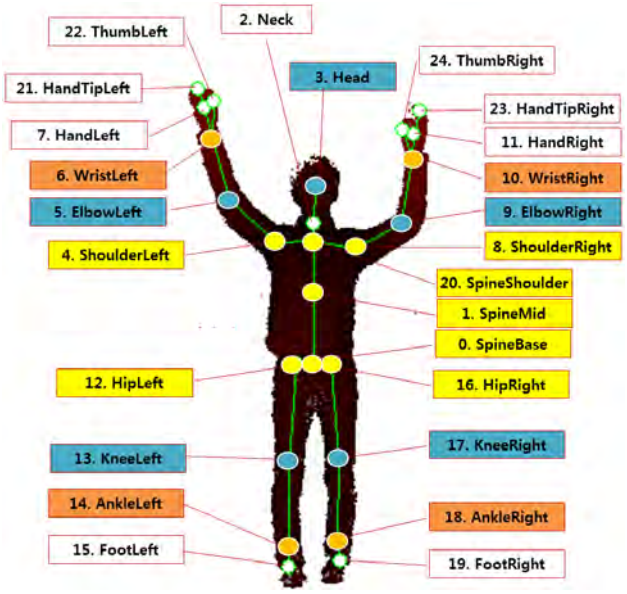

그림 2. 골격추적방법[5]에 의해 검출된 관절 정보들

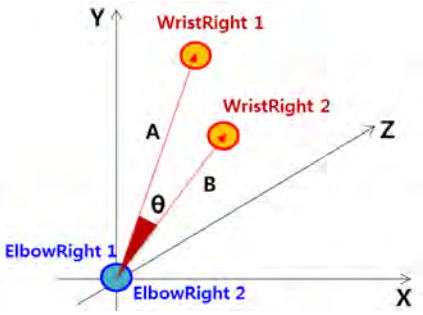

그림 3. 자세간 매칭거리 산출 방법

논문에서는 이들 추출된 골격 정보 중 몸통관절, 1차 관절, 2차 관절의 3차원 위치좌표들의 시퀀스를 자세 기술자로 저장한다. 1차관절은 몸통관절과 연결된 5개의 관절(양팔꿈치, 양무릎, 머리)을 의미하며 하며, 2차관절(양손목, 양발목)은 1차관절과 연결된 4개의 관절을 말한다.

핵심자세 검출 모듈은 자세기술자 시퀀스를 분석하여 질의 동작에서의 대표적인 자세에 해당되는 핵심자세(key pose)들을 검출한다. 핵심자세는 동작을 구성하는 몇 개의 자세 그룹에서의 대표적인 자세로서 클러스터링 기법을 통하여 추출이 가능하다.

핵심자세 매칭 모듈은 검출된 핵심자세들 각각을 안무 모션캡처 데이터베이스에 저장된 안무 동작들의 자세들과 매칭하고 두 자세가 얼마나 유사한 지에 대한 척도인 매칭거리를 자세매칭거리 매트릭스에 저장한다. 두 자세간의 매칭거리를 계산하는 방법은 그림 3과와 같다. 그림과 같이 3차원 공간에서 두개의 오른쪽 아래팔 분절은 부모관절의 위치를 원점으로 하는 2개의 벡터 A와 B로 표현 가능하다. (1차관절의 부모관절은 몸통관절이고, 2차관절의 부모관절은 1차관절이다.) 이 때 두 벡터의 A와 B가 이루는 벡터의 사이각(0도~180도)을 이용하여 두 분절간의 매칭거리를 $(1-\cos\theta)/2$ 로 계산한다. 예를 들어 두 분절(벡터 A와 B)의 방향이 완전히 일치하면(즉,θ=0) 두 분절간

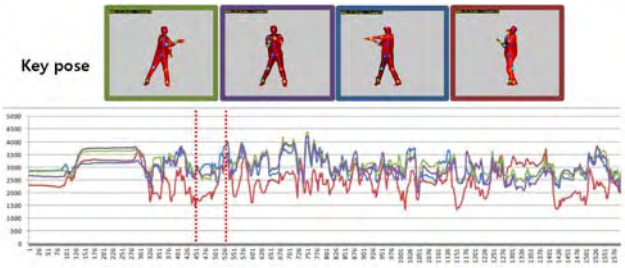

그림 4. 핵심자세들과 댄스곡의 자세들간의 매칭거리

의 매칭거리는 0이 되고 두 분절의 방향이 정반대 방향이면(즉,Θ=180) 두 분절간의 매칭거리는 1이 된다. 자세 기술 모듈에서 기술된 9개의 관절에 대하여 9개의 분절이 존재할 수 있으며, 각 분절마다 동일한 방법으로 매칭거리가 산출된다. 따라서 최종적으로 9개의 분절에서 산출된 매칭거리의 합이 두 자세간의 최종 매칭거리로 결정한다.

후보구간 선정 모듈은 자세매칭거리 매트릭스를 분석하여 핵심자세들이 모여있다고 판단되는 복수개의 후보구간을 선정한다. 후보구간 선정 모듈의 역할은 고속으로 후보구간을 검색함으로써 안무 동작 검색 모듈의 고확장성을 확보하기 위한 방법이다. 자세매칭거리 매트릭스는 그림 4와 같이 도시 될 수 있다. 4개의 핵심자세 각각과 데이터베이스에 저장된 안무 자세들(x축)과의 매칭거리(y축)가 도시되어 있다. 후보 구간을 선정하기 위해서 빨간색 점선으로 표시된 타임 윈도우(time window)를 순차적으로 스캔하면서 타임 윈도우 내에서 각 핵심자세의 최소 매칭거리들의 합을 그 구간의 구간매칭거리로 설정한다. 데이터베이스에 존재하는 모든 곡에 대하여 스캔을 모두 마치면 산출된 구간매칭거리값이 작은 순서로 정렬을 하고, 상위 n개의 구간을 후보구간으로 선정한다.

핵심동작 매칭모듈은 핵심동작들과 후보구간들을 모두 비교하여 가장 유사한 후보 구간을 최종 검색된 안무 구간으로 출력한다. 핵심동작은 검출된 핵심자세를 중심으로 전후 자세들을 포함하는 약 2~3초에 해당되는 지역적 동작(local motion)이다. 안무 구간을 검색함에 있어 하나의 프레임에서 획득한 자세 정보는 여러 프레임으로 구성된 동작 정보에 비해서 그 분별력이 떨어진다. 핵심동작은 핵심자세를 중심으로 하는 자세의 변화정보를 포함하기 때문에 그 분별력이 훨씬 크다. 즉, 후보구간 선정 모듈의 역할이 자세 간의 매칭거리를 간단한 방법으로 빠르게 산출하는 것인 반면에 핵심동작 매칭 모듈의 역할은 동작 간의 매칭거리를 정확하게 산출하는 것이다. 하나의 핵심동작과 하나의 후보 안무구간(동작)의 매칭거리는 Dynamic Time Warping(DTW) 방법을 통하여 산출한다. 이 때 전통적인 DTW 방법이 아닌 unbounded DTW 방법을 사용하여 후보 안무구간 내에서 핵심동작과 가장 유사한 부분열을 추출하고 그 매칭거리를 산출한다 [6]. 같은 방법으로 나머지 핵심동작에 대해서도 부분열과 매칭거리를 산출할 수 있으며, 최종적으로 핵심동작과 후보 안무 구간과의 매칭거리는 이들 매칭거리들의 합이다.

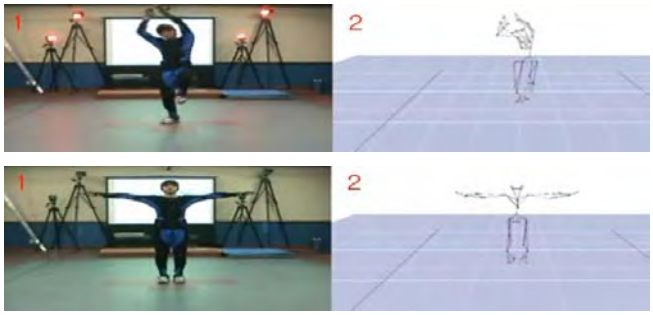

그림 5. K-POP 댄스 모션캡쳐 촬영현장(좌)과 캡쳐된 모션캡쳐 데이터 샘플(우)

그림 6. 키넥트 카메라로 수집한 포인트 안무 샘플 영상

3. 실험 결과

제안한 시스템의 성능평가를 위해 총 100곡의 K-POP 댄스 곡에 대한 모션캡쳐 데이터베이스를 구축하였다. 모션캡쳐 전용 스튜디오에서 전문 댄서의 춤동작을 대상으로 캡쳐하였다. 댄서가 모션캡쳐 전용 수트를 입고, 총 37개의 적외선 마커를 몸에 부착한 후 전용 스튜디오에서 춤을 추면 총 12대의 적외선 카메라가 마커의 3D 위치를 캡쳐하였다. 총 644,135 프레임, 약 358분 분량의 데이터들을 수집하였다. 그림 5는 모션캡쳐 촬영현장과 캡쳐된 모션캡쳐 데이터 샘플이다.

마이크로소프트 키넥트2 카메라를 사용하여 테스트에 사용될 3D 깊이정보 데이터를 수집하였다. 각 곡마다 2개의 포인트 안무를 선정하여 총 200개의 포인트 안무에 대한 춤 동작을 수집하였다. 검색 대상인 포인트 안무의 평균 퍼포먼스 시간은 약 5.6초였다. 그림 6은 수집된 포인트 안무의 샘플 영상이다.

테스트 데이터의 질의 동작이 입력되면 앞서 언급한 일련의 과정을 거쳐 200개의 후보 구간을 선정하였다. 선정된 후보 구간들 중에 검색하고자 하는 구간(즉, 정답구간)이 포함된 경우는 97.5% 였다. 표 1은 매칭결과의 순위별 정확도(precision)이다. 본 안무 검색 시스템은 입력된 질의 동작과 가장 유사한 안무 구간들의 리스트를 유사도가 높은 순서대로 정렬하여 출력한다. 표 1에서와 같이 검색대상 구간이 검색결과 리스트의 상위 10위 내에 포함되는 경우는 92%였으며, 상위 20위 내에는 95%가 포함되었다.

표 1. 매칭결과의 순위별 정확도(precision)

순위	후보구간선정 정확도 (%)	핵심동작매칭 정확도 (%)
rank 1	43.5	73
rank 5	63	90.5
rank 10	72	**92**
rank 20	79	**95**
rank 100	90	-
rank 200	**97.5**	-

그림 7. 질의동작기반 안무 DB 검색 프로그램 스크린 샷

K-POP 안무 동작의 복잡성과 퍼포먼스의 다양한 정도를 고려해 볼 때, 이러한 결과는 동작검색에 있어 매우 경쟁력있는 성능치라고 판단된다. 평균응답시간은 4.7초 였다.

4. 결론

본 논문에서는 대용량의 K-POP 안무 동작 데이터베이스에서 원하는 안무 구간을 검색하기 위한 방법으로 검색하고자 하는 안무 동작을 저가형 3D 카메라로 직접 생성하여 검색할 수 있는 직관적인 인터페이스를 가지는 시스템을 제안하였다. 복잡한 K-POP 댄스 동작을 고속으로 검색하기 위해 핵심 자세 및 핵심 동작 단위로 매칭을 수행하는 방법을 적용하였다. 또한 고속으로 후보구간을 선정하고 한정된 후보에 대해서만 정밀 매칭을 수행하는 2단계 접근방법으로 속도 향상을 도모하였다. 앞으로는 검색 정확도 개선을 위해 관절 위치 추적에 강인한 포즈 기술 방법의 개발에 집중하고 근사근접이웃탐색(approximate nearest neighbor search)방법과 fast DTW 방법을 적용하여 응답속도를 향상시킬 계획이다.

Acknowledgment

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2015년도 문화기술연구개발지원사업의 연구결과로 수행되었음.

참고문헌

[1] Deng, Zhigang, Qin Gu, and Qing Li. "Perceptually consistent example-based human motion retrieval." In Proceedings of the 2009 symposium on Interactive 3D graphics and games, pp. 191-198. ACM, 2009.

[2] Essid, Slim, Xinyu Lin, Marc Gowing, Georgios Kordelas, Anil Aksay, Philip Kelly, Thomas Fillon et al. "A multimodal dance corpus for research into real-time interaction between humans in online virtual environments." In ICMI Workshop On Multimodal Corpora For Machine Learning. 2011.

[3] Chao, Min-Wen, Chao-Hung Lin, Jackie Assa, and Tong-Yee Lee. "Human motion retrieval from hand-drawn sketch." Visualization and Computer Graphics, IEEE Transactions on 18, no. 5 (2012): 729-740.

[4] Hu, Min-Chun, Chi-Wen Chen, Wen-Huang Cheng, Che-Han Chang, Jui-Hsin Lai, and Ja-Ling Wu. "Real-Time Human Movement Retrieval and Assessment With Kinect Sensor." Cybernetics, IEEE Transactions on 45, no. 4 (2015): 742-753.

[5] Shotton, Jamie, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore. "Real-time human pose recognition in parts from single depth images." Communications of the ACM 56, no. 1 (2013): 116-124.

[6] Tormene, Paolo, Toni Giorgino, Silvana Quaglini, and Mario Stefanelli. "Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation." Artificial intelligence in medicine 45, no. 1 (2009): 11-34.