

PC 카메라와 모션 캡처를 이용한 3차원 체험형 댄스 콘텐츠 개발

3D Actual Experiencing Dance Contents Development using PC Camera and Motion Capture Device

김중남, Jongnam Kim, 이승찬, Seungchan Lee, 정영기, Youngkee Jung
호남대학교 컴퓨터공학과

요약 본 논문에서는 댄스를 디지털콘텐츠와 접목시켜 누구나 쉽고 재미있게 즐길 수 있는 체험형 콘텐츠를 개발하고자 한다. 개발된 콘텐츠에서는 마우스나 키보드 등의 정적 입력 인터페이스를 벗어나 체험자의 능동적 참여를 위해 영상처리기술을 통해 PC 카메라로부터 얻은 영상에서 체험자의 모습을 분리하고 댄스 모션이 적용된 3D 모델과 같은 공간에 출력하여 두 모델간의 동작을 비교함으로써 같은 동작을 취하고 있는지를 판단하게 된다. 이때 체험자에게 어떠한 센서나 마커도 부착하지 않고 강건한 환경에서의 체험자 영상만을 추출하기 위한 방법과 댄스 모델과 체험자의 움직임 비교 방법을 새로이 제안하였다.

핵심어: Interactive Game, Dance Contents, Motion Capture, Motion Comparison

1. 서론

인터넷의 급속한 확대와 유무선 통신 인프라 등의 확충으로 디지털콘텐츠산업은 지식정보 사회의 핵심 산업으로 급부상하고 있으며 세계 유수의 선진국들은 시장을 선점하기 위해 디지털콘텐츠 산업을 중요한 전략사업으로 선정, 집중적인 투자를 진행하고 있다[6]. 하지만 국내 디지털콘텐츠 산업 중 가장 큰 규모인 게임 분야에서도 대부분이 온라인 게임에 치중되어 있어 여러 가지 사회병리학적 문제를 야기하고 있으며 다양한 콘텐츠 이용의 욕구가 늘고 있다. 특히 마우스(mouse)나 키보드(keyboard) 등 기존 인터페이스 방식을 기반으로 한 게임과 온라인 중독, 지나친 몰입이 맞물려 게임에 대한 부정적인 이미지를 확산시키고 사회 문제를 일으키는 주요 원인으로 인식되게 하였다[7].

위 문제를 해결하기 위해 기존 인터페이스 방식에서 벗어나 몸 전체의 동작을 인터페이스로 하는 콘텐츠 제작이 필요하다. 특히 체험이 가능하고 활동량이 많으며 가족 간의 오락문화를 발전시킬 수 있는 내용과 콘텐츠 이용자에게 흥미를 유발할 수 있어야 한다. 위 항목들을 충족하는 가장 적합한 아이템으로 댄스를 꼽을 수 있다. 댄스는 남녀노소를 불문하고 많은 사람들이 즐기는 엔터테인먼트 요소로서 수많은 다이어트 프로그램, UCC(User Created Contents) 동영상, 오락 프로그램의 요소로 활용되고 있다.

이를 반영하듯 최근 다양한 종류의 댄스 콘텐츠가 출시되

고 있다. '클럽 오디션'과 'iDance', 'Eye Toy'가 대표적이다[8] [9] [10].



(a) 오디션

(b) 아이댄스



(c) 아이토이

그림 1. 기존 출시된 댄스 콘텐츠

지금까지의 댄스 관련 콘텐츠들은 단순히 댄스 동영상을 보여주는 교습용 콘텐츠이거나 음악에 맞춰 간단한 키를 입력하는 것만으로 댄스를 즐기는 게임이 대부분이다. 이들은 다양한 장르의 댄스를 배울 수 있다는 것과 쉬운 조작법으로 멋진 춤을 출 수 있는 장점이 있지만 만들어진 콘텐츠의 활용 범위가 작고 쉽게 흥미를 잃을 수 있으며 단지 가상 모델의 춤을 보는 즐거움뿐이다.

이런 점에 착안하여 본 논문에서는 모션캡처 장비이용, 영상처리 및 3차원 그래픽스 기술을 활용한 콘텐츠 제작 기술과 사용자가 직접 참여하는 댄스를 접목시켜 많은 사람들에게 흥미를 제공하고 누구나 쉽고 재미있게 즐길 수 있는 3차원 체험형의 댄스 콘텐츠를 제작 하고자 한다.

2. 전체 시스템 개요

제안된 체험형 댄스 콘텐츠는 크게 세 부분으로 분류된다. 시스템의 각 부분은 아래와 같다.

- 모션캡처 장치를 이용한 댄스 모션 데이터 획득과 이를 활용한 3D 댄스모델 생성 부.
- 입력된 PC Camera영상에서 배경과 체험자를 분리하는 영상 분리 부.
- 댄스 모델과 체험자의 동작을 비교하는 동작 비교 부.

그림 2 는 본 논문에서 제안한 체험형 댄스 콘텐츠의 시스템 구성을 나타내고 있다.

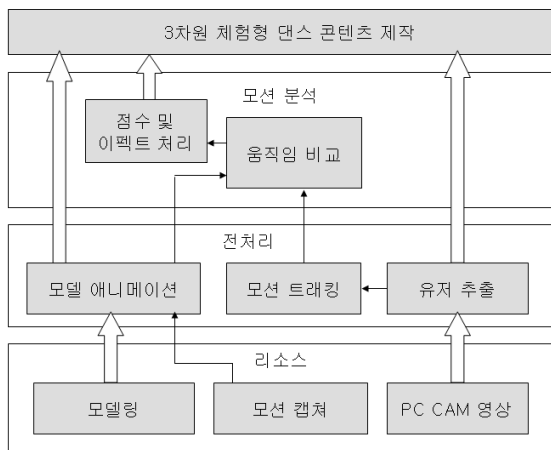


그림 2. 전체 시스템 구성도

개발된 콘텐츠에서는 3D 댄스 모델과 체험자의 모습을 가상공간의 무대에 함께 위치시켜 체험자가 보다 쉽게 동작을 따라 할 수 있으며 댄스 모델과 체험자의 동작을 비교함으로써 본 콘텐츠로의 몰입을 증대시켰다.

2.1 모션캡처를 이용한 3D 댄스 모델 생성

본 연구에 사용된 모션캡처 시스템은 8대의 적외선 카메라가 장착된 디지털 방식의 Eagle 광학식 모션캡처 시스템을 이용하였다. 이 시스템은 사람의 동작을 초당 60 프레임 이상 캡처함으로써 데이터 손실 없이 빠른 시간에 실감 있는 디지털 동작을 획득하기에 적합하다.

그림 3 은 댄스모델 생성을 위한 모션캡처 과정이다. 댄스 장르의 모션캡처를 위한 4m x 4m 캡처공간과 8대의 카메라 셋팅이 끝나면 카메라 보정(Calibration)을 하는데 바디 모션캡처를 위한 카메라 3D Residuals 파라미터 평균값은 0.45로 설정했다. 이 값이 0.6이상 이 되면 데이터에 Gap과 Spike 데이터가 발생하여 후처리 작업 시간이 많이 발생하므로 재 보정작업을 실시했다.

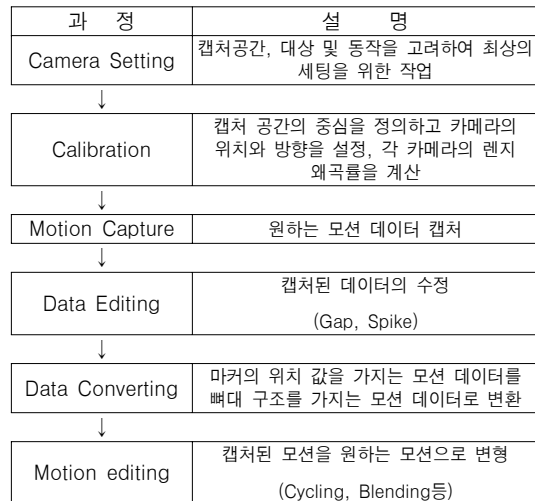
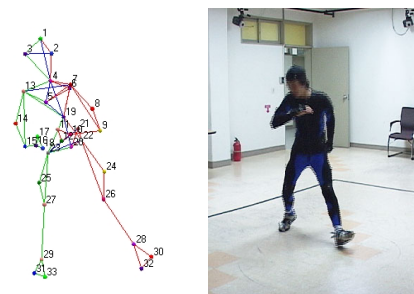


그림 3. 모션 캡처 과정

모션캡처 과정에서 생성된 디지털 모션 데이터는 33개의 3차원 공간상의 X, Y, Z 좌표를 가지는 점이다. 이 데이터는 Gap과 Spike 필터링 등의 에디팅 과정을 거치고 데이터 변환과정에서 뼈대 구조를 가지는 모션 데이터로 변환을 한 후에 3차원 캐릭터에 적용하여 활용한다.

캐릭터 디자인 과정이 끝나면 캐릭터가 모션데이터에 맞게 동작할 수 있도록 뼈대를 생성해주는 과정을 거친다. 이 과정에서는 3ds MAX의 Character studio에 있는 바이페드(Biped) 기능을 사용하였다. 3ds MAX 모델링 툴에서 지원하는 애니메이션 옵션은 뼈대 속성이 있는 모션 포맷인 BVH(Biovision Hierarchical)를 이용하여 모션캡처 시스템에서 생성되는 데이터를 사용할 수 있다[5].



(a) 마커 위치

(d) 모션 캡처

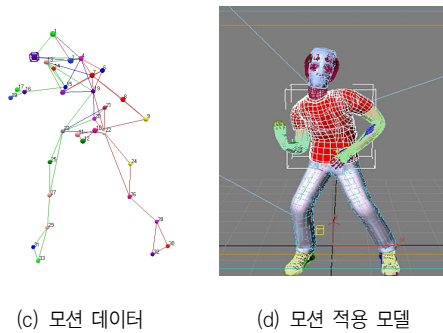


그림 4. 3D 댄스 모델 생성 과정

2.2 체험자 영상 분리

배경 영상에서 체험자만의 영상을 분리하기 위해 기본적으로 배경으로 지정된 영상과 현재 영상의 차영상을 통해 얻게 된다. 하지만 크로마키 등의 환경을 갖추지 않고 강건한 환경에서의 체험자 영상만을 분리하기 위해서는 여러 가지 문제점을 해결해야 한다.

첫째, 보편화된 PC Camera는 자동 조광보정 기능이 기본적으로 내장되어 있기 때문에 영상내의 명암분포가 변하게 될 경우 전체 영상의 밝기와 색상을 변경하게 된다. 또 움직이는 물체가 없다 하더라도 한 픽셀에 대해 여러 프레임에서의 일정한 값이 보장 되지 않기 때문에 차영상에서 전역에 걸쳐 노이즈가 심하게 발생하게 된다.

둘째, 배경영상에서 체험자의 모습과 유사한 명암이나 색이 겹치게 될 경우 배경으로 간주하여 체험자의 영상에 홀(hole)이 발생하게 된다. 감시 카메라에서처럼 대략의 형태를 얻고자 함이 아니며 가상공간으로의 표현에 이질감을 줄이기 위해 최대한 완전한 형태에 가까워야 한다.

셋째, 본 콘텐츠는 가상공간에서의 댄스 모델 출력 등 많은 3D 연산을 필요하게 되므로 체험자 영상 추출 부분이 실시간으로 이루어져야한다. 때문에 레이블링(Labeling) 등의 처리 시간이 많이 걸리는 작업은 피하고 되도록 연산 속도가 빠른 알고리즘을 택해야 한다.

본 논문에서는 위 문제들을 해결하기 위한 체험자 영상 분리 방법을 그림 5 와 같이 제안한다.

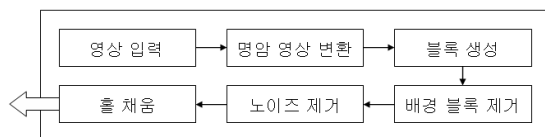


그림 5. 체험자 분리 절차

입력 영상에 대해 명암 영상으로 변환 후 몇 개의 블록으로 지역화를 시킨다. 지역화 된 블록에 대해 배경 영상의 블록과 템플레이트 매칭을 하여 배경 블록과 체험자 블록을

구분한다. 분류된 체험자 블록의 차영상을 구하고 노이즈 제거 후 주변 픽셀을 참조한 체험자 영역의 홀을 채우게 된다.

먼저 영상 잡음과 밝기 변화에 둔감하며 빠른 속도로 배경과 체험자 영역을 구분하기위해 320 x 240 영상을 20 x 20개의 블록으로 나눈 후 배경 영상과 입력 영상 블록의 템플레이트 매칭을 통해 농담의 유사 정도를 비교하게 된다.

식 (1)~(3)은 농담정규화 매칭법 (Normalized Gray-level Correlation)을 나타낸다 [1].

$$r = \frac{a \cdot b}{|a| \cdot |b|} = \frac{\sum_{i=0}^M a_i b_i}{\sqrt{\sum_{i=0}^M (a_i)^2 \sum_{i=0}^M (b_i)^2}} \quad (1)$$

여기서 $a(i) = g(i) - \bar{m}$, $b(i) = t(i) - \bar{t}$, M 은 템플레이트 크기이다. 식 (2)는 평균밝기를 빼서 정규화 시킨 두 영상을 식 (1)의 두 벡터 비교식을 이용한 유사도를 검사를 나타낸다.

$$r = \frac{(M \sum gt - (\sum g) \sum t)}{\sqrt{[(M \sum g^2 - (\sum g)^2)] [(M \sum t^2 - (\sum t)^2)]}} \quad (2)$$

식 (2)에서 $\sum g$, $\sum g^2$, $\sum gt$ 를 제외한 나머지 부분들을 미리 계산해 놓은 방법으로 식 (3)을 얻을 수 있다.

$$Score = [\max(r, 0)]^2 \quad (3)$$

그림 6 은 비교된 블록 유사도의 예를 나타내고 있다.

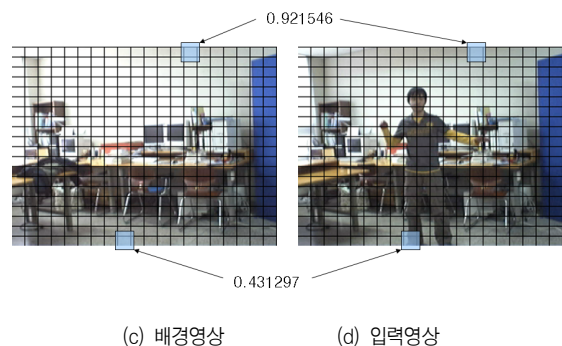


그림 6. 블록의 농담 비교

입력 영상의 블록에 체험자가 존재하지 않을 경우 배경 영상과의 유사도는 1에 가깝게 된다. 반면 체험자가 존재한

부분의 블록은 1의 값에서 크게 멀어지게 되어 해당 블록에 체험자가 있다는 판단을 할 수 있다. 따라서 배경으로 분류된 블록은 더 이상의 처리가 필요하지 않으므로 배경 영상의 블록에 1/누적 프레임 수의 가중치만큼 업데이트를 하게 된다.

남아있는 체험자 후보 블록에서 차영상을 얻게 되는데 이때의 임계값은 다음과 같은 방법을 사용하였다.

$$T = 1 - r(|m_1 - m_2|) \quad (4)$$

r 은 식 (3)에 의한 결과이고 m_1, m_2 는 배경과 입력 영상 블록의 밝기 평균이다. 최소, 최대의 한계값을 가지며 밝기 평균의 차에 유사도를 적용함으로써 블록이 비슷할수록 낮은 값을 선택할 수 있다.

그림 7은 배경블록과 노이즈가 제거된 결과이다.



그림 7. 제거된 배경 블록과 노이즈

모폴로지 연산을 통해 대부분의 노이즈가 제거되고 홀이 채워지지만 배경과 유사한 체험자 영역이 겹치게 될 경우 더 큰 홀이 발생하여 여전히 남아있게 된다.

홀을 채우기 위해 홀의 밝기 값을 파악하게 되는데 조사된 픽셀의 밝기와 유사한 밝기이며 하나 이상의 픽셀들로 연속되어 있다면 해당 픽셀들의 위치를 저장하게 된다. 배경 블록의 같은 위치에 밝기를 조사하여 홀 위치보다 더 넓은 지역으로 연속되게 분포되어 있다면 배경과 겹침에 의한 홀이라 간주하고 홀을 채우게 된다. 마지막으로 체험자 영상 내의 노이즈를 위해 모폴로지 연산을 한번 더 수행한다.

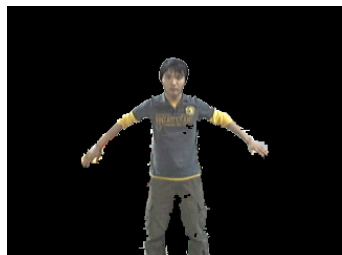


그림 8. 홀 채우기 결과

2.3 체험자와 댄스 모델간의 움직임 비교

체험자가 본 콘텐츠에 몰입할 수 있도록 체험자와 댄스 모델의 움직임을 비교하고 각 부위별 유사도를 이펙트로 출력 하게 된다.

동작 비교 절차는 그림 9와 같다.

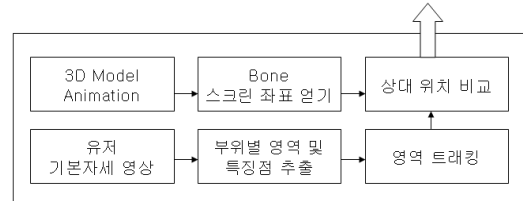


그림 9. 동작 비교 절차

댄스 모델의 각 Bone들을 정면에서 2D 좌표로 투영하여 체험자 영상과 동작을 비교할 수 있는 대상을 생성하였다. 이는 실루엣 비교방식에 비해 연산량이 적고 각 영역에 따른 동작 유사도를 판단할 수 있다. 하지만 다른 동작임에도 불구하고 비교 영역에 체험자 영상이 존재할 수 있으므로 더 정확한 동작 비교를 위하여 머리, 손, 발의 영역은 체험자의 초기 T 자세에서 추적하여 댄스 모델의 Bone 위치와 비교 데이터로 사용한다.

동작 비교에 앞서 체험자는 실시간으로 댄스 모델과 같은 동작을 취하기 어렵기 때문에 체험자의 동작은 항상 다른 동작으로 판단하게 될 것이다. 위 문제를 해결하고 조금 늦게 동작을 취하더라도 정확한 동작이라면 그에 따른 차등 점수를 적용시키기 위해 다음과 같이 적용하였다.

$$a = \begin{cases} +0.2 & \text{if } IC \subset A \\ -0.1 & \text{if } IC \subset (B - A) \end{cases} \quad (5)$$

여기서 A는 분리된 체험자 영상인 I_t 이고, B는 일정 프레임동안 a 의 누적 값인 $\sum_{k=t-1}^{t-n} a_k$ 이다. 그림 10은 식 (5)를 색으로 구분하여 나타내고 있다.



그림 10. 팔 동작의 색 구분

그림 10 에서 구부린 팔(푸른색)을 A, 편 상태의 팔(붉은색)을 B, 중첩된 영역(자주색)을 C라 하고 B에서 A로 동작하였다면 B, C는 초기에 모두 0.2의 값을 가지지만 다음 프레임에서 B에 해당하는 영역은 존재하지 않기 때문에 0.1의 가중치를 감소하여 0.1이 되고 C는 0.4, A는 0.2가 된다. 이를 통해 누적된 가중치의 집합 군과 댄스 모델의 Bone 위치를 비교하여 늦게 동작한 체험자에 대해서도 알맞은 점수를 적용 시킬 수 있게 된다.

체험자와 댄스 모델을 비교하기 위해서는 두 물체간의 크기가 같아야 하는데 이를 위해 크기의 척도로 삼을 체험자의 얼굴, 손, 발 영역의 위치를 파악해야 한다. 콘텐츠의 진행초기에 체험자가 위치할 공간에 T 자세의 실루엣을 출력하여 체험자에게 기초 자세를 요구하게 된다. 이후 YCbCr 컬러 모델 변환을 통해 피부색 영역을 검출 [4]하여 얼굴과 손 영역을 찾게 되고, 꼭짓점 추출 영상처리 기법 [2] [3]을 통하여 발 영역의 특징점으로 추출한 후 이 점들을 추적하게 된다. 각 부위는 분리된 체험자 영상에서만 탐색을 시도하기 때문에 전역 탐색에 비해 훨씬 속도가 빠르며 프레임간의 관계를 통해 행동 범위를 예측할 수 있어 탐색 영역을 제한 할 수 있다.

댄스 모델의 크기를 조정하기 위해 그림 11 과 같이 체험자의 얼굴 중심 좌표와 최하단의 발 좌표의 길이를 기준으로 계산된 댄스 모델의 Bone 좌표를 조정하게 된다. 이는 체험자의 동작에서 양 발의 위치가 다른 영역보다 높기는 힘들기 때문이다.

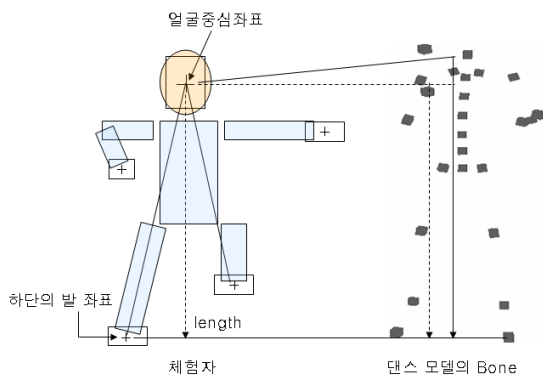


그림 11. 체험자를 기준으로 한 본(Bone)의 위치변환

크기가 같아진 두 물체간의 동작을 비교하고 이를 점수화 할 수 있도록 다음 두 가지의 기준을 정의하였다. 첫째는 본의 위치를 기준으로 한 사각영역에 포함되는 체험자 영역의 비율이며 둘째는 체험자 영상에서 추적된 주요 부위와 그에 해당하는 본의 상대비교 값이다.

체험자의 포함 비율은 검색 영역 중 0.1 이상의 가중치

값을 가진 픽셀들에 대해서이며 검색 영역 크기의 1/5 보다 작은 비율의 영역은 계산 대상에서 제외시킨다.

$$Score = ratio \sum_{i=0}^M \sum_{j=0}^N T(x_i, y_j) \quad (6)$$

비율과 가중치에 따른 점수에 추적한 주요 부위의 자세를 점수화 하여 더하게 된다.

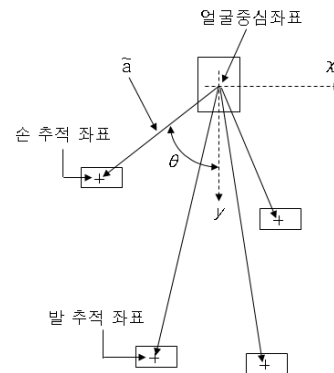


그림 12. 얼굴을 중심으로 한 손과 발의 벡터와 내적

그림 12 와 같이 체험자 영상과 댄스 모델의 얼굴 위치를 중심으로 각 손과 발의 벡터의 크기와 내적을 구한다. 구해진 값으로 서로 상이한 정도를 점수로 환산하여 그에 따른 이펙트 효과를 화면에 출력하게 된다.

그림 13 은 추적된 체험자의 주요 위치와 실루엣에 본을 대입한 비교 절차를 나타낸다.

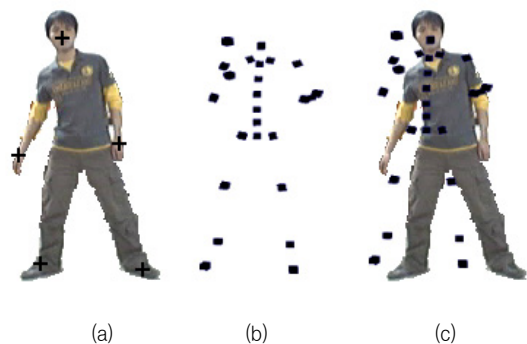


그림 13. (a) 추적된 체험자의 주요 부위 (b) 스케일링 된 본 (c) 본과 체험자의 대입 비교

3. 구현 및 실험

본 장에서는 제안된 시스템에 의한 실험결과를 나타낸다. 체험자의 영상을 입력받기 위한 PC 카메라는 Logitech QuickCam Pro 4000 모델을 사용하였으며 Pentium IV 3.0Ghz, 1G RAM, VGA Geforce 6600 의 환경에서 테스트

트 되었다.

그림 14는 밝기의 변화에 대응하고 체험자 영역의 많은 홀들이 제거됨을 보여주고 있다.

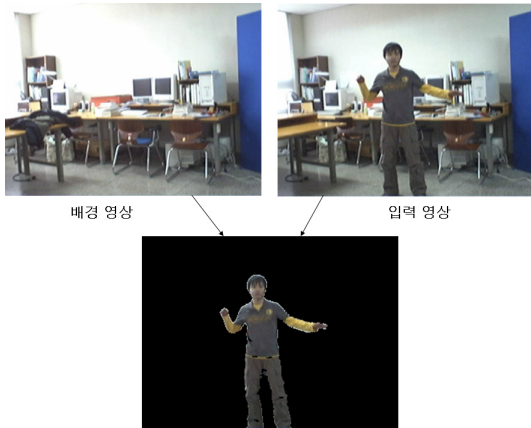


그림 14. 분리된 체험자 영상

분리된 체험자 영역 주변은 알파(Alpha)채널을 0으로 하여 투명처리 한 후 가상공간의 우측에 출력하고 좌측에는 댄스 모델을 출력하게 된다. 그림 15는 체험자가 바라 볼 구현된 결과 화면이다.



그림 15. 구현 결과

그림 16은 개발된 콘텐츠의 진행 모습을 보여주고 있다.

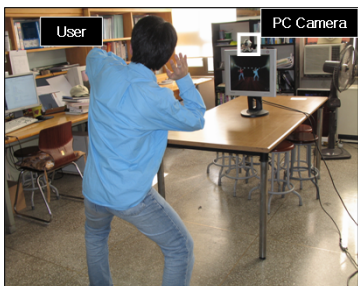


그림 16. 시연 모습

4. 결론

본 논문은 기존에 출시되어 있는 댄스 콘텐츠들의 단순하고 지루한 진행방식에 대한 해결책을 제시하고자 체험자가

직접 몸으로 체험하면서 댄스를 즐길 수 있는 새로운 방식의 댄스 콘텐츠를 목표로 제작 되었다.

단순한 키 조작이 아닌 몸으로 직접 체험할 수 있는 게임 환경의 제공으로 체험자에게 보다 많은 흥미를 안겨줄 수 있을 것으로 기대된다. 또한 모션캡처 기술의 적용으로 3D 캐릭터에 역동성과 현실성이 더욱 부여되었을 뿐만 아니라 적용된 모션 캡처 기술을 응용하여 아케이드 게임, 무술, 요가, 유아 교육용 체험기 등 다양한 문화적 동작에 대한 모션 서비스를 실시 할 수 있고 댄스 모델의 동작을 함께 따라 추는 배틀 형식의 온라인 콘텐츠로 개발하여 여러 유저들의 모습을 감상 할 수 있고 커뮤니케이션 가능한 개인 교습 장치로 활용 가능 할 것으로 전망된다.

제안한 방식에 의해 홀을 제거하고 주요 부위의 위치를 추적하였지만 복잡한 배경과 빠른 동작에서의 홀 채움 오류나 옆모습과 뒷모습에서의 추적 오류가 많아 이러한 다양한 분야에 응용 가능하고 콘텐츠의 질을 더욱 끌어올리기 위해서 완전한 체험자 분리 및 더욱 섬세한 동작 비교가 가능하도록 개선해 나가야 한다.

참고문헌

- [1] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing 2/E", Prentice Hall, 2003.9
- [2] Beymer, D., McLauchlan, P., Malick, J., "A real-time computer vision system for measuring traffic parameters,"Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 12, pp.495-501, 1997.
- [3] 오종안, 정영기, "능동카메라 환경에서의 특징기반의 이동물체 추적", 대한전자통신학회, 2002
- [4] 하태진, 우운택, "Video see-through HMD 기반 증강현실을 위한 손 인터페이스", HCI2006 학술대회 발표 논문집, 2006
- [5] 김남호, "모션캡처 시스템을 이용한 모바일 3D 댄스 콘텐츠 제작 연구", 한국콘텐츠학회논문지, 2006
- [6] 최동진, 김강진, 박윤식, 김재운, "2005년도 국내 디지털 콘텐츠산업 시장조사 보고서", 한국소프트웨어진흥원, 2006
- [7] 유승호, 신영철, 정의준, 양소연, 김영선, 게임몰입증(게임중독)의 현황과 대처방안, 문화관광부, 2001
- [8] Club Audition
<http://www.clubaudition.com>
- [9] iDance
<http://www.idance.co.kr>
- [10] Eye Toy Play
<http://www.eyetoyplay.com>