

포즈 인식을 이용한 가상 비행 체험 콘텐츠

박재완, 조병수, 이철우
전남대학교 전자컴퓨터공학과
e-mail: cyanlip@naver.com

Virtual Flight Experiment Contents using Pose Recognition

Jae-Wan Park, Byeong-Su Jo Chil-Woo Lee
Dept of Electronics and Computer Engineering, Chonnam National University

요 약

본 논문은 사용자의 포즈를 인터페이스로 사용하는 가상 비행 체험 콘텐츠에 대하여 기술한다. 사용자의 포즈를 인식하기 위해서 제스처를 구성하는 상반신의 포즈를 식별하여야 한다. 본 논문에서 기술한 콘텐츠는 한정된 공간에서 사용자의 움직임을 인식하고 가상공간에 아바타를 이용하여 표현하고 있다. 그러므로 사용자는 가상공간에서 정의된 포즈를 사용하여 가상 비행을 체험할 수 있고 인식된 포즈는 OS-Value 이벤트를 이용하여 가상 비행 체험 콘텐츠에서 인터페이스로 활용이 가능하다.

1. 서론

인간의 동작을 인식하여 컴퓨터와 상호작용하는 HCI에 대한 연구는 많은 방면에서 응용이 가능하다. 그러나 인간의 신체는 복잡한 구조를 가지고 있고 그 동작변화가 매우 빠르게 일어나기 때문에 인간의 동작을 정확히 인식한다는 것은 매우 어려운 일이다. 그리고 한정된 공간에서 상호작용에 필요한 모든 입력을 인간의 동작으로 사용하는 것은 사용자 입장에서도 불편하다.

본 논문에서는 사용자의 상반신 포즈를 입력으로 사용하는 가상 비행 체험 콘텐츠에 대하여 기술한다. 한정된 공간에서 사용자의 움직임을 가상공간에 표현하는 것은 한계가 있다. 예를 들면 좁은 공간에서 사용자의 위치에 따라 아바타의 움직임을 표현하는 것은 제한적이다. 그러므로 본 논문에서는 사용자의 상반신 포즈를 이용하여 인간의 욕망 중 하나인 자유비행을 소재로 하는 가상 비행 체험에 대한 시나리오를 가지고 있다. 사용자는 가상의 공간에서 가상의 아바타를 통해 인간의 비행을 표현함으로써 공간의 제약을 벗어나 콘텐츠를 체험할 수 있다. 신체의 일부인 팔을 날개로 대체하고 사용자의 행동이 가상공간에서 하나의 도구로 사용되는 것을 체험하고 정의된 포즈에 따라 아바타의 애니메이션이 실행된다. 사용자는 아바타의 애니메이션을 보

고 명령이 수행되는 것을 확인할 수 있다.

신체의 포즈를 인식하기 위해 다음과 같은 대표적인 연구들이 있다. Haritaoglu et al.[1]은 단일 실루엣 영상을 분석하여 6개의 신체 부위를 검출하였다. 그래서 실루엣 영상의 수직, 수평 히스토그램과 데이터베이스 히스토그램과의 비교를 통해 유사도가 가장 높은 포즈로 결정하는 방법을 제안했다. Iwasawa et al[2]은 적외선 카메라로 획득한 열 영상을 분석해 머리 끝, 손끝과 같은 주요 신체 부분을 찾아 사람의 포즈를 인식하는 방법을 제안했다. 3D 기반 방법은 하나의 카메라를 이용해서 촬영한 단일시점 영상으로부터 인체의 3차원 포즈를 예측하는 방법과 다수의 카메라를 사용해 3차원 포즈를 복원하는 방법으로 구분할 수 있다. Micilotta et al[4]은 단일 영상에서 3차원 상반신 포즈 추정을 위해 adaboost[5]를 사용하였다. 확장된 adaboost를 이용해 얼굴, 몸통, 양 손을 검출하고, 검출된 각 부분의 확률적 조합으로 사람의 위치를 결정하였다. 일단 상반신의 2차원 조합이 결정되면, 실루엣과 사용자의 에지 맵을 매칭해 3차원 포즈를 추정하였다. 그리고 입력 영상의 실루엣과 모든 데이터베이스 사이의 중복 비율을 계산하여 매칭 점수를 계산한다. 상위 10%의 매칭 점수를 갖는 실루엣 데이터베이스 집합에서 에지 맵 데이터베이스의 에지 맵과 비교하

여 3차원 포즈를 추정하였다. D.M. Gavrilu et al[6]은 4개의 카메라를 설치한 후 카메라 파라미터 추정을 통해 3차원 정보를 얻었다. 이 정보를 이용해 3차원 인체 모델을 구성하였다. 그리고 각 카메라에서 얻은 2차원 영상에 투영시켜 chamfer 매칭[7]을 통해 3차원 포즈 모델의 각 부분들을 순서대로 맵핑하였다. 하지만 이 방법은 사각형 모양으로 각 모서리에 4대의 카메라를 설치해야 하기 때문에 장소 제약에 대한 단점이 있다.

3D 기반의 포즈 인식 방법은 일반적인 환경에서 복잡한 3차원 움직임 포즈도 인식할 수 있고 관절의 자유도를 구현하고 있기 때문에 구체적인 포즈를 추정할 수 있지만 3차원 포즈를 인식하기 위해서는 수많은 파라미터가 사용되기 때문에 파라미터를 조정하는 것이 어렵고, 수행시간이 오래 걸린다. 그러므로 실시간으로 인체의 3차원 포즈를 인식하기 위해서는 소수의 신체 부분만을 검출하거나 사전에 정의하는 포즈의 개수를 제한해야 한다.

3차원 인체 포즈 인식 방법의 사용은 신체 부위의 모양정보(shape)와 관절 사이의 한계각도 정보를 선행지식으로 사용할 수 있는 장점이 있으며 이러한 정보를 이용하여 신체 부위의 겹침 현상에 대한 문제점을 해결할 수 있다.

2. 콘텐츠의 구성

본 콘텐츠의 구성은 다음과 같다. 아바타는 시작지점에서 사용자의 입장을 기다리고 있다. 사용자가 가상공간에 입장하고, 아바타가 사용자를 인지하는 것에 따라 아바타 애니메이션이 실행된다. 그리고 사용자가 아바타에 매칭되면 사용자의 팔 움직임에 따라 아바타가 동작하게 된다.

사용자가 일정 빠르기의 날개짓을 하면 시작지점에서 날아오르는 아바타 애니메이션이 실행되고, 아바타는 비행 상태를 유지한다. 그리고 사용자의 날개짓 속도에 따라 가속과 감속이 구분된다. 비행만으로는 사용자에게 몰입감을 주기가 어려우므로 장애물을 피하거나 공중에 떠있는 동전을 획득하는 게임구간을 추가하였다. 그리고 착륙지점에 정확히 도달하면 착륙 애니메이션이 실행되고, 그렇지 못하면 착륙에 실패하는 것으로 본 콘텐츠는 마무리 된다.

본 콘텐츠는 정의된 포즈를 통해 아바타의 동작을 명령하고, 가상공간 내에서 사용자의 동작에 대한 반응 효과를 캐릭터 애니메이션을 통해 보여줌으로써 콘텐츠에 대한 몰입도와 재미를 더했다.

본 논문에서 사용되는 애니메이션의 종류는 그림 [1]과 같다.

3. 포즈 정의

본 논문에서 기술하는 비행 체험 콘텐츠에서 사용되는 포즈는 다음 그림[2]와 같다. 그림[2]에서의 포즈들은 콘텐츠의 시나리오에 따라 시작지점에서 날아오르기 위해 사용되는 제스처와 비행을 유지하는 제스처, 좌·우 이동에 관련된 제스처 등에 사용된다. 콘텐츠에서 사용되는 애니메이션은 그림[1]과 같이 6개이다.

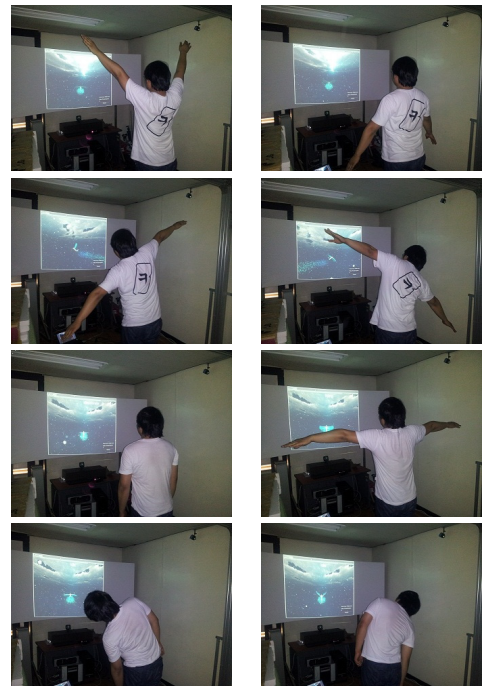


그림 2. 사용되는 포즈 정의

4. 실험 내용

인체의 3차원 포즈를 추정하기 위해서는 입력된 영상에서 배경을 제거하고 인체의 형상을 정확히 분할해야 한다. 하지만 영상에서 인체를 분할하는 것은 복잡하고 어려운 과정이다. 본 논문에서는 포즈의 특징을 정의하고 인식하는 것에 중점을 두고 있기 때문에 정확한 포즈 식별을 위하여 키넥트를 이용한다.

본 논문에서는 키넥트를 이용한 신체 Joint를 획득하고 그 Joint를 이용하여 각 포즈에 대한 개별적인 포즈 템플릿을 구성하였다. Joint들은 회전축을 기준으로 이동하므로 주요관절정보들만 이용하여 포즈 템플릿을 구성할 수 있다.

Index	관절 벡터
V_1	HIP_CENTER -> SHOULDER_CENTER
V_2	SHOULDER_CENTER -> RIGHT_SHOULDER
V_3	RIGHT_SHOULDER -> RIGHT_ELBOW
V_4	RIGHT_ELBOW -> RIGHT_HAND
V_5	SHOULDER_CENTER -> LEFT_SHOULDER
V_6	LEFT_SHOULDER -> LEFT_ELBOW
V_7	LEFT_ELBOW -> LEFT_HAND
V_8	SHOULDER_CENTER -> HEAD
V_9	HIP_CENTER -> RIGHT_HIP
V_{10}	HIP_CENTER -> LEFT_HIP

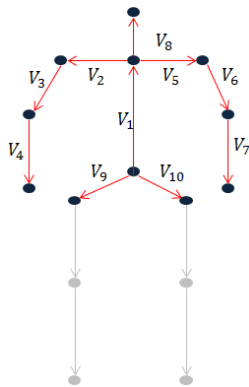


그림 3. 포즈 추정에 사용되는 주요 관절

각 포즈 템플릿은 11개의 관절에 대한 관절 벡터로 구성된다. 즉 하나의 포즈 템플릿은 10개의 벡터 특징을 갖는다.

$$P_i = [V_1, V_2, V_3, \dots, V_{10}] \quad (1)$$

사용자가 정의된 8가지 포즈 중 임의의 포즈를 취하면 사용자의 포즈와 가장 가까운 포즈 템플릿을 계산해서 구한다. 포즈 번호를 구하기 위한 방법은 아래 식으로 표현된다.

$$PoseNum = \operatorname{argmin}(|P_i - X|^2) \quad (2)$$

식(2)에서 P_i 는 미리 정의된 포즈 템플릿을 의미하고 X 는 현재 사용자의 신체 관절 정보를 의미한다.

5. 포즈인식 실험결과

사용자가 비행 상태를 유지하고 있는 동안 가속, 비행유지, 감속, 좌우 회전등의 이벤트가 발생하게 된다. 본 논문에서는 단위시간동안 인식된 포즈들의 위치 차이를 고려하여 동작을 감지하였다. 포즈들의 위치 차이는 식(3)과 같다.

n 은 단위시간, $|P_t - P_{t-1}|$ 은 이전 포즈와 현재 포즈의 위치 차이를 나타낸다.

$$Value_{def} = \sum_{t=1}^n (|P_t - P_{t-1}|) \quad (3)$$

표 1에서는 5명의 사용자들에 대한 포즈를 인식한 결과를 보이고 있다. 본 논문에서 인식을 통해 감지된 포즈들은 실험 결과 8개의 포즈 중 7개의 포즈가 90%이상의 인식률을 보였다. 하지만 포즈를 취할 때 정면 혹은 측면에서 몸이 틀어진 경우 다른 포즈로 인식하거나 인식하지 못하는 오류가 있었다.

표 1. 포즈 인식 실험 결과

포즈번호	인식률	포즈번호	인식률
POSE 1	95%	POSE 5	95%
POSE 2	95%	POSE 6	87%
POSE 3	93%	POSE 7	92%
POSE 4	92%	POSE 8	95%

6. 결론

본 논문에서는 사용자의 포즈를 인터페이스로 사용하는 가상 비행 체험 콘텐츠에 대하여 기술하였다. 사용자는 정의된 포즈를 통해 아바타의 동작을 명령할 수 있고, 가상공간 내에서 사용자의 동작에 대한 반응을 캐릭터 애니메이션을 통해 보여줌으로써 콘텐츠에 대한 몰입도와 재미를 더했다. 그리고 실시간으로 사용자의 포즈에 대한 반응을 보여주고 그래픽 프로그램 모듈이 포함된 상호작용 콘텐츠를 개발하였다.

Acknowledgement

"본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음" (NIPA-2012-H0301-12-3005)

본 연구는 지식경제부 및 정보통신산업진흥원의 지능로봇을 위한 3D 센싱 및 비전 기반 사람/물체 인식 기술 개발 지원사업의 연구결과로 수행되었음 (NIPA-2011-C7000-1001-0007)

"본 연구는 지식경제부 주관 지역SW융합사업의 지원을 받았습니다."

참고문헌

[1] I. Haritaoglu, D. Harwood and L.S. Davis, "Ghost: A human body part labeling system using silhouettes," Proc. of the 14th International Conference on Pattern Recognition, pp.77-82, 1998.

[2] S. Iwasawa, K. Ebihara, J. Ohya, S. Morichima, "Real-Time Estimation of Human Body Posture from Monocular Thermal Images," Proc. of 1997 IEEE Computer Society Conference on computer Vision and Pattern Recognition, pp.15-20, 1997.

[3] D. E. Goldberg, Genetic Algorithm in search, optimization, and machine learning, Addison-Wesley, 1989.

[4] A. R. Micilotta, E. J. Ong and R. Bowden, "Real-time Upper body detection and 3D pose estimation in monoscopic images," Proc. of the 9th European Conference on Computer Vision, LNCS, vol.3953, pp.139-150, 2006.

[5] P. Viola and M. Jones, "Robust real-time object detection," In Proc. of IEEE Workshop on Statistical and Computational Theories of Vision, 2001.

[6] D. Gavrilu and L. davis, "3-D model-based tracking of humans in action: a multi-view approach, in Proc. of IEEE Conference on Computer Vision and Pattern Recognition, san Francisco, 1996, pp. 73-80.

[7] Borgefors, Gunilla, "Hierarchical Chamfer Matching : A Parametric Edge matching Algorithm," IEEE Trans. Pattern Anal. Mach. Intell., vol.6, pp. 849-865, 1998.

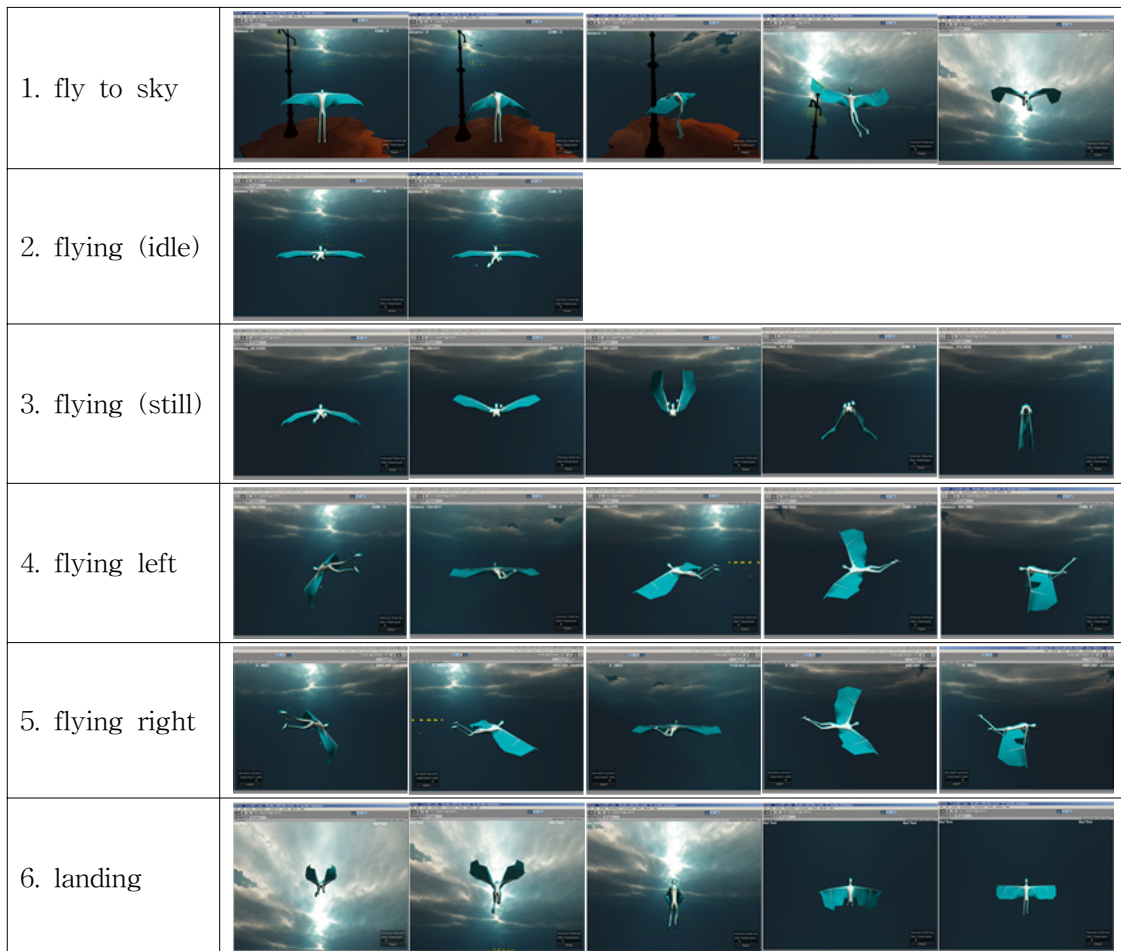


그림 1. 콘텐츠에서 사용된 애니메이션 종류