

비전 기반 신체 제스처 인식을 이용한 상호작용 콘텐츠 인터페이스

(Interface of Interactive Contents using Vision-based Body Gesture Recognition)

박재완*, 송대현*, 이칠우**

(Jae Wan Park, Dae Hyun Song, Chil Woo Lee)

요약

본 논문은 비전 기반 신체 제스처 인식 결과를 입력인터페이스로 사용하는 상호작용 콘텐츠에 대해 기술한다. 제작된 콘텐츠는 아시아의 공통문화요소인 도깨비를 소재로 사용하여 지역 문화에 친숙하게 접근할 수 있도록 하였다. 그리고 콘텐츠를 구성하는 시나리오는 도깨비와의 결투장면에서 사용자의 제스처 인식을 통해 결투를 진행하므로 사용자는 자연스럽게 콘텐츠 시나리오에 몰입할 수 있다. 시나리오의 후반부에서는 사용자는 시간과 공간이 다른 다중의 결말을 선택할 수 있다. 신체 제스처 인식 부분에서는 키넥트(KINECT)를 통해 얻을 수 있는 각 신체 부분의 3차원좌표를 이용하여 정지동작인 포즈를 활용한다. 비전기반 3차원 신체 포즈 인식 기술은 HCI(Human-Computer Interaction)에서 인간의 제스처를 전달하기 위한 방법으로 사용된다. 특수한 환경에서 단순한 2차원 움직임 포즈만 인식할 수 있는 2차원 포즈모델 기반 인식 방법에 비해 3차원 관절을 묘사한 포즈모델은 관절각에 대한 정보와 신체 부위의 모양정보를 선행지식으로 사용할 수 있어서 좀 더 일반적인 환경에서 복잡한 3차원 포즈도 인식할 수 있다는 장점이 있다. 인간이 사용하는 제스처는 정지동작인 포즈들의 연속적인 동작을 통해 표현이 가능하므로 HMM을 이용하여 정지동작 포즈들로 구성된 제스처를 인식하였다. 본 논문에서 기술한 체험형 콘텐츠는 사용자가 부가적인 장치의 사용 없이 제스처 인식 결과를 입력인터페이스로 사용하였으며 사용자의 몸동작만으로 자연스럽게 콘텐츠를 조작할 수 있도록 해준다. 본 논문에서 기술한 체험형 콘텐츠는 평소 접하기 어려운 도깨비를 이용하여 사용자와 실시간 상호작용이 가능케함으로써 몰입도와 재미를 향상시키고자 하였다.

■ 중심어 : | 제스처 인터페이스 | 상호작용 콘텐츠 |

Abstract

In this paper, we describe interactive contents which is used the result of the inputted interface recognizing vision-based body gesture. Because the content uses the imp which is the common culture as the subject in Asia, we can enjoy it with culture familiarity. And also since the player can use their own gesture to fight with the imp in the game, they are naturally absorbed in the game. And the users can choose the multiple endings of the contents in the end of the scenario. In the part of the gesture recognition, KINECT is used to obtain the three-dimensional coordinates of each joint of the limb to capture the static pose of the actions. The vision-based 3D human pose recognition technology is used to method for convey human gesture in HCI(Human-Computer Interaction). 2D pose model based recognition method recognizes simple 2D human pose in particular environment. On the other hand, 3D pose model which describes 3D human body skeletal structure can recognize more complex 3D pose than 2D pose model in because it can use joint angle and shape information of body part. Because gestures can be presented through sequential static poses, we recognize the gestures which are configured poses by using HMM. In this paper, we describe the interactive content which is used as input interface by using gesture recognition result. So, we can control the contents using only user's gestures naturally. And we intended to improve the immersion and the interest by using the imp who is used real-time interaction with user.

■ keywords : | Gesture Interface | Interactive Contents |

I. 서 론

제스처 인터페이스기술(Gesture Interface Technology)은 최근 미국 매사추세츠공과대학(MIT)이 발표한 '2011년 10대 유망

기술'[1] 중 하나로 선정되었다. 현재 제스처 인터페이스 기술은 가상현실, 스마트 감시 시스템, 항공 교통 모니터링 시스템, 게임, 스포츠와 같은 엔터테인먼트 등 여러 분야에서 응용되고 있다. 더욱이 최첨단의 제작기술과 표현기술의 발전은 디지털 콘텐츠

* 학생회원, 전남대학교 전자컴퓨터공학과

** 종신회원, 전남대학교 전자컴퓨터공학부 교수

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-2010-C7000-1001-0007)

접수번호 : #2012-05-29-0023

접수일자 : 2012년 05월 29일

심사완료일 : 2012년 06월 20일

교신저자 : 이칠우 e-mail : leecw@jnu.ac.kr

시장을 급속도로 성장시켜 사람과 컴퓨터간의 효과적인 상호작용을 위한 더욱 편리하고 자연스러운 사용자 인터페이스 (User Interface)를 요구하고 있다. 따라서 입력 장치를 사용하지 않고 컴퓨터와 상호작용하기 위한 인간-컴퓨터 상호작용 기술의 필요성은 점차 증대되고 있으며 특히 제스처 인식 기술과 멀티모달 인터페이스에 대한 연구의 필요성이 함께 높아지고 있다.

본 논문은 신체 제스처 인식을 입력인터페이스로 사용하는 콘텐츠에 대해서 기술한다. 논문에서 제안하는 콘텐츠는 프로젝션 컴퓨터를 통해 대형 스크린에 투사되고, 사용자는 대형 스크린 앞에서 제스처를 사용함으로써 자연스런 상호작용이 가능하다. 제작된 콘텐츠는 아시아 공통문화요소인 도깨비를 소재로 사용하고 있으며 사용자는 콘텐츠를 체험함으로써 아시아 및 지역 문화에 대한 이해와 함께 콘텐츠를 즐길수 있게 한다.

콘텐츠에 등장하는 도깨비는 한국, 중국(인도), 일본의 도깨비를 참고하여 디자인하였다. 한국의 도깨비는 사람을 홀리거나 장난을 치기도 하지만 특별한 악의는 없으며, 때로는 자신에게 먹을 것을 대접한 사람에게 큰 돈을 가져다주어 부자로 만들어 주기도 하는 매우 순박한 존재이다. 일본의 도깨비는 ‘오니’라 불려지며, 원시인 복장에 이마에 뿔이 달려 있으며, 손에는 뜯어 박힌 철퇴를 들고 있는 모습을 하고 있다. 일본의 도깨비 오니는 한국의 도깨비와는 다르게 굉장히 무서우며, 악령의 속성을 지니고 있다. 중국(인도)의 도깨비는 ‘귀’, ‘야차’라 불려지며, 야차는 또한 야차팔대장, 십육대야차장으로 불리며, 본래 놀라게 한다는 뜻으로 인도의 ‘베다’에서 나오는 신적 존재였지만, 불교에서는 불교를 지키는 신으로 묘사되고 있다[2].

사용자의 제스처 인식을 이용한 입력인터페이스는 콘텐츠를 체험하기 위해서 신체 이외의 디바이스를 사용하지 않아도 될 뿐만 아니라 콘텐츠의 일부로서의 캐릭터에 사용자의 동작을 매핑하여 사용자에게 자연스러운 몰입감을 부여하는 장점이 있다. 그리고 스테이지를 공략하기 위해서 비전 기반 제스처 인식 시스템을 적극 활용하여 별도의 디바이스를 요구하지 않는 신체의 동작을 통해 도깨비와 결투 장면을 구현하였다. 그리하여 사용자는 양 손과 몸의 움직임으로 이루어진 신체 제스처를 이용하여 콘텐츠를 진행할 수 있으며 콘텐츠에 몰입할 수 있는 체험효과를 받을 수 있다.

제작된 콘텐츠에서 미리 정의된 제스처는 사용자가 콘텐츠에서 요구하는 제스처를 사전에 미리 숙지하고 있는 경우, 제스처에 의미를 부여하기 위해 사용자가 용도에 맞춰 사용이 가능하다.

II. 전체 시스템 구조

본 논문에서 기술하는 콘텐츠를 구동하기 위한 하드웨어 시스템은 다음 그림1과 같다.

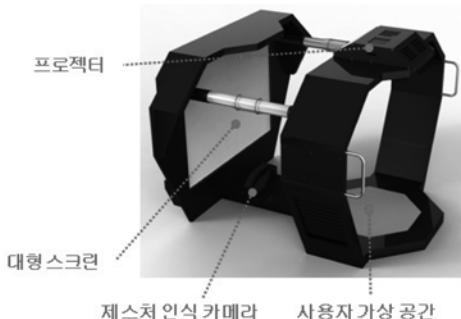


그림 1. 전체 하드웨어 시스템 구조

사용자의 제스처를 입력받을 수 있는 사용자 가상공간, 제스처를 인식하는 KINECT카메라와 콘텐츠를 체험할 수 있는 프로젝터, 대형스크린으로 구성되어 있다. 그리고 소프트웨어 시스템으로는 제스처 인식 모듈과 체험형 콘텐츠를 구동할 수 있는 데스크톱 PC를 추가로 가지고 있다.

일단, 사용자가 가상공간에 입장하게 되면 시스템에서 사용자를 감지하여 콘텐츠가 실행된다. 그리고 콘텐츠를 진행하기 위해 요구되는 이벤트는 제스처 인식을 이용한 제스처 입력인터페이스로 대체된다.

III. 콘텐츠 구현

제작된 콘텐츠는 크게 세 부분으로 나뉜다. 첫 번째 부분은 사용자가 가상공간으로 진입하였을 때, 사용자를 대신할 캐릭터를 선택하는 장면이 실행된다. 두 번째 부분은 세 번째 장면으로 이동하는 관문을 지키는 외다리 도깨비와의 결투장면이 실행된다. 세 번째 부분은 시·공간을 이동할 수 있는 혼천의를 조작하는 장면이다.

사용자를 인지하면 다음 그림2에서와 같이 첫 번째 장면이 실행된다. 사용자는 양 손을 앞으로 움직여 탈을 집어드는 제스처를 사용하여 아기도깨비 탈을 선택할 수 있다. 사용자는 도깨비감투를 머리에 착용함으로써 콘텐츠에서 사용자와 캐릭터의 동작을 매칭하는 이벤트를 실행시킨다.



그림 2. 사용자가 하드웨어 시스템에 입장했을 때 발생되는 도깨비감투 선택 장면 (한 손을 이용하여 원하는 도깨비감투를 이동시킬 수 있고, 사용자가 양손을 앞으로 뻗는 제스처에 맞춰 도깨비감투가 선택됨)

이 장면에서는 사용자가 양 손을 앞으로 뻗는 제스처를 실행하면 도깨비감투가 머리에 착용되면서 콘텐츠는 두 번째 장면으로 진행된다. 두 번째 장면은 혼천의를 지키는 관문인 외다리도깨비와의 결투이다.



그림 3. 혼천의를 지키는 외다리 도깨비와의 결투 장면 (손과 발을 이용한 공격)

사용자가 외다리도깨비를 공격할 때, 그림3과 같이 근접한 외다리도깨비에게 손과 발을 이용한 제스처를 사용하고, 외다리도깨비가 사용자를 공격할 때, 사용자는 좌·우 회피동작으로 이루어진 제스처를 통하여 공격을 무마시킬 수 있다.



그림 4. 시·공간을 이동할 수 있는 혼천의를 양손 제스처를 사용하여 조작하는 장면 (사용자는 혼천의를 조작하여 원하는 지역과 시간을 체험할 수 있음)

사용자가 관문을 지키는 외다리도깨비를 물리치게 되면, 콘텐츠의 장면은 그림4와 같이 시·공간을 이동할 수 있는 혼천의로 이동된다. 사용자는 혼천의를 돌리는 제스처를 통하여 시·공간을 선택하고, 시·공간을 선택한 후, 줌인(Zoom-In) 제스처를 통하여 선택된 시·공간으로 이동이 가능하다.

IV. 제스처 인식

1. 관절각의 추정

인체의 3차원 포즈를 추정하기 위해서는 입력된 영상에서 배경을 제거하고 인체의 형상을 정확히 분할해야 한다. 하지만 영상에서 인체를 분할하는 것은 매우 복잡하고 어려운 과정이다. 본 논문에서는 포즈의 특징을 정의하고 인식하는 것에 중점을 두고 있기 때문에 인체 형상을 분할하는 문제를 간단히 하기 위해 Microsoft사의 Kinect[9] 카메라와 NUI Skeleton API를 사용해 인체 영역을 분할하고 인체의 관절각 정보를 획득하였다. NUI Skeleton API는 Kinect 센서 앞에 있는 사람을 인식해 신체의 주요 20개 부위에 해당하는 관절 좌표와 회전 정보를 제공

해 준다.(표 1) 인체의 각 관절은 (x, y, z)의 3차원 좌표 값을 갖는다.

표 1. Kinect에서 제공하는 20개 신체 관절

번호	관절 이름	번호	관절 이름
1	HEAD	11	SPINE
2	SHOULDER_CENTER	12	HIP_CENTER
3	SHOULDER_RIGHT	13	HIP_RIGHT
4	SHOULDER_LEFT	14	HIP_LEFT
5	ELBOW_RIGHT	15	KNEE_RIGHT
6	ELBOW_LEFT	16	KNEE_LEFT
7	WRIST_RIGHT	17	ANKLE_RIGHT
8	WRIST_LEFT	18	ANKLE_LEFT
9	HAND_RIGHT	19	FOOT_RIGHT
10	HAND_LEFT	20	FOOT_LEFT

NUI Skeleton API에서는 다음과 같은 과정을 통해 인체의 3차원 관절 정보를 얻는다. 먼저 Kinect의 적외선 카메라로부터 영상이 들어오면 깊이 영상이 생성되고 깊이 정보를 이용해 사람과 배경을 분리한다. 배경과 분리된 인체 영역은 학습된 데이터와 깊이 영상의 특징에 의해, 관절의 관심 영역에 위치한다고 추정되는 신체의 각 부분으로 나뉘어 레이블링 된다. 레이블링 된 신체의 각 부분은 다시 3차원 공간으로 재 투영된다. 각각의 신체 부분의 공간적인 분포에 따라 신뢰 가중치를 부여한 후 각 관절의 3차원 위치가 결정된다[10].

2. 특징 추출

본 논문에서는 총 28개의 포즈를 정의하였으며, 각 포즈별로 개별적인 포즈 템플릿을 구성해 포즈를 인식하였다.

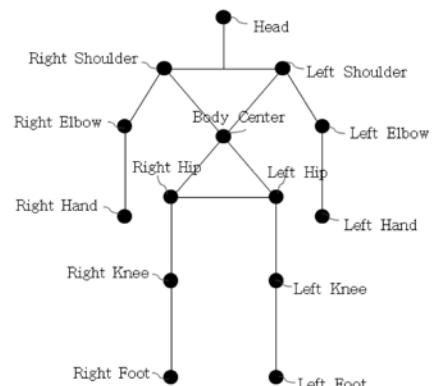


그림 5. 실험에 사용한 14개의 신체 관절

각 포즈 템플릿은 14개의 관절에 대한 관절 벡터의 모음으로 정의할 수 있다. 각 포즈에 대한 특징을 추출하기 위해 NUI Skeleton API에서 제공하는 20개의 인체 관절 중 왼쪽 손목, 오

른쪽 손목, 왼쪽 발꿈치, 오른쪽 발꿈치, 어깨 중심점 그리고 엉덩이 중심점을 제외하고 14개 관절(머리, 오른쪽 어깨, 왼쪽 어깨, 오른쪽 팔꿈치, 왼쪽 팔꿈치, 오른 손, 왼 손, 몸 중심, 오른쪽 엉덩이, 왼쪽 엉덩이, 오른쪽 무릎, 왼쪽 무릎, 오른 발, 왼 발)을 사용했다. 제외시킨 6개의 관절은 포즈가 달라져도 큰 변화를 갖지 않거나 손, 발과 같은 관절과 매우 가까워 해당 관절 값의 사용 유무가 포즈를 정의함에 있어 큰 의미를 갖지 않는 관절이다. 획득한 14개 신체 관절(그림 5)의 3차원 좌표(x,y,z)를 얻은 후 14개 신체 관절의 3차원 위치 좌표 정보를 바탕으로 각 포즈에 대한 개별적인 포즈 템플릿을 구성하였다. 각 포즈 템플릿은 14개 관절에 대한 관절 벡터로 구성된다. 즉, 하나의 포즈 템플릿은 14개의 벡터 특징을 갖는다.

$$P_i = [V_1, V_2, V_3, \dots, V_{14}] \quad (1)$$

식(1)의 P_i (i=1 to 14)는 각 포즈에 해당하는 포즈 템플릿을 의미하고 V_i 는 관절 벡터를 의미한다.

3. 포즈 인식 방법

앞 절에서 설명한 것과 같이 28가지 유형의 포즈를 인식하기 위해서 각 포즈에 해당하는 템플릿을 구성해야 한다. 제안하는 포즈 시스템은 사용자의 포즈를 각각 인식할 수 있게 구성하였다. 이 시스템은 사용자가 들어오면 칼리브레이션 및 배경 제거 후 각 사용자에 대한 3차원 좌표 정보를 획득한다. 사용자가 정의된 28가지 포즈 중 임의의 포즈를 취하면 사용자의 포즈와 가장 가까운 포즈 템플릿을 계산해서 구한다. 포즈 번호를 구하기 위한 방법은 아래 식(2)로 표현한다.

$$\text{PoseNum} = \arg \min (|P_i - X|^2) \quad (2)$$

식(1)에서 P_i 는 미리 정의된 포즈 템플릿을 의미하고 X 는 현재 사용자의 신체 관절 정보를 의미한다. 포즈는 사용자에 따라 약간의 오차가 발생 할 가능성이 있기 때문에 템플릿을 구성하는 관절 관계의 특징에 대해 5%의 오차 범위를 두고 포즈를 인식했다.

4. 제스처 인식 방법

인간이 사용하는 제스처는 정지동작인 포즈들의 연속적인 동작을 통해 표현이 가능하다.

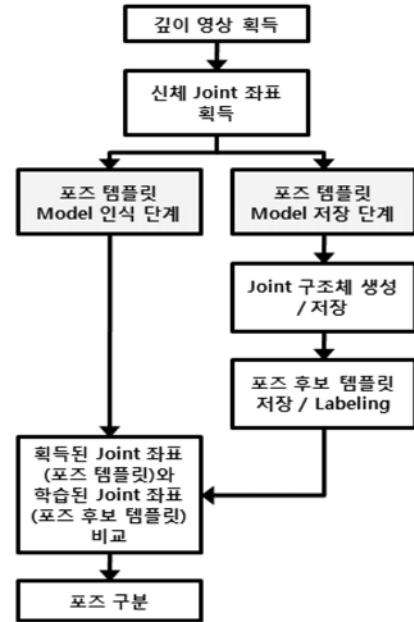


그림 6. 포즈 인식 과정

그리므로 포즈 인식 과정에서는 키넥트(KINECT)를 통해 각 신체 관절 부분의 3차원 좌표를 얻는다. 그리고 획득한 신체 관절에 대한 3차원 좌표는 그림5와 같이 정지동작인 포즈 템플릿으로 구성된다. 그 후, 미리 학습된 포즈 템플릿과 획득된 포즈 템플릿을 비교하여 제스처 인식에 사용될 포즈를 구분한다. 정지동작인 포즈들로 구성된 제스처는 시계열 정보에 대해 효과적으로 인식이 가능한 HMM을 사용하였다.

제스처 인식 단계에서는 인식된 포즈 영상들로부터 제스처를 인식한다. 본 논문에서는 제스처 인식을 위해 HMM을 사용하였다. HMM은 시간적으로 제약을 받는 정보의 구조를 모델링하는데 뛰어난 모델이다. Simple Markov Model만으로 모델링하기 힘든 실세계의 문제를 통계적 매개변수로 접근할 수 있게 해준다. 상태 전이 매개 변수는 순차적인 일련의 사건 발생을 모델링한다. 그리고 관측 심볼 확률 분포는 각 사건의 특징을 유한 개의 심볼로 대응시킨다.

HMM은 이러한 두 가지 확률 과정의 결합으로 이루어져 있고, 이 기준에 따라 생성된다. 생성된 HMM은 학습 데이터를 이용한 학습을 통해 적절한 제스처 모델을 구성한다.

인식과정에서는 인식하고자 하는 제스처와 학습이 끝난 후 생성된 모든 HMM의 제스처 모델을 비교한다. 그리하여, 가장 유사하다고 판단되는 제스처 모델을 선택하고 결과를 확률로 나타낸다. 상태전이확률을 이용하여 각 샘플들을 이용한 전처리 단계는 잡음제거와 대표점 추출 과정을 거쳐 궤적을 체인코드로 변환하는 과정으로 각 포즈의 번호를 코드로 사용한다.

학습은 각 제스처별로 이루어지며, 해당 제스처의 HMM 모델에 학습결과를 적용한다. HMM의 학습 과정은 각 숫자별로 신체

포즈를 이용하여 은닉 마르코프 모델을 구성하는 과정으로, EM 알고리즘의 하나인 Baum-Welch 알고리즘을 이용한다. HMM의 인식 과정은 전처리 모듈을 통해 변환된 체인코드가 어느 숫자의 은닉 마르코프 모델에서 나타날 확률이 높은지를 판단하는 과정으로, 각각의 숫자 모델에 전향(Forward) 알고리즘을 적용하여 가장 높은 확률을 보이는 숫자 모델을 최종 인식 결과로 출력한다. HMM은 아래와 같은 요소로 구성된다.

- N : 상태의 수,
 $S = \{S_1, S_2, \dots, S_N\}$: 상태의 집합,
 q_t : 시간 t 의 상태
- M : 관측 심볼의 수,
 $V = \{v_1, v_2, \dots, v_M\}$: 관측 심볼의 집합
- $A = \{a_{ij}\}$: 상태 전이 확률 분포
 $a_{ij} = P(q_{t+1} = S_j | q_t = S_i), 1 \leq i, j \leq N$
: 상태 i 에서 상태 j 로 전이할 확률
- $B = \{b_j(k)\}$: 관측 심볼 확률 분포,
 $b_j(k) = P(v_k | q_t = S_j), 1 \leq j \leq N, 1 \leq k \leq M$
: j 에서 심볼 v_k 를 관측할 확률
- $\pi = \{\pi_i\}$: 초기 상태 확률 분포,
 $\pi_i = P(q_1 = S_i), 1 \leq i \leq N$: 초기 상태가 i 일 확률

HMM을 이용한 제스처 인식방법은 다음 그림7과 같다. 학습을 위해 미리 정의 해놓은 포즈를 이용하여 상태천이 확률변수를 생성한 후, 인식할 포즈의 상태천이 확률변수를 이용하여 제스처를 인식한다.

제스처를 구성하는 포즈의 개수가 N 이라고 할 때, HMM의 학습과정에서는 N 개의 포즈들에 대한 체인코드를 생성하고, 제스처를 구성하는 N 개의 포즈들의 순차적인 정보를 이용하여 $N+N(N-1)/2$ 개의 상태천이 확률변수를 생성한다.

예를 들어 제스처를 구성하는 포즈가 4개일 경우 10개의 유효한 상태천이 확률변수를 생성하게 된다. 제스처 인식과정에서는 학습된 제스처를 구성하고 있는 포즈들의 상태천이 확률변수와 순차적 포즈로 이루어진 제스처 프레임의 상태천이 확률변수를 이용하여 제스처를 인식한다.

각 제스처는 시작포즈를 가지고 있고 제스처에 해당하는 상태천이 확률이 무효할 경우에는 “IDLE” 상태를 가지고 있으며, 상태천이 확률이 유효한 시점부터 각 제스처에 해당하는 포즈를 전부 만족할 경우 제스처가 종료된 것으로 판단한다. Left-Right HMM 구조 모델은 포즈 사이에 잡음이 생겨도 미리 학습한 상태천이 확률이 유효하므로 실시간 인터페이스로 사용하기에 무리가 없다.

그림8은 HMM을 사용하여 최대 유사값을 가지는 제스처를 인식하는 과정을 나타낸다.

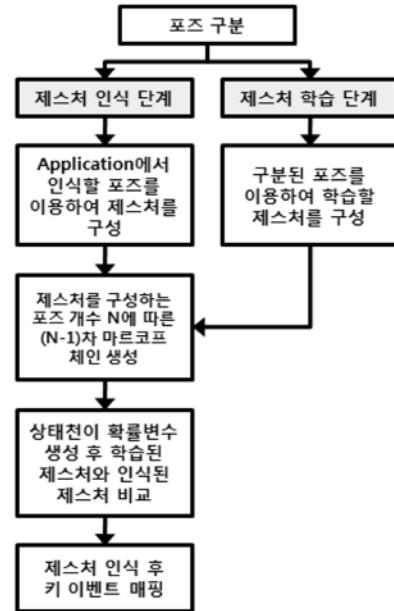


그림 7. 제스처 인식 과정

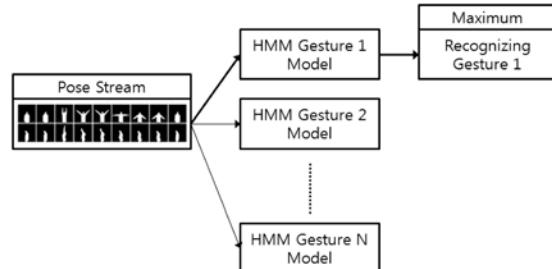


그림 8 최대 유사값을 가지는 제스처 인식 과정

V. 콘텐츠 실험 결과

제작된 콘텐츠에서는 캐릭터의 행동을 조작하기 위해 사용자의 동작을 다음과 같이 인식한다. 사용자의 왼팔을 인식하여 캐릭터의 왼손을 움직이고, 사용자의 오른팔을 인식하여 캐릭터의 오른손을 움직인다.

이러한 기본동작들로 구성된 제스처를 사용하여 제작된 콘텐츠에서는 세 개의 장면에 따라 그림7, 8, 9와 같이 입력인터페이스를 위한 제스처를 정의하였다.

도입장면에서는 도깨비감투를 선택하기 위해 한 손을 좌, 우로 이동시키는 제스처와 도깨비감투를 착용하는 제스처를 사용한다. 도깨비 결투장면에서는 손과 발을 이용한 공격제스처와 특수 아이템을 획득하는 제스처를 사용한다. 그리고 혼천의를 조작하는 장면에서는 한손을 좌·우로 돌리거나 양손을 시계·반시계방향으로 돌리거나 줌·인·아웃 제스처를 사용한다.

본 논문에서는 총 13개의 제스처를 정의하였고 제스처에 필요한 포즈의 개수로 키포즈와 주변포즈를 포함하여 각 제스처마다 3~6개의 포즈로 구성하였다.

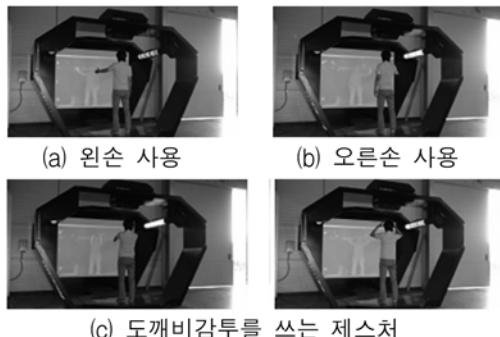


그림 9. 도입장면에서 사용하는 제스처

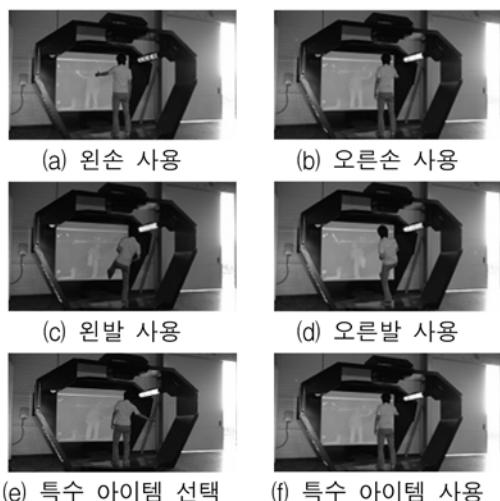


그림 10. 도깨비 결투 장면에서 사용하는 제스처

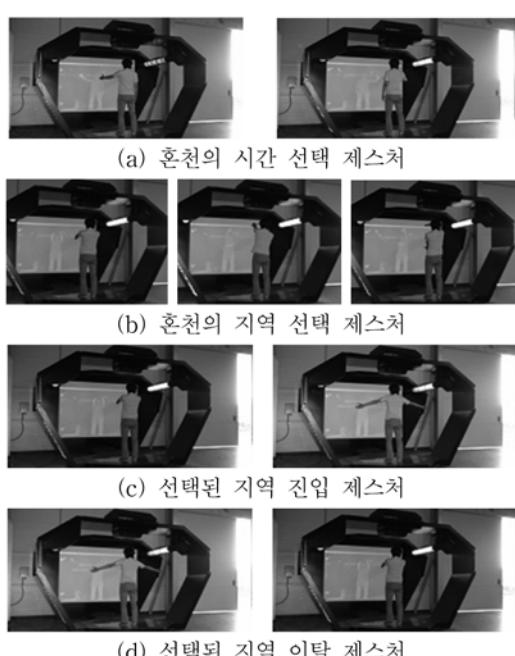


그림 11. 혼천의 장면에서 사용하는 제스처

표 2. 제스처를 구성하는 포즈의 개수에 따른 인식률

구성 포즈 개수 제스처 종류	3 pose	4 pose	5 pose	6 pose
gesture 1	92%	88%	78%	78%
gesture 2	87%	86%	66%	82%
gesture 3	74%	84%	78%	74%
gesture 4	82%	84%	64%	34%
gesture 5	74%	78%	58%	42%
gesture 6	92%	94%	84%	78%
gesture 7	94%	96%	86%	82%
gesture 8	66%	62%	32%	14%
gesture 9	58%	68%	42%	34%
gesture 10	80%	82%	74%	74%
gesture 11	78%	78%	76%	64%
gesture 12	88%	80%	80%	68%
gesture 13	90%	80%	78%	72%

제스처를 구성하는 포즈 영상의 수에 따른 제스처 인식률은 표 2과 같다. 동일한 제스처라고 할지라도 제스처에 포함되는 포즈의 개수가 많을수록 인식률이 낮아지는 결과를 보였다. 그러므로 각 제스처에 필요한 적절한 포즈들을 이용하여 학습하는 과정이 필요하고, 가능하다면 구분이 가능한 포즈들을 되도록 많이 정의하여 포즈가 겹치지 않아야 한다는 것을 알 수 있었다.

VI. 콘텐츠 실험 결과

본 논문에서 기술한 콘텐츠는 양 손의 동작을 기반으로 구성된 신체 제스처 인식을 입력인터페이스로 사용하였다. 그리고 사용자가 제작된 콘텐츠를 체험하는데 있어서 콘텐츠 속의 캐릭터와 사용자의 행동을 매핑하여 몰입도와 재미를 느낄 수 있도록 하였다.

하지만 제작된 콘텐츠를 효과적으로 체험하기 위해서는 사용자의 제스처를 입력받을 수 있는 하드웨어 시스템을 갖추고 있어야 한다. 본 논문에서는 제스처 인식을 위해 제스처를 구성하는 포즈를 인식하는 방법을 사용하였다. 인체에서 특징이 되는 14개 관절의 3차원 좌표 값을 이용해 각 포즈에 대한 포즈 템플릿을 구성하고, 현재 포즈와 포즈 템플릿의 거리를 비교해 포즈를 인식했다. 그리고 3차원 좌표 값만을 이용해 포즈를 인식했기 때문에 비교적 간단하지만 높은 인식률로 정의한 포즈를 인식할 수 있었다. 이렇게 인식된 포즈들을 이용하여 제스처를 구성하고 인식된 제스처를 콘텐츠의 인터페이스로 사용하였다.

본 콘텐츠에서는 요구하는 제스처를 사용하기 위해서 사용자가 미리 제스처를 학습하여 이벤트를 발생시킬 수 있지만, 미리 학습

되지 않은 제스처나 정의되어지지 않은 제스처는 사용하지 못하는 단점이 있다. 그리고 일반적인 콘텐츠의 성격이나 체험방식에 따라 요구되는 제스처의 종류가 모두 같지 않다. 즉, 제작된 콘텐츠에서 정의하지 않은 다양한 제스처를 사용하는 다른 콘텐츠의 목적에 적합한 제스처 인식 모듈을 제작하기에 어려움이 있다. 그러므로 앞으로는 제스처 인식 기술을 해당 콘텐츠에 최적화 시킬 뿐만 아니라 학습과 인식과정을 개선하여 일반적인 제스처에서도 쉽게 반응하는 제스처 인식 모듈을 제작하여 제스처 인식 결과 개선 및 인식 시간을 단축하여야 할 것이다.

참 고 문 헌

- [1] MIT, "10 Emerging technologies 2011," *Technology Review*, May, 2011.
- [2] <http://www.culturecontent.com>
- [3] George Caridakis and Kostas Karpuzis, "SOMM: Self organizing Markov map for gesture recognition," *Pattern Recognition Letters* 31, pp.52-59, 2010.
- [4] K. Morrison and S.J. McKenna, "An Experimental Comparison of Trajectory-based and History-based Representation for Gesture Recognition," *5th International Workshop on Gesture and Sign Language based Human-Computer Interaction*, Genoa, Italy, pp.152-163, 2003.
- [5] 박아연, 이성환, "이산 은닉 마르코프 모델에 기반한 3차원 전신 제스처 인식", 한국정보과학회 컴퓨터 비전 및 패턴인식연구회 추계 워크샵 발표 논문집, pp.154-156, 2004.
- [6] Antonios Oikonomopoulos, Ioannis Patras, Maja Pantic and Nikos Paragios, "Trajectory-Based Representation of Human Actions," *Human Computing*, LNAI 4451, pp. 133 - 154, 2007.
- [7] L. R. Rabiner, "A tutorial on hidden Markov Models and selected applications in speech recognition," *Proceedings of the IEEE*, 77(2):257-286, 1989.
- [8] A. Bobick and J. Davis, "The Recognition of Human Movement using Temporal Templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(3):257-267, March 2001.
- [9] Microsoft Corp. Redmond WA. Kinect for Xbox360
- [10] J. Shotton, A. Fitzgibbon, M. Cook, M. Finocchio, R. Moore, A. Kipman and a. Blake, "Real-Time Human Pose recognition in Parts from single depth Images," *CVPR*, 2011.
- [11] D. M. Gavrila and L.S. davis, "3-D model-based tracking of humans in action: a multi-view

approach," *Proc. of IEEE Computer Vision and Pattern Recognition*, 2006.

- [12] I. Cohen and H. Li, "Posture and gesture Recognition using 3D body Shapes Decomposition," *Proc. of IEEE Computer Society Conference on Computer vision and Pattern Recognition*, 2005.

저 자 소 개

박 재 완



2007 호남대학교 정보통신공학과 학사
2009 전남대학교 전자컴퓨터공학과 석사
2009~현재 전남대학교 전자컴퓨터공학과 박사과정
<주관심분야 : HCI, 테이블탑 디스플레이, 게임 제스처 인터페이스>

송 대 현



2007 전주대학교 멀티미디어공학과 학사
2009 전남대학교 전자컴퓨터공학과 석사
2009~현재 전남대학교 전자컴퓨터공학과 박사과정
<주관심분야 : 전통문화 콘텐츠, 멀티미디어, 증강현실>

이 칠 우



1992 동경대학교 전자공학과 박사
1992~1995 일본 이미지정보과학연구소 수석연구원
2002~2003 미국 NC A&T State University 방문교수
1996~현재 전남대학교 전자컴퓨터공학과 교수
<주관심분야 : 컴퓨터 비전, 지능형 휴먼 인터페이스, 디지털 콘텐츠, 컴퓨터그래픽스>