

# 게임을 위한 제스처 기반의 인터페이스

박혜선<sup>o</sup>, 강현<sup>\*</sup>, 정기철<sup>\*\*</sup>, 김은이<sup>\*\*\*</sup>, 박민호<sup>\*\*\*\*</sup>, 김향준<sup>\*</sup>  
<sup>\*</sup>경북대학교 컴퓨터공학과 인공지능연구실, <sup>\*\*</sup>승실대학교 승실대 미디어 학부,  
<sup>\*\*\*</sup>건국대학교 멀티미디어 공학부, <sup>\*\*\*\*</sup>경북대학교 전산정보원  
 {hspark, hkang, khkim}@ailab.knu.ac.kr, kcjung@comp.ssu.ac.kr, eykim@konkuk.ac.kr, mhpark@knu.ac.kr

## A Gesture-based interface for interactive computer games

Hye Sun Park<sup>o</sup>, Hyun Kang<sup>\*</sup>, Keechul Jung<sup>\*\*</sup>, Eun Yi Kim<sup>\*\*\*</sup>, Min Ho Park<sup>\*\*\*\*</sup>, and Hang Joon Kim<sup>\*</sup>

<sup>\*</sup>Dept. of Computer Engineering, Kyungpook National Univ., Daegu, South Korea

<sup>\*\*</sup>School of Media, College of Information Science, Soongsil Univ., Seoul, South Korea

<sup>\*\*\*</sup>School of Internet and Multimedia, Konkuk Univ., Seoul, South Korea

<sup>\*\*\*\*</sup>Computer Center, Kyungpook National University

### 요 약

본 논문에서는 사용자의 연속적인 제스처들을 실시간으로 제스처를 적출하고 인식하는 게임 인터페이스를 제안한다. 제안된 인터페이스는 동영상에서 사용자의 자세를 예측하는 포즈 추정 모듈과 연속된 포즈 심벌열로부터 제스처를 적출하고 인식하는 제스처 인식 모듈로 구성되어 있다. 사용자의 자세는 영상에 나타난 머리, 양손의 좌표를 취하여 대표되는 자세를 클러스터링을 통해 구하였다. 제스처를 적출하기 위해서 연속된 포즈 심벌열로부터 가제스처들을 생성한다. 생성된 가제스처의 인식엔진의 인식값을 기준으로하여 제스처를 판별한다. 제안된 논문에서 정의한 제스처를 인터페이스로 사용하여 실제 게임인 퀘이크II에 적용해 봄으로써, 컴퓨터 게임 안에서, 제안된 인터페이스가 타당함을 증명하였다.

### 1. 서론

게임을 할 때, 사용자가 장치의 구속 없이 단지 자신의 몸짓이나 손, 얼굴 등을 이용한 의사 표현으로 게임을 진행할 수 있다면, 사용자는 게임을 좀 더 직접적으로 보다 쉽게, 접근할 수 있을 것이다. 특히 퀘이크II같은 1인칭 액션게임의 경우, 이런 제스처 인터페이스는 사실감을 더해줄 것이다. 제스처 인터페이스를 성공적으로 구현하기 위해서는, 사용자의 의미없는 몸짓들로부터 제스처를 적출하여 인식하는 방법이 제공되어야 한다. 본 논문에서는, 입력 영상에서 제스처의 시작점과 끝점을 찾는 것을 제스처 적출이라고 정의한다.

제스처 적출 방법은 시공간상에서 발생하는 변이문제 뿐만 아니라 연속적으로 입력들어오는 제스처에 대한 분절(co-articulation) 문제 때문에, 어려운 작업으로 알려져 있다. 제스처 적출의 이전 연구를 보면, Iba[2]은 수정된 은닉 마르코프 모델을 사용하여 제스처 기반의 로봇 컨트롤 시스템을 구현하였고, 이현규의 연구[3]에서는 제스처의 유무를 가리는 입계치 유사도를 계산하는 모델을 가지고 제스처를 적출하는 시스템을 구현하였다. Nishimura[4]은 또한 continuous dynamic programming(CDP)을 이용하여 제스처 적출 방법을 제안하였다.

제안된 논문에서는 이전의 연구들과는 달리, 입력들어온 제스처들을 적출하는 동시에 인식하는 제스처 적출 방법을 제안한다. 입력영상으로부터 제스처가 될만한 부분을 모두 생성한다. 이를 가제스처라 부르며, 이 가제스

처의 인식값을 구하여 이를 기준으로 제스처를 비제스처들로부터 판별해 낸다. 이를 퀘이크II의 인터페이스로 적용하여 보았다. 그림 1은 시스템에 사용된 제스처들을 나타낸다. 7개의 제스처이며 2~5개의 포즈들로 구성되어 있다. 그림 2는 인터페이스의 전체 구성도이다.

다음 2장에서 포즈 추정 모듈에 관한 내용이 기술되며, 3장에서 제스처의 적출과 인식, 4장에서 실험결과가 기술된다. 그리고 5장에서 결론을 맺는다.

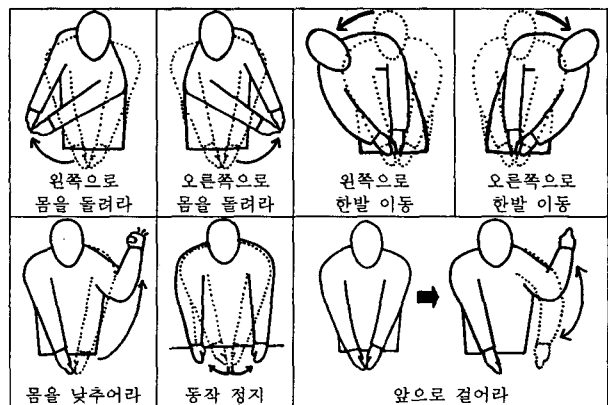


그림 1. 7개의 정의된 사용자 제스처 명령어

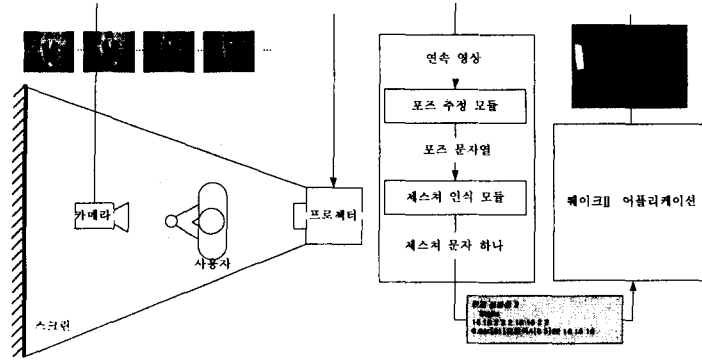


그림 2. 제안된 인터페이스 구조

2. 포즈 추정 모듈

사용자의 자세를 표현하기 위해서, 사용자의 신체 일부분의 위치를 검출한다. 이들의 대표적인 자세를 구하기 위해서 클러스터링을 한다. 포즈 추정 모듈은 영상 처리 서브 모듈과 포즈 분류 서브 모듈로 구성된다. 영상 처리 서브 모듈에서는 스킨 칼라 모델, 모폴로지 연산, 연결 성분 분석, 휴리스틱등의 영상 처리 기술들을 활용하여 각 프레임 안에서 신체 부분들의 위치를 검출한다. 사용된 휴리스틱은 다음과 같다. 1) 상자 크기가 임계값보다 커야한다. 2) 검출된 상자 중에, 가장 큰 상자가 머리부분이다. 3) 영상에서 오른쪽에 검출된 상자는 오른손, 왼쪽에 검출된 상자는 왼손이다. 4) 영상에서 상자가 하나만 존재할 때는, 머리와 양손이 겹쳐진 상태이다. 5)상자가 두 개 존재할 때는, 위의 상자는 머리이고 아래 상자는 양손이 겹쳐진 상태이다. 포즈 분류 서브 모듈에서는 k-means 클러스터링 기술을 사용하여, 포즈를 추정한다. 제안된 논문에서 k는 32로 결정되고 유클리디안 거리를 이용하여, 검출된 신체 부위들의 절대적 위치가 결정된다[5]. 그림 3은 포즈 추정 모듈 단계의 영상 처리 서브 모듈 결과 그림이다. 그림 3-라 그림에서 검정 상자는 머리, 빨강 상자는 왼손, 연두 상자는 오른손이다.

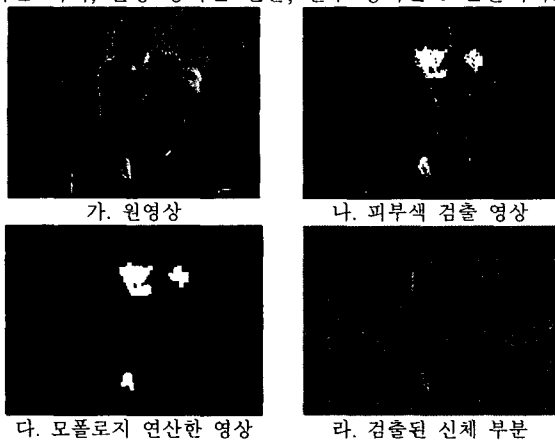


그림 3. 포즈 추정 서브 모듈 단계 결과 영상들

3. 제스처 인식 모듈

본 논문에서는 적출과 인식을 동시에 수행하는 제스처 적출 방법을 제안한다. 제스처 인식 모듈은 다시 두 개의 서브 모듈을 가진다: 제스처 적출 모듈과 제스처 인식 모듈. 제스처 적출은 두단계로 이루어 지는데 후보 컷을 검출하고 심볼열과 특징 벡터 열로부터 가제스처를 생성하는 후보 컷 생성단계와 그때 가제스처의 인식값을 사용하여 정의된 제스처와 분별하는 제스처 판별 단계로 이루어진다. 후보 컷 생성 단계에서는 제스처의 속력, 정적 제스처, 그리고 제스처 궤적의 곡률과 같은 문맥 정보를 사용하여 후보 컷을 검출한다.

가제스처들은 연속 포즈열에서 후보 컷들로 나누어진 부분들에 의해서 생성된다. 즉 가제스처는 생성된 후보 컷들로부터 제스처의 시작점과 끝점이 될 수 있는 포즈열의 부분이다. 검색의 범위를 정하기 위해서  $n \times n$  배열로 구성된 슬라이딩 윈도우를 사용한다. 슬라이딩 윈도우의 크기는 실험적으로 구해진다. 윈도우 안에서 모든 가제스처들에 대한 인식값이 계산될 때, 적출기는 가장 높은 인식 값을 가진 가제스처를 현재 입력된 사용자의 제스처로 결정한다. 슬라이딩 윈도우 안에서 의미 없는 제스처들만으로 포함될 수 있기 때문에, 선택된 가제스처의 인식값은 임계값(경험적으로 결정한 값)보다 높아야 한다. 슬라이딩 창안에서 제스처가 판별된 후 다음 제스처를 적출하기 위해 슬라이딩 창이 움직인다. 이전 제스처의 끝점이 다음 슬라이딩 창에 대한 시작점이 된다. 효율성을 계산하기 위해, 가제스처들은 두 개의 인접하는 슬라이딩 창을 포함하고 있는 겹쳐진 블럭을 제외한 슬라이딩 창 안에서 인식된다.

적출된 가제스처의 심볼열에 대한 제스처인식값을 위해서 제스처 인식 모듈의 반환값을 사용한다. 본 논문에서는 제스처 템플릿을 이용한 다이나믹 프로그래밍을 적용하여 제스처를 인식한다. 이 방법은 심볼 삽입, 삭제와 대체를 포함한 2 개의 문자열의 값들을 매칭하는 것을 기본으로 한다. 이 때, 하나의 문자열은 하나의 가제스처며 또 다른 하나는 제스처 템플릿이다. 제안된 매칭 알고리즘안에서는 식(1)같이 재귀적 특성을 정의한다.

$$dist_{ij} = cost(i,j) + \min\{\alpha \times dist_{i-1,j}, \beta \times dist_{i,j-1}, \gamma \times dist_{i,j-1}\} \quad (1)$$

식에서  $dist_{ij}$  는  $i$  번째 심볼과  $j$  번째 심볼을 읽을 때의 최소 매칭 값이고,  $\alpha, \beta, \gamma$  는 심볼의 삽입, 대체, 삭제에 위한 상수이다.  $cost(i, j)$  는  $i$  번째 심볼과  $j$  번째 심볼 간의 거리를 반환하는 함수이다.  $cost(i, j)$  를 계산하기 위해,  $i$  번째 클러스터의 중심과  $j$  번째 클러스터의 중심간의 유클리디안 거리가 사용된다. 이처럼, 제스처는 입력 연속 영상과 각 제스처 템플릿 간의 거리들을 얻는 재귀적 특성을 사용해, 인식기는 가장 짧은 거리를 가진 제스처에 대해 가장 높은 값을 반환하고, 이 제스처를 인식한다.

4. 실험 결과

실험된 영상은 고정된 상업용 USB 카메라를 사용하여 획득된, 24 비트 RGB 컬러 영상으로 320\*240 크기의 영상들이다. 실험은 팬티엄 IV 1.3Ghz PC 에서 비주얼 c++를 사용하여 구현하였다. 그림 4 는 제안된 인터페이스로 퀘이크 게임을 하는 모습이다.

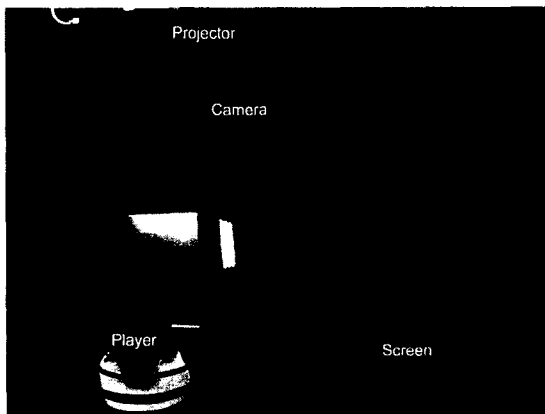


그림 4. 제안된 인터페이스를 사용한 퀘이크 게임 실험

제안된 인터페이스에서 사용된 새로운 적출 방법의 적출능력을 평가하기 위해 피실험자 40 인으로부터 평가 데이터를 얻었다.

제스처 적출에는 제스처가 아닌 패턴을 제스처라고 하는 삽입 에러, 제스처를 찾지 못하는 삭제에러(deletion error), 다른 제스처와 혼동하는 대체에러(substitution error)와 같은 세 가지 형태의 에러가 나타난다. 이를 위해 우리는 이현규의 연구[3]에서 제안한 제스처 적출을 계산하기 위한 측정법인 식(1)에서 정의된 검출비를 사용하여 제안된 인터페이스의 강력함과 신뢰성을 입증한다.

$$\text{검출비} = \frac{\text{바르게 인식한 제스처}}{\text{입력받은 총 제스처수}} \quad (1)$$

표 1 은 식 1 에 따른 제안된 인터페이스에 대한 제스처 인식 결과이다.

제안된 적출 방법 안에서, 제스처 적출 수행과 속도는 슬라이딩 창 크기에 따라 결정된다. 검출비와 처리 속력

은 슬라이딩 창 크기가 6 일 때 가장 높은 값을 가지며, 수행속도도 빨랐다.

표 1 . 제스처 인식 결과

	입력받은 총 제스처	바르게 인식한 제스처	검출비
왼쪽으로 몸을 돌려라	200	193	96.50
오른쪽으로 몸을 돌려라	200	195	97.50
왼쪽으로 한발 이동	200	194	97.00
오른쪽으로 한발 이동	198	191	96.46
앞으로 걸어나라	166	157	94.58
몸을 낮추어나라	163	149	91.41
동작 정지	160	150	93.75
총 수	1287	1225	95.31

5. 결론

제안된 논문에서는, 우리는 가제스처 인식 결과들을 사용하여 사용자의 의미없는 움직임들로부터 제스처들을 적출하는 제스처 적출 방법을 사용한 시각 기반의 게임 인터페이스를 제안한다. 제안된 인터페이스에서, 인식 기반의 적출은 우선 카메라 입력들로부터 가제스처들을 생성하고 가제스처들의 인식값을 사용하여 제스처들을 판별한다. 제안된 인터페이스의 주요한 잇점은 인식 엔진의 독립된 구조와 제스처 공동 분절과 제스처 혼동에 대한 효과적인 적출 방법이 제안된 것이다. 제안된 시각 기반의 인터페이스는 어떤 부가적인 하드웨어 장치도 사용하지 않고 단지 직관적인 시각만을 이용한 것이다. 우리는 제안된 이 인터페이스의 효율성을 증명하기 위해, 제안된 논문에서 정의된 7 개의 제스처를 사용하여, 퀘이크II 게임에 적용해 보았다. 그 결과, 시각 기반의 인터페이스는 사용자로 하여금 직관적이며 게임에 있어서 적용할 만 하였다. 현재 우리는 제안된 인터페이스가 퀘이크II에 보다 적합하게 적용될 수 있도록 제스처 명령들을 좀 더 확장하고 보다 강력한 인터페이스가 될 수 있도록 연구 중이다.

참고문헌

- [1] Iba, J.M.V. Weghe, C.J.J. Paredis, P.K. Khosla, An architecture for gesture-based control of mobile robots, *IEEE/RSJ International Conference on Intelligent Robots and Systems* Vol. 2, 851-857, 1999.
- [2] H.K. Lee, J.H. Kim, An HMM-based threshold model approach for gesture recognition, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 21 (10), 961-973, 1999.
- [3] T. Nishimura, R. Oka, Towards the integration of spontaneous speech and gesture based on spotting method, *In Proceedings of the 1996 IEEE/SICE/RSJ International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 433-437, 1996.
- [4] H. Kang, C.W. Lee, K. Jung, H.J. Kim, Recognition-based Gesture Spotting in vision-based Human Computer Interaction, *International Conference on Automatic Face and Gesture Recognition*, to be submitted, 2003.