

HMM 을 이용한 제스처 기반의 게임 인터페이스

장상수^o, 박혜선, 김상호, 김항준
경북대학교 컴퓨터공학과 인공지능연구실
{ ssjang, hspark, shkim, khkim}@ailab.knu.ac.kr

A Gesture-based Game Interface using HMM

Sang Su Jang^o, Hye Sun Park, Sang Ho Kim and Hang Joon Kim
Dept. of Computer Engineering, Kyungpook National Univ., Daegu, South Korea

요 약

본 논문에서는 컴퓨터 액션 게임 중에 하나인, 퀘이크 II 게임을 위한 제스처 기반의 인터페이스를 제안한다. 제안된 인터페이스는 연속된 입력 영상열로부터 제스처를 검출하고 인식하기 위해 HMM 을 사용한다. 먼저 제스처를 검출하기 위해 입력 영상열로부터 포즈 심볼열을 추출하여 사용한다. 인식하기 위해 사용된 HMM 은 추출된 포즈 심볼을 입력받아, 상태 확률값을 계산하여 계속적으로 갱신한다. 이때 갱신되는 상태 확률값 중에 각 제스처에 속하는 특정상태의 확률값이, 미리 정의된 임계값과 비교하여 초과하면 검출되고 인식된다. 현재 제안된 시스템은 실제 퀘이크 II 게임에서 키보드버튼과 마우스를 통해 입력되는 명령어들 중에서 게임을 진행하기 위해 먼저 필요한 움직임과 시점 변환에 관계되는 명령어들을 13 개의 제스처로 표현하고 이 제스처 명령어를 검출하고 인식한다.

1. 서론

3 차원 그래픽스 기술과 음향 기술의 발전으로 날이 갈수록 게임의 박진감과 현실감이 더해지고 있다. 하지만 컴퓨터 게임의 경우, 문서편집이나 간단한 사용자 인터페이스를 위해 고안된 키보드와 마우스로 게임 인터페이스로 사용하고 있고, 게임기의 경우 조이스틱이나 버튼들로 게임 인터페이스의 기능을 구현하고 있지만 보다 현실감 넘치는 게임을 구현하기에는 한계가 있다. 따라서 마우스, 키보드, 조이스틱 등의 주변 장치를 사용하지 않고 사용자가 직접 게임의 가상 현실 속에서 자연스러운 몸 동작으로 게임을 진행할 수 있는 새로운 비전기반의 게임 인터페이스의 개발이 진행되고 있다.

현재 다양한 제스처 기반 인터페이스는 HCI(human-computer-interaction) 시스템으로 개발되고 있고, 일부는 게임에도 적용되고 있다. Soshi Iba[1]는 손 제스처를 HMM 으로 인식하여, 로봇을 제어하는 시스템을 개발하였다. Freeman 과 Weissman[2]은 손을 추적하여 TV 를 제어하는 시스템을 제안하였고, Yang et al.[3]은 제스처를 사용하여, 원격으로 로봇을 제어하였다.

본 논문에서는 카메라 한 대로 입력 받은 영상열에서 사용자의 제스처를 검출하고, 인식하는 제스처 기반의 인터페이스를 제안하고, 이를 이용하여 퀘이크 II 게임에 적용하였다. 시스템 안에서 제스처 인식을 위해 HMM 을 사용하였고, 제스처의 인식 결과를 향상 시키기 위해서 전형적인 HMM 의 일부를 개선하여 사용하였다.

2 장에서는 제안된 인터페이스의 개요를 간단하게 소개하고, 3 장에서는 시스템의 핵심인 제스처 인식기에 대해 상세하게 설명한다. 그리고 4 장에서는 제안된 인터페이스의 실험결과를 보여주고, 마지막으로 5 장에서 결론을 맺는다.

2. 시스템 개요

제안된 인터페이스 시스템은 카메라, 프로젝터, 스크린 그리고 프로세싱 장치로 구성된다. 프로세싱 장치에는 두개의 프로세서인 “제스처 인식기”와 “게임 제어기”로 이루어져 있다. 그림 1 은 제안된 인터페이스 시스템의 구조도를 보여준다. 제스처 인식기는 카메라에서 입력 받은 영상열로부터 제스처를 검출하고 인식한다. 게임 제어기는 인식된 제스처를 명령어로서 입력하여 게임을 제어한다. 이를 위해 퀘이크 II 게임 내의 입력처리 모듈인 gamex86.dll 을 수정하여 인식기로부터 인식된 결과를 입력받아 사용하였다.

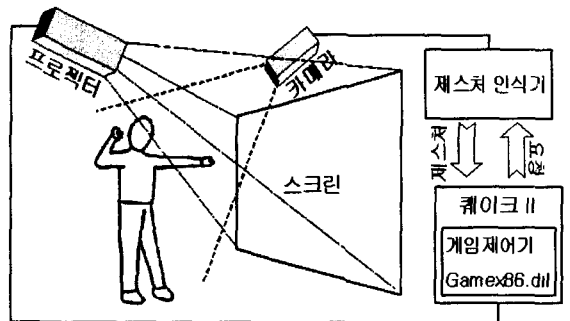


그림 1. 시스템 구조도

3. 제스처 인식기

제스처 인식기는 포즈 심볼열을 분류하고 사용자가 수행하고 있는 제스처 명령어가 13 개의 제스처 명령어 중에 어느 것인지 판단한다. 제스처 인식기는 특징 추출, 포즈 분류 그리고 제스처 판별로 구성된다.

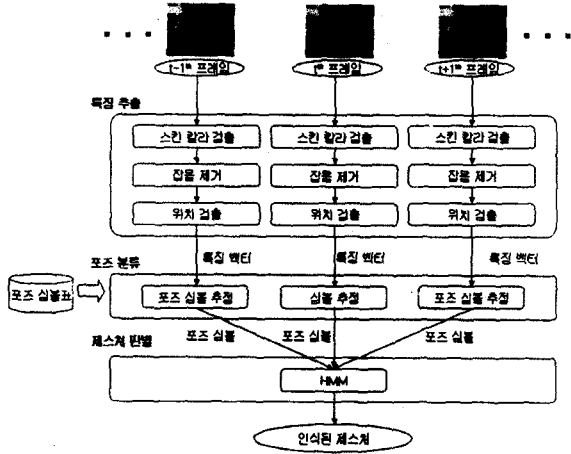


그림 2. 제스처 인식기

3.1 특징 추출

사용자의 제스처를 인식하기 위해서는 연속된 입력 영상에서 특징을 추출해야 한다. 제안된 논문에서는 사용자의 머리와 양손의 위치들을 특징으로 하고 있다. 따라서 먼저 스킨칼라의 검출을 통해 머리와 양손을 찾고 그 위치들을 추출하는 작업이 필요하다. 본 연구에서는 [4]에서 제안된 스킨칼라 모델을 이용하여 스킨칼라 영역을 검출하였다. 스킨칼라의 분포는 채색 공간에서 좁은 영역에 밀집되어 있고, 2D-가우시안 분포로 근사화 되어질 수 있다. 그런데 실제 조명의 상태 등 외부 환경에 따른 잡음이 발생되므로 추출된 스킨 칼라 영역 안에 생겨난 틈과 잡음으로 인해 잘못된 검출된 영역은 간단한 연산을 통해 제거한다. 이렇게 추출된 영역 중 가장 큰 영역은 머리, 다음으로 큰 2 개의 영역을 양손으로 한다. 이때 두 손이 교차되지 않는다는 전제하에, 양손으로 추출된 영역 중에서 보다 오른쪽에 가까운 영역은 오른손, 왼쪽에 가까운 영역은 왼손으로 한다. 이렇게 머리와 양손 영역이 정해지면 정해진 영역의 중심좌표가 각각 머리와 양손의 위치, 즉 특징벡터로 추정된다.

3.2 포즈 분류

포즈 분류기는 각 프레임마다 추출한 특징벡터들을 입력 받아, 이를 미리 오프라인 상태에서 정의된 포즈 심볼 테이블 안에 입력 받은 특징 벡터와, 비교하여 가장 차이가 작은 포즈 심볼을 찾아 해당하는 포즈 심볼로 분류한다. 이때 비교 방법은 유클리디안 디스턴스를 사용하였다. 제안된 논문에서 각 제스처들은 다른 제스처와 확연하게 구분될 수 있는 "특정 포즈"를 가지고 있고, 모든 제스처는 같은 포즈로써 시작하고 그것과 같은 포즈로 끝난다고 가정한다. 각 제스처들은 시작포즈와 특정 포즈 사이에 해당하는 포즈, 특정 포즈 그리고 시작포즈와 같은 끝포즈들로 구성된다.

3.3 HMM 을 이용한 제스처 판별

제스처를 검출하고 인식하기 위해 인식기는 포즈 심볼을 입력받는다. 인식기는 매번 포즈 심볼이 주어질 때마다 미리 정의된 13 개의 제스처 중 하나를 사용자가 수행하고 있는지 판별한다. 만약 사용자가 제스처를 수행하고 있다고 판단되면 케이크 II 는 인식기가 인식한 제스처 명령어를 입력 받는다. 제안된 인터페이스는 제스처를 판별하기 위해 HMM 을 사용한다. 사용된 HMM 은 전형적인 HMM 과는 달리, 3 가지 중요한 차이점을 가진다.

첫번째 차이점은 13 개의 제스처가 하나의 HMM 안에 모두 모델링되어져 있다. 이러한 HMM 은 좀 복잡해 보이지만 공동-분절화(co-articulation)문제를 보다 쉽게 풀 수 있다. 제안된 HMM 은 그림 3 과 같다. 그림 3 에서 가장 중간에 있는 상태는 사용자의 준비 자세이고, 이는 모든 제스처의 시작포즈이자 마지막 포즈에 해당한다. 중간에 있는 상태와 각각 가장 멀리 떨어져 있는 각 13 개의 상태는 특정포즈이고 이는 각 제스처와 관련된 경로를 판별하기 위해 사용된다.

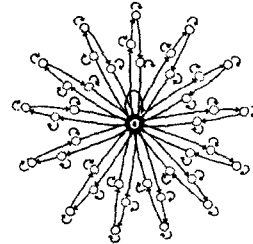


그림 3. 제스처 판별을 위한 HMM

두번째 차이점은 HMM 의 입력으로, 연속된 포즈 심볼을 사용한다는 점이다. 사용된 HMM 은 초기 상태의 확률 $s^0 = (s_k^0)$ 을 가지고 시작하고, 식 (1) 과 같이 각 포즈 심볼의 상태 확률값은 갱신된다.

$$s'_k = \left(\sum_{i=0}^{k-1} s_i^{k-1} \times a_{ik} \right) \times b_{nk}$$

$$s''_n = \frac{s'_n}{\sum_{k=0}^{K-1} s'_k}$$

식 (1)

- $S = \{s_k\}$: S 는 상태 확률 벡터, s_k 는 상태 k 의 상태확률.
- $K =$ 상태의 수 $M =$ 포즈의 수.
- $A = \{a_{ij}\}$: A 는 상태 전이 확률 분포에 대한 $K \times K$ 행렬.
 a_{ij} 는 상태 s_i 에서 s_j 로 전이확률.
- $B = \{b_{ij}\}$: B 는 관찰 포즈 심볼 확률 분포에 대한 $K \times M$ 행렬.
 b_{ij} 는 상태 s_i 에서 v_k 의 포즈 심볼이 나올 확률.

마지막 차이점은 사용된 HMM 이 각 제스처와 관련된 경로를 구별하기 위한 몇 개의 특정상태를 가진다는 점이다. 만약 어떤 제스처에 포함된 특정상태가 실험에 의해 미리 정의된 임계치보다 높으면, HMM 은 그 제스처를 인식한다. 그러므로 제스처 인식기는 가장 높은 상태 확률값의 특정상태를 가진 제스처를 선택한다.

4. 실험 결과들

본 연구에서는 제스처를 사용하여 일인칭 액션게임의 하나인 케이크 II 게임을 제어한다.

제안된 시스템은 사용자와 3m 떨어진 곳에서 3m 위에 45 도

각도로 설치된 디지털 비디오 카메라를 입력 장치로 사용했다. 사용된 비디오 카메라는 smilecam, Su-320 갈라 비디오 카메라이다. 이는 24bit RGB color 로 320*240 의 해상도를 가지고 초당 약 10 프레임으로 캡처한다. 출력장치로는 사용자 뒤에 설치된 프로젝터를 사용하였다. 그림 4 은 제안된 시스템의 실험 환경을 나타낸다.

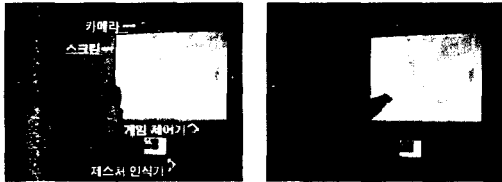


그림 4. 시스템 실험 환경

사용된 제스처 명령어는 실제 퀘이크 II 게임 안에서 빈번하게 사용되는 명령어 13 개로 그림 5 와 같다.

공격	달리기	앞으로 걷기	뒤로 걷기	점프/ 위로
왼쪽 돌기	오른쪽 돌기	왼쪽 한걸음	오른쪽 한걸음	아래로
위로 보기	아래로 보기	가운데보기	합계 : 13 동작 명령어 : 10 관점 변경 명령어 : 3	

그림 5. 퀘이크 II 게임에 적용된 제스처 명령어

제안된 시스템은 한 대의 카메라로부터 입력 받은 영상열로부터, 사용자의 머리와 양손의 위치를 특징벡터로 추출하였다. 특징 벡터 추출은 그림 6 와 같다.

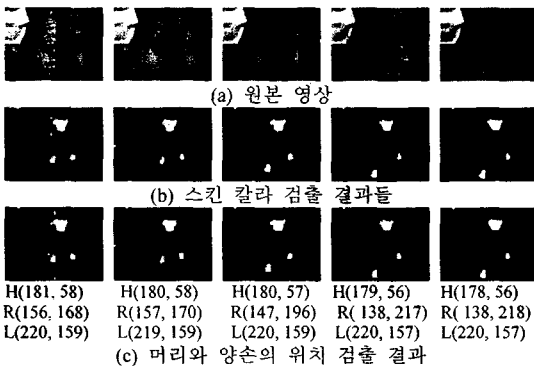


그림 6. 특징 추출 예제

그림 6 에 (a)와 같은 입력 영상들로부터 (b)와 같이 스킨 칼라 영역을 검출한다. 검출된 영역들의 중심 좌표들은 제안된 시스템의 특징 벡터들로 사용된다.

특징 추출과정에서 제안된 시스템의 일반화를 위해 같은 사람이 같은 제스처를 여러 번 수행하고 다른 사람이 같은 제스처를 여러 번 수행하는 작업을 반복하여 약 50,000 개의 데이

터 영상을 수집하였다. 이러한 실험에 사용된 영상들은 일정한 조명아래, 카메라 한 대로 찍은 영상이다.

추출된 특징 벡터들은 오프라인 상태에서 미리 정의된 포즈 심볼 테이블 중에 하나의 포즈 심볼로 분류된다. 그림 7 은 포즈 심볼 테이블을 알아보기 쉽게 그림으로 표현한 것이다.

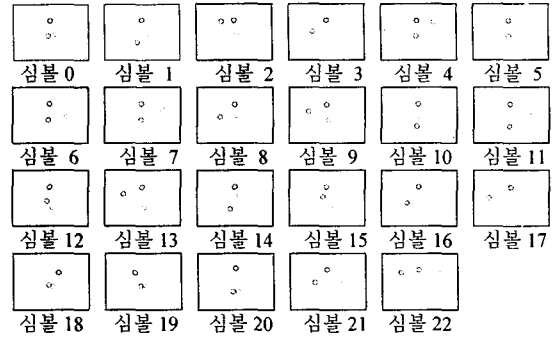


그림 7. 포즈 심볼 테이블

마지막으로 표 1 은 제스처 인식결과이다. 실험 결과 제안된 시스템은 퀘이크 II 게임의 실제 명령어에 해당하는 13 개의 제스처 명령어에 대해서 약 93.08%의 인식성공률을 보여 주었다.

표 1. 각 제스처의 인식 결과

제스처 명령어	(%)	제스처 명령어	(%)
공격 (A)	84	위로 보기 (LU)	88
앞으로 걷기 (WF)	96	아래로 보기 (LD)	96
뒤로 걷기 (BP)	86	가운데 보기 (CV)	98
왼쪽 돌기 (TL)	98	점프/ 위로 (JU)	90
오른쪽 돌기 (TR)	96	앞기/ 아래로 (C/D)	96
왼쪽 한걸음 (SL)	94	달리기 (R)	94
오른쪽 한걸음 (SR)	92		

5. 결론

본 논문에서는 HMM 을 사용한 제스처 기반의 컴퓨터 게임 인터페이스를 제안하였다. 제안된 인터페이스는 연속된 영상열로부터 13 개의 제스처를 검출하고 인식하여, 인식한 제스처를 퀘이크 II 명령어로 변환한다. 제스처 검출과 인식을 위해 모든 제스처가 포함된 하나의 HMM 을 사용하였다. 이 HMM 은 일반적으로 제스처 후보로써 미리 선택된 유한개의 포즈 심볼들을 입력으로 하는 일반적인 HMM 과는 달리 연속적인 포즈 심볼을 그대로 입력 받는다. 실험결과, 전형적인 HMM 인식기보다 더 나은 인식결과를 산출할 수 있었다.

참고 문헌

- [1] S. Iba, J.M.V. Weghe, C.J.J. Paredis, P.K. Khosla, "An architecture for gesture based control of mobile robots", *Intelligent Robots and Systems*, Vol. 2, pp. 850-857, 1999.
- [2] W.T. Freeman, and C.D. Weissman, "Television control by hand gestures", In *Proceedings of International Workshop of Face and Gesture Recognition*, pp. 179-183, 1995
- [3] J. Yang, Y. Xu and C.S. Chen, "Gesture Interface: Modeling and Learning", *Robotics and Automation*, pp. 1747-1752, 1994
- [4] Jie Yang, Waibel, A., "A real-time face tracker", *Applications of Computer Vision*, 1996. WACV '96, Vol 15, no. 1, pp. 142-147, 1996.