

가상현실 3D 오브젝트와의 상호작용을 위한 MCSVM 기반 손 제스처 인식

김윤제¹, 고택균², 윤민호³, 김태영⁴ *

서경대학교 컴퓨터공학과

e-mail : {forteas2003¹, pigpeg², src8655³, tykim⁴ *}@skuniv.ac.kr

Hand Gesture Recognition Method based on the MCSVM for Interaction with 3D Objects in Virtual Reality

Yoon-Je Kim¹, Tack-Kyun Koh², Min-Ho Yoon³, Tae-Young Kim⁴ *

Department of Computer Engineering, Seo-Kyeong University

요 약

최근 그래픽스 기반의 가상현실 기술의 발전과 관심이 증가하면서 3D 객체와의 자연스러운 상호작용을 위한 방법들 중 손 제스처 인식에 대한 연구가 활발히 진행되고 있다. 본 논문은 가상현실 3D 오브젝트와의 상호작용을 위한 MCSVM 기반의 손 제스처 인식을 제안한다. 먼저 다양한 손 제스처들을 립모션을 통해 입력 받아 전처리를 수행한 손 데이터를 전달한다. 그 후 이진 결정 트리로 1차 분류를 한 손 데이터를 리샘플링 한 뒤 체인코드를 생성하고 이에 대한 히스토그램으로 특징 데이터를 구성한다. 이를 기반으로 MCSVM 학습을 통해 2차 분류를 수행하여 제스처를 인식한다. 실험 결과 3D 오브젝트와 상호작용을 위한 16개의 명령 제스처에 대해 평균 99.2%의 인식률을 보였고 마우스 인터페이스와 비교한 정성적 평가 결과에서는 마우스 입력에 비하여 직관적이고 사용자 친화적인 상호작용이 가능하다는 점에서 게임, 학습 시뮬레이션, 설계, 의료분야 등 많은 가상현실 응용 분야에서의 입력 인터페이스로 활용 될 수 있고 가상현실에서 몰입 도를 높이는 데 도움이 될 수 있을 것이다.

1. 서론

최근 그래픽스 기반의 가상현실 기술이 발전하고 다양한 기기들이 보급됨에 따라 가상현실에 대한 사람들의 관심이 주목되고 있으며 가상현실의 3D 객체와 상호작용을 위한 사용자 친화적인 인터페이스(NUI: Natural User Interface)를 지향하는 연구 및 개발이 활발하게 진행되고 있다[1].

키넥트 출시 이후로 깊이 카메라를 이용하여 손 제스처를 인식하는 연구들이 다수 진행 되어왔다[2-4]. 영상 기반의 기존 연구는 카메라를 통하여 얻어진 낮은 해상도의 깊이 값 정보를 이용하여 인식을 수행하므로 손을 이용한 상호작용을 제공하는데 한계가 있다. 때문에 키넥트를 사용한 대부분의 연구는 손 제스처 보다는 몸의 동작을 인식하는 것이 대부분이며 손을 이용하더라도 아주 간단한 손 동작을 인식한다.

이러한 한계를 해결할 수 있는 립모션, 리얼센스, MYO 등 손에 대한 정확한 움직임을 감지할 수 있는 센서 기반의 다양한 장치들이 등장하면서 손을 이용

한 3D 객체와의 상호작용에 대한 연구가 최근 활발히 시도되고 있다. 관련 연구로는 직관적인 손동작을 고려한 립모션 기반 게임 인터페이스의 연구[5], 립모션 기반으로 손 제스처를 윗놀이의 사용자 인터페이스로 활용하는 연구[6], 3인칭 슈팅 게임의 현실감 개선을 위한 립모션 기반 인터페이스의 연구[7] 등이 진행되어 왔다. 이러한 연구들은 손 제스처의 종류가 부족하고 복잡하지 않은 손 제스처만 정의하여 사용하는 것이 대부분이다.

본 논문은 기존 연구보다 더 다양한 손 제스처를 인식하여 가상현실 3D 오브젝트와 상호작용이 가능한 인터페이스를 제안한다. 본 논문의 손 제스처 인식 방법은 립모션을 통해 입력 받아 전처리한 손 데이터를 이진 결정 트리로 1차분류를 하고 분류된 손 데이터를 리샘플링 후 체인코드를 생성하고 히스토그램으로 특징 데이터를 구성한다. 그리고 난 후 MCSVM(MultiClass Support Vector Machine) 학습을 통해 2차 분류를 수행하여 최종적으로 제스처를 인식한다. 본 논문의 검증을 위하여 이를 적용한 응용 프로그램을 개발하고 손 제스처의 인식률을 실험하는 정량적 분석과 기존 인터페이스와 비교하는 정성적

* 교신저자

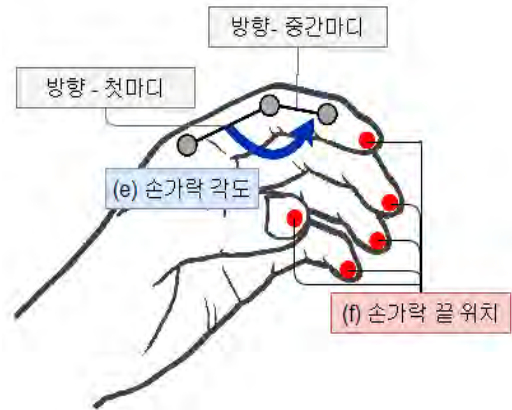
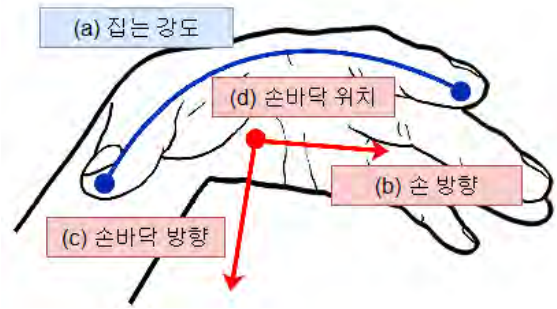
분석으로 나누어 실험한다.

본 논문의 구성은 다음과 같다. 2장에서 손 제스처 분류 및 입력에 대해 언급하고 3장에서는 제스처 인식 과정에 대하여 설명한다. 4장에서는 실험환경 및 실험결과를 언급하고 5장에서 향후 연구 방향을 기술하고 결론을 맺는다.

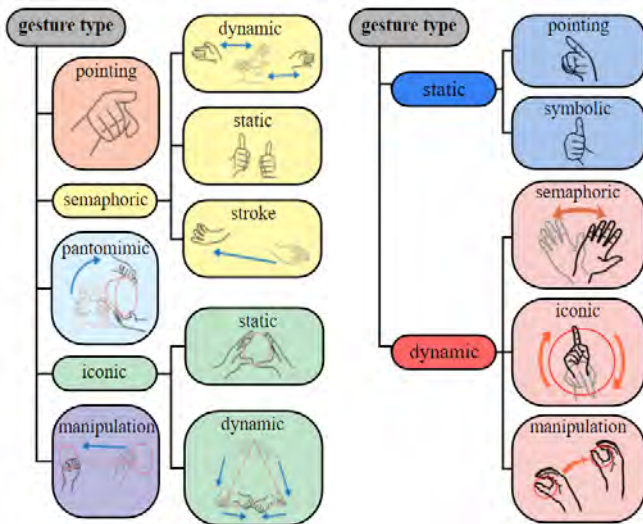
2. 손 제스처 분류 및 입력

2.1 손 제스처 분류

본 논문은 기존연구[8]에서 분류한 손 제스처의 유형(그림 1)(a) 보다 전반적인 손 제스처 유형을 파악하기 쉽도록(그림 1)(b)와 같이 제안한다. 먼저 제스처 유형을 Static(정적)과 Dynamic(동적)으로 분류하는데 이때 Static은 시작부터 끝까지 포스처와 위치가 변하지 않는 제스처를 뜻하고 Dynamic은 변하는 제스처를 뜻한다. Static은 다시 Pointing(방향), Symbolic(기호 및 상징)으로 나뉘고 Dynamic은 Semaphoric(특정 의미 전달), Iconic(객체의 모양과 크기), Manipulative(객체 조작)으로 나뉜다.



(그림 2) 립모션 손 데이터



(a) 참고 논문 (b) 본 논문에서 제안

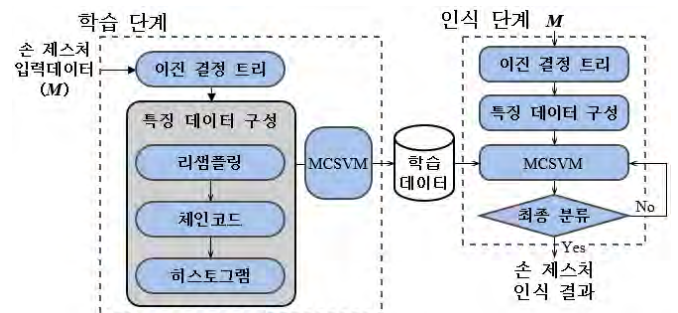
(그림 1) 손 제스처 유형 분류

2.2 손 제스처 입력정보

립모션은 적외선 카메라를 이용한 깊이 센서를 통해 다양한 손 데이터를 3차원 정보로 제공한다. 제공되는 데이터는 손의 각 관절의 위치와 방향에 대한 정보 및 손의 상태에 대한 정보들이 있으며, 센서에 촬영될 때 손 가림 현상이나 오인식으로 해당 정보가 오지 않는 경우도 있다. 그러나 인식되지 않더라도 립모션 내장 API에서 대략적인 정보를 추측하여 손 관절들을 형성한다[9]. (그림 2)는 손 데이터 중에서 본 논문에서 손 제스처 분류를 위하여 사용하는 데이터들이다.

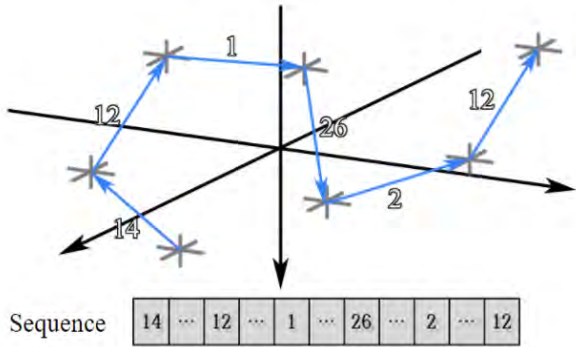
3. MCSVM 기반 손 제스처 인식 방법

본 논문에서는 크게 학습 단계와 인식 단계로 나누어 손 제스처를 인식한다(그림 3). 학습 단계에서는 립모션에서 입력된 손 데이터를 전처리를 통해 오인식된 값들을 제거한 후 이진 결정 트리로 1차 분류한다. 그 후 특징 데이터를 구성하고 MCSVM[10]을 통해 학습데이터를 생성한다. 그리고 인식 단계에서는 손 제스처의 특징 데이터를 학습된 MCSVM을 통해 손 제스처를 인식한다.



(그림 3) 제스처 인식 과정

특징 데이터 구성 방법은 입력 받은 손 제스처 궤적의 구간을 일정하게 나누는 리샘플링을 하고 손 제스처 궤적을 방향에 따라 27 체인코드를 생성한 후 이에 대한 각 구간의 히스토그램으로 특징 데이터를 구성한다(그림 4).



(그림 4) 시간 흐름에 따른 방향 벡터의 코드화

다수의 제스처를 분류하기 위해 SVM에서 확장된 MCSVM 방식 중 각각의 클래스마다 경계선을 찾는 방법인 일대다(one vs rest) 방법을 사용한다. 분류해야 할 제스처에 대하여 1차 이진분류의 4가지 제스처 종류(양손/한손, 정적/동적)에 따라 MCSVM을 각각 학습한다. 그리고 난 후 입력한 손 제스처의 특징 데이터를 4개의 학습된 MCSVM들 중 해당하는 MCSVM을 통해 가장 높은 유사도를 갖는 제스처 레이블로 2차 분류한다.

4. 실험

본 방법은 Intel Core I5 7500 프로세서, GTX1060 그래픽 카드, Visual Studio 2017 환경에서 개발되었다. 게임 개발을 위하여 유니티 2017 엔진을 사용하였고 손 인식을 위하여 립모션을 사용하며 MCSVM은 Accord-Framework 3.7의 라이브러리를 사용하였다.

본 방법을 검증하기 위해 여러 가지의 블록을 제어할 수 있는 Virtual Block이라는 응용프로그램을 개발하였다(그림 5). 사용자는 가상환경에서 손으로 원, 네모, 별 모양을 그려 3차원 객체를 생성하고 잡기, 쌓기, 던지기 등 블록을 조작할 수 있다. 이외에도 중력을 반전시켜 블록을 공중에 띄우고 내리거나 블록의 색상을 변경하는 등 3D 물체와 상호작용이 용이하다. 본 실험은 Virtual Block에서 정의된 손 제스처 중 16가지 제스처의 인식률에 대해 정량적 평가를 하였고 기존의 마우스를 이용한 인터페이스와 비교하는 정성적 평가를 하였다.



(그림 5) Virtual Block 실행화면

4.1 정량적 평가

실험에서 먼저 3명의 실험자가 각 명령 제스처를 90번씩 학습시킨다. 그리고 실험자 별로 각 명령마다 90번씩 반복 실험한다. <표 1>은 실험에 사용한 16종류의 제스처와 각각의 인식률을 나타낸다.









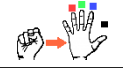

실험 결과를 보면 전반적으로 95% 이상으로 높은 인식률을 보였다. 특히 양손 제스처는 두 가지의 제스처끼리만 비교하기 때문에 100%의 결과를 보였으며 블록 생성의 경우 두 제스처의 모션이 흡사하기 때문에 서로 오인식을 하는 경우가 발생하였다.

<표 1> 제스처에 따른 인식률

분류	명령	인식률	제스처
Pointing	이동	99.7%	
Symbolic	배경 시간 변경	99.8%	
Semaphoric	블록 크기 조절	100%	
	중력 반전/복원	98.8%	
Iconic	모든 블록 삭제	100%	
	색상 옵션 열기/닫기	98.3%	
	블록 생성 구체/정육면체	98.6%	
	도움말 열기/닫기	98.6%	

실험을 하면서 실시간으로 인식하는 도중에 실험자의 의도하지 않은 동작들이 명령 제스처와 유사할 경우 명령이 실행되는 오작동이 일어남을 확인할 수 있었다. <표 2>는 오작동을 일으키는 유사 제스처를 정리한 표이다. <표 2>의 유사 제스처를 인식시키면 명령 제스처 중에 가장 비슷한 제스처로 인식된다. 이를 방지하기 위해 유사 제스처들을 학습시켜서 명령 제스처와 유사 제스처의 분류가 가능하도록 하였고 그 결과 약 15% 이상 인식률이 향상되는 것을 확인할 수 있었다.

<표 2> 유사 제스처

분류	제스처	유사 제스처
Pointing		
		
Symbolic		
Semaphoric		
		

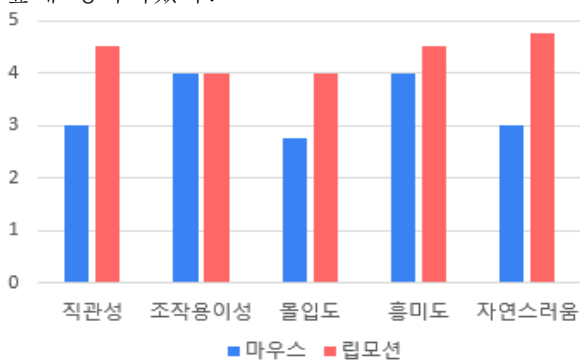
4.2 정성적 평가

정성적 평가를 하기 위하여 20대의 학생 10명에게 사용자 입력 인터페이스(마우스, 립모션을 이용한 손 제스처)를 달리하여 Virtual Block을 체험한 후 <표 3>의 기준에 따라 설문을 실시하였다.

<표 3> 정성적 평가 기준

평가 기준	점수
직관성	0-5
조작의 용이성	0-5
몰입도	0-5
흥미도	0-5
자연스러움	0-5

(그림 6)은 <표 3>의 기준에 따라 실험한 결과를 그래프로 나타낸 것이다. 설문 결과 직관성, 몰입도, 자연스러움 부분에서 제스처 인터페이스가 보다 더 높은 점수를 보였다. 마우스로 버튼을 클릭하여 명령을 실행하는 방식과 달리 제스처 인터페이스는 제스처에서 사용하고자 하는 명령의 의미가 연상되기 때문에 직관성과 자연스러움에서 많은 차이를 보였다. 몰입도 에서는 마우스 인터페이스가 사용자와 3D 객체 사이에 마우스라는 중간단계를 거치는 반면 제스처 인터페이스는 직접 3D 객체와 상호작용을 하여 더 높게 평가되었다.



(그림 6) 정성적 평가 결과

5. 결론 및 향후 과제

본 논문은 가상현실 3D 오브젝트와의 상호작용을 위한 MCSVM 기반 손 제스처 인식 방법을 제안하였다. 립모션을 통해 얻은 손 데이터를 전처리 과정을 통해 이상치를 제거하고 이진 결정 트리로 분류한 후 MCSVM을 통하여 학습한 정보를 이용해 다양한 손 제스처 인식이 가능한 인터페이스를 지원함으로써 다양한 가상현실 응용에 확장이 가능함을 보여주었다. 향후 립모션 자체의 인식 한계를 해결하고자 딥러닝(Deep Learning)과 같은 인공지능 기술을 적용해 보고자 한다.

Acknowledgement

이 논문은 2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. NRF-2017R1D1A1B03029834)

참고문헌

- [1] 양희덕, “딥러닝과 제스처 인식 기술”, 방송과 미디어, 제22권 제1호, pp. 67-74, 2017. 1.
- [2] 최재인, 유종원, 김태영 "스켈레톤 기반 제스처 인식을 통한 3D FPS 게임 인터페이스 개발", 한국 HCI학회 학술대회, pp. 464-467, 2012
- [3] 조선영, 변혜란, 이희경, 차지훈, “키넥트 센서 기반 슈팅 게임을 위한 팔 제스처 인식”, 정보과학회 논문지 : 소프트웨어 및 응용, 제39권 제10호, , p. 796-805, 2012
- [4] 최한석, “3D 콘텐츠 제어를 위한 키넥트 기반의 동작 인식 모델”, 한국콘텐츠학회논문지, 제14권 제1호, , pp. 24-29, 2014. 1.
- [5] 김민재, 허정만, 김진형, 박소영, 장준호, "직관적인 손동작을 고려한 립모션 기반 게임 인터페이스의 개발 및 평가", 한국컴퓨터게임학회 논문지, 제27권 제4호, pp. 69-75, 2014
- [6] 성재훈, 김태영, "손 제스처 상호작용을 통한 립모션 기반 윗놀이 게임", 한국차세대컴퓨팅학회 논문지, 제13권 제1호, pp. 77-81, 2017
- [7] 신인호, 천동훈, 박한훈, “슈팅 게임의 현실감 개선을 위한 립모션 기반 인터페이스 구현”, 한국HCI학회 논문지, 제11권 제1호, , pp. 5-10, 2016. 5
- [8] R. Aigner, D. Wigdor, H. Benko, M. Haller, D. Lindbauer, A. Ion, S.Zhao et al, "Understanding Mid-Air Hand Gestures: A Study of Human Preferences in Usage of Gesture Types for HCI," Microsoft Research TechReport. MSR-TR-2012-111, 2012.
- [9] Leap Motion, developer.leapmotion.com
- [10] L.Zhibin and J. Lianwen, “LATTICESVM A new method for multi-class Support Vector machines,” Int'l. Joint Conf. on Neural Networks, pp. 727-733, 2008.