

# 적외선을 이용한 다기능 제스처 인식 포인터

## Multifunctional Gesture Recognition using Infrared Light Emitters

이규화 Kyu-Hwa Lee, 진윤석 Yoon-Suk Jin, 이진규 Jin-Kyu Lee,  
최영규 Young-Kyu Choi, 박준 Jun Park  
홍익대학교 컴퓨터공학과 Interactive Media Lab.

### 요약

본 연구는 기존의 프레젠테이션의 단순한 포인터 및 클릭 기능을 넘어선 다기능 제스처 인식 포인터의 구현이 주 목적이다. 주된 연구 내용으로는 적외선 발광 다이오드에서 나오는 빛을 인지하여 포인터의 궤적 인식을 하는 기술과 이를 응용한 어플리케이션의 개발에 있다. 본 하드웨어에는 기본적인 마우스 클릭 기능 외에도 적외선 발광 다이오드를 탑재하여 PC 와 연결된 카메라가 이를 인지하도록 한다. 장점으로는 카메라가 항상 PC 와 가까이 있을 필요가 없으므로 어느 장소이든 관계없이 적외선 투과 필터가 장착된 카메라 하나만 설치할 수 있는 곳이면 구현이 가능하며 적외선을 이용하므로 가시광선의 영향을 적게 받아 효율적으로 영상 처리를 할 수 있다는 데에 있다. 제스처 인식을 통하여 수많은 여러 가지 기능을 실행할 수 있으며, 보다 사용자와 인터랙티브한 작업이 가능해진다. 구현된 하드웨어와 소프트웨어 시스템은 프레젠테이션 및 게임 컨트롤에 사용되었다.

**핵심어:** HCI, Gesture Recognition, Infrared LED, pointer

### 1. 서론

현대 사회에 들어 컴퓨터-사용자간의 상호작용을 만족시킬 수 있는 기기들의 필요성이 증가함에 따라 이를 위한 연구가 많이 진행되고 있다 [1,2]. 여러 컨퍼런스에서 버튼 클릭 등의 마우스 기능과 기본적인 레이저 기능을 탑재한 프리젠테이션의 사용은 필수가 되었으며 이를 게임 분야에서도 활용하기 위한 시도도 다수 진행되고 있다 [3,4]. 현재 실제로 쓰이고 있는 대부분의 기기들은 단순히 마우스 버튼 클릭이나 미리 정의된 키보드의 기능키를 누르는 등의 기본적인 기능만이 제공되었다. 이런 기본적인 기능을 넘어선 새로운 제어기기에 대한 연구와 이를 응용하여 회의, 게임 등에 실제로 적용시킬 수 있는 방법에 대한 내용이 본 논문의 핵심이다. 실제로 구동이 되는 하드웨어를 이용하여 많은 사용자들에게 테스트를 실시하였으며 그에 대한 결과를 정리하였다.

본 연구는 하드웨어적인 측면과 소프트웨어적인 측면으로 나뉘어서 생각할 수 있는데, 이렇게 나뉘어서 생각하는 이유는 각각의 측면에 따라서 확장, 응용할 수 있는 부분들이 구분되기 때문이다. 예를 들어서 전자의 경우에는 기존의 컨트롤 디바이스보다 사용자에게 더욱 직관적이고 손쉽게 사용할 수 있고, 후자의 경우에는 개선된 하드웨어를 이용하여 의료, 복지, 교육 등의 넓은 분야에 새로운 UI나 편리한 커뮤니케이션 방식을 가지는 응용 어플리케이션들을 구현할 수 있기 때문이다.

리케이션들을 구현할 수 있기 때문이다.

하드웨어적인 측면에서 크게 적외선을 방출하는 발광부, 필터를 통과한 빛을 받아 해석하는 수광부로 나뉘어진다. 발광부는 간단한 구동회로와 배터리, 적외선 발광 다이오드만으로 이루어지므로 적은 예산으로도 구현할 수 있다는 장점이 있다. 수광부의 경우 시중에서 쉽게 구할 수 있는 표준적인 PC 카메라에 적외선 투과 필터를 장착하여 처리하게 되는데 이 방법은 적은 컴퓨팅 자원으로도 사용자의 움직임을 감지할 수 있다는 장점이 있다.

소프트웨어적인 측면에서 실험에 행하여진 제스처 인식은 Hidden Markov Model(HMM)을 적용시켜 개발한 프로그램을 사용하였다 [8,9]. 이러한 사용자와 시스템의 직관적이고 동적인 커뮤니케이션 도구는 앞서 설명했던 것처럼 기존의 다른 입력 방식들보다 직관적이고 편리하므로 여러 분야에 응용될 수 있다.

본 연구를 통해 프레젠테이션뿐 아니라 가정에서도 활용할 수 있는 다기능 입력 장치로서의 역할을 하는 시스템의 구현이 가능해질 것이며 이는 앞으로 여러 가지 응용 기반 기술의 확장을 통하여 보다 효율적이고 상호 협력적인 컴퓨팅 환경을 구축하는데 도움이 될 것이다. 사용하기 어렵지 않으면서도 낮은 컴퓨팅 환경에서 높은 정밀도와 빠른 속도로 사용자의 움직임을 포착하여 사용자가 여러 가지 기능을 손쉽게 구현할 수 있도록 하는 것이 최종 목표이다.

## 2. 하드웨어

### 2.1 하드웨어 구조

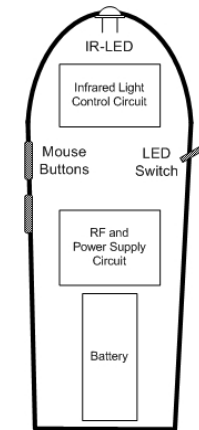


그림 1 하드웨어 구성

실제 동작을 위한 설계는 그림 1에 도시되어 있는 바와 같다. 그림에서 알 수 있듯이 적외선 발광 다이오드를 전면부에 배치하였다. 적외선 발광 다이오드의 위치를 인식하는데 있어 정확도를 높이기 위하여 다이오드의 적외선 발광 범위와 발광 강도(high radiant intensity)의 특성을 최우선으로 고려하였다. 본 실험에서는 발광 각도 40°, 발광 강도 40mw/sr, 그리고 950nm 파장의 빛을 주로 방출하는 특성을 가지고 있는 다이오드를 사용하였다. 이에 대하여 적외선을 인식하기 위한 장치로는 적외선 투과 필터를 씌운 640 \* 480 화소의 해상도를 지닌 일반적인 PC 카메라를 이용하였다. 마지막으로 전원을 절약하기 위하여 다이오드의 전원을 켜고 끌 수 있도록 ON/OFF 스위치를 부착하였다. 그림 2는 실험에 사용한 포인터의 내부 구조를 보여준다. 본 실험에 사용한 기기의 프로토타입은 시중에서 시판되고 있는 일반적인 프레젠테이션에 적외선 발광 회로를 삽입하고 스위치를 추가하여 필요할 때만 동작할 수 있도록 하였다.

그림 3에서 알 수 있듯이, 적외선 발광 다이오드의 각도가 약간 아래로 향하므로 사용자가 굳이 포인터를 리모컨 쓰듯이 손목을 구부릴 필요가 없다. 카메라와 사용자 사이의 거리는 2m로 제한하였으나 (실험적으로 이 거리에서는 인식이 매우 높았다) 신뢰 거리를 늘리기 위해서는 다른 종류의 발광 다이오드를 사용하여 해결할 수 있다.

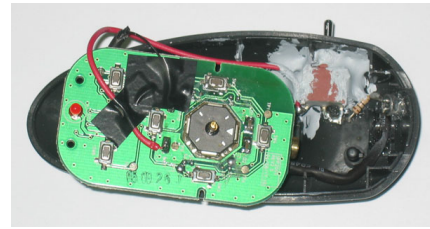


그림 2 내부 구조도



그림 3 실험에 사용된 포인터

## 3. 적외선 LED의 위치의 추적과 제스처 인식

### 3.1 적외선 LED의 위치 추적

위치 추적은 카메라가 사용자를 향하게 하고 포인터의 적외선 LED 스위치를 켜는 것으로 시작한다. 준비가 되면 위치 추적 프로그램을 구동하여 카메라를 통하여 들어온 프레임 해석하여 화면상의 2차원 좌표를 계산한다. 좌표의 계산은 그림 4에서 볼 수 있듯이 여러 번의 테스트를 통한 최적의 임계값인 200을 기준으로 그보다 밝은 화소들의 평균 중심 좌표를 구하는 방법을 사용하였다. 즉 좌표  $(X_0, Y_0)$  부터  $(X_n, Y_n)$ 까지의 픽셀들이 임계값에 의해서 구분된 적외선 LED 영역이므로 이 좌표들의 평균 중심 좌표를 구하는 것이다. 그리고 이렇게 구한 좌표들 간의 차이를 비교하여 위치 벡터를 구한다. 이 위치 벡터를 마우스 커서의 움직임에 대응시켜 사용자가 실시간으로 확인할 수 있도록 한다. 적외선 LED의 인식은 포인터의 각도에 따라 영향을 받게 되는데 각도와 인식률 사이의 관계에 대한 실험 결과를 표 1에 정리하였다.

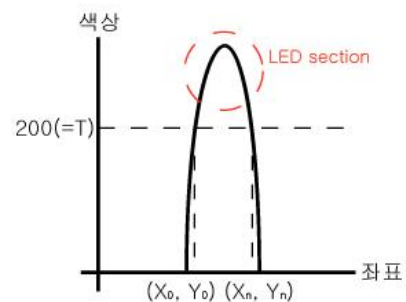


그림 4 임의의 임계값( $T = 200$ )을 이용한 LED 영역의 구분

표 1 적외선 LED와 카메라의 각도에 따른 인식률

각도	인식률
0	0.0
45	0.3
90	1.0
135	0.2
180	0.0

### 3.2 위치 추적

사용자의 제스처 패턴을 해석하고 인식하기 위하여 Konstantin Boukreev 의 신경망을 이용한 제스처 학습 라이브러리를 이용하여 심볼을 인식하는 프로그램을 개발하였다 [5,6]. 신경망을 학습시킨 알고리즘은 Back-propagation 방식으로써 입력층(input layer), 은닉층(hidden layer), 출력층(output layer)에 각각 32, 32개, 29개의 layer를 할당하였으며, 패턴인식의 오차율(error rate)은 2%, 즉 학습 결과 2%의 오차율이 발생하도록 하였다. 여기에서 출력층(output layer)은 프로그램이 검증할 수 있는 제스처 패턴의 수이므로 지금은 테스트를 위해서 임의의 수인 29개의 제스처 패턴을 정한 것이다.

절차는 다음과 같다.

1. 마우스의 기본 경로들을 기록한다.
2. 기준 좌표들 사이의 경로를 평활화(Smoothing)한다.
3. 각 점들을 각도로 나타내어지는 벡터정보로 바꾼다.
4. 벡터의 sine, cosine 값을 계산한다.
5. 신경망의 입력으로 4에서 계산한 값을 넘겨준다.
6. 출력 신경망의 벡터에 softmax 함수를 적용한다.
7. 가장 합당한 값을 찾아내어 검증한다.

그림 5는 학습된 제스처 패턴을 테스트하기 위해 만들어진 프로그램의 설정 화면이다. 그림에서 볼 수 있듯이 각각의 제스처 패턴에 따라 임의의 프로그램 실행 파일을 맵핑해 놓음으로써 사용자의 제스처가 인식이 되는지 테스트해 볼 수 있다.

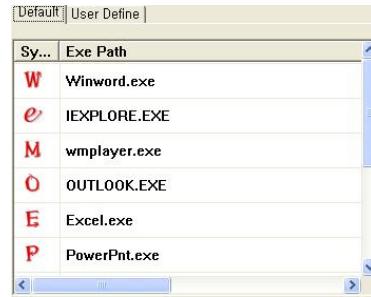


그림 5 제스처에 따라 실행할 프로그램의 정의

LED 포인터의 단순한 움직임과 사용자가 의도하는 제스처 패턴의 움직임은 구분되어야 하는데 이 때 적용된 것이 Message Hooking이다. 그림 6을 통해 Message Hooking의 처리 과정을 볼 수 있는데 Hook Chain을 구성하고 있는 Hook Procedure는 시스템에서 원하는 Message가 발생하면 Target Window에게 알림으로써 모니터의 역할을 수행한다. 이를 적용한 것은 단순한 움직임과는 다르게 제스처 패턴의 움직임은 포인터의 특정한 메시지(마우스 오른쪽 버튼)를 발생시킴으로써 두 가지의 움직임을 구분해야 하기 때문이다.

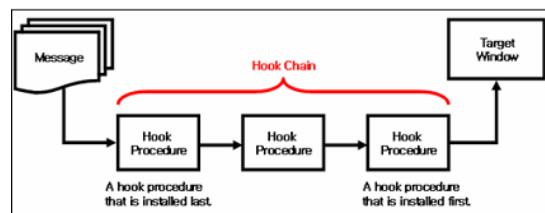


그림 6 Message Hook의 처리 과정

## 4. 실험 결과 및 응용 사례

### 4.1 프레젠테이션 및 원격회의

단순히 마우스 이동에 관련된 기능들뿐 아니라, 행동 인식은 프레젠테이션을 함에 있어서 매우 효율적인 방법을 제공해주었다. 개발한 프로그램에는 11개의 미리 정의된 인식에 대응한 액션들과 추가로 사용자 정의 가능한 액션이 10개있다. 29개의 출력층을 통해서 29개의 제스처 인식이 가능하지만 21개의 제스처로 구분한 이유는 임시적인 테스트 목적으로 제작했기 때문이다. 그러나 학습의 정도에 따라 더욱 여러 가지 종류의 액션을 취할 수 있을 것이며 프로그램의 수정으로 액션 수의 추가는 얼마든지 가능하다. 만약 액션 수가 늘어난다면 각각의 액션을 구분하는 차이가 커져야 하므로 더욱 정확한 학습을 필요로 하게 될 것이다. 그림 7에 본 포인터를 들고 취한 제스처를 인식하는 장면이 나와있다. 본 예에서는 사용자가 적외선 LED를 통해 'W' 제스처를 행할 경우 문서 편집기 애플리케이션을 실행시키는 장면을 보여주고 있다.

위에서 언급한 제스처 인식과 마우스 이벤트 등의 기능들 외에도 프레젠테이션 시 사용자가 화면에 글자들을 직접 기록하는 것도 가능하였다. 글자 인식부분은 현재 아직 구현되어 있지 않지만 이 기능만으로도 원격 회의 등에서 매우 유용하게 쓰일 수 있다는 것을 알 수 있었다. 그러나 심볼 인식 기능은 원활히 수행되므로 현재까지의 기능만으로도 매우 다양한 작업을 수행할 수 있다. 본 기기를 사용한 사용자들이 주로 지적한 부분으로는 기존의 프레젠테이션용 포인터 또는 자이로스코프 마우스에 비하여 직관적인 인터페이스를 가지고 있다는 점, 그리고 상대적으로 더욱 세밀하게 움직이는 것이 가능하다는 점 등이 있었다. 특히 직관적인 인터페이스라는 특성은 본 기기가 여러 연령층에 걸쳐 사용될 수 있다는 점에서 주목할 만하다.



그림 7 제스처 인식 장면

## 4.2 엔터테인먼트

또 다른 적용 가능한 영역으로는 게임이 있다. 마우스와 조이스틱이 사용자에게 친근한 인터페이스와 편리성을 제공하지만 이런 기기들의 가장 큰 단점들 중 하나는 상호 작용을 위한 공간이 마우스 패드나 테이블 등에 제한되어 있으며 사용자가 바라보는 모니터 등의 화면과 분리되어 있다는 점이다. 그러나, 제시된 기기를 이용하면 사용자들은 마치 화면을 직접 보며 컨트롤 할 수 있는 효과를 낼 수 있다. 그림 6에서는 사용자가 3-Ball 게임을 하는 장면을 보여주고 있다. 이 게임을 해본 사용자들은 시스템이 직관적이며 컨트롤에 익숙해지기까지 시간이 거의 들지 않았음을 지적하였다. 이러한 특성은 게임에 익숙하지 않은 사람들이 같이 참여하는데 있어 진입 장벽을 낮추는 계기를 마련할 것이다.



그림 8 게임에 적용시킨 예

## 4.3 의료 및 교육

예전부터 행해지고 있는 재활훈련 등은 훈련 요원이 있는 특정한 공간에서 훈련을 받아야 하므로 시간 및 공간적 제한을 받을 수 밖에 없었다. 그리고 그 이후로 오랜 기간 동안 훈련을 받아야 한다면 이에 추가로 발생하는 비용 또한 적지 않을 것이다.

움직임을 인식하여 의사 결정에 이용하는 또 다른 가능성으로는 규칙적인 운동과 관련된 부분이 있다. 현대인들 사이에서는 점차 웰빙 문화가 급속도로 퍼져 나가고 있는데 에어로빅이나 웨이트 트레이닝에 관한 운동 기구들이 다른 어떤 때보다 성황을 누리고 있고, 이와 관련된 멀티미디어 동영상 콘텐츠들은 많은 사람들에게 수익을 얻고 있다. 하지만 이러한 것들도 역시 시간적, 공간적인 제약이 심하고 비용이 많이 들어간다는 단점이 있다. 하지만 이러한 훈련 상황이나 교육 등에 관한 시뮬레이션을 간단한 패키지 식으로 만들어서 응용할 수 있다면 위에서 언급한 몇 가지 문제점들을 해소할 수 있을 것이다.

위에서 언급한 일련의 방법들을 수행하기 위해서는 먼저 사용자의 움직임에 따른 이동 패턴의 인식을 행하게 되는데, 예를 들어 에어로빅을 할 때 포인터를 옷에 부착하고 이를 바탕으로 행동 정확도를 평가하고자 한다면 다음 절차를 따르면 된다.

- 1) 기본적인 장비들을 설치한 후, 프로그램을 시작하고 사용자를 등록한다.
- 2) 카메라로 입력되는 사용자의 영상을 분석한다. 이 때 사용자의 영상은 다수의 적외선 LED들을 구분하고 서로의 관계 등을 정의해주어야 하므로 클러스터링이나 패턴 매칭 등의 알고리즘등이 필요하다.
- 3) 분석된 사용자의 영상과 저장해 놓은 템플릿 영상과의 패턴 유사도를 계산한다.
- 4) 사용자에게 계산된 유사도를 보여주고 재실행 여부를 묻는다.

기존의 특정 마커 등을 표지로 삼는 방식은 대상이 어느 정도 충분한 크기를 가지고 있어야 하며 정확히 카메라에 대한 각도를 유지해야 인식할 수 있다는 제약이 있었다. 반면에 적외선 LED를 사용한 방법에서는 가시 광선의 차단으로 인해서 적외선 영역을 확실하게 추출할 수 있으므로 기존에 비하여 인식률이 높고 또한 크기가 매우 작으므로 마크 방식에서 존재하던 제약 사항들을 보다 효율적으로 감소시킬 수 있다는 장점이 있다.

## 5. 결론

본 기기는 여러 방향의 충분히 많은 수의 사람들에게 테스트 되었으며 실내에서는 안정적인 결과를 보여주었으나 실외에서는 태양에서 나오는 강한 적외선 때문에 많은 양의 노이즈가 발생하는 문제가 있음을 밝혀준다. 실외에서도 사용을 하려면 후드를 씌우거나 각도를 조정하는 등의 방법이 필요하다. 그 밖에 적외선 LED가 발광하는 각도의 한계로 인하여 사용자가 움직일 수 있는 범위에 제한이 생기는 문제가 제기되었는데 이는 더 넓은 발광각 특성을 가지는 소자로 교체함으로써 해결할 수 있다. 그러나 일반적으로 LED의 발광 각도가 커질수록 강도가 약해지는 경향이 있어 노이즈에 대한 처리가 어려워지므로 소자를 선택함에 있어 발광 각도와 발광 강도의 두 가지 요소를 동시에 고려하여 이에 대한 많은 실험을 행해야 한다. 그러나 본 연구에서는 개발된 기기를 사용하여 새로운 영역에 적용시키는 방법 및 결과에 대한 고찰이 주요 주제이므로 발광소자들의 특성에 대한 더욱 깊은 조사는 차후의 과제로 남겨두었다.

개발된 본 기기의 특성 중 높은 기능 확장성과 여러 디바이스에의 응용 가능성은 중요한 특성 중 하나이다. 예를 들어, 본 기기를 응용하여 가정에서 많은 종류의 가전 제품의 제어에 사용할 수 있으며 한 개뿐 아니라 여러 개의 적외선 발광 다이오드를 적용하여 더욱 많은 종류의 작업을 실현할 수도 있다. 사용자는 개별적인 고유의 심볼들을 정의하여 기존의 단순히 실내등이나 TV, 세탁기 등의 가전 제품을 켜고 끄는 등의 기능을 넘어 세밀한 기능을 하는 것도 가능하다. 또한 노안 등 여러 가지 이유로 글씨가 잘 안 보이는 고령층이나 안구 질환을 겪고 있는 환자들에게 본 기기는 직관적인 제스처를 통하여 글씨를 모두 읽을 필요가 없이 전자 디바이스들을 제어하는 것이 가능해진다. 이는 일반적으로 제스처로 기억하는 것이 일반적인 컨트롤러의 버튼 위치를 기억하는 것보다 쉽다는 점에서 기인한 결과이다.

또 다른 응용 가능성으로 손을 잘 움직이지 못하는 장애인들에게 커뮤니케이션 도구로서의 기능을 제공하는 것이 있다. 본 논문에서 제작한 다기능 제스처 인식 포인터를 팔이나 옷 등에 부착하여 사용자 간에 게임을 하거나 커뮤니케이션을 하는 것이 가능하다. 조이스틱 등 기존의 비디오 게임 컨트롤러를 사용하는 것에 비하여 몸을 직접 움직여 활동에 참여하는 것이 훨씬 쉬울 뿐만 아니라 즐거움을 느끼기 쉬울 것이다. 여기에 덧붙여, 본 포인터를 제작하는데 내구성 등을 고려한 기계적인 요소는 필요하지 않고 단순히 손안에 들어갈만한 크기의 포인터와 필터를 씌운 PC 카메라와 낮은 성능의 처리 시스템이 필요하므로 생산 비용 또한 다른 조이스틱이나 컨트롤러에 비하여 무척 낮다. 만약 소음이 심한 협업 환경에서 응용한다면 사용자들은 카메라가 장착된 콘솔에 제스처를 통하여 다른 작업자에게 간단한 메시지나 요구사항 등을 전달할 수 있을 것이다.

과거의 다기능 포인터와 비교해서 사람들이 언급한 가장 크게 두드러진 점으로는 빠른 적응성이었다. 이러한 특성은 어린이나 기타 기계에 익숙하지 않은 사용자가 사용할 때 특히 좋은 점이 될 수 있다. 공중에서 사용이 가능한 자이로스코프 마우스와도 비교 테스트를 해봤는데, 대부분의 사람들은 커서를 원하는 대로 정확히 움직이기 위해 너무 많은 노력을 기울여야 하여야 한다고 지적하였다.

우리 연구팀은 카메라 하드웨어와 직접적인 관련이 있는 부분인 인식 속도를 높이는 동시에 인식률에 대한 신뢰성을 유지하는 방법을 연구하고 있다. 이러한 개선 사항에는 인식 속도를 30 FPS (Frames Per Second) 까지 늘리는 점과 후드 등을 이용하여 외부 광원에 의한 간섭을 최소화 하는 것이라 할 수 있겠다. 이러한 개선 사항들을 통하여 우리는 3차원 공간에서 사용자의 제스처를 정교하게 인식하는 것이 가능해지길 기대하고 있다. 또한 여기에 3개의 적외선 LED를 부착하여 더욱 많은 종류의 작업을 할 수 있게 되길 기대하고 있다. 이러한 기능들을 실현하기 위하여 Perspective-n-Points 해법 [7]을 기본 알고리즘으로 삼아 3차원 좌표를 추적하는데 이용할 계획이다. 본 연구는 실질적인 여러 환경에서의 테스트를 통하여 다양한 환경에서의 적응이 가능하고 정확한 인식이 가능하도록 계속 연구가 진행될 예정이다.

## Acknowledgements

본 연구는 21세기 프론티어 연구개발사업의 일환으로 추진되고 있는 정보통신부의 유비쿼터스컴퓨팅 및 네트워크 원천기반기술개발사업의 지원에 의한 것임

## 참고문헌

- [1] Grant Wallace, Otto J. Anshus, Peng Bi, Han Chen, Yuqun Chen, Douglas Clark, Perry Cook, Adam Finkelstein, Thomas Funkhouser, Anoop Gupta, Matthew Hibbs, Kai Li, Zhiyan Liu, Rudrajit Samanta, Rahul Sukthankar, and Olga Troyanskaya. "Tools and Applications for Large-Scale Display Walls," IEEE Computer Graphics and Applications, July 2005.
- [2] J. Rekimoto and M. Saitoh, "Augmented Surfaces: A Spatially Continuous Work Space for Hybrid Computing Environments," Proc. ACM CHI '99, Pittsburgh, PA, May 15-20, 1999, pp. 378-85.
- [3] D.J. Sturman, D. Zeltzer, "A survey of glove-based input", IEEE Computer Graphics and Applications, Volume 14, Issue 1, Jan. 1994 Page(s):30 - 39

- [4] D.H. Wilson and A. Wilson, "Gesture Recognition Using the XWand", Technical Report CMU-RI-TR-04-57, Robotics Institute, Carnegie Mellon University, April 2004
- [5] Daw-Tung Lin, "Spatio-temporal hand gesture recognition using neural networks," Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence
- [6] <http://www.codeguru.com/Cpp/misc/misc/mouseandcursorhandling/article.php/c3819>
- [7] M. A. Fischler, R. C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Comm. of the ACM, Vol 24, pp 381-395, 1981.
- [8] Y. Ephraim and N. Merhav, Hidden Markov processes, IEEE Trans. Inform. Theory, vol. 48, pp. 1518-1569, June 2002.
- [9] Lawrence R. Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE, 77 (2), p. 257-286, February 1989.