

실시간 손동작 인식을 위한 동작 평면 추정

정 승 대[†] · 장 경 호^{††} · 정 순 기^{†††}

요 약

손동작을 인식하는 연구가 오랫동안 이뤄져 왔지만 대개의 시스템들이 값비싼 깊이 카메라를 사용하거나 여러 개의 카메라를 사용해 분석하는 등 그 비용이 크며 작동이 가능한 작업 공간이 지극히 제한적이었다. 본 논문에서는 가전제품을 원격 제어하기 위한 목적으로 두 개의 회전 모터를 사용해 작업 공간을 확대하고 저렴한 일반 카메라를 사용해서 효율적으로 손동작을 인식하기 위한 시스템을 제안한다. 이 시스템은 입력된 카메라의 자세 정보와 이미지상의 2차원적 손가락 위치 정보를 이용하여 3차원 궤적을 추정하고 이를 동작 평면으로 투영시켜 의미 있는 선형 동작 패턴으로 복원한다. 또한 본 논문에서는 개발된 시스템을 테스트하여 주어진 목적에 맞는 정확도를 가지는 작업 영역을 정의한다.

키워드 : 카메라, 비전, 영상추적, 동작인식, 원격제어

Motion Plane Estimation for Real-Time Hand Motion Recognition

Seung Dae Jeong[†] · Kyung Ho Jang^{††} · Soon Ki Jung^{†††}

ABSTRACT

In this thesis, we develop a vision based hand motion recognition system using a camera with two rotational motors. Existing systems were implemented using a range camera or multiple cameras and have a limited working area. In contrast, we use an uncalibrated camera and get more wide working area by pan-tilt motion. Given an image sequence provided by the pan-tilt camera, color and pattern information are integrated into a tracking system in order to find the 2D position and direction of the hand. With these pose information, we estimate 3D motion plane on which the gesture motion trajectory from approximately forms. The 3D trajectory of the moving finger tip is projected into the motion plane, so that the resolving power of the linear gesture patterns is enhanced. We have tested the proposed approach in terms of the accuracy of trace angle and the dimension of the working volume.

Keywords : Camera, Vision, Tracking, Motion Recognition, Remote Control

1. 서 론

오늘날 우리는 다양한 역할을 가진 스마트 기기가 주변을 둘러싼 유비쿼터스(Ubiquitous) 환경에서 살고 있다. 그리고 이러한 환경에서 기기들이 인간에게 적절한 서비스를 제공하기 위해서 '어떤 기기에게 명령을 내릴 것인가?'하는 선택의 문제와 '어떻게 인간의 의도를 전달하는가?'하는 방법의 문제가 자연스럽게 발생했다. 초기에는 조작하고자 하는 기기에다가가서 부착되어 있는 버튼이나 다이얼을 직접 조작함으로써 문제를 해결했으며, 오래지 않아 이동을 최소화하

고 편의를 도모하기 위해 유선을 확장하여 연결한 스위치가 발명되었다. 그리고 현대에 이르러 대다수의 주변 기기들은 적외선이나 무선 주파수를 이용한 리모컨을 채택하여 원격에서 리모컨의 조작으로 원하는 기기에 명령을 내릴 수 있도록 발전했다.

리모컨은 서로 다른 코드의 신호를 발신하여 어떤 기기의 어떤 기능인지를 구분하게 되는데, 별도의 외부 장치이므로 따로 휴대전원을 공급해 줘야 하며 분실할 우려도 많았다. 게다가 늘어가는 기기의 개수만큼 리모컨의 개수도 늘어나고, 이를 해결하기 위한 통합 리모컨도 등장했지만 그 레이아웃은 통합된 만큼 복잡해졌다. 이런 리모컨의 단점을 보완하기 위해 최근에 음성, 비전, 그 외의 각종 센서를 이용한 다양한 기술들을 사용하는 연구가 활발히 진행되고 있다. 그 중 비전(Vision)정보는 인간 의도를 파악하기 위해 활용하기 좋은 정보로서 입력자의 위치, 표정, 눈의 깜빡임, 눈동자의 위치, 몸 전체의 자세 및 동작, 손의 모양과 손으

* 본 연구는 2단계 BK21 경북대학교 정보기술인력양성 사업단의 지원을 받았음.

† 준 회원 : 경북대학교 전자전기컴퓨터학부 박사과정

†† 정 회원 : 경북대학교 전자전기컴퓨터학부 연구 교수

††† 종신회원 : 경북대학교 컴퓨터공학과 교수

논문접수 : 2009년 5월 25일

심사완료 : 2009년 6월 23일

로 그리는 동작 등의 정보를 비전으로 얻을 수 있다.

본 논문에서는 리모컨을 사용하는 인간의 행동에 주목하였다. 초창기 리모컨은 적외선의 범위가 좁고 약해서 기기를 향해서 사용해야 작동했지만 현재는 리모컨의 성능이 대폭 향상되어 방안의 대부분의 방향에서 작동하므로 굳이 기기를 정확하게 향하지 않아도 작동한다. 그러나 그 인간은 여전히 자신이 조작하고자 하는 기기를 향해 리모컨을 조작하는 것이 자연스럽다.

인간이 관심이 있는 대상을 향하여 행동들을 취하게 된다는 사실에 착안하여 여러 연구자들이 인간의 주시 방향을 파악하기 위한 연구들을 진행하였다[1-4]. 하지만 카메라를 이용한 동작 인식의 경우 대부분의 연구가 인간이 카메라를 바라보거나 특정 위치를 바라보아야만 입력이 가능하다는 단점을 가진다[5-8]. 이러한 단점을 해소하기 위해 웨어러블 기술을 활용한 방법과 다수의 카메라를 사용하는 방법이 연구되었다. 이 중 웨어러블 기술은 영상을 촬영하는 카메라를 인간에게 직접 장착하여 바라보는 지점을 입력 받는 방법으로 장소나 방향에 구애받지 않고 비교적 자유롭게 입력을 할 수 있어 최근에도 관련 연구가 많이 발표되며 좋은 결과를 보여 주고 있지만 인간의 몸에 거주장스러운 무언가를 부착하게 된다는 단점이 있다[9-10].

특정한 위치에서 입력을 할 것으로 제한하고 동작을 인식하는 연구들을 살펴보면 완전하게 2차원적인 평면에 제한된 방향을 바라보며 하는 동작을 판단하는 연구들이 대부분이다. 이를 좀 더 자유롭게 하기 위한 방법으로 고가의 Depth 카메라나 2대 이상으로 스테레오 카메라를 사용하여 완벽한 3차원 위치를 찾는 연구들도 있다[11-12]. 한정된 방향을 극복하였지만 입력 위치는 여전히 제한적이며, 고가의 장비들을 사용하고 초기 설정이 복잡하다는 단점을 가진다.

인간이 직관적으로 의사를 전달하려면 전달될 대상을 바라보면서 원하는 기능에 대한 상식적인 손동작을 행하는 것이 자연스럽다. 이를 위해서는 인간이 바라보는 방향에 관계없이 손동작을 인식하고 이를 분석할 수 있는 방법이 필요하다. 본 논문에서는 실시간으로 동작하고 비교적 자유로운 입력 범위에서 인간의 3차원적 손동작을 인식하기 위해 동작 평면이라는 새로운 개념을 제시하며, 또한 모터가 장착된 단일 카메라를 이용하여 입력 대상 영역을 방안 전체로 확대할 수 있도록 하였다.

본 논문의 구성은 다음과 같다. 2절에서 관련연구와 본 논문이 제안하는 시스템의 개요를 살펴보고, 3절에서 움직임이 있는 영역을 추출하기 위한 배경 분리와 피부 영역 분리 방법을 살펴본다. 4절에서는 모터의 회전각, 바다면, 인간의 지시방향의 상호 관계를 설명하고, 이로부터 동작 평면을 추정하는 방법을 살펴본다. 5절에서 실험을 통한 성능 평가 결과를 제시하며, 끝으로 6절에서는 결론을 맺는다.

2. 관련 연구 및 시스템 개요

2.1 관련연구

비전 기술에서 입력으로 가장 적합한 인간의 행동 중 하

나는 손동작이다. 손동작에 관한 연구는 오래전부터 진행되어 왔지만 손동작의 대상이 되는 사물은 특정화되어 있는 경우가 대부분으로서 컴퓨터의 포인팅 입력[13-14], 사람 사이의 커뮤니케이션을 보조하기 위한 동작[8], 로봇을 조종하기 위한 동작[3, 11] 등 사람의 상식적인 행동 양식인 아닌 작위적인 패턴을 따르는 입력들이었으며 항상 카메라를 바라보고 동작을 수행해야 했다.

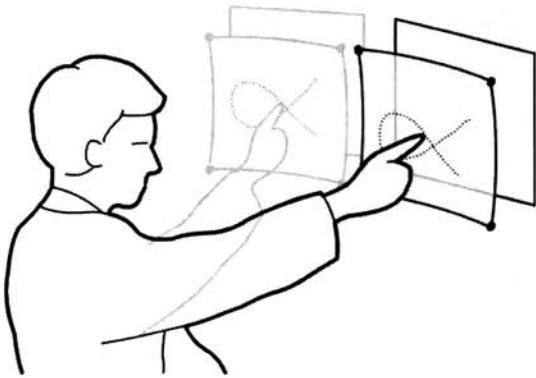
Xiyang Wang 등[5]은 카메라는 항상 손의 정면을 향한 상태에서 파티클 필터를 사용하여 손가락과 손을 추적한 후, 손가락의 2차원적 위치를 키보드나 마우스의 입력을 대체하는 것으로 사용했다. Odest Chadwicke Jenkins 등[6]은 정면에 설치된 단일 카메라에 손과 발이 겹치지 않도록 동작해서 실루엣 정보를 추출하여 동작 데이터베이스에서 추출된 동작을 찾고 다음 동작을 예측하여 로봇 조종에 사용하였다. 하지만 평면적 실루엣의 한계 때문에 좌우로 팔을 벌리고 어색한 동작을 취할 수밖에 없다. Narut Soontranon 등[8]도 정면의 카메라를 향하여 수화동작을 취하고 이를 인식하여 청각 장애자와 일반인 사이의 대화를 원활하게 하기 위해 활용하였다. MIT Media Lab의 Pfänder 시스템[12]은 단일 카메라에서 얻은 영상을 이용하여 머리와 손의 2차원 영상과 색의 통계 모델을 적용해 머리와 어깨, 손가락 끝을 찾고 카메라 정면에서 얻어진 팔꿈치 위와 아래의 팔 길이를 운동학적으로 분석하여 3차원 동작을 인식하였다.

다수의 카메라를 사용해 3차원 경로를 추적[12]하거나 레이저 센서를 이용한 측정 기법[14]에 관한 연구도 있지만, 미리 카메라를 정확하게 설정하여야 하기 때문에 장비들의 위치는 항상 고정되어 있어 작업 범위는 카메라의 FOV(Field of View)에 따라 제한 될 수밖에 없다. Satoshi Yonemoto 등[11]은 스테레오 카메라를 사용하여 영상을 획득한 후, 각 영상에서 색상정보를 이용하여 영역을 분리하고 상단영역을 머리, 하단과 좌우 영역을 양손으로 인식한 다음, 3차원 위치를 계산하여 아바타를 제어하지만 결국 자신의 데스크 한쪽 면에서 벗어나지는 못했다.

하지만 웨어러블 카메라는 앞서의 연구들과는 다르게 공간과 방향의 제한에 대해 상당 부분 해결할 수 있다. Yang Liu 등[10]은 스테레오 카메라가 장착된 헬멧을 쓰고 깊이의 경계면을 따라 실루엣을 계산하여 손가락 형태를 찾고 손가락 끝이 어떤 물체와 겹쳐지는가를 판단하여 복잡한 배경에도 문제없이 각 물체에 대해 4가지 정도의 서로 다른 입력을 신뢰성 있게 사용할 수 있었다. Kazuhiko Sumi 등[9]이 참여한 Active Wearable Vision Sensor 프로젝트에서는 사용자의 방향과 경로를 인식하기 위해 헬멧에 스테레오 카메라를 장착하고 눈동자를 촬영하여 눈동자가 가리키는 방향을 입력이 이뤄지는 영역으로 판단하고 눈동자의 주시 지점을 이용한다.

2.2 시스템 개요

본 논문에서는 실시간으로 동작하고 자유로운 입력 범위에서 인간의 3차원적 손동작을 인식하기 위해 동작 평면이



(그림 1) 어깨를 축으로 한 구상의 동작 궤적과 지시 방향으로 확장된 평면의 관계

라는 새로운 개념을 제시하고 있다. (그림 1)과 같이 사람의 손동작은 어깨나 팔꿈치 그리고 손목을 축으로 움직이는 행동이 대부분이며, 앞으로 밀치는 동작이나 뒤로 당기는 동작을 제외한 일상생활에서 사용되는 대부분의 동작은 2차원 평면상에서 인식될 수 있다.

본 논문은 손동작을 인식하기 위한 동작 세그먼트 궤적을 효과적으로 표현하기 위해 손가락 끝이 가리키는 지시 방향에 대해 수직인 가상의 2차원 평면을 동작평면이라 정의한다.

제안된 시스템은 동작방향에 관계없이 손동작의 궤적을 입력으로 받아들이기 위해 입력대상을 추적하기 위한 카메라와 동작을 기록하기 위한 카메라를 천장에 장착하였으며, 방안의 모든 영역에서 손동작을 추적하기 위해 추적용 카메라는 광각렌즈를 사용하여 방안을 넓게 촬영하고 촬영된 영상에서 움직임이 있는 사람을 추적하여 입력 카메라가 바라볼 방향을 결정한다. 입력 카메라는 2개의 축으로 회전 가능한 모터에 장착되어 있어 방안의 어디든 자세하게 촬영할 수 있으며 촬영된 영상에서 손을 추적하여 동작 평면을

추정하고 그 궤적을 기록하여 투영한다.

3. 대상 영역 추출

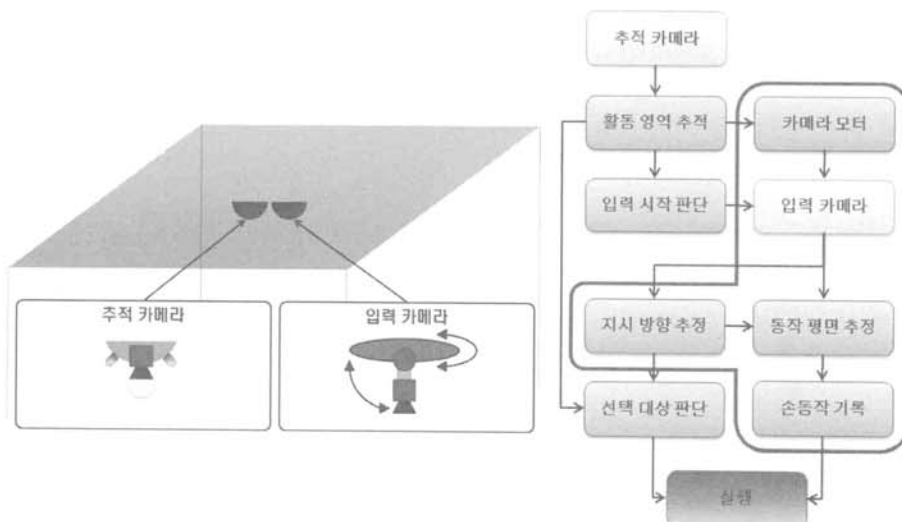
본 절에서는 손동작의 인식을 위해 입력대상이 되는 영역을 구분하기 위한 방법론을 살펴본다. 3.1절에서는 픽셀 색상 변화 시간을 이용한 활동 영역 분할방법을 설명하며, 3.2절에서는 추출된 활동영역에서 HSV 색상 모델을 사용하여 피부색 영역을 분할하는 방법을 살펴본다. 3.3절에서는 손끝 검출과정을 설명한다.

3.1 활동 영역 분할

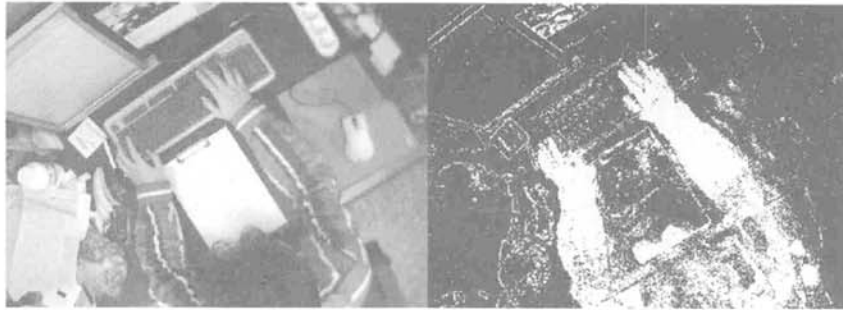
본 논문에서 관심영역은 사람 손동작의 움직임이다. 입력 공간인 방 안에서 대부분의 물건들은 인간의 손이 닿지 않는 한 움직임이 없다. 그러므로 움직임이 없는 배경 영역은 관심 대상에서 제외한다. 입력영역을 구분하기 위한 첫 번째 과정은 활동이 활발한 영역을 추출하는 것이다. 제안하는 시스템은 단일 카메라의 영상을 획득하고 있으며, 영상 획득 시 대상이 되는 손은 항상 움직임이 있다는 가정에서 픽셀타이머를 사용하여 변화가 일정시간 없는 픽셀을 배경으로 간주한다. 현재 픽셀의 RGB 색상 벡터를 p_{cur} 로 두고 바로 이전 프레임의 색상 벡터를 p_{prev} , 그리고 해당 위치의 배경 색상 벡터를 p_{back} 으로 두면 현재 픽셀의 타이머 값은 식 (1)과 같다.

$$t_n = \begin{cases} 0 & \text{where, } |p_{cur} - p_{back}| < T_{rgb}, \\ t_{n+1} + 1 & \text{where, } |p_{cur} - p_{prev}| < T_{rgb}, \\ 1 & \text{otherwise.} \end{cases} \quad (1)$$

식 (1)에서 T_{rgb} 의 값은 RGB 색상 성분의 차이에 따른 임계값이다. 수식으로부터 픽셀 타이머가 일정 시간의 임계



(그림 2) 전체 시스템 구성



(그림 3) 픽셀 변화 타이머를 이용한 활동 영역 분리

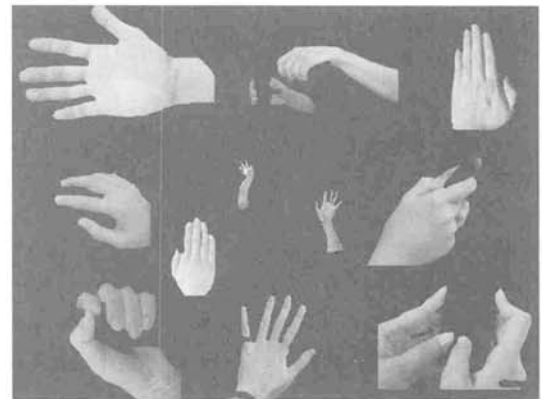
값 T_{time} 이상 유지된다는 것은 이전의 배경 색상과는 다른 색상이지만 일정 시간 이상 변화가 없었다는 뜻이므로 이때 현재 색상을 배경 색상으로 할당하여 배경을 갱신한다.

$$p_{back} = \begin{cases} p_{back} & \text{where, } t_n < T_{time} \\ p_{cur} & \text{otherwise.} \end{cases} \quad (2)$$

픽셀타이머를 이용하여 활동영역을 추출하는 이유는 피부색 검출 시, 피부색과 유사한 가구나 물건들을 제거하기 위한 목적으로 사용되며, 이러한 움직이지 않는 가구나 물건들을 효과적으로 제거할 수 있다. 그러므로 다음단계인 피부 영역 분할단계에서 효과적으로 피부영역을 추출해 낼 수 있다. (그림 3)은 픽셀 타이머를 이용하여 활동 영역을 분리한 결과이다.

3.2 피부 영역 분할

피부 영역을 인식하기 위해 많은 연구자들은 에지 검출방법보다 색상정보를 많이 사용한다. D. Chai 등[16]은 사람의 얼굴을 추출하기 위해 YCbCr 색공간의 Cb-Cr 성분을 이용하였고, Qiang Zhu 등[17]은 HSV 공간에서 인간의 피부를 검출하는데 적합한 모델들은 다양하게 검증하고 설명하였다. 본 논문에서는 피부영역의 추출을 위해 HSV 모델을 사용한다. 이 경우 H 색상 성분만으로 피부를 구분할 경우 피부가 아닌 유사 색상까지로 범위가 넓어진다는 단점이 있다. 하지만 조명이나 환경에 영향이 비교적 적으며, 3.1절에서 설명한 바와 같이 활동 영역을 찾아내고 이를 기반으로 피부영역을 추출함으로써 움직임이 없는 물체들에 대해서는

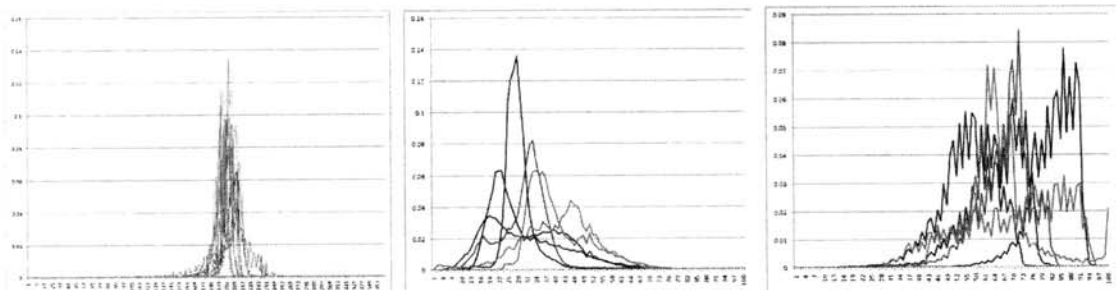


(그림 4) 피부색 임계값 학습에 사용된 여러 장의 이미지들

관심영역에서 배제할 수 있다. 피부 색상의 임계값을 찾기 위해 구글 이미지를 통해 다양한 환경에서의 손 영상과 직접 촬영되어 추가된 이미지까지 20여장의 이미지를 학습 데이터로 사용하였다. (그림 5)는 HSV 각 영역에 대한 히스토그램이다.

(그림 5)(a)의 H에 대한 히스토그램 그래프를 살펴보면 비교적 분산이 적어 피부색상의 임계 영역이 확실하다. 하지만 (그림 5)(b)와 (그림 5)(c)의 이미지 몇 장에 대한 S와 V의 히스토그램에서는 분산이 매우 큼을 알 수 있다. 이는 신뢰성 있는 임계값을 가지지 못하므로 다음 식 (3)을 사용해 S와 V의 값을 조합하여 임계 영역을 찾게 된다.

$$skinSV(s,v) = \omega_s s + \omega_v v \quad (3)$$



(a) H값 (180도 회전되어 있음)

(b) S값

(c) V값

(그림 5) HSV 각 영역에 대한 히스토그램



(그림 6) 피부 영역 분할

채도와 명도가 동시에 조명의 영향을 받으므로 이 두 값을 조합한 값에서 임계 영역을 찾는다면 비교적 조명에 영향이 적은 영역을 검출할 수 있다. 따라서 $skinSV(s,v)$ 는 선형 합의 형태를 나타내게 되고 식 (3)에서 분산을 최소화시키는 가중치 w_s 와 w_v 를 평가하여 그 계수를 사용해 다시 학습하고 임계값을 계산하게 된다. 학습된 임계값을 적용하여 영상으로부터 피부 영역을 분리해낸 결과는 (그림 6)과 같다.

3.3 손끝 검출

피부색 검출 수행 후 레이블링 기법을 적용하여 영역을 분류한 후 최종적인 입력 영역을 선별하기 위해 중요도를 계산하고 정렬했다. 각 영역을 중요한 입력 대상 순서대로 정렬하기 위해 사용된 대표값은 식 (4)와 같이 계산된다. 각 라벨 l 의 영역에 대한 대표값 $w(l)$ 은 해당 영역의 모든 픽셀에 대해서 $s(n)$ 은 피부일 경우 1의 값을 그 외는 0의 값을 가지며 $b(n)$ 은 배경일 경우 1의 값을 그 외에는 0의 값을 가진다. 상수 a 를 사용해 피부영역과 배경영역의 비중을 조절하는데 1을 초과하는 값을 사용하여 피부영역에 더 가중치를 주었다. 마지막으로 입력 영역을 선별하기 위해 $w(l)$ 이 특정 임계값 이상이 되는 영역 중 상위 3개의 영역을 입력 대상으로 선택한 후 그 중 영역의 전체 넓이에서 전경 영역의 비율이 높은 입력 대상을 선택하여 최종적인 입력 대상으로 판단한다. (그림 7)은 이와 같은 과정을 거쳐

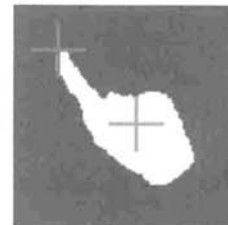


(그림 7) 피부 영역 분류와 배경 영역을 사용하여 선별된 입력 영역

선별된 입력 영역을 나타낸다.

$$w(l) = \sum_{n=0}^{all(l)} (as(n) + (1-b(n))) \quad (4)$$

최종 선택된 입력 영역에서 무게 중심을 계산하고 그 무게 중심에서 가장 먼 점을 손끝으로 판단하는데, 이와 같은 판단이 유효한 이유는 (그림 8)과 같이 손바닥 부분의 피부 영역이 둥글게 뭉쳐있기 때문에 무게 중심이 손바닥 중심에 가깝게 결정되어 단순히 거리가 가장 먼 점을 선택하여 손끝으로 판단해도 무리가 없기 때문이다.



(그림 8) 무게 중심에서 먼 픽셀을 손끝으로 결정

4. 동작평면과 손동작 인식

본 절에서는 추출된 손끝 정보를 이용하여 손이 가리키는 방향을 추정하여 실제 동작이 일어나는 동작평면을 결정하고 손동작의 궤적을 동작 평면으로 투영하는 방법과 제한 사항에 대해 살펴본다. 4.1절은 모터에 장착된 운동 카메라와 인간의 지시 방향에 대한 기하 관계를 설명하고, 4.2절은 기하 관계를 이용해 바닥 평면을 추정하고 바닥 평면에 대한 동작 평면을 추정하는 방법을 설명한다. 4.3절에서는 앞서 추정된 동작 평면으로 카메라에 촬영된 손동작을 투영하는 방법을 설명하고 4.4절에서는 동작 평면의 유사성을 비교 평가하기 위해 기록하는 방법을 설명한다.

4.1 카메라와 입력 대상의 기하 관계

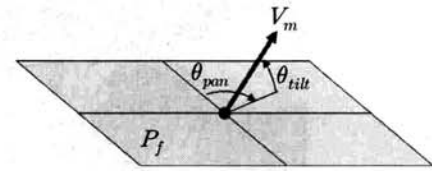
동작 평면을 추정하기 위해서는 먼저 카메라와 바닥 평면, 동작 평면이 이루는 기하 관계를 알아야한다. 두 개의 축으로 회전이 가능한 카메라는 모터에서 정확한 회전각을

제공받을 수 있기 때문에 별도의 회전각을 계산하거나 측정할 필요가 없다. 그리고 카메라의 월드 좌표에 대한 위치는 회전각이 정해져 있으므로 대상에 대한 거리에만 영향을 미치게 되는데 대상에 대한 거리가 측정된 동작에 미치는 영향은 바로 동작 궤적의 크기 비율에만 관여하게 되므로 이를 정규화 시킨다면 월드 좌표계에 대한 카메라의 위치를 알 필요가 없다. 따라서 실질적으로 카메라의 외부 파라미터를 보정하거나 측정할 필요가 없다. (그림 9)(a)에서 보듯이 카메라는 X축 회전과 Z축 회전이 가능한 모터에 장착되어 있다. 각 모터가 이루는 각도를 알 수 있고 카메라는 천장에 설치되어 있으므로 P_f 의 법선 벡터로 표현되는 바닥 평면에 대해 카메라에서 촬영되는 이미지 평면 P_c 는 (그림 9)(b)와 같이 나타나고 식 (5)와 같이 정의할 수 있다.

$$P_c = R_{\theta_z} R_{\theta_x} P_f \quad (5)$$

사람이 지시하는 방향 벡터 V_m 을 바닥 평면에 대한 각도로 표현하면 (그림 10)과 같이 pan 회전각과 tilt 회전각으로 분리해서 생각할 수 있다. 그런데 천장에 장착된 카메라에 촬영된 영상으로 tilt 각도를 파악하기 어려우며 일반적인 환경에서 대상을 지시할 때 tilt 각도의 중요성은 pan 각도에 비해 매우 작으므로 tilt 각도를 무시할 수 있다.

앞서의 제한 사항에 따라 tilt 각도를 무시하였을 때 바닥

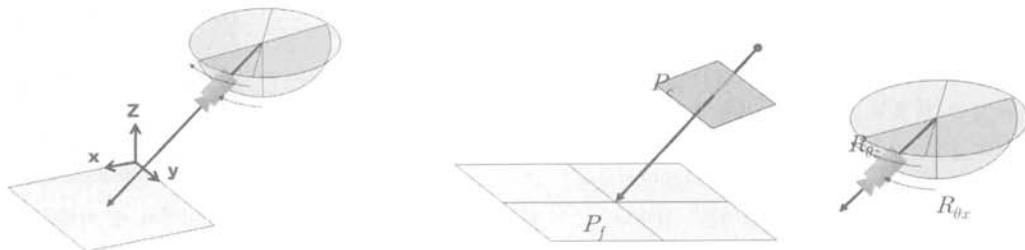


(그림 10) 바닥 평면과 지시 방향의 관계

평면에서 인간이 지시하는 방향 벡터 V_m 은 월드 좌표계에서 Z축에 해당하는 회전각 θ_{pan} 만 남게 된다. 그리고 동작이 기록되는 동작 평면 P_m 은 방향 벡터 V_m 에 직교하는 면이므로 식 (6)과 같이 나타낼 수 있다. 식 (6)에서 M_m 은 아무 변환 없이 바닥 평면 P_f 와 평행한 P_m 을 적절한 이동과 회전을 통해 바닥면과 직교할 수 있도록 만드는 변환으로 그림 11)의 (a)에서 (b)로 가기 위한 변환이다. 그리고 최종적으로 Z축을 기준으로 θ_{pan} 만큼 회전시키는 변환 R_{θ_p} 를 적용하면 (그림 11)(b)에서 (c)로 변환이 완료되고 손가락 지시방향에 맞는 동작 평면 P_m 을 얻을 수 있다.

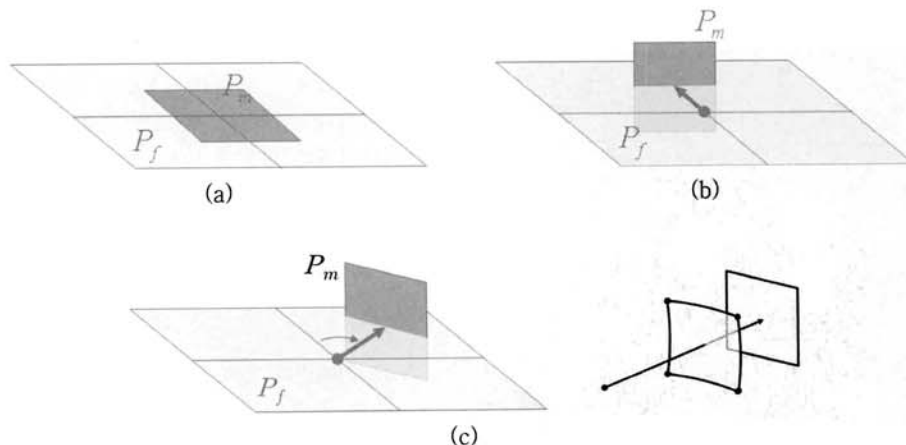
$$P_f = R_{\theta_p} M_m P_m \quad (6)$$

앞서의 식 (5)와 식 (6)의 관계에 의해 카메라 이미지 평면과 동작 평면의 관계는 식 (7)과 같이 정리된다.



(a) 두 개의 모터가 장착된 카메라와 바닥 평면 (b) 영상 평면과 바닥 평면의 관계

(그림 9) 카메라, 영상 평면과 바닥 평면의 관계



(그림 11) 바닥 평면과 동작 평면의 관계

$$P_c = R_{\theta_z} R_{\theta_x} R_{\theta_r} M_m P_m \quad (7)$$

그리고 위의 식 (7)을 사용하면 카메라 이미지 상에 나타난 동작의 궤적을 동작 평면으로 복원시킬 수 있다.

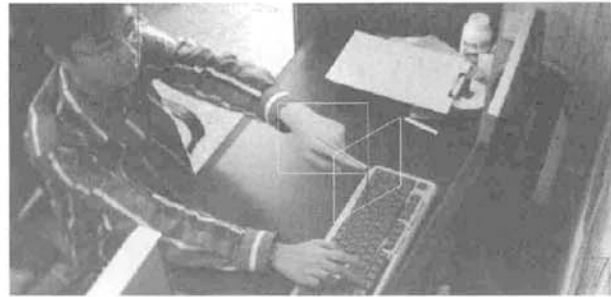
4.2 동작 평면 추정

앞서 4.1절에서 바닥 평면과 카메라의 관계를 보면 단순히 모터의 회전각으로만 이루어져 있다. 즉 카메라의 월드 좌표계에 대한 위치 관계는 표현되어 있지 않는데 이는 촬영 각도가 같을 때 위치가 다른 곳에 존재하더라도 2차원 카메라 이미지 상에서 크기 비율에만 영향을 미치지 않기 때문에 무시한 것이다.

동작 평면은 카메라 모터의 회전각 θ_x, θ_z 와 카메라 이미지 평면 P_c 상에서의 손가락이 지시하는 방향 벡터 V_c 에 의해 추정된다. 벡터 V_c 를 Z축에 대한 회전이 적용된 바닥 평면 P_f' 에 투영하여 바닥 평면에서의 지시 방향 V_f 를 계산하고 월드 좌표계의 X축을 향하는 정규 벡터 N_x 에 대해 V_f 가 이루는 각도 θ_f 를 구하면 우리가 찾고자 하는 동작 평면을 알 수 있게 된다. 각 벡터를 2차원 벡터로 표현하면 식 (8)과 같이 나타낼 수 있다.

$$V_f = \begin{bmatrix} 1 & 0 \\ 0 & \cos(\theta_x) \end{bmatrix} V_c \quad (8)$$

이렇게 계산된 θ_f 는 Z축에 대해 회전되어 있는 평면 P_f' 에 대한 각도이므로 여기에 카메라의 Z축에 대한 회전각 θ_z 를 빼면 월드 좌표계에 대한 지시 방향 벡터 V_m 을 얻는다. 최종적으로 동작 평면 P_m 은 V_m 에 직교하는 평면을 계산하면 (그림 13)과 같이 동작 평면을 추정할 수 있다.



(그림 13) 손가락 지시 방향으로 동작 평면(흰 사각형)을 추정

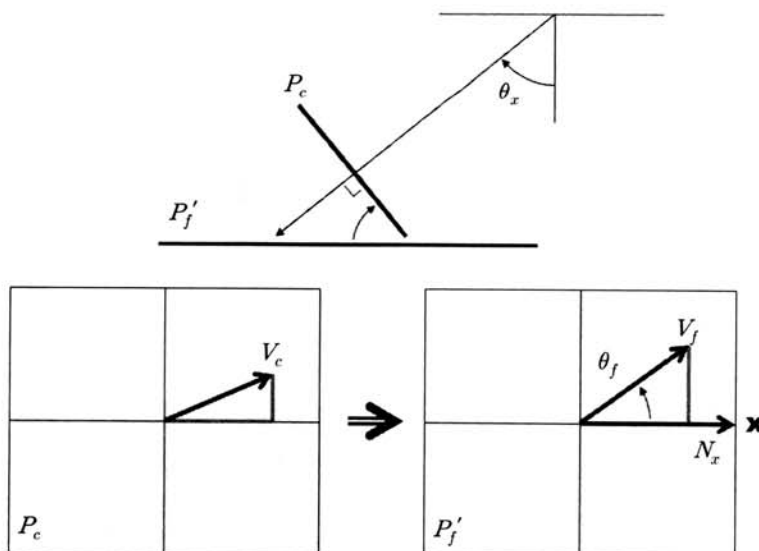
4.3 손동작을 동작 평면으로 투영

추정된 동작 평면과 카메라 이미지 평면 사이에 회전 관계는 4.1절의 식 (7)에 정의되었다. 카메라 이미지에 그려지는 손동작의 궤적은 동작 평면에 그려진 인간의 손동작 궤적이 카메라 이미지 평면으로 투영된 것이라 볼 수 있다. 실세계에 존재하는 카메라 평면 P_c 를 카메라 이미지 I_c 상에 투영시키는 변환을 T_c 라고 하면 식 (9)와 같다.

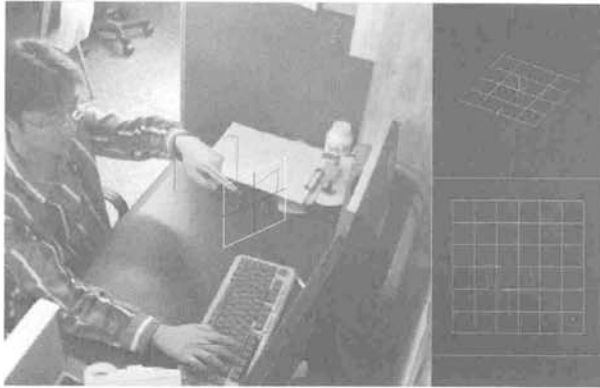
$$I_c = T_c P_c \quad (9)$$

최종적으로 카메라 이미지를 동작 평면으로 투영하기 위해 식 (10)과 같은 변환을 사용해서 복원하였다. 카메라 이미지상의 임의의 점은 식 (10)을 사용해 동작 평면상에 투영할 수 있으므로 카메라 이미지에 촬영된 동작 궤적을 동작 평면으로 투영하면 (그림 14)와 같이 손동작 궤적을 복원할 수 있다.

$$T_{m \rightarrow c} = R_{\theta_z} R_{\theta_x} R_{\theta_r} M_m, P_m = T_{m \rightarrow c}^{-1} T_c^{-1} I_c \quad (10)$$



(그림 12) 이미지 평면의 지시 방향을 바닥 평면으로 투영

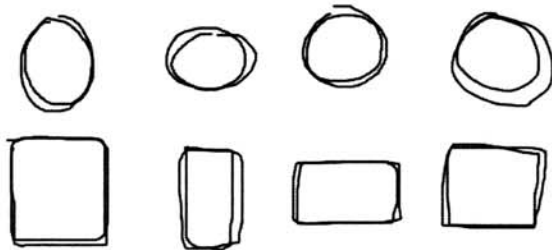


(그림 14) 손동작의 궤적을 동작 평면에 투영한 결과

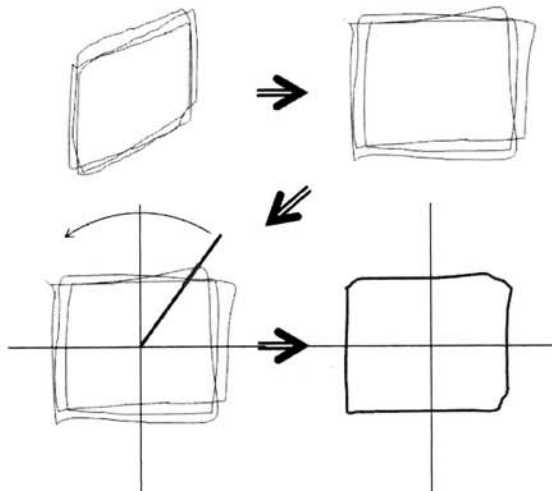
4.4 손동작 기록

인간이 그리는 손동작의 궤적이 어떠한 도형을 이룰 때 가로 세로 비율과 그 크기 비율에 관계없이 형태에 따라 같은 뜻을 가진다는 특징이 있다. (그림 15)는 서로 다른 시간에 원과 사각형을 허공에 그린 동작의 궤적이다.

그래서 근본적으로 손이 그리는 궤적은 가로 세로 비율에 관계없이 정규화 된 크기로 학습이 수행되고 비교 평가가 이루어져야 한다. 따라서 동작 평면에 투영된 동작 궤적은 가로 세로의 비율과 회전에 관계없이 비교가 가능해야 한다.



(그림 15) 인간이 허공에 원을 그리는 동작(위)과 사각형을 그리는 동작(아래)



(그림 16) 손동작 궤적의 정규화 과정

이를 위해 사용할 수 있는 정보들은 시작과 끝의 프레임 번호와 해당 프레임에서의 위치, 시작점과 끝점의 위치이다.

우선 최초 궤적의 학습을 위해 몇 번의 반복된 같은 동작의 궤적을 얻는다. 수집된 내용의 시작과 끝 프레임 안에서 가로의 위치 차이와 세로의 위치 차이를 계산한 후 이 두 차이를 1:1에 가깝게 정규화 시킨다. 그리고 중심 위치를 탐색한 후 중심에서 회전하면서 각 궤적의 거리를 기록한다. 이 때 반복된 몇 번의 동작들에 대해 평균 필터를 적용하여 궤적을 정리하며 최종적인 궤적은 회전각에 대해 정규화 된 거리를 기록하여 완성시킨다.

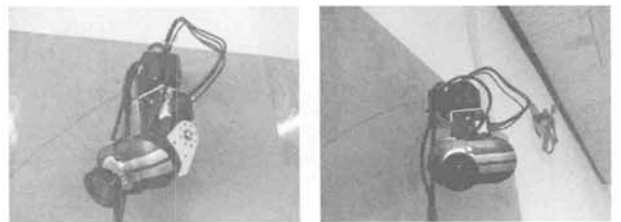
이러한 동작의 학습은 여러 위치와 방향에서 수행되고 다시 비교되는데 서로 다른 방향에서 수행된 궤적은 각 회전각에 대한 평균 거리와 분산으로 기록된다. 최종적인 손동작 μ 는 각 회전각 θ 에 대한 길이 l 과 분산 δ 를 가진 벡터의 집합으로 표현할 수 있으며 식 (11)과 같다.

$$\mu = \{(\theta, l, \delta)_t \mid 0 \leq t \leq 1\} \quad (11)$$

5. 실험 결과

실험은 임의의 방향으로 동작 평면을 추정하고 해당 동작 평면에 일정한 각도로 사선을 그었을 때 과연 얼마나 정확하게 복원되는가를 측정함으로써 동작 평면의 신뢰성을 확인하였다. 우선 약 3m 높이의 한쪽 벽면에 2개의 축을 가지는 입력 카메라를 (그림 17)과 같이 제작하여 부착하였다. 그리고 정확한 각도의 사선을 긋기 위해 (그림 18)과 같이 간단한 로봇손을 제작하였다.

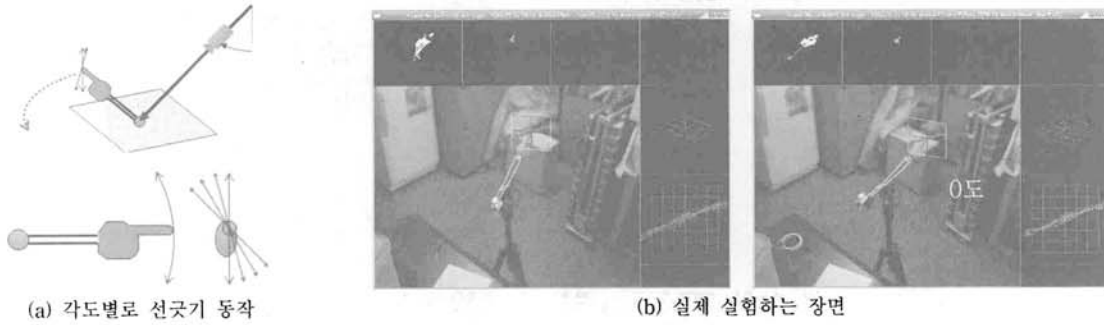
로봇손은 ± 10 도의 범위로 회전하여 선긋기를 수행하고 월드 좌표의 Z축을 기준으로 15도씩 회전해가며 모든 각도에서의 오차를 계산할 수 있게 도와준다. (그림 19)(a)와 같이 월드의 Z축으로 0도에서 90도 사이를 15도씩 회전해가며 X축에 대해 0도에서 90도 사이를 15도씩 회전한 선상에 선긋기를 수행하여 누적하고, 그 궤적에 근사한 선의 각



(그림 17) 벽면에 부착된 입력 카메라



(그림 18) 정확한 각도로 선을 긋는 로봇손



(그림 19) 선긋기를 실험하는 장면

<표 1> 각 각도별 최대, 최소 각도 오차

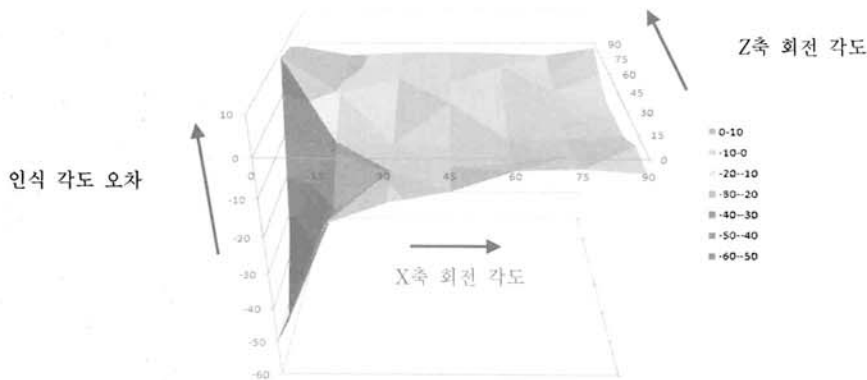
최소 / 최대	Z 0°	Z 15°	Z 30°	Z 45°	Z 60°	Z 75°	Z 90°
X 0°	-109.6 / 9.8	-29.9 / -2.8	-16.0 / -5.6	-15.2 / -0.8	-6.0 / 0.1	-5.6 / 0.4	-7.2 / 0.2
X 15°	-102.6 / -4.7	-32.9 / -8.8	-13.6 / -5.2	-11.9 / -5.2	-10.2 / -4.7	-7.5 / -2.9	-3.2 / -2.7
X 30°	-60.2 / -7.3	-8.4 / -7.7	-10.1 / -7.4	-13.2 / -5.1	-10.2 / -5.1	-8.7 / -5.5	-7.3 / -5.6
X 45°	-6.7 / 24.5	-1.1 / 0.1	-6.3 / -4.8	-5.6 / -0.5	-7.5 / -5.1	-7.6 / -5.2	-7.2 / -6.8
X 60°	-4.6 / 19.3	-1.7 / 2.5	-5.1 / 2.0	-4.0 / -1.1	-5.2 / -5.0	-5.3 / -4.8	-9.8 / -5.3
X 75°	-8.9 / 15.1	-4.8 / 6.3	-4.7 / 1.8	-5.1 / 0.0	-5.1 / 0.2	-8.0 / -2.3	-6.7 / -2.9
X 90°	-15.1 / 0.2	-16 / 1.1	-5.9 / 0.3	-7.2 / 1.9	-5.8 / 0.1	-6.9 / -0.2	-2.4 / -0.2

도 차이를 최소와 최대 차이로 나누어 <표 1>과 같이 기록하였다.

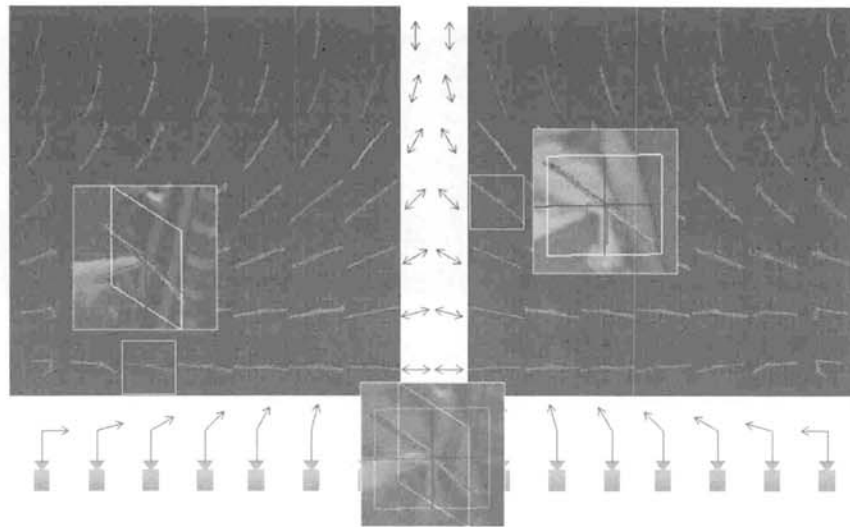
최대 최소 오차가 벌어질수록 그리고 그 오차 평균이 클수록 신뢰성 없는 각도가 된다. 그런데 Z축 회전이 0도일 때는 카메라 이미지 상에서 동작 평면의 폭이 거의 선에 맞힐 정도로 좁게 표시되므로 측정 결과에 신뢰성이 없지만 그 외의 각도에서는 15도에서의 몇 경우를 제외하고 0에서 -10도 내의 오차를 보이고 있어 매우 안정적으로 작동한다. 평균 오차 값만으로 볼 때 신뢰할 수 있는 영역은 (그림 20)과 같이 3차원 표면 그래프로 나타낼 때 연두색과 하늘색 영역이며 대다수의 경우가 10도 차이의 신뢰 범위 안에 들어감을 확인할 수 있다.

(그림 21)는 Z축의 회전 각도에 따라 X축으로 회전 시킨 선을 그었을 때 동작 평면을 사용하여 복원한 궤적의 결과를 모아놓은 것 중 서로 다른 각도에서 선긋기를 수행했지만 카메라 이미지 상에서는 동일한 궤적을 나타내는 일부의 경우이다. 이러한 경우 동작 평면을 사용해야 제대로 구분할 수 있다.

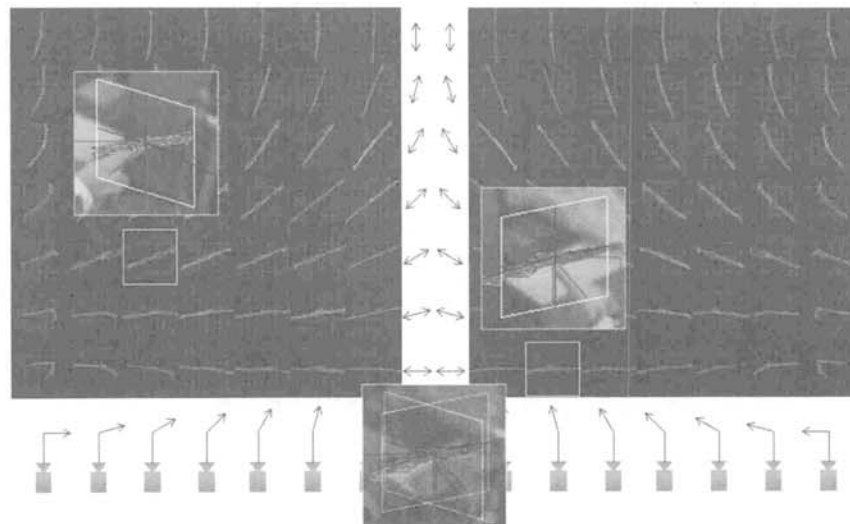
(그림 22)은 서로 다른 방향에서 취한 동작이 카메라 이미지 상에서는 구분하기에 애매하지만 최초 지시 방향을 기준으로 추정된 동작 평면을 사용할 경우 확실히 구분되는 한 예이다.



(그림 20) 신뢰할 수 있는 작업 영역

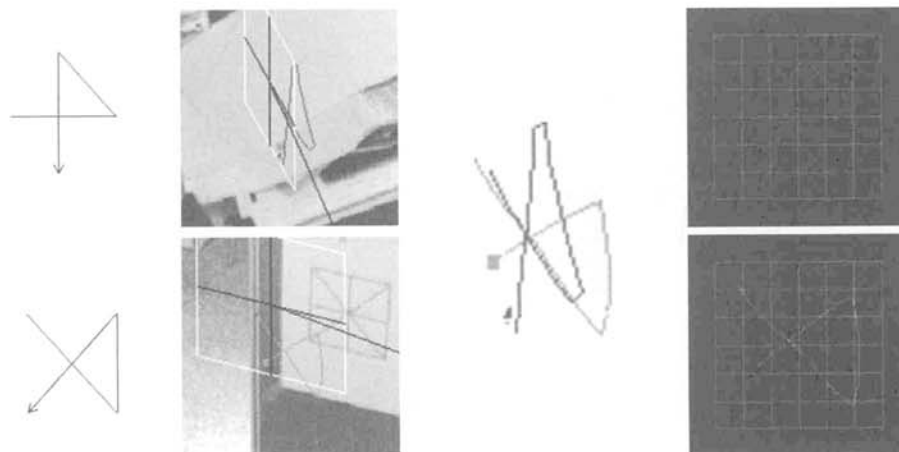


(a) Z축 15도, X축 0도와 Z축 90도, X축 45도



(b) Z축 30도, X축 30도와 Z축 105도, X축 0도

(그림 21) 서로 다른 각도이지만 같게 보이는 경우들



(그림 22) 카메라 이미지 상에서 구분하기 힘든 동작

6. 결론 및 향후 과제

본 논문은 모터가 장착된 입력 카메라를 회전시켜 방안에 입력 대상을 촬영하고 임의의 각도로 손동작을 수행할 때 올바르게 복원할 수 있도록 동작 평면의 개념을 도입하여 궤적을 투영하고 기록하는 방법을 제안하였다. 기존의 연구들이 한 방향에 대한 입력이나 좁은 작업 영역을 가진 것에 비해 상대적으로 넓은 공간의 영역을 입력 대상으로 단일 카메라를 사용하여 충분한 신뢰도를 가짐을 확인하였고, 약 82%의 작업 영역에 대해 높은 정확도로 궤적을 복원하였으며 카메라 보정과 같은 초기화 작업 없이도 신뢰성이 있음을 입증하였다.

최종 시스템은 펜티엄4 2GHz의 PC에서 약 17fps의 처리 속도를 보였으며, 가전기기를 수직적으로 가깝게 배치하지 않고 카메라를 사용자의 뒤쪽과 측면 천장에 부착하지만 않는다면 일반적으로 바로 적용할 수 있는 시스템을 제작하였다.

그러나 입력 대상을 추적하기 위해 활동 영역을 분할하는 방법의 적절함이 검증되지 않았고 손끝을 찾기 위해 단순히 무게중심에서 먼 지점을 선택하였기 때문에 정확하게 찾지 못하는 한계가 있다. 향후 보다 정확한 판단을 위해 Andrew Wu 등[15]의 A Virtual 3D Blackboard 시스템같이 경계 패턴을 이용한 손끝 판단이나 Xiying Wang 등[5]이 제안한 파티클 필터를 이용한 손바닥과 손끝 판단 기법등과 같은 연구가 필요하다.

그리고 카메라가 매우 높은 곳에 설치되거나 반대로 아주 낮은 곳에 설치된다면 특정 방향의 동작이 이미지 평면상에서 매우 작아져서 신뢰성을 잃게 되는 문제점도 가진다. 이는 향후 광각 카메라와의 협조 과정을 통해 해결할 수 있는 방안을 연구할 것이다.

참 고 문 헌

- [1] Paul P. Maglio, Teenie Matlock, Christopher S. Campbell, Shumin Zhai and Barton A. Smith, "Gaze and Speech in Attentive User Interfaces," Proceedings of the International Conference on Multimodal Interfaces, pp.1-7, 2000.
- [2] Barry Brumitt and JJ Cadiz, "Let There Be Light: Comparing Interfaces for Homes of the Future," Proceedings of INTERACT'01, pp.375-382, 2001.
- [3] Kai Nickel, Edgar Seemann and Rainer Stiefelhagen, "3D-Tracking of Head and Hands for Pointing Gesture Recognition in a Human-Robot Interaction Scenario," IEEE International Conference on Automatic Face and Gesture Recognition, pp.565-570, 2004.
- [4] Qi Jie Zhao, Da Wei Tu, Da Ming Gao and Ren San Wang, "Human-Computer Interaction Models and Their Application in an Eye-Gaze Input System," Proceedings of 2004 International Conference on Machine Learning and Cybernetics, pp.2274-2278, 2004.
- [5] Xiying Wang, Xiwen Zhang and Guozhong Dai, "Tracking of deformable human hand in real time as continuous input for gesture-based interaction," Proceedings of the 12th ACM International Conference on Intelligent user interfaces, pp.235-242, 2007.
- [6] Odest Chadwicke Jenkins, German González and Matthew Maverick Loper, "Tracking Human Motion and Actions for Interactive Robots," Proceeding of the ACM/IEEE international conference on Human-robot interaction, pp.365-372, 2007.
- [7] Andy Cassidy, Dan Hook and Avinash Baliga, "Hand Tracking using Spatial Gesture Modeling and Visual Feedback for a Virtual DJ System," Proceedings of the 4th IEEE International Conference on Multimodal Interfaces, pp.197-202, 2002.
- [8] N. Soontranon, Supavadee Aramvith and Thanarat H. Chalidabhongse, "Improved Face and Hand Tracking for Sign Language Recognition," Proceedings of the International Conference on Information Technology: Coding and Computing, Vol.2, pp.141-146, 2005.
- [9] Kazuhiko Sumi, Akihiro Sugimoto, Takashi Matsuyama, Masato Toda and Sotaro Tsukizawa, "Active Wearable Vision Sensor: Recognition of Human Activities and Environments," Proceedings of the International Conference on Informatics Research for Development of Knowledge Society Infrastructure, pp.15-22, 2004.
- [10] Yang Liu and Yunde Jia, "A Robust Hand Tracking and Gesture Recognition Method for Wearable Visual Interfaces and Its Applications," Proceedings of the Third International Conference on Image and Graphics, pp.472-475, 2004.
- [11] Satoshi Yonemoto, Hiroshi Nakano and Rin-ichiro Taniguchi, "Avatar Motion Control by User Body Postures," Proceedings of the eleventh ACM international conference on Multimedia, pp. 347-350, 2003.
- [12] Jakub Segen and Senthil Kumar, "Gesture VR: vision-based 3D hand interace for spatial interaction," Proceedings of the sixth ACM international conference on Multimedia, pp.455-464, 1998.
- [13] Le Song and Masahiro Takatsuka, "Real-time 3D Finger Pointing for an Augmented Desk," Proceedings of the Sixth Australasian conference on User interface, Vol.40, pp.99-108, 2005.
- [14] Álvaro Cassinelli, Stéphane Perrin and Masatoshi Ishikawa, "Smart Laser-Scanner for 3D Human-Machine Interface," CHI'05 extended abstracts on Human factors in computing systems, pp.1138-1139, 2005.
- [15] Andrew Wu, Mubarak Shah and Niels da Vitoria Lobo, "A Virtual 3D Blackboard: 3D Finger Tracking Using a Single Camera," IEEE International Conference on Automatic Face and Gesture Recognition, pp.536-543, 2000.

- [16] Douglas Chai and King Ngan, "Face Segmentation Using Skin-Color Map in Videophone Applications," IEEE Transactions on Circuits and Systems for Video Technology, pp.551-564, 1999.
- [17] Qiang Zhu, Ching-Tung Wu, Kwang-Ting Cheng and Yi-Leh Wu, "An adaptive skin model and its application to objectionable image filtering," Proceedings of the 12th annual ACM international conference on Multimedia, pp.56-63, 2004.



정 승 대

e-mail : sdjeong@vr.knu.ac.kr
 2005년 경북대학교 컴퓨터공학과(공학사)
 2007년 경북대학교 컴퓨터공학과(공학석사)
 2008년~현 재 경북대학교 전자전기컴퓨터
 학부 박사과정
 관심분야 : Human-Computer Interface,
 Augmented Reality, Computer
 Vision



장 경 호

e-mail : khjang@knu.ac.kr
 1998년 경성대학교 전산통계학과(이학사)
 2000년 경북대학교 컴퓨터공학과(공학석사)
 2005년 경북대학교 컴퓨터공학과(공학박사)
 2005년~현 재 경북대학교 전자전기컴퓨터
 학부 연구 교수

관심분야 : Image-Based Modeling and Rendering, Virtual
 Reality, Computer Graphics, Computer Vision.



정 순 기

e-mail : skjung@knu.ac.kr
 1990년 경북대학교 컴퓨터공학과(공학사)
 1992년 한국과학기술원 전산학과(이학석사)
 1997년 한국과학기술원 전산학과(공학박사)
 1997년~1998년 University of Maryland,
 Research Associate

2001년~2002년 IRIS, University of Southern California, Research
 Associate
 1998년~현 재 경북대학교 컴퓨터공학과 교수
 1999년~현 재 (주)아이디스 기술고문
 관심분야 : Virtual Reality, Artificial Intelligence, Computer
 Vision, Image Processing, Computer Graphics