
지능형 환경을 위한 실시간 신체 부위 추적 시스템

조명 및 복장 변화에 강한 신체 부위 추적 시스템

A real-time robust body-part tracking system for intelligent environment

정진기, Jinki Jung*, 조규성, Kyusung Cho**, 최진, Jin Choi**, 양현승, Hyun S. Yang***

요약 본 논문은 지능형 환경에 적합하도록 사용자의 자유를 제한하지 않는 실시간 신체 부위 추적 시스템의 구현에 관해 다룬다. 본 시스템은 기존의 신체 부위 추적 시스템과 달리 소매의 길이에 제약 없이 긴 소매, 반 소매 모두 인식할 수 있도록 함으로써 시스템의 generality 를 높였다. 정확한 인식을 위하여, 한 대의 카메라에서 얻어진 이미지에서 손, 머리, 발의 신체 부위를 따로 검출하고 각 신체 부위를 검출할 때 해당 부위에 적합한 특징점을 따로 선택하여 검출하였다. 검출된 2 차원의 신체 부위들을 calibrate 된 카메라를 이용하여 3 차원 자세로 변환하였다. 본 시스템은 실험을 통해 실시간(50fps) 성능으로 향상된 손 추적 성능을 보여주었다.

Abstract We proposed a robust body part tracking system for intelligent environment that will not limit freedom of users. Unlike any previous gesture recognizer, we upgraded the generality of the system by creating the ability the ability to recognize details, such as, the ability to detect the difference between long sleeves and short sleeves. For the precise each body part tracking, we obtained the image of hands, head, and feet separately from a single camera, and when detecting each body part, we separately chose the appropriate feature for certain parts. Using a calibrated camera, we transferred 2D detected body parts into the 3D posture. In the experimentation, this system showed advanced hand tracking performance in real time(50fps).

핵심어: *Body-part tracking, Intelligent environment, 3D posture*

본 논문은 2008 년 한국과학기술원 학술 연구비 지원에 의하여 연구되었음.

*주저자 : 한국과학기술원 전산학과 석사 과정 e-mail: jk@paradise.kaist.ac.kr

**공동저자 : 한국과학기술원 전산학과 박사 과정 e-mail: qtboy@paradise.kaist.ac.kr

**공동저자 : 한국과학기술원 전산학과 박사 과정 e-mail: jin_choi@paradise.kaist.ac.kr

***교신저자 : 한국과학기술원 전산학과 교수; e-mail: hsyang@kaist.ac.kr

1. 서론

지능형 환경은 지능을 가진 에이전트가 센서를 통해 인간의 언어, 행위, 의도를 이해하여 인간이 하려는 일을 도와주는 환경을 의미한다. 지능형 환경은 공간의 문맥을 인식하여 인간의 행위를 분석, 그들의 선호도를 알아내어 인간에게 필요한 서비스를 제공한다. 이를 위하여 지능형 환경은 인간의 위치와 자세를 감지할 수 있는 능력을 갖추어야 한다. 인간에게 서비스를 제공하기 위해선 사용자와 환경 사이의 상호작용 또한 중요한 이슈이다. 본 논문은 이에 맞춰 사용자의 위치와 행동을 인식하여 가상의 에이전트 들과 상호작용할 수 있는 신체 부위 추적 시스템을 제안한다.

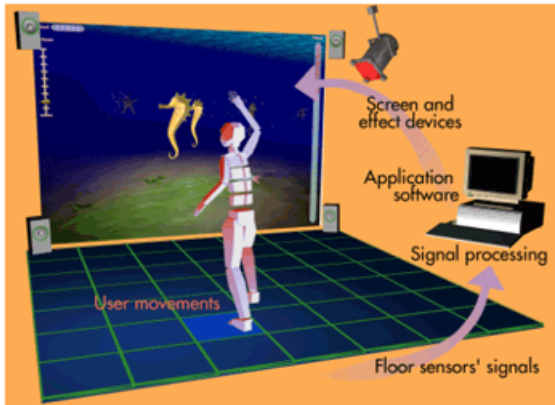


그림 2 지능형 환경의 구성도

본 논문에서 제안하는 시스템은 사용자의 자유를 제한하지 않으면서 3차원 자세를 추정하는 신체 부위 추적 시스템을 만드는 것을 목적으로 한다. 이 때 3차원 자세 추정은 한 대의 카메라로 이루어지며, 이는 실시간 성능(50fps 이상)을 보여야 한다. 먼저 사용자의 자유를 제한하지 않기 위하여 부수적인 장치(hand marker)를 제외, 조명이나 피복색에 관계 없이 추적을 할 수 있도록 하였다. 또한 기존의 신체 부위 추적 시스템과 달리 소매의 길이에 제약 없이 긴 소매, 반 소매 모두 인식할 수 있도록 함으로써 generality 를 높였다.

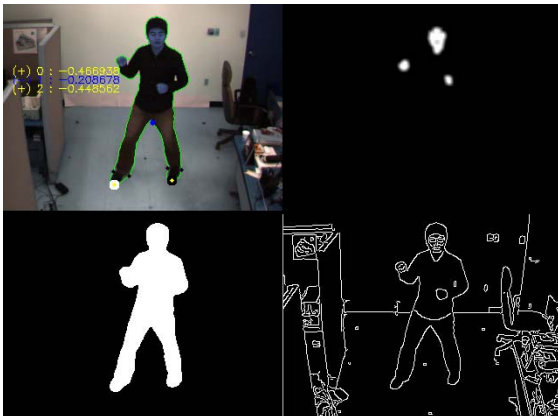


그림 1 사람 이미지에서 뽑을 수 있는 특징점들. 왼쪽 위부터 시계방향으로 Contour, Skin blob, foreground, edge 정보

더 정확한 제스처 인식을 위하여 한 대의 카메라에서 얻어진 이미지에서 손, 머리, 발 등의 신체 부위를 검출 시, 각 부위에 적합한 특징점을 따로 선택하였다. 모든 부위에 대하여 같은 특징점을 적용한다면, 각 신체 부위마다 가진 다른 성질(high variability of scale, shape, color 등)로 인해 검출기의 성능이 저하되기 때문이다.

본 논문의 구성은 다음과 같다. 2장에서 기존의 관련 연구들의 문제점과 개선할 점을 언급하고 3장에서 본 논문에서 제안하는 신체 부위 시스템을 상세히 설명한다. 4장에서 속도와 손 인식 정확도에 대한 성능 평가를 하고, 5장에서 결론을 맺는다.

2. 관련 연구

다른 센서나 마커를 사용하지 않고 오직 Vision 기술만을 이용하여 사람의 신체 부위를 인식하기 위한 노력은 꾸준히 있었다. 이러한 노력들은 크게 2D 신체 부위 추적 방식과 3D 신체 부위 추적 방식으로 나눌 수 있다. 본 논문에선 2D 방식의 대표적 연구로써 Wren[9], Siddiqui[4], 3D 방식 연구로써 Michilotta[5], Keskin[8]의 연구를 비교하였다. 각 연구의 장단점 및 제약사항은 아래 표와 같다.

논문	장점	단점	제약사항
Wren	전체 몸 에 대한 2D 트래킹 가능 트래킹 속도 성능 우수	손 겹침 현상 처리 불가	반 소매
Siddiqui	팔 인식 성능 뛰어나 트래킹 속도 성능 우수	손 인식 불가 검출이 400~700ms 걸림	반 소매
Michilotta	3차원 자세 추정 팔꿈치 위치 추정 가능	낮은 속도 성능 (8fps) 손 인식 성능 낮음	긴 소매, Stereo camera
Keskin	3차원 손 위치 추정 HMM을 이용한 제스처 인식	오프라인 시스템 사용자가 장갑 마커를 착용해야 함	긴 소매, Stereo camera

위의 표에 나와 있듯이 2D 방식은 속도 성능이 우수하지만 3D 방식은 속도 성능이 전반적으로 떨어져 있다. 또한 2D와 3D 방식 모두 전체 신체 부위 중 마커를 착용하지 않는 한, 손 인식에 있어서는 낮은 성능을 보인다. 그 이유는 손이 다른 신체 부위에 비해 빠른 움직임 속도와 넓은 움직임의 범위, 손과 손, 얼굴과 손의 겹침 현상 등의 문제점이 있기 때문이다. 또한 각 시스템마다 사용자가 입어야 할 옷을 반 소매, 긴 소매 등으로 제약사항을 걸어두었다.

손은 모든 제스처의 핵심으로써 사물을 지정, 변화, 생성할 때 쓰인다. 본 논문에서는 전체 신체 부위 중 손에 대한 검출, 추적을 강화시킴으로써 손의 다양한 움직임, 모양의 변화, 색에 관계 없이 손을 발견할 수 있도록 하였다. 이를

위해 검출과 추적을 동시에 실행하는 방식을 사용하였다. 또한 전체 팔 영역에 대한 손 인식에 있어 향상된 성능을 보이도록 Hand Energy Map 을 고안하여 사용하였다. 이는 이 전 연구의 제약 사항이었던 반 소매, 긴 소매와 상관 없이 손의 위치를 정확하게 찾도록 도와준다.

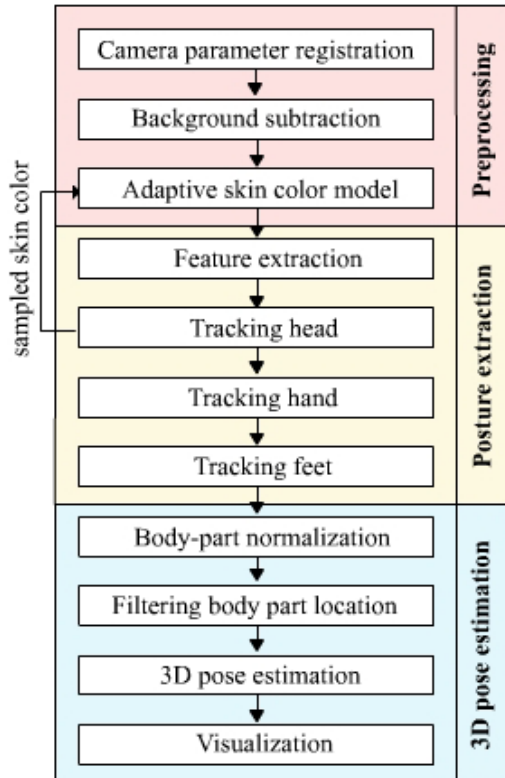


그림 3 신체 부위 추적 시스템 Overview

3. 신체 부위 추적 시스템

머리, 손, 발의 각 신체 부위를 효율적으로 추적하기 위한 시스템의 흐름은 다음과 같다. 각 단계의 결과는 다음 단계에 영향을 미친다.

3.1 Camera parameter registration

발의 3 차원 위치를 나타내기 위해서는 발의 3 차원 좌표가 실제 현실의 화면에 적절한 곳에 적절한 크기와 방향으로 합성되는 것이 일차적으로 필수적인 요건이라 할 수 있다. 이를 위해서는 카메라를 통해 입력되는 실제 환경에 대한 정보를 미리 얻고 그 정보를 밑받침으로 발의 위치나 크기를 변형시켜야 한다. 가상 3D 위치 정보를 현실 정보에 맞추는 과정을 registration 과정이라 하며 이를 위해서는 현실 배경의 공간에 대한 정보가 추가적으로 필요하다. 본 논문에선 tracking 단계 이전에 자체 제작한 틀을 이용하여 공간에 대한 정보로써 바닥의 정보를 뽑아내었다. 바닥의 정보를 자동적으로 얻을 수도 있으나 적은 수의 수동 입력만으로 필요한 만큼 정확한 정보를 제공하는 것이 가능하기 때문에 추가적인 제약을 가정하기로 한다.

바닥은 평면이기 때문에 평면을 특정하기 위해서는 같은 직선상에 있지 않은 최소 3 점의 좌표가 필요하다. 3 점의 화면상에서의 2 차원 좌표와 실제 공간상에서의 3 차원 좌표를 안다면 화면 안에서 평면을 특정 지을 수 있다. 측정 시 오차가 있기 때문에 최소한의 점만을 제공하게 될 경우 실험 결과에 영향을 줄 정도로 오차가 상당히 크다. 따라서 본 논문에선 총 9 점의 좌표를 제공하였다. 얻어진 9 점의 좌표로써 Zhengyou Zhang 의 방법[2]을 사용하여 internal parameter, external parameter 를 얻었다.

3.2 Background subtraction

카메라로부터 연속된 영상이 주어지면 우선 앞으로 추적과 인식의 대상이 될 객체들을 검출하고 배경과 분리한다. 특히 객체를 추적하거나 행동을 모델링 할 때 검출된 객체의 외향에 의존하기 때문에 정확하고 빠르게 객체를 검출하는 것이 중요하다. 하지만 입력으로 들어오는 원영상에는 조명의 변화나 카메라의 이동, 배경으로 취급되어야 하는 물건의 움직임 등의 문제가 생기기 때문에 원하는 대상만을 완벽하게 구분 짓는 일은 쉬운 일이 아니다. 따라서 본 논문은 실내에서 실험을 위하여 정리된 환경을 가정하고 실험하였다. 본 시스템은 원활한 제스처 인식을 위하여 화면에 움직이는 대상은 한 명의 사용자 뿐이라는 가정이 있다.

본 논문은 전경과 배경을 구분하는 방법으로써 Jin et al.[1]가 제안한 bottom up 방식의 quadtree-base hierarchical framework 을 적용한 MoG 방법을 사용하였다.

3.3 Adaptive skin color model

본 논문은 조명의 변화에 강인하게 피부색을 검출하기 위해서 adaptive skin color model 을 사용하였다. 조명 변화를 적용한 skin color model 을 만들기 위해 초기값으로 지정된 피부색 값을 이용하여 얼굴 및 손을 검출하고, 얼굴의 피부색 샘플을 뽑아내어 이를 학습시켰다. 본 논문에선 손과 얼굴을 성공적으로 뽑아내기 위하여 옷에 피부색이 포함되어 있지 않다는 가정을 두었다.

검은색과 피부색으로 이루어진 얼굴 이미지에서 피부색만을 뽑아내기 위해 K-means 클러스터링을 사용하여 얼굴을 주로 이루고 있는 피부색을 분리해낸다. 피부색이 뽑아지면 skin-color distribution 을 학습하기 위한 표본집합을 얻을 수 있다.

$$X = \{x_1, \dots, x_i, \dots, x_n\}, x_i = (h_i, s_i) \quad (1)$$

여기에서 n 은 샘플의 개수이고 각 샘플은 피부색의 Hue 와 Saturation 값으로 이루어져 있다. 피부색에 대한 확률적 모델을 만드는 방법으로써 이산 히스토그램이나 mixture of Gaussian 등의 통계적 모델을 비교 및 실험해

본 결과, Gaussian 모델이 시간과 성능 면에서 가장 좋은 결과를 나타냈다.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \Sigma = \frac{1}{n} \sum_{i=1}^n \{(x_i - \bar{x}), (x_i - \bar{x})'\} \quad (2)$$

표본 집합들에서 한번 x 과 시그마에 대한 값을 찾아내면, 새로운 피부색을 찾아낼 확률을 다음과 같이 계산할 수 있다.

$$P(X_{is_skin}) = \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} \exp\left\{-\frac{1}{2} \{(x_i - \bar{x}) \Sigma^{-1} (x_i - \bar{x})'\}\right\} \quad (3)$$

이 2 차 Gaussian 모델을 적용한 skin color model 은 조명의 변화나 개인의 피부색 차이에 관계 없이 좋은 피부색 검출 결과를 보여주었다.

3.4 Tracking head

본 논문은 다른 신체 부위 인식의 기초 특징점으로써 얼굴을 선택하였다. 그 이유는 얼굴의 텍스처가 가장 특징적이고 조명 변화에 대해 안정적인 특징점이기 때문이다. [7] 따라서 본 시스템은 초기에 특정 포즈를 잡을 필요 없이 한번이라도 얼굴이 화면 내에 검출이 되면 그때부터 모든 검출 및 추적 과정이 시작된다.

얼굴 검출 방식은 Cascade of boosted classifiers working with haar-like features[3]를 이용한 얼굴 검출기를 사용하였다. 얼굴 검출기에 사용할 텍스처는 두 눈과 볼 사이의 명암의 차이, 두 눈 사이의 코 영역에 대한 명암의 차이를 나타내는 두 특징점을 AdaBoost 를 이용하여 선택하였다.

한 번 얼굴이 검출되고 난 후엔 얼굴에서 색을 검출한다. 이는 두 가지 목적이 있는데 첫째로는 얼굴이 포함된 머리 부분을 추적하기 위하여 particle filter 를 적용할 초기값을 넣어주는 것이고 둘째로는 동적으로 피부색 모델을 만들기 위함이다.

얼굴 색에 대한 초기값이 정해지면 particle filter 로 예측을 하는데 이 때 각 particle 들의 성능을 측정하기 위한 척도로써 Mean Squared Error 를 사용한다. MSE 를 비교하여 가장 우수한 particle 을 고른 후, 이것이 임계값을 넘으면 안정적인 특징점으로 삼는다. Particle 이 임계값을 넘지 않으면 이를 사용하지 않고 얼굴 검출 단계로 넘어가 얼굴 텍스처를 찾는다.

3.5 Tracking hands

피부색 모델이 정해지면 이를 이용하여 foreground 에서 skin blob 들을 검사한다. 이 때 피부색과 비슷한 노이즈 값을 제거하기 위하여 모폴로지 연산(open 과 close)을 행한다. 연산 후엔 open 연산으로 인해 끊어진 blob 들을

잇기 위하여 blur 를 시킨다. 이는 blob 픽셀 수가 적은 noise 값들보다 blob 픽셀 수가 큰 true blob 들을 이어주는 역할을 한다.

처음에 Skin color 만 검출된 이미지를 받아오면 이를 Connected component method 를 사용, skin blob 들로 묶는다. Skin blob 들이 다 나오게 되면 이들 중 일정 임계값 이하의 크기가 작은 blob 은 noise 로써 취급하여 무시한다.

Threshold 값으로 걸러낸 Large blob 들의 갯수가 2 개 이하라면 손과 손, 손과 머리가 occlusion 이 일어난 상황이므로 검출 실패 메시지를 발생시킨다. 검출이 실패하면 기존에 가지고 있었던 constant velocity model 을 그대로 적용한 추적 정보를 유지한다. Hand blob 갯수가 3 개일 때가 두 손과 머리가 발견된 이상적인 상황이라 할 수 있다. 이 경우에 blob 의 위치가 얼굴의 위치와 같다면 그 blob 은 무시하고, 다른 두 blob 들 중에서 왼손과 오른손을 선택하도록 한다.

검출 단계에서 나온 두 손의 위치 값을 tracking 단계의 초기값으로 쓰인다. Tracking 단계에선 이전 2 단계의 프레임들 사용하기 때문에 3 프레임 동안 검출이 성공해야만 추적 모듈이 동작한다. Hand tracking 단계는 하나지만 두 손에 대하여 각자 따로 추적 한다. 추적 모듈이 시작되면 시간 t 에 해당되는 프레임에서 손에 해당되는 skin color blob 을 결정하기 위해 skin color 검출을 이용하는데 이 위치는 $t-1$ 시간의 state 로부터 정해진 hand state hypothesis 에 의해 결정된다.



그림 4 팔 영역에서의 손 모양 Edge 정보

사용자가 반팔을 입었을 경우에 이 전과 같은 알고리즘을 계속 적용하면, 손 검출기가 팔의 피부색을 검출하기 때문에 손의 위치가 엉뚱한 위치로 인식된다. 이와 같은 오류 때문에 기존의 시스템들은 긴 소매 옷, 혹은 반 소매 옷으로 인식에 대한 제약사항을 두었다. 본 논문은 전체 팔 영역 중 유독 손 부분만 고주파 정보(edge 영역)가 많다는 것을 착안하였다. 그림 3 의 그림을 보면 손이 어떤 모양이든지 항상 다른 영역에 비해 edge 정보가 많음을 알 수 있다.



그림 5 Skin blob 과 edge 정보를 합친 Hand Energy Map

본 논문에서는 이에 canny edge 검출기를 사용, edge 정보를 추출하여 이를 skin blob 정보와 합친 Hand Energy Map 을 만들었다. 이 HEM 을 이용하여 위 검출과 추적 알고리즘에서 손 영역의 blob 을 찾을 때 평가의 척도로써 사용하였다. High frequency 정보를 h, skin blob 의 정보를 s 라 했을 때 손 영역을 구하는 Hand Energy Map 은 (4)의 수식과 같다.

$$\max \left(\sum \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{(x-x_a)^2}{2\sigma_x^2} + \frac{(y-y_a)^2}{2\sigma_y^2} \right)} + h \right) \right) \quad (4)$$

3.6 Tracking feet

Foreground 영상에서 얻어진 사람의 윤곽선 정보의 아래 1/3 영역, 하반신에 해당하는 선을 얻은 후 윤곽의 무게 중심에서부터의 거리를 측정해보면 3 개의 local maximum 을 구할 수 있다. 이 3 개의 꼭지점에 해당되는 부분이 볼록점 혹은 오목점인데 이 그래프만으로는 이를 구별할 수 없다. 따라서 꼭지점 전 후에 있는 임의의 두 점을 선택, 외적을 취하여 얻은 벡터가 양의 값이면 볼록점, 음의 값이면 오목점으로써 분류한다. 이 중 최대 2 차 미분 값을 갖는 볼록점을 두 발로써 사용한다.

본 논문은 발의 위치에 이어 발의 각도를 구하기 위하여 발 윤곽선에서 어느 일정 범위의 2 점을 잡고 이를 이루는 내각을 2 등분 하는 각도를 발의 각도로써 사용하였다. 또한 두 발의 각도를 보간하여 몸이 바라보는 방향 벡터도 구할 수 있었다. 그림 6 에 몸의 3 차원 위치를 나타내는 녹색 십자 표시와 몸의 방향 벡터가 파란 화살표로 나타나 있다. 또한 이 화면에서 발의 위치 및 각도는 2 차원이 아닌 3 차원 좌표임을 볼 수 있다.

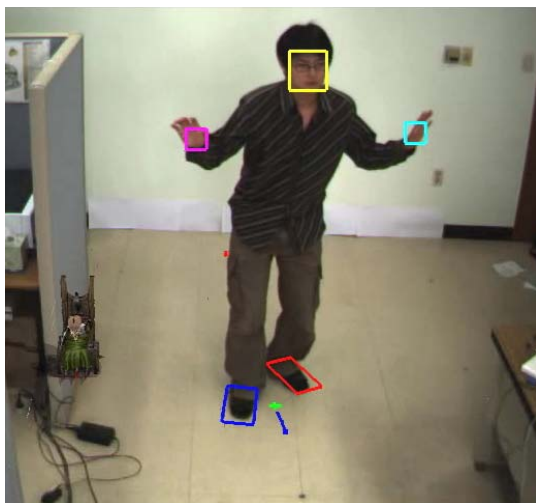


그림 6 실시간 신체 부위 추적 시스템 동작 화면

3.7 3D pose estimation

발의 3 차원 위치를 나타내기 위해서는 발의 3 차원 좌표가 실제 현실의 화면에 적절한 곳에 적절한 크기와 방향으로 합성되는 것이 일차적으로 필수적인 요건이라 할 수 있다. 이를 위해서는 카메라를 통해 입력되는 실제 환경에 대한 정보를 미리 얻고 그 정보를 밑받침으로 발의 위치나 크기를 변형시켜야 한다. 이를 위해서 현실 배경의 공간에 대한 정보(바닥 평면에 대한 정보)를 추가적으로 제공하였다.

얻어진 각 신체 부위의 위치를 정규화한 후에, 이를 카메라와 발 위치 간의 거리를 이용하여 3 차원으로 reconstruction 시킨다. 양 발의 위치(f_1, f_2)를 기준으로 바닥과 수직인 평면 π 를 구하기 위해 양 발 f_1, f_2 그리고 f_1 의 Z 값만을 수정한 f_3 로 이루어진 $X = (f_1, f_2, f_3)$ 의 right null space 를 구한다. 이를 구하기 위해 X 에 SVD 를 취한 뒤, SVD 의 결과를 $X = UDV^T$ 라 표현 했을 때 D 의 가장 작은 singular value 에 대응되는 V 의 column vector 가 π 의 값이 된다.

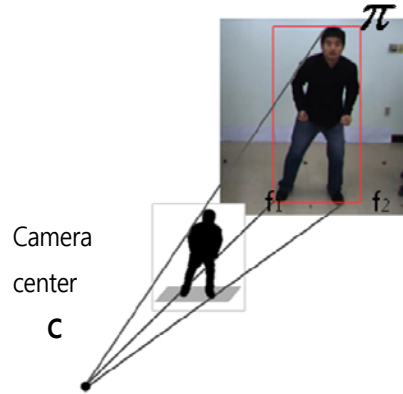


그림 7 3차원 자세 추정

카메라 중심값 C 로부터 화면 상에서의 손과 발, 머리 위치를 직선으로 이으면 그림 7 과 같다. 이 직선이 이전에 구한 평면 π 와 만나는 점을 계산하면 실제 사람의 3 차원 위치의 근사값을 추정할 수 있다.

4. 성능 평가

본 논문은 intel dual 2.20GHz CPU 를 가진 노트북을 실험 환경에서 시스템의 성능을 평가하였다. 성능 평가는 해상도 640 x 480 의 동영상에 대하여 측정되었다.

4.1 속도 성능

아래의 도표에 나와 있듯이 모든 신체 부위에 대한 검출 및 추적 과정이 dual 2.20GHz CPU 의 환경에서 총 40ms 밖에 걸리지 않았다. 이는 초당 프레임 수로 계산해봤을 때 대략 50 fps 로써 [4]의 10fps(Dual 3GHz Xeon), [5]의

16fps(2.8 GHz CPU)에 비해 월등히 빠른 속도이다. [6]의 48 fps (2.8 GHz CPU)와 비슷한 성능이다.

추적값	속도
배경 추출	14.4 ms
머리	3.5 ms
손	2.35 ms
발	2.82 ms
총 합계	23 ms

제한한 시스템이 타 시스템보다 빠를 수 있었던 이유는 매 단계를 거치면서 계산 영역을 좁혀갔기 때문이다. Background subtraction 단계 이후엔 연산 범위를 foreground 영역으로 좁혔고, 각 신체 부위를 검출 및 추적 할 때에는 검색 범위를 각 부위의 움직임 수 있는 최대의 영역에 맞게 줄였다. 특히 가장 시간이 오래 걸리는 이미지와 이미지 간의 연산의 수는 최대한 줄이고, 연산이 필요할 때엔 연산이 가능한 최소 영역 만을 범위로 잡았다. 또한 연산 시간을 최대한 줄이기 위해 constant velocity model, particle filter, Adaboost 등을 사용하였다.

4.2 손 추적 성능

모든 신체 부위 중에서 손은 오직 피부색만을 특징점으로 사용하기 때문에 추적 시 문제 상황이 자주 발생한다. 본 논문은 전체 추적 성능 중에 손을 성능 측정의 지표로 삼고 문제 상황을 3 가지로 정의하였다. 첫 번째는 Self occlusion 상황으로, 손과 손이 서로 겹쳤다가 따로 떨어지는 상황이다. 두 번째는 Crossing arms 상황이고 두 팔이 완전히 겹쳤다가 다시 원래대로 돌아가는 상황이다. 세 번째는 손과 색이 같은 얼굴에 손이 붙었다가 떨어지는 Occlusion with face 의 상황이다. 모두 skin blob 이 하나가 되었다가 두 개로 분리되는 상황이다.

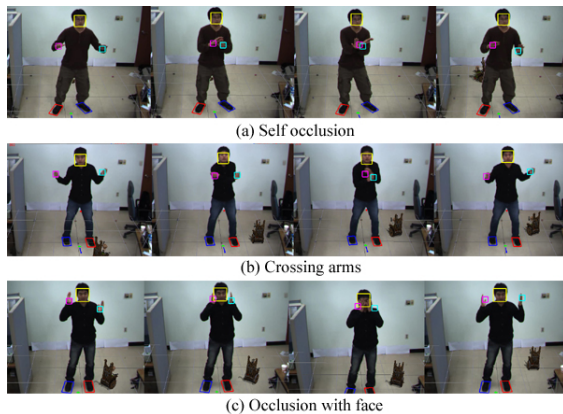


그림 7 긴 소매 상황에서의 추적 성능

그림 8 은 긴 소매일 때의 추적 성능을 나타낸다. 위에서 정의한 3 가지 문제 상황에 대하여 좋은 성능을 나타내고

있다. 두 손이 가까워지는 상황에서 검출을 하기 때문에 예측되는 추적 정보 값이 더 정확해 지기 때문이다. 또한 얼굴과 손이 붙는 경우에는 검출이 실패하므로 이전 정보를 그대로 유지하다가 얼굴과 손이 떨어질 때 다시 검출을 통해 손 위치를 추적함을 볼 수 있다.

그림 9 는 반 소매일 때의 추적 성능을 보여준다. 반 소매 일 때는 팔 영역 모두가 skin blob 으로써 그 영역 내에서 손 영역을 추적하는 것이 더 어려워진다. 하지만 그림에서 보듯이 3.5 장에서 제안하였던 Hand Energy Map 을 이용하였을 때, 긴 소매와 크게 추적 성능 차이가 나지 않을 정도로 우수한 손 위치 추적 성능을 보여준다.

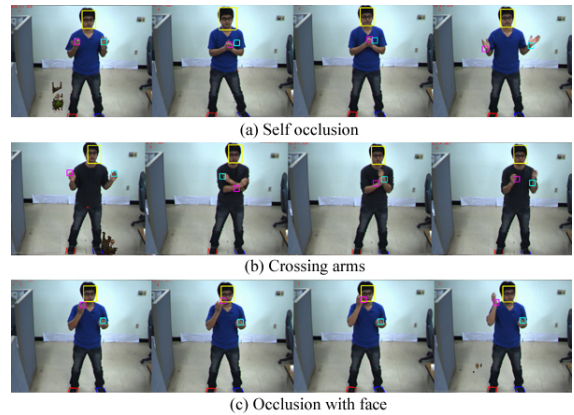


그림 6 반 소매 상황에서의 추적 성능

5. 결론

본 논문은 한 대의 카메라 환경에서 실시간으로 머리, 손, 발의 각 신체 부위의 3 차원 위치를 얻어내는 신체 부위 추적 시스템에 대하여 다루었다. 본 신체 부위 추적 시스템에선 각 신체 부위마다 다른 특징점을 사용하였다. 머리 부분은 얼굴에 대한 텍스처가 다른 특징점들에 비해 잘못된 검출(false detection)이 잘 나지 않기 때문에 특징점으로써 텍스처를 사용하였고, 손 부분은 넓은 움직임의 범위와 다양한 자세, 빠른 움직임 때문에 skin color 를 사용하였으며, 발 부분은 손에 비해 그 움직임이 크지 않고 항상 땅과 맞닿아 있으므로 윤곽선을 사용하였다. 본 논문에선 각기 다른 파트에 다른 특징점을 사용함으로써 검출 속도와 추적 성능을 더 높일 수 있었다. 또한 다른 시스템에서 문제 되었던 손에 대한 추적 성능을 향상 시켰음을 실험을 통하여 볼 수 있었다.

본 시스템은 손과 머리의 3 차원 근사값을 얻었지만 타 3D 방식의 단점이었던 처리 속도 문제를 해결하기 위해 검색 범위 축소, Adaboost, particle filter 등의 기법을 사용함으로써 빠른 수행 속도(50 fps)를 보일 수 있었다.

위에서 얻어낸 신체 부위의 위치를 neural network, HMM 등의 학습기에 학습을 시킨다면 더 향상된 제스처 인식이 가능할 것이다. 또한 본 논문에서는 아직 3D human model 을 사용하지 않고 있지만 확률적 모델을 도입하여 현재 포지션에 대하여 가능한 body configuration 을 계산한다면 실시간으로 팔꿈치의 3 차원

위치 및 전체 몸의 3 차원 자세까지 추정할 수 있으리라 기대한다.

Acknowledgement

본 연구는 문화체육관광부 및 정보통신연구진흥원의 IT 핵심기술개발사업의 일환으로 수행하였음. [과제관리번호: 2008-S-025-01, 과제명: 디지털 교과서 및 u-러닝 활성화를 위한 요소 기술 개발].

참고 문헌

- [1] Jin Choi, Yong-il Cho, Kyusung Cho, Su-jung Bae and Hyun S. Yang, "A View-based Multiple Objects Tracking and Human Action Recognition for Interactive Virtual Environment" , The International Journal of Virtual Reality, Vol. 7, No. 3, IPI Press, pp. 71~76, 2008.
- [2] Zhang,Z., "A flexible new technique for camera calibration", Transactions on Pattern Analysis and machine Intelligence, Vol.22, pp1330-P.1334,2000
- [3] R. Lienhart. "Cascade of boosted classifiers working with Haar-like features". In Proc.of 2nd Intl. Conf. on Aut.Face and Gesture Recognition, pages 236{241. 2002.
- [4] M. Siddiqui and G. Medioni. "Robust real-time upper body limb detection and tracking". In VSSN '06: Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks, pages 53-60, New York, NY, USA, 2006. ACM.
- [5] E.-J. O. Antonio S. Micilotta and R. Bowden. "Real-time upper body detection and 3d pose estimation in monoscopic images". Computer Vision - ECCV 2006, 3953:139-150,2006.
- [6] A. Micilotta and R. Bowden. "View-based location and tracking of body parts for visual interaction". In In Proc. of BMVC, pages 849-858, 2004.
- [7] Y. Ishii, H. Hongo, K. Yamamoto, and Y. Niwa. "Face and head detection for a realtime surveillance system". Pattern Recognition, International Conference on, 3:298-301, 2004.
- [8] C. Keskin, A. Erkan. Real time hand tracking and 3d gesture recognition for interactive interfaces using hmm. In ICANN/ICONIPP, 2003.
- [9] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentl. Pfinder: Real-time tracking of the human body. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19:780-785, 1997.