

AdaBoost 학습 알고리즘과 칼만 필터를 이용한 손 영역 탐지 및 추적

김병만*, 김준우**, 이광호***

An Application of AdaBoost Learning Algorithm and Kalman Filter to Hand Detection and Tracking

Byeong-Man Kim *, Jun-Woo Kim **, Kwang-Ho Lee ***

요 약

웨어러블 컴퓨터의 개발로 인해 인간과 컴퓨터간의 전통적인 인터페이스는 점차 사용하기 불편하게 되었고 이는 새로운 형태의 인터페이스에 대한 요구로 이어지게 되었다. 본 논문에서는 이러한 추세에 맞추어 디지털 카메라를 통해 인간의 제스처를 인식하는 새로운 인터페이스를 연구하였다. 카메라를 통해 손 제스처를 인식하는 방법은 빛과 같은 주변 환경에 영향을 받기 때문에 탐지기는 덜 민감해야 한다. 최근에 Viola 탐지기는 얼굴 탐지에 좋은 결과를 보여 주었으며, 이는 적분 이미지로부터 추출한 하일 특징을 이용한 Adaboost 학습 알고리즘을 사용하였다. 본 논문에서는 이 방법을 손 영역 탐지에 적용하였으며 피부색을 이용한 고전적인 방법들과 비교 실험을 수행하였다. 실험 결과는 빛과 같은 방해 요소가 있는 환경에서 Viola 탐지기가 피부색을 이용한 탐지 방법보다 더 견고함을 보여 주었다.

Abstract

With the development of wearable(ubiquitous) computers, those traditional interfaces between human and computers gradually become uncomfortable to use, which directly leads to a requirement for new one. In this paper, we study on a new interface in which computers try to recognize the gesture of human through a digital camera. Because the method of recognizing hand gesture through camera is affected by the surrounding environment such as lighting and so on, the detector should be a little sensitive. Recently, Viola's detector shows a favorable result in face detection, where Adaboost learning algorithm is used with the Haar features from the integral image. We apply this method to hand area detection and carry out comparative experiments with the classic method using skin color. Experimental results show Viola's detector is more robust than the detection method using skin color in the environment that degradation may occur by surroundings like effect of lighting.

▶ Keyword : AdaBoost Algorithm, Kalman Filter, Hand Detection, Hand Tracking

• 제1저자 : 김병만

• 접수일 : 2005.05.18, 심사완료일 : 2005.08.10

* 금오공과대학교 컴퓨터공학부 교수 ** 금오공과대학교 소프트웨어공학과 석사, *** 목포대학교 컴퓨터교육과 부교수

1. 서론

유비쿼터스(Ubiquitous)라는 용어는 물이나 공기처럼 시공을 초월해 '언제 어디에나 존재한다'는 뜻의 라틴어(語)로, 사용자가 컴퓨터나 네트워크를 의식하지 않고 장소에 상관없이 자유롭게 네트워크에 접속할 수 있는 환경을 말한다. 유비쿼터스의 등장으로 많은 분야들이 새로운 변화를 시작하거나 재조명 받기 시작했다. 그중 하나가 HCI(Human Computer Interface)분야이다. 유비쿼터스 환경, 특히, 입는 컴퓨터(Wearable Computer)에서는 마우스나 키보드 등 기타장비를 사용한 접근에 많은 한계를 가지게 되어 새로운 방식의 인터페이스가 절실히 요구되고 있다.

사람들은 의사소통의 수단으로 말, 행동, 문자 등 여러 가지 방식을 통하여 자신의 의사를 표현한다. 실제 심리학자들의 연구결과 [1]에 의하면 문자나 말인 언어수단은 실제 7%에 불과하고 기타 다른 방식, 얼굴표정이 55%, 제스처나 목소리 톤과 같은 준언어(Para-language)가 38%에 해당하고 있다. 지금까지 인간과 컴퓨터와의 인터페이스는 키보드, 마우스, 기능 버튼 정도가 대부분을 담당해 왔다. 보다 편리하고 다양한 인터페이스를 제공하기 위해 음성인식, 감정인식, 제스처 인식 등 다양한 연구가 이루어지고 있다. 이 중 제스처 인식의 한 분야인 손 영역 인식은 실생활에서 사물을 지정하거나 손동작으로 여러 가지 빠른 의사를 전달할 수 있으며 원거리에서도 사용할 수 있다는 장점 때문에 현재 많은 연구들이 진행되고 있다[2][3][4].

손 영역 획득방법에는 데이터 글로브를 착용하는 방식과 카메라를 이용하는 비전 방식으로 나눌 수 있다. 데이터 글로브는 광센서나 굴절 계측기를 글로브에 달아 손가락 하나 하나의 움직임까지 상세하게 알아낼 수 있는 장점이 있다. 하지만 사용을 하기 위해서는 착용해야 하며 장비의 가격이 비싼 단점들이 있다[2]. 그리고 카메라를 이용하는 비전 방식은 데이터 글로브보다 장착이 쉽고, 원거리에서도 획득이 가능하다는 장점이 있다.

카메라를 이용하는 비전 방식은 사용하는 카메라의 종류에 따라 세분할 수 있는데, 이 중 열 감지 적외선 카메라 방식은 사람이 항상 일정 체온을 유지 한다는 특징을 이용하는 방법으로, 실제 측정값은 거리 등을 감안하여 체온보

다 낮은 값을 임계값으로 이용하여 손 영역을 획득한다[3]. 다른 카메라에 비해 조명이나 배경에 영향을 받지 않는 장점이 있다. CCD 카메라를 이용하는 경우는 카메라의 개수에 따라 분류할 수 있다. 카메라를 두개 혹은 그 이상을 이용하는 방법과 하나의 카메라를 이용하는 방법으로 나눌 수 있다. 여러 개의 카메라를 이용하는 방식은 사람처럼 두 눈으로 물체가 멀리 있고 가까이 있음을 확인할 수 있는 방식으로 스테레오 카메라를 이용하는 방법이다[4]. 이러한 방식은 카메라를 나란히 왼쪽, 오른쪽에 두어 각각 카메라에 찍힌 영상의 변위를 이용하여 물체의 원근감을 알 수 있지만 카메라가 더 필요하고 두 영상의 변위나 제어에 필요한 계산량이 많고 복잡한 단점이 있다. 마지막으로 카메라 한 대를 이용하는 방법은 가장 일반적으로 사용할 수 있으며 값이 싼 장점이 있다. 하지만 조명과 배경의 영향을 많이 받고 정면 영상 밖에 사용할 수 없기 때문에 필요한 정보 획득이 어렵다.

한편, 손이나 손가락 등의 국지적인 영상 영역의 추적과는 달리 [5]에서는 이동 물체의 전체 영역을 대상으로 특징을 추출하고 최소 사각 영역을 연속적으로 검출하여 이동 물체를 추적하고자 하였고, [6]에서는 얼굴의 평균적 영상과 윤곽선, 그리고 색채 특징으로부터 예측맵을 구성함으로써 자동 얼굴 인식의 전단계인 얼굴 탐색 방법을 제안한 바 있다.

Viola[7]는 Papageorgiou[11]에게 영향을 받아 일반적인 물체 탐지 방법을 제안하였다. 이 방법은 빠르면서 복잡한 환경이나 기타 외부 변수에도 강한 성격을 가지고 있으며, 다양한 배경과 낮은 해상도에서도 좋은 성능을 보여준다. Viola의 방법은 인텔 펜티엄 700MHz의 컴퓨터에서 초당 384 x 288 픽셀이미지의 15프레임을 처리한다. 현재 이 방법을 얼굴 영역 탐지에 적용한 사례는 있으나 손 영역 탐지에 적용한 사례는 없다. 따라서 본 논문에서는 이 방법을 손 영역 탐지에 적용하여 기존의 피부색 특징을 이용한 손 영역 탐지 방법과 성능을 비교하여 보았다.

본 논문의 구성은 다음과 같다. 2장에서는 Viola 탐지기에 사용된 AdaBoost와 추적을 위한 칼만 필터를 알아보고, 3장에서는 AdaBoost를 이용한 손 영역 탐지 및 칼만 필터 설계, 4장은 실험 및 테스트의 결과를 보여주며, 마지막 장은 결론 및 향후 연구에 대해서 다룬다.

- 최상의 특징들의 선형 결합으로 강력한 최종 분류기를 훈련한다.

II. 관련 연구

탐지기로 사용할 Viola 얼굴 탐지기에 대해 알아보고, 탐지된 영역의 누적된 정보를 이용하여 다음 영역을 예측하여 추적하는 칼만 필터에 대해 알아본다.

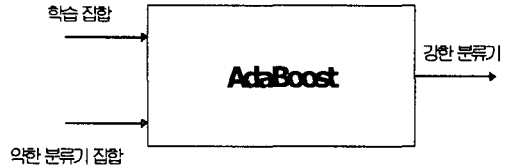


그림 1. AdaBoost의 기본 개념
Fig. 1 Basic concept of AdaBoost

2.1 AdaBoost 알고리즘

Viola 탐지기는 AdaBoost를 이용하여 얼굴 탐지에 강인하면서 빠른 좋은 결과를 얻었다. 적분 이미지로 표현된 이미지에서 하알 특징을 추출하고 AdaBoost 알고리즘을 이용하여 분류를 하게 된다. AdaBoost는 약한 분류기를 캐스캐이드 형태로 결합하여 최종적으로 강한 분류기를 만든다.

Viola가 제시한 방법 이전에 Pavlovic(8)은 픽셀 값을 기준으로 한 약한 분류기를 결합하여 설정된 임계값을 통해 positive와 negative로 분류하였다. 이 방법은 최적의 얼굴 구조를 표현하는 픽셀 집합을 훈련 데이터로부터 학습을 통해 선택한다. 이 방법은 주요 특징 표현에 좋은 결과를 보였지만, 만약 24 x 24 픽셀의 이미지를 고려할 때 576개의 픽셀 하나하나를 계산해야 하고 각 윈도우에서 적용해야 할 576개의 약한 분류기를 결합해야 최종적인 강한 분류를 할 수 있다. 이미지가 보다 크다면 탐색 크기는 물론, 분류기의 결합 역시 비례적으로 점점 커지게 된다. Viola는 이러한 문제점을 해결하기 위해, 즉, 좀 더 효율적인 특징 추출과 계산을 위해 하알 특징을 사용하였다.

윈도우 내에 많은 특징 집합은 적분 이미지 표현을 통해 간단히 계산할 수 있다. 그러나 특징 집합을 모두 적용한다면 픽셀 수 보다 훨씬 많은 특징들을 계산해야 하기 때문에 많은 시간을 필요로 하게 된다. 이러한 문제점을 해결하기 위해 Viola는 (그림 1)에 보인 Adaboost(Adaptive Boosting)라 불리는 알고리즘을 사용하였다. 이 알고리즘은 다음과 같은 두 가지 주요 목적을 가진다.

- 관심 대상에 가능한 충분히 표현하면서 가장 적은 수의 특징 집합을 선택한다.

AdaBoost는 훈련 에러와 전체 에러를 현저하게 줄이며, 간단한 통계적인 학습기의 조합을 이용하는 효율적인 부스트 알고리즘이다. AdaBoost의 장점과 단점을 살펴보면 다음과 같다.

▶ 장점들 :

- 선행지식이 필요 없음 :
AdaBoost 알고리즘은 훈련 데이터집합과 특징 집합의 두 가지 입력을 필요로 한다. 그리고 학습도중 가장 잘 표현된 특징이 자동 선택되므로 관심 대상에 대한 선행 지식이 필요하지 않다.
- 적응적인 알고리즘 :
학습의 각 단계에서 분류기가 positive와 negative 샘플들을 테스트한다. 어떤 샘플을 분류하기 어려워 잘못 분류했다면 잘못 분류된 샘플의 판별력을 높이기 위해 다음 단계에서 보다 높은 가중치를 부여하게 된다. 그래서 쉽게 분류할 수 있는 샘플들은 첫 단계에서 분류되고 다음 단계에서는 이의 가중치를 줄여 분류기가 좀 더 힘든 샘플들에 초점을 맞추어 학습할 수 있게 한다.
- 훈련하는 에러는 이론적으로 거의 0에 가까이 수렴 :
Freund와 Schapire의 증명(9)에 따르면, positive와 negative의 유한집합이 주어지고 충분한 단계를 거치면 훈련 시 발생하는 에러는 거의 0에 도달함을 보인 바 있다.

◆ 단점들 :

- 약한 분류기와 데이터에 의존적 결과 :
최종 탐색의 성능은 훈련 데이터를 통해 학습하기 때문에, 사용된 훈련 데이터에 대한 의존성이 강하다. 좋지 못한 샘플을 사용하여 최적의 특징이 선택되지 않는다면 성능은 그 만큼 저하될 것이다. 따라서 집합의 크기와 내부 분류 변수는 중요한 요소가 된다. 또한, 결합된 기본 분류기의 타입에 따라 결과에 영향을 미친다. 좋은 성능을 발휘하기 위해서는 적절한 반복 횟수를 정하는 것 이외에는 임의적인 선택이 아닌 필수적인 기본 함수들을 사용하는 것이 좋다.
- 아주 느린 훈련 :
각 순환단계에서 특징과 샘플 집합의 크기에 비례하여, 알고리즘은 샘플의 특징을 계산할 시간이 필요하게 된다. 수천의 positive와 negative의 샘플과 20개의 특징 집합을 가지고 훈련시킨다면 많은 시간이 소요된다. 계산 시간은 두 집합의 크기에 비례하여 증가한다. 실제 Viola-Jones 탐색기는 3000개의 샘플을 실험하였는데 훈련 시간은 약 일주일이 소요되었다.

2.2 칼만 필터

칼만 필터는 이동 중인 물체 추적, 노이즈 제거 필터링, 모형 예측 등에 다양하게 사용되고 있다. 칼만 필터는 순환적 선형구조를 가지고 있고, 알고리즘이 단순하고 수렴성이 좋으며, 이전의 스펙트럼 분석 방식에서 벗어나 시간 영역에서 직접적인 처리가 가능하므로 디지털 컴퓨터에서 사용하기에 적합한 방법이다. 칼만 필터 알고리즘의 일반적인 내용은 Barreto[15]의 연구를 참조하면 되므로, 본 논문에서는 실험에 사용한 이산 칼만 필터에 대한 설명을 한다.

우선, 칼만 필터에서는 프로세스를 수식 (2.1)과 (2.2)의 선형 상태 방정식으로 모델링한다.

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \dots\dots\dots (2.1)$$

$$z_k = Hx_k + v_k \dots\dots\dots (2.2)$$

여기에서, x 는 모델의 시스템 상태 변수, A 는 모델의 전이 행렬(State Transition Martrix), B 는 선택적인 사항인 전이 행렬, z 는 측정 값, H 는 측정값과 상태 값을 연결시켜 주는 행렬, w , v 는 평균이 0이고 각각 공분산 Q , R 로 표시되는 독립적인 백색 잡음 벡터이다.

칼만 필터는 피드백 제어로 프로세스를 추정하게 된다. 필터는 프로세스 상태를 어떤 시간에 추정을 하고, 노이즈가 있는 측정으로부터 피드백을 얻는다. 그리고 칼만 필터 방정식은 시간 갱신(time update) 방정식과 측정 갱신(measurement update) 방정식 두 그룹으로 나뉜다. 시간 갱신 방정식은 다음 시간 단계를 위해 이전 평가를 얻어 현재의 상태와 에러 공분산을 예측한다. 측정 갱신 방정식은 피드백을 담당하며, 새로운 측정을 이전 추정에 포함시켜 개선된 다음 평가를 얻는다. 시간 갱신 방정식은 예측 방정식으로 볼 수 있고, 반면 측정 갱신 방정식은 정정 방정식으로 볼 수 있다.

이산 칼만 필터 시간 갱신 방정식은 수식 (2.3)과 (2.4)와 같다.

$$\hat{x}_k^- = A \hat{x}_{k-1} + Bu_{k-1} + w_{k-1} \dots\dots\dots (2.3)$$

$$P_k^- = AP_{k-1}A^T + Q \dots\dots\dots (2.4)$$

여기에서, \hat{x}_k^- 는 시간 k 에서의 측정 이전 (priori) 추정 상태를, \hat{x}_k 는 시간 k 에서의 측정 이후 (posteriori) 추정 상태를, P 는 에러 공분산을 나타낸다.

이산 칼만 필터 측정 갱신 방정식은 수식 (2.5)과 (2.6), 그리고 수식 (2.7)에 나타내었다.

$$K_k = P_k^- H^T (H P_k^- H^T + R)^{-1} \dots\dots\dots (2.5)$$

$$\hat{x}_k = \hat{x}_k^- + K_k (z_k - H \hat{x}_k^-) \dots\dots\dots (2.6)$$

$$P_k = (I - K_k H) P_k^- \dots\dots\dots (2.7)$$

여기에서, K_k 는 칼만 이익(Kalman gain)으로 상태값과 에러 공분산을 갱신하는데 사용된다.

(그림 2)에서 보는 바와 같이 칼만 필터는 예측-정정의 과정을 계속 반복한다. 예측 단계에서는 수식 (2.3)과 (2.4) 각각을 이용하여 상태값과 에러 공분산을 예측하고, 정정 단계에서는 수식 (2.5)를 이용하여 먼저 칼만 이익을 계산하고 이와 실제 측정값 z_k 를 이용하여 수식 (2.6)에 의해 상태값을 갱신한다. 그리고 수식 (2.7)을 이용하여 에러 공분산을 수정한다.

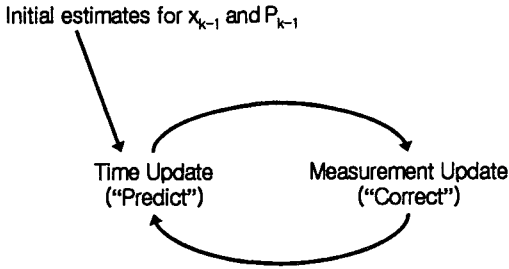


그림 2. 이산 칼만 필터 사이클
Fig. 2 Discrete Kalman filter Cycle

III. 추적을 위한 AdaBoost를 이용한 손 영역 추적

전체 시스템은 (그림 3)과 같이 먼저 카메라로부터 얻어진 전체 영상 이미지에서 AdaBoost 알고리즘을 사용하여 손 영역을 탐지한다. 손 영역을 찾으면 칼만 필터를 초기화하고 찾은 영역을 지정한다. 이 다음부터는 칼만 필터에 따라 손 영역 위치를 예측하고 이 예측된 손 영역 내에서 AdaBoost 알고리즘을 사용하여 손가락을 탐지한다. 만약 예측된 손 영역 내에서 손가락을 탐지할 수 있으면 계속해서 칼만 필터에 따라 모델링 변수들을 수정해 나간다. 손 영역 내에서 손가락을 찾을 수 없으면 전체 이미지에서 손 영역을 다시 찾아 위의 과정을 반복한다.

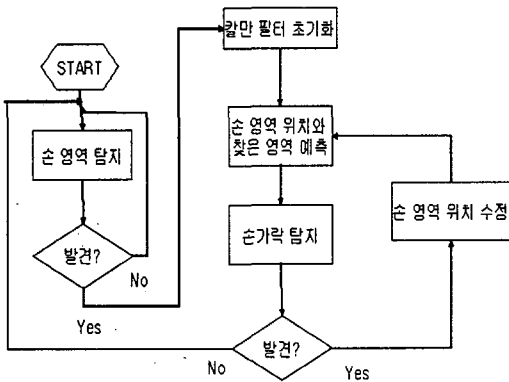


그림 3. 전체 시스템 구조
Fig. 3 Overall diagram of the system

3.1 하알 특징을 이용한 손 영역 탐지

이미지 탐지 방식은 이미지 중심 방식과 기하학 중심 방식으로 나눌 수 있다. 이미지 중심 방식은 관심 대상 얼굴이나 손등이 들어있는 집합을 positive로 여러 학습 알고리즘을 이용하여 대상을 탐지하는 방법이다. 그리고 특징 기반 방식이라고도 불리는 기하학 중심 방식은 객체의 독특한 특징이나 구조로, 다른 외부 요인에 의해서 변화하지 않는 것을 두며, 어떤 특징을 사용하느냐에 따라 탐지의 성능이 크게 좌우 될 수 있다.

일반적으로 손 영역 추적의 특징으로 가장 많이 알려진 방법은 피부색 색상을 이용하는 방법이다. 카메라를 통해 들어오는 영상은 RGB색상 모델로 붉은색, 녹색, 푸른색의 빛의 삼원색이 각자 값들로 들어가 혼합되어 색상을 만들어 낸다. 이 때문에 원하는 색상을 추출하는데 어려움이 따른다. 일반적으로 실제 사용하는 방법은 RGB 칼라 모델을 HSI 칼라 모델로 변환을 하여 사용한다. HSI 칼라 모델은 HSV와 같은 모델로 색채 H, 채도 S, 명도 I(V)들의 값을 나타낸다. 명도 부분을 색상에서 제외시킴으로서 조명에 비교적 강하게 된다. 그리고 이 색상 모델에서는 가우시안 분포의 확률적인 방법을 통해 임계값보다 큰 값을 가지면 피부색으로 인정하여 손 영역으로 탐지하여 필요한 인터페이스로 사용하게 된다.

그러나, 피부색과 유사한 색상, 즉 노란색, 혹은 붉은 색 계통의 옷을 입고 있을 때나 주위에 있을 경우, 적절한 임계값을 구하기가 어려워져 피부색 영역을 탐지하기 어렵게 된다. 그리고 피부색 특징을 이용하게 되면 원하는 손 영역 외에 다른 손이나, 얼굴 등도 같은 피부색으로 탐지를 하게 된다. 이로 인해 관심 영역과 비 관심 영역의 구분을 필요로 하는 문제점 또한 발생한다. 그리고 차차 웨어러블 컴퓨팅 (Wearable Computing) 등 외부 환경에서도 사용하게 된다면 장갑 착용 등의 제약을 받게 된다.

피부 색상 특징을 이용하여 손 영역을 탐지할 경우의 단점을 극복하기 위해 여러 추가적인 방식들을 사용하는 연구들 [10][11][12]이 있지만 본 연구에서는 피부 색상이 아닌 다른 특징을 사용하여 손 영역을 탐지하는 방법을 사용하였다.

손 영역 제스처만 살펴보는 경우 손 모양의 형태를 사용하는 템플릿을 이용할 수도 있지만 손이 딱딱한 물체가 아닌 부드러운 유기체이기 때문에 경계가 뚜렷하지 않아 손 형태를 추출하는 방법이 쉽지 않고, 사람마다 손 모양이나 크기가 다양하여 탐지에 어려움이 따른다. 이러한 이유로 본 연구에서는 하알 특징을 이용한 손 영역 탐지를 시도하여 보았다. 하알 특징도 어떤 면에서는 모양 기반으로 한

특징이기도 하지만, 사각형의 특징을 이용하여 경계 및 영역간의 차이로 관심 대상을 탐지하기 때문에 경계에 대해서도 융통성을 가지며 기타 우리가 흔히 알지 못하는 다른 적절한 특징들도 사용할 수 있다. 그리고 하알 특징 추출 시 배경 지식을 필요로 하지 않을 뿐만 아니라 쉽고 간단하게 연산을 할 수 있으며 성능이 빠른 장점이 있다.

본 연구에서는 Viola가 얼굴인식에 사용했던 방법을 손 영역 탐지에 적용하려 하였다. 그러나 손 영역은 얼굴에 비해 자유도가 높아 움직임이나 기울어짐이 크다. 때문에 기본 하알 특징만을 이용할 경우 원하는 인식률을 얻기가 어렵다. 본 논문에서는 기울어진 대상에 대해서도 인식률을 높이기 위해 Rainer Lienhart가 제안한 추가의 하알 특징들(12)을 사용하였다. 즉, (그림 4)에서 보는 바와 같은 기본 하알 원형들을 45도로 기울인 하알 특징들을 추가로 사용하였다.

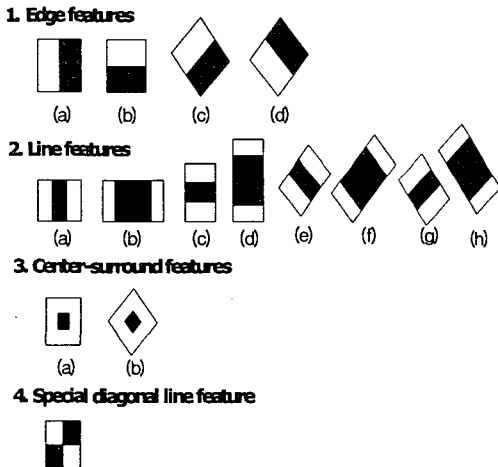


그림 4. 추가로 사용한 하알 유사 특징들
Fig. 4 Additional Haar features

3.2 관심 영역 설정

시스템에 맞는 관심 영역 설정이 필요하다. [13]은 국지영역 컨텍스트(local context)와 객체 중심(object-centered) 탐지에 관하여 다루었다. 이 논문에서는 얼굴에 대한 관심 대상을 얼굴 자체만 보는 객체 중심과 지역의 영역을 포함한 대상, 즉 얼굴과 그 주변 형태들, 예를 들어 사람의 상반신에 해당하는 영역을 대상으로 한 국지영역 컨텍스트로 분류하였다.

[13]의 결과에 따르면 단순히 탐지율만 보았을 때는 국지영역 컨텍스트가 더 좋은 성능을 보인다. 하지만 객체 중심 탐지는 대상을 직접 탐지할 수 있으며 비교적 정확한 위치를 탐지해 낼 수 있는 장점이 있다. 그리고, 국지영역 컨텍스트는 저해상도에서는 성능이 좋은 반면 고해상도의 이미지에 대해서는 배경에 따라 오판율(false positive)이 증가하며, 뒷 배경이 복잡한 경우 인식률이 떨어진다. 또한 포인트 지정처럼 세밀한 영역을 가리키기 위해서는 세부 관심 대상에 대한 새로운 탐지가 필요하다.

본 논문에서는 이 두 방법을 적절하게 사용하였다. 포인터 지정을 위한 손 영역 추적 시스템을 보면, 시스템을 사용하더라도 지속적인 추적은 필요치 않다. 마우스 사용 예로 든다면, 시스템은 키보드나 다른 작업을 하고 있다가 포인터 지정을 위해 마우스를 사용하는 순간부터 추적을 하면 될 것이다. 즉, 추적을 시작하기 위해 진입 단계를 필요로 하게 된다. 본 시스템에서는 진입 단계를 위해서는 국지영역 컨텍스트 방법을 사용하여 손 영역을 탐지하고, 세밀한 추적을 위해서는 손가락 영역을 관심 대상으로 하여 추적하였다.

일반적으로 포인터를 지정할 때 검지 손가락 하나를 사용한다. 그러나 하나의 손가락만을 특징으로 지정하기에는 추출되는 하알 특징수가 너무 적고 대상을 적절하게 표현할 수 없어 유사 대상에 대한 오판율이 높아진다. 따라서 본 시스템에서는 (그림 5)와 같이 두 손가락(검지와 중지)을 이용하여 포인팅한다고 가정하여 실험하였다. 이러한 포인팅 방식을 사용한 이유는 기존 포인팅 방식보다 더 많은 특징을 추출할 수 있어 더 좋은 성능을 얻을 수 있고, 사람이 사용할 때 불편함이 적기 때문이다. 그리고 향후 클릭 등의 추가 동작으로 쉽게 확장할 수 있기 때문이다.

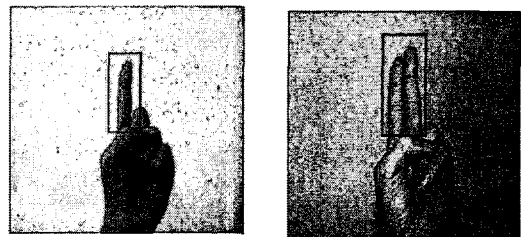


그림 5. 기존 포인터 지정 방식과 실험에 사용한 방식
Fig. 5 Traditional pointing method and proposed pointing method

3.3 칼만 필터 설계

실제로 이동물체를 추정할 경우 실제 시스템과 일치하는 모델을 구축하는 것은 불가능하다. 따라서 최대한 비슷한 모델을 선택해야 한다. 어떤 형태의 운동이라도 충분한 구간으로 자른다면 이 구간 내에서는 등속운동을 한다고 가정할 수 있다. 본문에서는 실제 시스템과 추정 모델과의 오차를 시스템 노이즈로 고려하고 칼만 필터를 설계하였다.

선형 운동을 가정했을 경우는 2절의 상태 방정식과 측정 방정식으로 시스템을 모델링할 수 있다. 하지만, 실세계의 많은 시스템은 선형식으로 정확히 표현할 수 없다. 이런 경우, 수식 (3.1)과 (3.2)의 비선형 방정식을 이용하여 시스템을 표시할 수 있다.

$$x(k+1) = F(x(k)) + w(k) \dots\dots\dots (3.1)$$

$$z(k) = G(x(k)) + v(k) \dots\dots\dots (3.2)$$

직각 좌표계의 비선형 상태 방정식은 선형 시스템으로 표현되더라도 오차가 크게 포함되지 않으므로 어느 정도 모델링 오차를 감수하여 선형화 된 시스템으로 표현할 수 있다. 함수 F, G를 등속 운동 모델로 적용하기 위한 선형화는 테일러 급수의 1차항을 취해서 유도한 자코비 행렬을 사용한다. 자코비 행렬 A(k), H(k)는 수식 (3.3) 및 (3.4)와 같은 상태 천이행렬로 표현된다. 본 논문에서도 이러한 행렬을 사용하였다.

$$A(k) \cong \frac{\delta F(X)}{\delta X} = \begin{bmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 1 \end{bmatrix} \dots\dots (3.3)$$

$$H(k) \cong \frac{\delta G(X)}{\delta X} = \begin{bmatrix} \frac{f}{z(k)} & 0 & 0 & 0 \\ 0 & 0 & \frac{f}{z(k)} & 0 \end{bmatrix} \dots\dots (3.4)$$

여기서, Δt는 샘플링 시간이고, f는 카메라의 초점거리, z(k)는 시간 k에서의 물체의 z 좌표값을 의미한다.

추적을 위한 칼만 필터의 시스템 노이즈와 관측 노이즈의 공분산 행렬로는 다른 응용분야에서 사용하던 수식 (3.5)로 표현되는 행렬 Q와 수식(3.6)으로 표현되는 행렬 R을 사용하였다.

$$Q = \begin{bmatrix} \sigma_{w1}^2 & 0 & 0 & 0 \\ 0 & \sigma_{w2}^2 & 0 & 0 \\ 0 & 0 & \sigma_{w3}^2 & 0 \\ 0 & 0 & 0 & \sigma_{w4}^2 \end{bmatrix} = \begin{bmatrix} 60 & 0 & 0 & 0 \\ 0 & 90 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 60 \end{bmatrix} \dots\dots\dots (3.5)$$

$$R = \begin{bmatrix} \sigma_{v1}^2 & 0 \\ 0 & \sigma_{v2}^2 \end{bmatrix} = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} \dots\dots\dots (3.6)$$

IV. 실험 및 결과

4.1 데이터 집합

AdaBoost의 학습에 있어서 데이터는 중요한 요소가 된다. 데이터의 종류와 양이 충분히 지원되어야 한다. 데이터 집합은 두 가지가 필요하다. 하나는 관심 대상을 포함하는 positive 데이터 집합이고 다른 하나는 관심 대상이 없는, 즉 배경 이미지인 negative 데이터 집합이다.

손 영역 탐지에서는 표준 데이터가 없으므로, 다음과 같이 나누어 보았다. 손의 생김새를 손가락이 긴 손가락, 짧은 손가락, 마디가 굵은 손가락, 가는 손가락으로 나누어 데이터를 획득하였다. positive 샘플은 남녀 각각 4명씩, 300만 화소 디지털 카메라를 이용하여, 거리는 3가지로 나누어 가까이, 보통, 먼 거리, 각각 100장씩 수집하였다. 2400장 중 잘못 찍은 몇 이미지를 제외한 2283장 중 학습데이터로 1783을 나머지 500 장과 캡을 이용하여 추가한 500장, 1,000장을 테스트 데이터로 이용하였다. negative 데이터는 주위에 배경과 기타 자연 배경을 바탕으로 2000장을 수집했다.

4.2 손 영역 및 손가락 탐지 성능

학습은 OpenCV 소프트웨어[14]를 사용하였다. 손 영역과 손가락 영역을 분리하여 학습시켰으며, 각 이미지에서 해당 영역을 찾아 관심 영역으로 등록시켜 하일 특징을 추출하게 하였다. 학습 알고리즘은 Discrete AdaBoost를 이용하였고, 윈도우 사이즈는 24 x 24를 이용하였다. 학습 수

행 시간은 CPU 2.4GHz, 1GB 메모리 시스템으로 약 5일 정도 소요되었다.

성능 비교 결과는 <표 1>과 같이 손 영역은 총 23 단계의 캐스케이드와 특징 수 18개로 구성되었고 탐지율은 91.2%, 2.4%의 오판율을 보였으며, 손가락 영역은 총 18단계 캐스케이드와 특징 수 7개로 이루어졌다. 탐지율과 오판율은 전체 이미지 영역에서 수행한 결과 97.4%와 68%이다. 오판율이 높은 것은 전체 이미지에서 검출할 경우 배경에 따라 유사 형태가 많이 잡히기 때문이다. 하지만 본 시스템에서는 손가락 영역의 탐지는 손 영역으로 예측된 제한된 공간에서 탐지기 때문에 손 영역 내에서 발생하는 오판율이 실제 오판율이라고 볼 수 있으며 결과는 2.8%이다.

실험 결과, 여전히 겹치는 문제는 해결하지 못함을 확인할 수 있었다. 즉 손가락과 얼굴이 중첩이 될 경우 이를 제대로 손가락으로 판단하지 못했다. 얼굴 탐지인 경우는 손 영역보다 면적도 넓으며 내부에 중요 특징들이 하알 특징으로 존재하기 때문에 겹치는 문제에 영향을 받지 않는다. 하지만 손가락 영역 내부에 존재하는 하알 특징이 적어 적절한 표현을 찾지 못하기 때문에 판단 오류가 일어난 것으로 여겨진다.

표 1. 손 영역과 손가락 영역의 구성과 성능 비교
Table. 1 Comparative structure and performance of hand/finger regions

	캐스케이드 단계	특징수	탐지률	false positive
손 영역	23	18	91.2%	2.4%
손가락 영역	18	7	97.4%	68% (2.8%)

4.3 추적 성능

(그림 6)은 추적을 수행하는 실제 예이며 내부 사각형은 탐지 및 측정된 영역이며 외부 사각형은 측정 영역들을 바탕으로 예측 영역을 속도 등을 고려하여 확장한 영역이다. (그림 7)은 추적의 결과를 그래프로 표시한 것이다. (그림 7)에서 실선은 AdaBoost를 이용하여 손 영역을 추적한 그래프이며, 점선은 피부색 특징을 이용한 방법의 그래프이다. x축은 시간축이고 y축은 원 이미지의 영역과 예측의 이미지 차이를 나타낸 것이다. 처음에는 오차 범위가 크다가 차차 오차 범위가 적어지면서 추적을 하게 된다. 그리고 피부색 특징을 이용한 방법에서는 도중에 빛의 변화에 의해서 예측에 문제가 생기는 것을 볼 수 있다.

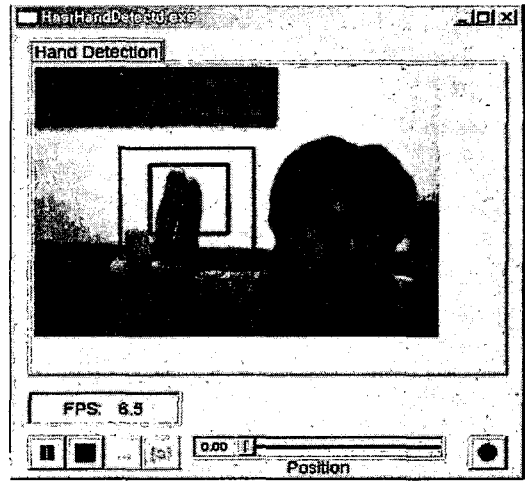


그림 6. 손 영역 탐지 실행의 예
Fig. 6 Example of hand detection

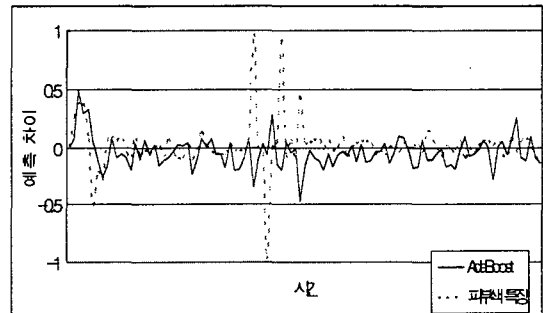


그림 7. 손 영역 추적 오차
Fig. 7 Hand tracking error

V. 결론 및 향후 과제

Viola 얼굴 탐지에 사용된 AdaBoost와 칼만 필터를 이용하여 손 영역을 탐지하고 추적하였다. Viola 탐지기는 적분 이미지로 표현한 이미지에서 하알 특징을 추출하고, AdaBoost 알고리즘으로 학습 및 탐지하는 방식으로 발견율이 높고 탐지 속도가 빠른 것으로 알려져 있다. 이러한 방법을 손 영역 탐지 및 추적에 적용해 본 결과, 피부색 특징을 이용한 추적 방법의 단점들, 즉 조명의 의한 색상 변화에 따른 임계값 설정이나 얼굴과의 구분, 그리고 사용용

위해서는 피부를 노출해야 하는 제약들을 부분적으로 해결할 수 있었다. 물론 장시간의 학습시간과 많은 데이터의 양을 다루어야 하는 점, 그리고 여전히 물체의 겹침에 대한 처리는 할 수 없는 문제점들이 있다. 본 제안 방법은 앞으로 손 영역과 다른 영역의 겹침 해결, 손동작 인식의 추가 등에 관한 연구가 이루어진다면 빠른 탐지 성능으로 인해 새로운 인터페이스로 사용하기에 충분하리라 여겨진다.

참고문헌

- [1] Albert Mehrabian, "Communication without words", Communication : Concepts and Process, Prentice-Hall Inc, 1971.
- [2] Melissa H. Welsh and David L. Akin, "The Effects of Extravehicular Activity Gloves on Human Hand Performance" Society of Automotive Engineering, 2001.
- [3] Kenji Oka and Yoichi Sato, "Real-Time Fingertip Tracking and Gesture Recognition", IEEE Computer Graphics and Applications, Vol. 22 (6), pp. 64-71, 2002.
- [4] Yuliang Zhu, "Hand Detection and Tracking in an Active Vision System", http://www.cs.yorku.ca/LAAV/projects/motionRecognition/file_HandTracking.pdf, 2003.
- [5] 강현중, 이광형, "이동 객체 감시를 위한 실시간 객체 추출 및 추적시스템", 한국컴퓨터정보학회논문지, 제10권 제2호, pp.59-68, 2005. 5.
- [6] 조경식, 구자영, "예측맵을 이용한 얼굴탐색의 가속화 기법", 한국컴퓨터정보학회논문지, 제8권 제2호, pp.31-36, 2003. 6.
- [7] Paul Viola and Micheal Jones, "Robust real-time object detection", Second International Workshop on Statistical Learning and Computational Theories of Vision Modeling, Learning, Computing and Sampling, July 2001.
- [8] Pavlovic V. and Garg A., "Efficient Detection of Objects and Attributes using Boosting", IEEE Conf. Computer Vision and Pattern Recognition, 2001.
- [9] Yoav Freund and Robert E. Schapine, "A decision-theoretic generalization of on-line learning and an application to boosting", Computational Learning Theory : Eurocolt pp. 23-37, 1995.
- [10] D. Saxe and R. Foulds, "Toward robust skin identification in video images", IEEE Intl. Conf. on Automatic Face and Gesture Recognition, 1996.
- [11] B. D. Zarit, B. J. Super, and F. K. H. Quek, "Comparison of Five Color Models in Skin Pixel Classification", Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 1999.
- [12] Rainer Lienhart and Jochen Maydt, "An Extended Set of Harr-like Features for Rapid Object Detection", IEEE ICIP 2002, 2002.
- [13] Hannes Kruppa, Modesto Castrillon-Santana and Bernt Schiele, "Fast and Robust Face Finding via Local Context", Joint IEEE Intl. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2003.
- [14] Open Computer Vision Library, <http://sourceforge.net/projects/opencvlibrary/>.
- [15] J. Barreto, P. Menezes, and J. Dias, "Human-Robot Interaction based on Harr-like Features and Eigenfaces", International Conference on Robotics and Automation '04, 2004.

저자 소개



김병만

1987 서울대학교 컴퓨터공학과 학사
1989 한국과학기술원 전산학과 공학 석사
1992 한국과학기술원 전산학과 공학 박사
1992년~현재 금오공과대학교 컴퓨터공학부 교수
<관심분야> 인공지능, 정보검색, 소프트웨어 검증



김준우

2003 금오공과대학교 컴퓨터공학부 소프트웨어공학전공 졸업
2005 금오공과대학교 소프트웨어공학과 석사
2005~현재 (주)인포큐어 근무
<관심분야> 패턴인식, 인공지능



이광호

1987 서울대학교 컴퓨터공학과 학사
1989 한국과학기술원 전산학과 공학 석사
1996 한국과학기술원 전산학과 공학박사
1996년~현재 목포대학교 컴퓨터교육과 부교수
<관심분야> 인공지능, 영상처리, 알고리즘