

실시간 영상에서의 휴먼 검출 및 얼굴 분류

김건우*, 남미영**, 한종욱*

요약

본고는 휴먼 객체 검출 및 분류를 위한 것으로서, 입력된 동영상에서 배경 이미지와의 차분 영상을 통해 객체 영역을 검출하고, 검출된 객체 영역에서 얼굴 즉 헤드 영역을 검출하는 방법에 대해서 설명한다. 실시간으로 녹화된 동영상에서 사람이 움직이는 위치와, 크기 등이 아주 다양하며, 또한 한 사람이 아닌 여러 사람 객체를 검출하기 위하여 다중의 사람객체 검출기를 이용한 캐스케이드 사람 객체 추출 방법을 제안한다. 얼굴 크기 등을 고려하여 헤드 영역의 shape 를 기반으로 하여 1차 검출을 수행하고, 검출되지 않은 영역에 대하여 히스토그램 기반의 얼굴 영역을 검출한다. 또한 중복된 영상에 대해 베이지안 얼굴 검출기를 통해 인중함으로써 성능을 향상시킬 수 있다.

I. 서론

최근 영상을 기반으로 한 보안 시스템에 대한 관심이 높아지고 있다. 이미 많은 CCTV 가 설치되어 다양한 용도로 활용되고 있다. 실시간으로 입력되는 영상 혹은 비디오 영상에서 오브젝트 분석은 움직이는 물체를 감지하고, 각 프레임에서 이러한 객체를 추적하며, 객체의 분석에 따른 behavior를 인식하는 것이다. 이러한 오브젝트 추적은 다양한 목적으로 활용된다. 사람 객체 인식, 서베일런스 시스템에서의 모니터링 기능, 비디오 인덱싱, 제스처 인식 등의 상호작용, 실시간에서의 트래픽 모니터링, 비디오 기반 패스 계획기능과 위험 상황을 인지하여 정보를 제공하는 자동차 네비게이션 등의 기능에도 적용될 수 있다^[1]. 실시간, 비디오 영상에서의 오브젝트 추적은 이처럼 많은 용도로 활용되어가고 있다.

객체 추적은 가장 간단하게는 영상에서 오브젝트의 궤적을 추정하는 것으로 정의할 수 있다. 이는 다른 프레임으로부터 추정되는 영역을 라벨링하는 것으로서, 트래킹을 수행하고자 하는 도메인에 따라 많이 적합하게 응용하여야 한다. 이러한 오브젝트의 추적 및 검출은 실제세계에서 2D 영상에서 검출해야 한다는 것과, 조명 및 다른 실내외적인 환경에 따른 잡음, 객체 움직임의 복잡성 중복되는 현상, 윤곽선의 복잡성, 실시간 처리를

위한 지연 등으로 인해 어려움이 있다^[1]. 따라서 미리 오브젝트의 크기와 수를 미리 정하고 추적하는 방법이 있다. 이것은 그 오브젝트의 유사성을 이용하여 다른 오브젝트와 구분하여 추적하기도 하고, 한 개의 오브젝트를 중심으로 추적하며 유사도에 따라 비교하기도 한다.

II. 관련 연구

오브젝트 트래킹의 방법은 학습에 의한 모델로서 확률 밀도 함수를 이용하는 방법^[2, 3, 4], Active appearance models^[5], 베이지안 알고리즘^[6]을 이용하여 구성하는 방법 등이 있다. 오브젝트 트래킹을 위한 특징 추출로서 컬러 정보를 이용하거나^[7], 특징점을 추출하거나, canny 에지 정보^[8]를 이용하기도 한다. 또한 사람 모델링을 통해 아다부스트 기법을 적용하여 수행하기도 한다. 그러나 이는 특정한 크기의 학습모델을 필요로 하고, 이를 위해 정규화 등의 과정을 수행하여야 한다. 실시간에서는 시간적인 부분이 문제가 될 수 있다.

실시간 시스템에서 오브젝트를 검출하는 방법으로 배경 영상과의 차분 영상을 통해 영역을 추출하는 방법과, 시프트 알고리즘과 같이 특징 포인트를 추출하여 검출하는 방법이 있다. 시프트 알고리즘은 오브젝트의 위

* 한국전자통신연구원(kimgw,hanjw@etri.re.kr)

** 인하대학교(nammiyoung@gmail.com)

치 크기 등에 강인한 방법으로서 동일한 물체를 검출하는 데는 효과적이다.

이렇듯 객체 추적은 많은 연구가 이루어지고 있으며, 사람 객체 추적에 대한 관심도 높아지고 있다.

특히 범죄, 수상한 움직임, 사람 추적 등의 보안 시스템의 부차적인 기능으로 각광을 받고 있다. 앞서 말한 바와 같이 동영상에서 사람 객체를 추적하는 것은 움직이는 속도, 주위 환경 등의 변화량이 다양하여 어려움이 있다. 실시간 시스템에서는 오브젝트의 크기, 방향 등이 일정하지 않으며, 사람의 경우 서로 겹치거나 간격차이가 나지 않을 정도로 인접해 있는 경우도 있으므로, 쉬운 문제가 아니다. 대부분의 사람 객체 추적은 하나의 scene (프레임, 장면)에서 단일의 사람을 추적하는 연구가 이루어지고 있으며^[9], 트래킹에 대한 연구^[10] 사람의 모션을 모델링 하는 영역^[11, 12]에 대한 부분이 이루어지고 있다. 다중의 사람 객체 검출 및 추적에 대하여는 다중의 카메라에서 입력되는 영상을 기반으로 한 연구가 되고 있다.

지도학습 방법을 이용하여 검출하기도 하며, Viola et al.^[13]에 의해 제안된 haar-like wavelet 방식을 응용하기도 한다. 이 방법은 얼굴의 특징 벡터를 추출하기 위하여 haar wavelet을 응용하여 haar-like feature를 생성함으로써 사람 영역을 검출하기도 하나 이는 학습 데이터에 민감하게 된다.

본 고에서는 동영상으로 입력받아 오브젝트를 검출하고, 사람의 헤드 영역을 추정함으로써 사람 객체를 검출하는 방법에 대해서 논의한다. 다중의 사람 객체를 검출하며, 크기, 포즈에 강인한 알고리즘을 제안하며, 중복된 정면 영상에 대하여도 각 사람의 객체를 검출한다. 이를 위하여 캐스케이드 방식의 사람 객체 검출 방법을 적용하며, 크게 두 단계로 이루어진다. 첫번째 단계는 사람 객체를 추출하고, 헤드 영역을 검출하기 위하여 오메가 shape를 검출한다. 이는 얼굴의 정면, 뒷면, 45도 정도의 오브젝트에서 헤드 영역의 윤곽적 Ω 모양을 검출한다. 그러나 가방 손잡이 등과 같이 유사한 모양을 가진 오브젝트도 존재할 수 있다. 따라서 히스토그램 기반의 사람의 기하학적 모델을 평가하여 헤드 영역을 검증한다.

Ω 모양의 헤드 영역이 검출되지 않을 경우 작은 영상, 혹은 정보가 많이 부족한 영상으로 판단하여, 두번째 검출기로 넘긴다. 두번째 검출기에서는 어깨와 헤드

영역에 대한 기하학적 모델을 기반으로 검출한다. 옆모습 및 작은 영상에 대한 검출을 위하여 적용한다.

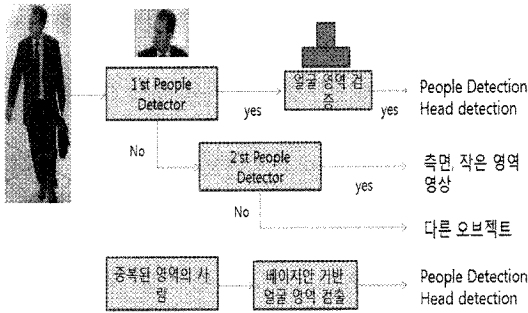
검출된 오브젝트 영역에 대해 가로 세로 비율을 적용하여 검진 영상에 대하여 얼굴 인증 알고리즘을 수행한다. 이는 미리 얼굴데이터로 학습된 베이지안 얼굴 인증기를 적용한다.

다중의 검출기와 캐스케이드 방식을 적용함으로써 오브젝트 영역 검출의 오류로 인한 사람 객체 검출의 성능을 향상시키고, 중복된 부분에 대한 얼굴 검출, 윤곽선 검출에 따른 오브젝트 정규화 등의 문제를 해결함으로써 실시간 환경에서의 수행 속도를 향상시키며, 적극적으로 수행할 수 있도록 한다.

따라서 본 고에서는 3장에서는 오브젝트 검출과 헤드 영역 검출에 대한 시스템 구조를, 4장에서는 움직임 검출 및 헤드 영역 검출을, 5장에서는 중복 부분에 대한 얼굴 인증 알고리즘을, 6장에서는 구현 및 실험에 대해 설명한다.

III. 객체 검출 및 헤드 영역 검출

CCTV 영상에서의 움직임을 검출하고, 움직임 영역을 기반으로 하여 오브젝트를 검출하고, 사람임을 인증하기 위하여 기하학적 방법과 오메가 모양 검출과 학습 기반의 얼굴인증 알고리즘을 적용한다. 720*576 크기의 프레임에서 전체 영역을 기반으로 하여 얼굴 영역을 검출하는 것은 실시간, 동영상 시스템에서의 검출에는 적합하지 못하다. 크기에 무관하게 검출하기 위해서 다중 해상도 분석을 수행해야 하므로, 실시간 모델에서는 적용의 효율성이 떨어진다. 사람 영역을 모델링하여 검출하는 방법이 있지만, 크기를 정규화하여야 한다. 따라서 본 논문에서는 효과적인 사람의 후보 영역을 검출하기 위해 배경 영상을 생성하여 위해 캐스케이드 검출 방법을 제안한다. 이전 프레임과의 차영상을 이용하여 검출할 수 있으나, 이는 움직임이 많지 않을 경우에 오브젝트 영역을 검출하기 어려우므로, 배경 영상과의 차를 이용한다. 중복된 영역은 하나의 영역으로 검출이 될 가능성이 있으므로, 가로 길이 및 세로길이의 비율을 이용하여 중복될 가능성이 있는 영역은 외형기반 학습 모델인 베이지안 얼굴 인증기를 적용하며, 기본적으로 얼굴 영역의 shape는 오메가 영역을 검출하며, 다른 오브젝트 영역과 비교하여 정확한 검출을 위하여 히스토그



(그림 1) 오브젝트 검출 흐름도

램 기반의 기하학적 모양을 검출한다. 그리고 작은 오브젝트 영역에서의 얼굴 검출을 위하여 기하학적 모델을 기반으로 하여 검출한다. 제안하는 시스템의 구조는 그림 1과 같다.

3.1 Context Aware based classifier construction

얼굴의 정면 영상과 측면 영상을 위하여 라벨링된 영상을 기준으로 하여 기하학적 모양을 기준으로 context를 판단한다. 일반적인 정면의 사람 기준을 라벨링된 영역에서 가로 비율이 세로 비율보다 30% 이하로 판정하며, 가로 비율에 비해 세로 비율이 일정 비율 이상이 되면, 중복된 얼굴이 포함되어 있는 현상으로 판단하여, 얼굴 검출 및 shaper 검출을 수행한다. 베이저안을 이용하여 얼굴 검출을 수행하는 경우, 실시간에서 수집되는 다양한 영상을 기반으로 하여 검출기를 계층적으로 수행한다.

Algorithm 1.

```

Input:
n ; /* no. of inputs
K ; /* no. of desired cluster
1, 2,... k ; /*initial mean of each cluster
Output:
Modified 1, 2,... k ;
/*updated mean of each cluster
Begin:
(1) Repeat step (2) and (3) until no change in i
(2) Classify n samples according to equation (i)

$$dismin = (\mathbf{x}_k - \mu_i)^T \sum_i^{-1} (\mathbf{x}_k - \mu_i)$$

(3) Recompute i
(4) Return i
End.
    
```

K-means 알고리즘을 이용하여 3개의 영역으로 분류한다. 사용되는 영상은 실시간에서 수집된 밝은 영상은 다음과 같으며, 이러한 영상을 기준으로 하여 분류하였다. 밝은 영상 보통 사무실 환경에서의 일반 영상, 어두운 영상을 기준으로 분류하였다. 또한, 20*20 크기의 영상을 기준으로 하여 픽셀의 값들을 일차원 vector로 표현하여 클러스터링을 수행한다.

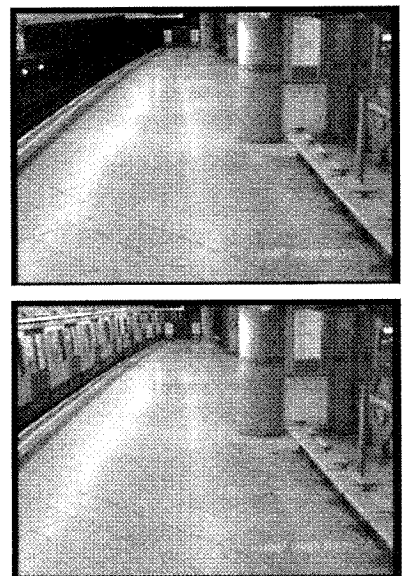
클러스터링 시 다음의 방법을 이용하여 클러스터간의 중심에 따른 상황을 판별하도록 하였다.

3.2. 배경 영상 검출

동영상에서 사람을 검출하기 위하여 CCTV가 설치되어 있는 움직임이 없는 부분을 검출하기 위하여 배경 영상을 추출하였다. 지하철이 오고 가지 않을 때의 영상으로써, 움직임이 많이 발생하게 되는 배경 영역을 masking 함으로써 다음과 같은 배경 영상을 생성하였다.

두 영상과 다른 프레임을 비교하여 비교적 많은 변화가 일어난 지하철 부분을 검은색 필드로 masking 하였다.

동영상의 앞 부분의 300개의 프레임에서 평균영상을 검출함으로써 배경 영상을 생성하였다. 프레임의 뒤로 갈수록 사람의 움직임 등이 불규칙적으로 움직이기 때



(그림 2) 배경 영상

문에 앞의 프레임만으로 배경 영상을 생성하였다.

3.3 First people detector

사람 영역 검출 및 헤드 영역을 검출하기 위하여 object 추적을 위한 오메가 shape를 검출한다. 수행하는 방법은 다음과 같다.

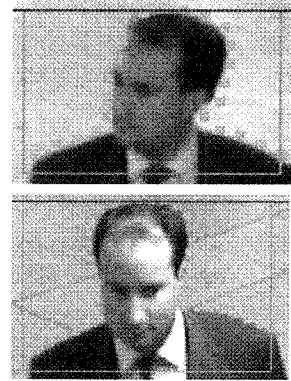
오메가 영역의 검출은 사람의 헤드 영역의 윤곽선이 Ω 모양임을 이용하여 검출하는 방법으로서, 실시간 시스템에서 빠르게 헤드 영역을 검출하는데 효과적이다. 정확한 shape를 검출하기 위하여 그레이 차분 영상에서 가우시안 필터링을 통해 일차적으로 잡음을 제거하여 shape 영역의 검출을 부드럽게 수행하였다.

배경과의 차분 영상은 다음의 식에 의한 구한다. 각각 r, g, b 값들의 차의 평균으로 그레이 차분 영상을 구한다. 차분 영상을 구하는 과정에서 발생하는 노이즈를 제거하고, 라벨링 시 발생하는 오브젝트 내의 홀을 축소시키기 위해 일차적으로 차분 그레이 영상을 생성하여 가우시안 필터링을 수행하였다. 이후, 경계값을 조정하여 이진화하고, 모폴로지를 수행하여 라벨링을 수행한다. 수행한 영역에서 오메가 shape를 검출함으로써 성능을 강화시킬 수 있었다.

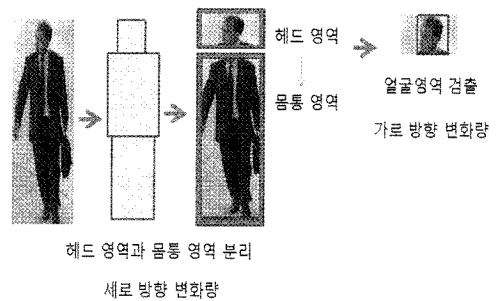
그림 4는 사람 영상의 헤드 영역에서 오메가 shape를 검출한 것을 보여준다. 이 영역을 기반으로 사람 객체의 영역을 검증한다.

그림 4에서 80도 가까이 회전된 영상과 정면 영상 검출 결과를 보여주며, 영상의 크기에 따른 검출 성능도 뛰어나다.

실시간 영상 및 실내-외에서 획득되는 영상은 조명 환경 등의 변화가 심하기 때문에 앞서 설명한 배경 영상에 따른 차분 영상에서 검출하며, 필터링을 수행함으로써 검출 성능을 향상시킬 수 있었다. 여러 프레임간의 변화가 많이 일어나지 않는 부분을, 변화량이 발생하지 않는 부분을 배경으로 선택함으로써 현재 상황을 반영하도록 한다.

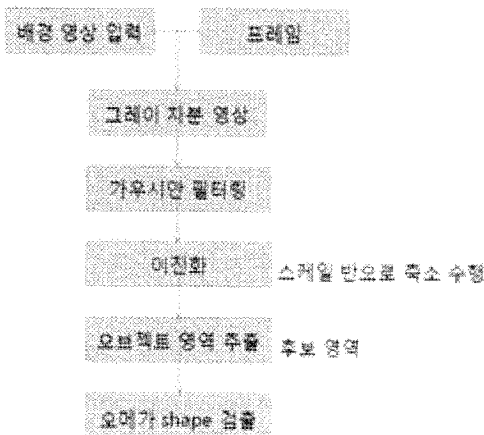


(그림 4) (a) 옆 모양 얼굴 (b) 정면 얼굴 모양

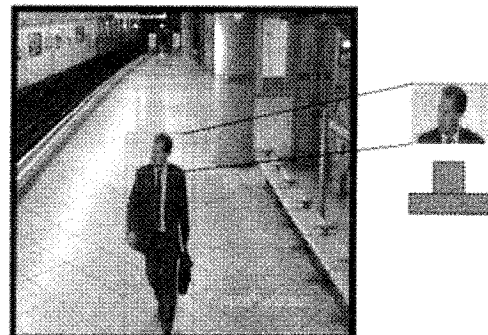


헤드 영역과 몸통 영역 분리
세로 방향 변화량

(그림 5) 두번째 검출기에서의 헤드 영역 검출



(그림 3) 첫번째 객체 검출기



(그림 6) 헤드 영역 검출

3.4 Second People Detector

두번째 사람 객체 추적 기법은 shape 가 아닌 기하학적 모델, 모형을 기반으로 하여 검출하는 방법이다. 이 방법은 haar 웨이블릿이나, 학습기를 통해 사람 객체 패턴 자체를 이용하는 경우로서, 몸통 부분과 헤드 영역의 상관관계를 이용하여 사람을 객체를 검출한다.

히스토그램 기반으로 기하학적 헤드 영역을 검출하며, 몸통과 헤드영역은 세로 방향의 히스토그램 변화량을 기반으로 하여 검출한다.

헤드 영역 검출 시 다른 영상과의 오류를 예방하기 위하여 히스토그램의 변화량을 기준으로 한다. 변화가 심하게 분포되는 경우는 헤드 영역이 아닌 부분으로 인식한다.

3.5 중복 영역에 대한 얼굴 영역 인증

분류된 조명 상황에 따라 베이지안 검출기를 구성한다. 실시간 반영을 위하여 3개의 조명 상황으로 분류된 영상들을 통해 다음과 같은 방법으로 검출기를 구성한다.

본 고에서 제안하는 방법으로 수직으로 중복된 영상에 대해 베이지안 검출기에 의한 얼굴 검출 방법을 제안한다. 동영상에서 인위적으로 검출한 얼굴 이미지와 FERET 얼굴 이미지를 기반으로 하여 학습하였으며, 40 * 40 영상을 기준으로 2*2, 3*3, 4*4, 5*5 의 크기로 벡터화하여 평균값을 추출하여 특징 벡터로 사용하였다.

추출된 특징 벡터들은 베이지안 분류기를 통해서 얼굴과 비얼굴로 분류한다. 베이지안 공식의 기본적인 형태는 식(1)과 같다[14].

$$p(\omega_i|x) = \frac{p(x|\omega_i)p(\omega_i)}{p(x)} \quad (1)$$

여기서 얼굴과 비얼굴에 대한 각각의 판별 함수, $g_f(x)$, $g_n(x)$ 를 아래의 식(4)과 같이 얻을 수 있다.

$$g_f(x) = \ln p(x|\omega_f) + \ln p(\omega_f) \quad (2)$$

$$g_n(x) = \ln p(x|\omega_n) + \ln p(\omega_n) \quad (3)$$

얼굴과 비얼굴의 두 종류로 분류하기 때문에 한 개의 판별식, $g(x)$ 을 사용할 수 있으며, 이는 식(4)와 같이 정의할 수 있다.

$$g(x) = g_f(x) - g_n(x) \quad (4)$$

주어진 패턴, x 는 식 (4)의 판별식을 사용해서 아래의 식(5)와 같은 방법으로 얼굴과 비얼굴로 분류할 수 있다.

$$x \Rightarrow \begin{cases} \omega_f, & \text{if } g(x) > 0 \\ \omega_n, & \text{else} \end{cases} \quad (5)$$

3.6 Human Face Classification

확률 변수를 다변량 정규 분포로 가정하면 아래의 식(8)와 같이 나타낼 수 있다. 각 판별식의 첫 번째 항인 마할라노비스 거리의 제곱을 제외한 나머지 항들은 x 와는 무관한 항이므로, 얼굴과 비얼굴 표본에 대한 마할라노비스 거리, rf 와 rn 의 제곱과 상수, f 과 n 으로 나타내어 식(6)과 같이 표현할 수 있다.

$$r_f^2 = (x - \mu_f)^T \sum_{f=1}^{n-1} (x - \mu_f)$$

$$r_n^2 = (x - \mu_n)^T \sum_{n=1}^{n-1} (x - \mu_n)$$

$$\alpha_f = -\frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\sum f| + \ln P(\omega_f)$$

$$\alpha_n = -\frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\sum n| + \ln P(\omega_n) \quad (6)$$

f 과 n 은 패턴 x 와는 무관하기 때문에 얼굴에 대한 판별식은 마할라노비스 거리, r_f 와 r_n 에 의해서만 영향을 받는다. 따라서 입력 패턴의 얼굴과 비얼굴에 대한 판단은 식(7)과 같은 방법으로 이루어진다.

$$x \Rightarrow \begin{cases} \omega_f, & \text{if } r_f^2 < r_n^2 + r \\ \omega_n, & \text{else} \end{cases} \quad (7)$$

여기서 rf , rn , 그리고 r 는 각각 아래의 식(8)과 같다.

$$r_f^2 = (x - \mu_f)' \sum_f^{-1} (x - \mu_f)$$

$$r_n^2 = (x - \mu_n)' \sum_n^{-1} (x - \mu_n)$$

$$\tau = -\ln \left| \sum_f \right| + \ln \left| \sum_n \right| + 2 \ln \frac{p(\omega_f)}{p(\omega_n)} \quad (8)$$

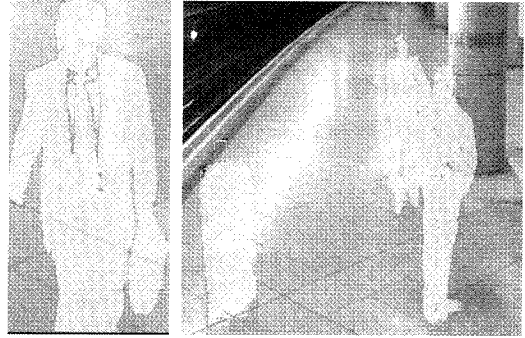
3.7 Face-like Nonface Classification

비얼굴이 얼굴과 달리 모델링하기 어려운 것은 특징을 나타낼 수 없다는 것이다. 얼굴이라는 대상을 제외한 모든 것이 비얼굴이기 때문에, 이 모든 것들의 특징을 뽑아서 비얼굴의 모델을 만들기는 어려운 일이다. 따라서 여기서는 얼굴 표본에 대한 마할라노비스 거리를 기준으로 얼굴과 유사한 비얼굴을 모델링 한다. 이런 비얼굴을 선택하는 방법은 아래의 과정으로 이루어진다.

- m_f 개의 얼굴 표본 집합을 T_f 라 하고, T_f 의 공분산 행렬인 f 를 생성한다.
- 모든 f, T_f 에 대해서 f 에 대한 마할라노비스 거리인 d_{r_i} 를 계산한다.
- 출력값을 $\{max d_{r_i} | 1 \leq i \leq m_f\}$ 를 δ 로 설정한다.
- m_n 개의 비얼굴 표본 집합을 T_n 이라하고, 각 표본, $n_i T_n$ 에 대해서 d_{r_i} 를 계산한다.
- $d_{r_i} < \theta$ 인 n_i 를 얼굴과 유사한 비얼굴 표본 집합인 T_i 로 저장한다.
- 비얼굴의 공분산 행렬을 λ 라 하고, 이를 T_i 로부터 생성한다.

이런 방식으로 선택된 집합의 각 요소들은 얼굴과 일정한 마할라노비스 거리, 안에 있는 얼굴과 유사한 비얼굴의 집합으로 구성된다. TI을 비얼굴 집합으로 사용하면 아래의 식(9)와 같은 방식으로 패턴을 분류할 수 있다.

$$x \Rightarrow \begin{cases} \omega_f, & \text{if } r_f < \theta \text{ and } r_f^2 < r_n^2 + \tau \\ \omega_n, & \text{else} \end{cases} \quad (9)$$



(그림 7) 배경 영상과의 차분 영상에서의 오브젝트 영역

다음은 Boosting 방법에 의해 얼굴 학습하는 방법을 기술한다.

- 학습을 위한(training) 데이터베이스에서 비얼굴 데이터들에 대해 고차원의 적은 샘플들로 구성된다.
- 현재의 데이터베이스에 대하여, 베이지안 분류기로 학습한다.
- 얼굴이 없는 이미지를 얼굴 검출기로 테스트를 한다. 얼굴로 잘못 인식하는 비얼굴 데이터를 수집한다.
- 이렇게 하여, 수집된 비얼굴 데이터를 새로운 비얼굴 클러스터로 구성한다. 단계 2로 가서 계속하여, 수행한다. 반복에 통해서 계속하여 클러스터를 생성한다.

IV. 구현

본 고에서 제안한 시스템은 리눅스 레드햇 10.2 운영 체제에서 개발되었으며, 동영상에서 처리된 프레임은 총 5472 개의 영상이다. 또한, 첫번째 이종의 경계값을 적용하여 shape를 검출할 경우와, 기하학적 모델을 적용할 경우 경계값을 다르게 조정하여 검출하였다. 다음 그림은 배경 영상과 프레임간의 차분 영상을 추출한 결과이다. 배경 영상은 앞서 설명한 masking과 지하철 영역이 표시된 배경을 적용하여 수행하였다. 사용한 영상이 전반적으로 밝고 조명 빛이나 지하철 내부에서의 빛의 반사 등이 심한 편이다.



1718 프레임 1720 프레임



1725프레임 1944 프레임

(그림 9) 객체 검출 및 오메가 shape 검출

검출된 영상을 기반으로 하여 라벨리를 수행한다. 라벨링은 8-이웃 경계선을 검출하였으며, 라벨링된 영상은 사람 오브젝트의 비율을 추정하여 분리한다.

그림 8에서 빨간색 박스는 오브젝트 영역을 검출한 것이며, 이 영역에서 오메가 shape 를 검출한다. 오메가 영역을 추출하여 얼굴의 shape 영역으로 등록한다. 학습 기반의 알고리즘으로 얼굴 영역을 검출하려면 얼굴의 크기 등을 고려하여야 하므로, 최소 3단계 정도는 분석을 해야 하며, 실시간 영상에서는 얼굴 크기가



(그림 10) 검출된 결과

14*14 정도에서 60* 60 정도이므로 이를 커버하기 위해서 8단계 정도로는 해야 하므로, 오메가 모양을 검출함으로써 동영상, 실시간 검출 시간 및 성능을 강화시킬 수 있다.

그림 9는 오브젝트 영역은 빨간 박스처럼 하나로 검출되었는데, 이를 하나의 영역 오브젝트로 보고 사람 객체 및 헤드 영역을 검증하기 위하여 영역 내에서 오메가 shape 를 검출한 결과 영상이다.



(그림 8) 오브젝트 영역 검출 및 오메가 shape



(a) 중복 영역



(b) 얼굴 영역 검출

(그림 11) 중복 영역 검출



(그림 12) 검출 예

(표 1) comparison of result for cascade classifier

Method		Accept	Reject
Single	Classifier	346	57
Cascade	Classifier	396	9

첫번째 검출기에서 검출되지 않은 오브젝트 영역에 대해 두번째 검출기에서 검출하도록 한다. 다음의 그림 10은 한 프레임에서 두 개의 오브젝트 박스를 검출하고, 왼쪽의 영역에서는 오메가 shape로 검출된 영상이고, 오른쪽 영역에서는 두번째 검출기에서 검출된 영역이다.

다음의 그림 11은 중복된 영상을 보여준다. 이 영역에서 베이지안 기반의 얼굴 인증기를 통해 검출된 얼굴 데이터가 (c)번에 나와 있으며, 검출된 박스 영역은 왼쪽 그림과 같다.

멀리 있는 작은 영상과 가방 손잡이 부분이 검출되지 않는다. 윤곽선만으로 평가하면 가방 손잡이 부분이 유사한 패턴으로 나오기는 하나, 히스토그램 분포로써 헤드 영역이 아닌 것으로 인식한다.

중복된 영역으로 판별된 영상에서의 얼굴 검출 성능은 다음과 같다. 5472개의 프레임에서 검출하였으며, 48개의 프레임에서 중복된 정면 얼굴 영상이 존재하며, context based Bayesian 얼굴 검출기를 이용하여 검출하였을 경우, 정면 얼굴이 포함된 32개의 프레임에서 얼굴 검출이 가능하였다. 총 5472개의 프레임 중 false alarm의 수는 10개이다.

조명 상황이 계속하여 변화하고, 밝은 영상이어서 사람 전체 영역보다 얼굴 부분의 경계를 찾는 것에 어려움 있어, 추출된 오브젝트 영역에서 얼굴 영역을 검출하여 검증하도록 하였다.

베이지안 얼굴 검출기를 이용하여 오프라인 이미지를 테스트 한 결과는 다음과 같다.

베이지안 얼굴 검출기는 실시간에서의 성능 향상을 위하여 조명 상황에 따라 다양한 형태로의 영상으로 분류한 후 각각의 분류기를 구성한다.

실시간으로 수집한 403장의 영상에서 하나의 검출기로 검출하였을 경우 다음의 결과를 얻을 수 있었다.

사람 객체의 헤드 영역의 추출을 위하여 베이지안 검출기를 적용하였더니 CCTV 저해상도 영상에서 성능 및 속도가 저하되었다. 따라서, 다중의 오브젝트 검출, 추적을 위하여 다중의 검출기를 캐스케이드 방식으로 적용함으로써 검출성능을 높일 수 있었다.

V. 결 론

객체 추적 기술은 배경 영상과의 차분 영상, 베이저안, 아다부스팅 기법 등의 외형기반 학습 방법 등 다양한 방법으로 연구되고 있다. 최근 들어 다중에 카메라를 이용하여 다수의 사람을 트래킹하는 연구도 되고 있으며, 정보를 이용하여 위협인지 등의 보안 사항으로 적용하기에 적합하다. 따라서 본 고에서는 이러한 기능을 수행하기 위해 배경 영상을 생성하여 차분영상을 통해 오브젝트 영역을 검출하고, 오메가 윤곽선을 기반으로 하여 사람객체를 추출한 후, 윤곽선으로 추출되지 않은 영역에 대해 히스토그램 기반의 기하학적 모델을 적용하여, 통계학적 얼굴 인증 기술을 통해 정면의 중복 문제를 해결하고자 하였다. 이에 전체적으로 얼굴 영역을 검출하거나 학습기반의 알고리즘에 비해 학습 데이터의 수집 등의 어려움을 다소 해결 할 수 있었으며, 오브젝트 검출 시 발생하는 오류 등에서도 사람 객체를 추출할 수 있도록 하였다.

향후, 실시간 추적이 좀 더 자연스러울 수 있도록 속도 및 알고리즘 개선의 노력이 필요하며, 조명 등의 빛에 대한 영상의 왜곡에 대한 문제를 해결할 수 있도록 전처리 및 모델링에 대한 연구가 필요할 것으로 보인다.

참고문헌

- [1] Alper Yilmaz, Omar Javed and Mubarak Shah, "Object Tracking: A Survey," ACM Computing Surveys, Vol. 38, No. 4, Article 13, 2006.
- [2] ELGAMMAL, A., DURAISWAMI, R., HARWOOD, D., AND DAVIS, L., "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance" *Proceedings of IEEE 90*, Vol7, pp.1151 - 1163, 2002
- [3] PARAGIOS, N. AND DERICHE, R., "Geodesic active regions and level set methods for supervised texture segmentation," *Int. J. Comput. Vision Vol.46*, No.3, pp.223 - 247, 2002.
- [4] COMANICIU, D., RAMESH, V., AND MEER, P. Kernel-based object tracking. *IEEE Trans. Patt. Anal. Mach. Intell.* 25, pp.564 - 575, 2003.
- [5] PARK, S. AND AGGARWAL, J. K., "A hierarchical bayesian network for event recognition of human actions and interactions." *Multimed. Syst.10*, pp. 164 - 179, 2004.
- [6] EDWARDS, G., TAYLOR, C., AND COOTES, T., "Interpreting face images using active appearance models," In *International Conference on Face and Gesture Recognition*, Pp.300 - 305, 1998.
- [7] PASCHOS, G. "Perceptually uniform color spaces for color texture analysis: an empirical evaluation," *IEEE Trans. Image Process. Vol 10*, pp.932 - 937, 2001.
- [8] CANNY, J. "A computational approach to edge detection," *IEEE Trans. Patt. Anal. Mach. Intell.* 8, 679 - 698, 1986
- [9] T. Olson and F. Brill., "Moving object detection and event recognition algorithms for smart cameras," *Proc. DARPA Image Understanding Workshop*, pages 159 - 175, 1997.
- [10] Ramanan, D. A. Forsyth, and A. Zisserman. Tracking people by learning their appearance. *PAMI*, 29:65 - 81, 2007
- [11] Sigal and M. J. Black. Measure locally, reason globally: occlusion-sensitive articulated pose estimation. *CVPR 2006*.
- [12] C. Sminchisescu, A. Kanaujia, and D. N. Metaxas. "BM3E: Discriminative density propagation for visual tracking." *PAMI*, 29:2030 - 2044, 2007.
- [13] Paul A. Viola, Michael J. Jones, Daniel Snow, "Detecting Pedestrians Using Patterns of Motion and Appearance" *ICCV 2003*, pp.734-741, 2003.
- [14] C. Liu, "A Bayesian Discriminating Features Method for Face Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.25, no.6, (2003), pp.725-740
- [15] http://www.elec.qmul.ac.uk/staffinfo/andrea/avss-2007_d.html



김 건 우 (Geon Woo Kim)

정회원

1998년 2월: 경북대학교 컴퓨터학과 졸업

2000년 2월: 경북대학교 컴퓨터학과 석사

2010년 2월: 경북대학교 컴퓨터학과 박사

2000년 2월 ~ 현재: 한국전자통신연구원 선임연구원

<관심분야> 지식보안, 영상보안, 네트워크보안



남 미 영 (Mi Young Nam)

정회원

2001년 2월: 신라대학교 전자계산학과 석사

2006년 2월: 인하대학교 컴퓨터정보공학과 박사

2008년: 인하대학교 BK21 박사후과정

<관심분야> Biometrics, Neural Networks, Face Recognition, Genetic Algorithm, Evolutionary System, Biological Signal Processing and Biomedical System



한 종 옥 (Jong Wook Han)

정회원

1989년 2월: 광운대학교 전자공학과 졸업

1991년 2월: 광운대학교 전자공학과 석사

2001년 2월: 광운대학교 박사

1991년 3월 ~ 현재: 한국전자통신연구원 책임연구원/융합서비스보안연구팀장

<관심분야> 융합보안, 물리보안