

Haar 웨이블릿 특징과 피부색 정보를 이용한 실시간 얼굴 검출 및 추적 방법

박 영 경[†] · 서 해 종^{**} · 민 경 원^{***} · 김 중 규^{****}

요 약

본 논문은 Haar 웨이블릿 특징과 피부색 정보를 이용한 실시간 얼굴 검출 및 추적 방법을 제안한다. 검출과 추적이 각기 다른 특징들을 이용해 이루어졌던 기존 방법과는 달리 본 논문에서는 피부색 정보와 Haar 웨이블릿 특징을 검출과 추적에 동시에 이용하고 두 특징들을 효과적으로 융합함으로써 빠르고 강한 추적을 가능하게 한다. 제안된 방법은 검출과 추적에 동일한 특징을 이용함에도 불구하고 표본화에 기반을 둔 Condensation 알고리즘의 특징으로 인해 검출 방법만을 적용했을 때 검출하지 못하는 얼굴의 회전이나 가려짐 등의 문제를 효과적으로 해결한다. 특히, 얼굴의 위치와 함께 크기 변화를 효과적으로 추적하기 때문에 얼굴의 위치 및 크기를 정확하게 알아야 하는 얼굴 인식이나 표정 인식 등의 다양한 어플리케이션에 이용되기에 용이하다. 제안된 방법은 복잡한 배경 및 다양한 얼굴 자세 등의 변화에 대한 테스트를 통해 효율성을 검증한다.

키워드 : Haar, CONDENSATION, AdaBoost, 크기 투영, 피부색

A Real-Time Face Detection/Tracking Methodology Using Haar-wavelets and Skin Color

Young Kyung Park[†] · Hae Jong Seo^{**} · Kyoung Won Min^{***} · Joong-Kyu Kim^{****}

ABSTRACT

In this paper, we propose a real-time face detection/tracking methodology with Haar wavelets and skin color. The proposed method boosts face detection and face tracking performance by combining skin color and Haar wavelets in an efficient way. The proposed method resolves the problem such as rotation and occlusion due to the characteristic of the condensation algorithm based on sampling despite it uses same features in both detection and tracking. In particular, it can be applied to a variety of applications such as face recognition and facial expression recognition which need an exact position and size of face since it not only keeps track of the position of a face, but also covers the size variation. Our test results show that our method performs well even in a complex background, a scene with varying face orientation and so on.

Key Words : Haar, Condensation, AdaBoost, Amplitude Projection, Skin Color

1. 서 론

얼굴은 개인의 고유성뿐만 아니라 감정을 반영하는 특징을 가지고 있기 때문에 사람을 인지하고 이해하기 위해 이용될 수 있는 가장 유용한 패턴이다. 얼굴 인식과 표정 인식은 이러한 얼굴 처리 기술의 대표적인 예로서 사람의 신원을 파악하고 감정 상태까지 추정할 수 있어 최근 많은 관심을 불러 모으고 있는 지능형 홈이나 인간형 로봇에 활용

되기에 가장 적절한 기술이라 할 수 있다. 이러한 얼굴 인식과 표정 인식을 수행하기 위해서는 영상에서 얼굴을 찾고 추적하는 기술이 먼저 선행되어야 한다. 특히, 실시간 얼굴 인식 및 표정 인식을 수행하기 위해서는 얼굴의 크기나 자세 같은 파라미터를 실시간으로 정확하게 추정하는 것이 요구되기 때문에 빠르고 정확한 얼굴 검출 및 추적은 무엇보다도 중요한 요소 기술이라 할 수 있다.

현재까지 얼굴 검출 및 추적에 관한 많은 연구가 이루어져 왔고 많은 방법들이 제안되어 왔다. 먼저 얼굴 검출은 크게 네 가지의 유형으로 나누어 볼 수 있다. 첫 번째 유형은 지식 기반 방법으로서 전형적인 얼굴에 대한 인간의 일반적인 지식을 기반으로 하는 방법으로 얼굴을 구성하는 눈,

[†] 준 회원 : 성균관대학교 정보통신공학부 박사과정
^{**} 준 회원 : 성균관대학교 정보통신공학부 석사과정
^{***} 정 회원 : 전자부품연구원 디지털미디어 연구센터 책임연구원
^{****} 정 회원 : 성균관대학교 정보통신공학부 교수
 논문접수 : 2006년 3월 7일, 심사완료 : 2006년 5월 23일

코, 입 간의 위치, 거리 등이 이에 해당한다. 두 번째로 불변 특징 접근 방법이 있다. 이 방법은 자세나 조명등의 변화에서도 얼굴 검출에 용이한 구조적 특징을 찾는 방법으로 주로 얼굴 국부 화에 이용된다. 세 번째 유형은 템플릿 매칭 방법으로 몇몇 얼굴의 기본 형태를 얼굴의 전체 및 부분에 따라 저장하고 입력 영상과의 비교를 통해 일정 수준 이상의 상관관계를 가질 경우 얼굴로 판단하는 방법이다. 네 번째 유형은 최근 가장 널리 연구되고 있는 방법으로 외형 기반 방법이다. 외형 기반 방법은 나머지 세 가지 유형들과 비교해 학습을 통해 얼굴을 모델링 한다는 점에서 큰 차이를 가지는데 최근 강력하고 우수한 기계 학습 알고리즘들이 많이 소개 되고 있고 자세의 변화에 따른 얼굴의 특성도 모델링 하기가 용이한 등 다른 방법들과 비교해 많은 장점을 가지고 있다. 외형 기반 방법에 적용되는 대표적인 방법으로는 neural network(NN)[2]나 support vector machine(SVM)[3], adaptive boosting (AdaBoost)[4, 5]가 있다. 특히, 이들 중 AdaBoost와 SVM은 최근 얼굴 검출과 관련해 많은 연구가 이루어지고 있는 학습 알고리즘으로서 우수하고 효과적인 검출 성능을 보이는 것으로 알려져 있다.

얼굴 추적은 동영상에서 얼굴의 위치 및 크기 변화를 추적하는 것으로 크게 머리를 추적하는 방법과 얼굴의 주요 특징을 추적하는 방법으로 나눌 수 있다. 특징을 추적하는 방법은 눈이나 입과 같은 얼굴내의 현저한 특징들을 추적하며 일반적으로 이러한 방법들은 각 특징마다 개개의 독립된 추적기가 요구된다[6]. 이와 달리 머리를 추적하는 방법은 머리의 전체적인 정보를 이용하는 것으로 색상, 형태 및 모델 등에 기반을 둔 다양한 방법이 있다. 이 중 색상이 기반의 방법은 조명변화에 강인하지 못하고 가려짐을 다루지 못하기 때문에 다른 추적 방법과 결합되는 형태로 많이 이용된다. 대표적인 예로서 Birchfield[7]는 타원형을 갖는 얼굴의 경계에서 경사도를 측정하고 그 타원의 내부의 색상정보를 이용했다. 이들 추적 방법들의 대부분은 Kalman 필터[8]나 Condensation 알고리즘[9]에 의해 추적 과정이 이루어지는데 최근에는 임의의 확률 분포를 모델링 할 수 있는 Condensation 알고리즘이 많이 이용되고 있다. 지금까지 소개한 바와 같이 검출과 추적은 각기 독립적인 프로세서로서 제안된 대부분의 방법들은 검출과 추적의 효율적인 결합방법 보다는 각각의 성능개선에 관련된 연구가 주로 진행되어 왔다. 그러나 실생활에 응용되기 위해서는 모든 과정이 자동적으로 이루어지는 것이 필요하고 이를 위해서는 검출과 추적을 효과적으로 결합하는 것이 요구된다. 본 논문에서는 이러한 요구를 만족시키기 위해 얼굴의 특징으로 훈련된 Haar 웨이블릿과 피부색을 검출과 추적에 동시에 이용함으로써 효율을 높이고 검출과 추적을 효과적으로 결합하는 방법을 제안한다. 본 논문에서는 검출과 추적에 있어 성능이 우수하다고 이미 알려져 있는 AdaBoost와 Condensation 알고리즘[9]을 사용한다. 비록 이 두 알고리즘들에 대한 많은 연구들이 소개되어 왔지만 우리가 제안하는 방법은 두 방법들을 효과적으로 결합함으로써 각 방법들이 가지고 있는 단점을 극복하고자 하는데 있다. 제안되는 방법의 특징은 크게 세 가지

로 나뉜다. 먼저 AdaBoost기반의 얼굴 검출 방법이 강인한 피부색 검출과 크기 투영에 의한 얼굴 후보 검출 및 효과적인 영상 검색 방법을 이용해 실시간 검출이 가능하도록 개선된다. 두 번째로 AdaBoost에 의해 훈련된 Haar 웨이블릿 특징과 피부색 밀도를 Condensation 알고리즘의 관측모델에 효과적으로 적용하고 얼굴 크기 파라미터를 동적 모델에 적용함으로써 빠르고 정확한 얼굴 추적을 가능하게 한다. 세 번째, 실제 얼굴 인식이나 표정 인식 등의 실시간 어플리케이션에 적용되기 위해 추적에 실패하는 상황을 자동으로 판단하는 방법과 얼굴의 크기 변화를 실시간으로 추적하기 위한 방법을 제시한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 얼굴 검출 방법을 설명한다. 3장에서는 얼굴 검출과 추적을 융합한 실시간 얼굴 추적 시스템을 설명한다. 이후 4장에서 실험 결과를 보인 후 5장에서 결론을 맺는다.

2. 피부색과 크기 투영에 의한 얼굴 검출의 속도 및 정확도 개선

얼굴 검출은 위치, 조명, 크기 등에 상관 없이 영상 내의 얼굴이 존재할 경우 얼굴의 존재 여부 및 위치를 알아내는 기술로서 모든 얼굴 처리 관련 시스템에서 선행되어야 할 요소 기술이다. 특히, 얼굴 추적을 위해서는 추적해야 할 얼굴의 영역을 결정하는 초기화가 반드시 요구되는데 정확한 얼굴 검출을 통해 초기화를 자동적으로 수행할 수가 있다. 그러므로 본 논문에서는 실시간 얼굴추적시스템에 적용되기에 적합한 얼굴 검출 방법을 제안한다. 제안되는 얼굴 검출 방법은 피부색 필터, 크기 투영에 의한 얼굴 후보 영역 선정 및 AdaBoost에 의한 얼굴 검증의 세 단계로 나뉜다.

2.1 피부색 영역 검출

피부색 필터를 얼굴 후보 영역의 검출 및 얼굴의 크기 추정을 위한 전처리 과정으로 이용함으로써 검출 시간의 단점을 해결하고 검출의 정확도를 높일 수 있다.

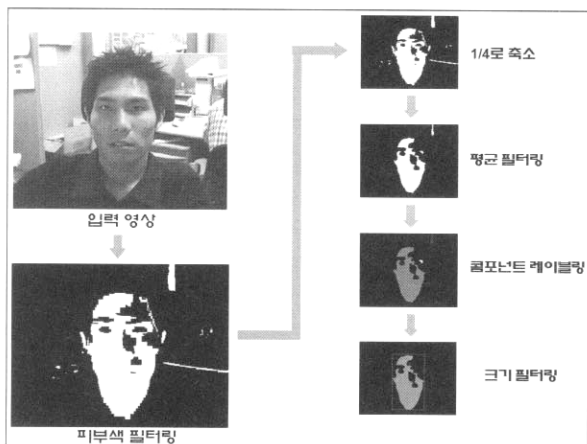
얼굴 검출 과정에서의 문제점 중 하나는 얼굴이 아님에도 불구하고 얼굴과 비슷한 형상이 얼굴로 판단되어 검출된다는 것이다. 뿐만 아니라 영상 검색에 기반을 둔 얼굴 검출 과정은 검색 영역의 크기에 따라 현저한 검출 시간의 차이를 가져온다. 이러한 문제점을 해결하기 위해 본 얼굴 검출 시스템에서는 입력영상에 피부색 필터를 적용시켜 피부색이 아닌 영역을 제거한 후 검출을 수행한다. 이러한 전처리 과정은 얼굴 검출 과정에서의 검색 영역을 줄임으로써 검출 시간을 단축시키고 피부색 필터링 과정을 통하여 초과 검출의 가능성을 줄여 보다 정확한 얼굴 검출이 이루어지도록 하고 불필요한 연산을 제거함으로써 검출 속도를 향상시킨다. 피부색 필터의 성능은 피부색 모델에 의해 결정되는데 본 논문에서는 Garcia[10]에 의해 제안된 YCbCr 색상 모델 기반의 피부색 모델(식(1))을 적용했다. 적용된 피부색 모델은 다양한 인종과 조명환경을 가진 훈련 영상 집합으로부터

생성된 것으로 환경에 비교적 강한 특성을 가지고 있다.

$$\left\{ \begin{array}{l} \text{if } (Y \leq 128) \theta_1 = -2 + \frac{256-Y}{16}; \theta_2 = 20 - \frac{256-Y}{16}; \theta_3 = 6; \theta_4 = -8 \\ \text{if } (Y > 128) \theta_1 = 6; \theta_2 = 12; \theta_3 = 2 + \frac{Y}{32}; \theta_4 = -16 + \frac{Y}{16} \\ \left\{ \begin{array}{l} Cr \geq -2(Cb + 24); Cr \geq -(Cb + 17); Cr \geq -4(Cb + 32) \\ Cr \geq 2.5(Cb + \theta_1); Cr \geq \theta_3; Cr \geq -0.5(Cb - \theta_4) \\ Cr \leq -\frac{Cb - 220}{6}; Cr \geq -1.34(Cb - \theta_2) \end{array} \right. \end{array} \right. \quad (1)$$

피부색 모델을 적용하기 위해 입력 받은 RGB 영상은 YCbCr 영상으로 변환된다. 변환된 YCbCr 영상에 각 픽셀마다 피부색 필터를 적용하면 (그림 1)과 같이 수많은 검출된 피부색 클러스터를 형성하게 된다. 이 가운데 잡음 요소는 제거하고 주요 피부색 영역들만을 검출하기 위해 피부색 필터가 적용된 영상을 1/4로 축소하고 평균 필터를 적용시킨다. 이러한 과정은 실제 동일한 영역임에도 불구하고 조명등과 같은 기타 요인들로 인해 분리되어 검출된 클러스터들을 연결시키는 효과가 있다. 평균 필터가 적용된 영상에는 일반적으로 한 개 혹은 여러 개의 검출된 피부색 영역이 나타나는데 이 각각의 연결된 피부색 영역들은 얼굴 후보 영역이 되며 각 영역에 얼굴 검출 알고리즘을 효과적으로 적용하기 위해서는 각 영역의 좌표 및 크기를 알아야 한다. 이를 위해 컴포넌트 레이블링을 적용함으로써 각각의 연결된 피부색 영역을 추출하고 좌표 및 크기를 계산한다. 추출된 각 영역에 대해서는 크기 필터를 적용해 일정 크기 이하의 피부색 영역은 제거한다. 그림은 피부색 영역을 검출하는 과정을 보여준다.

피부색 영역이 결정되면 강한 조명 효과로 인해 얼굴의 일부가 후보 영역에 포함되어 있지 않은 경우를 위해 영역을 상하좌우로 각각 0.1배로 확장시킨다. 이 과정은 얼굴의 일부가 후보영역에 포함되지 않은 경우 포함되게 하기 위한 과정이다. 0.1배로 확장된 후보 영역은 일반적으로 열

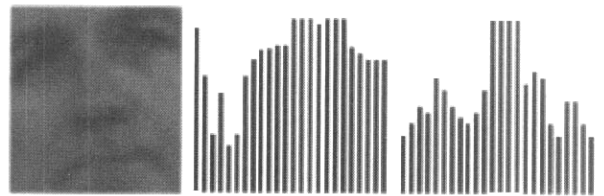


(그림 1) 피부색 영역의 검출 과정

굴보다 일정 비율만큼 큰 영역이 된다. 그러므로 확장된 후보 영역의 작은 변의 길이 a를 기준으로 결정된 비율로 영상을 축소하고 정확한 얼굴을 찾기 위해 영상을 검색하면서 검출 알고리즘을 적용한다. 본 논문에서는 24x24크기에서 얼굴 검출을 수행하기 때문에 축소 비율은 a/(24x1.5)로 결정된다.

2.3 크기 투영을 이용한 얼굴 후보 검출

입력 영상에서 피부색 영역 검출을 수행하고 나면 검출된 피부색 영역은 추정된 얼굴 크기에 따라 축소되고 24x24크기의 마스크로 얼굴후보영역을 검색하게 된다. 각 검색 단계에서 마스크를 통해 얻어지는 24x24영상에 대해 얼굴 검출을 수행하게 되는데 이때 얼굴의 수직방향과 수평방향의 변화 정도를 이용한 크기 투영[2]의 상호 상관 계수 값들을 이용해 판별함수에 적용하여 얼굴인지 아닌지를 미리 검사하게 된다. 이것은 일반적인 얼굴이 가지는 특성을 반영한 것으로써 (그림 2)는 일반적인 얼굴에 대한 수평, 수직 방향의 크기 투영을 보여준다. 얼굴 영상 자체로는 측정할 수 없었던 특징 즉, 코 부분이 다른 부분보다 그레이 레벨 값이 더 크게 나타나는 것을 수평, 수직 방향으로 영상을 투영한 그림을 통해서 알 수 있다. 이러한 크기 투영에 의한 얼굴 검출 방법은 주 검출 방법인 AdaBoost에 의한 검출 방법에 비해 정확도는 떨어지지만 시간이 적게 걸리며 이용하는 얼굴의 특성이 AdaBoost에 의한 검출 방법과 차이가 있어 AdaBoost에 의한 검출 오류를 줄일 수 있다. 이러한 점은 크기 투영을 주 검출 과정의 전 단계로 이용함으로써 얼굴검출시스템의 속도와 정확도를 개선시킬 수 있음을 말해준다. 특히, 크기 투영은 다음 절에서 소개되는 Haar 웨이블릿 특징 값을 쉽게 구하기 위한 누적 영상으로부터 간단하게 구할 수 있어 불필요한 추가 연산을 요구하지 않는다는 장점이 있다.



(그림 2) (a) 얼굴 영상, (b) 수평방향 크기 투영, (c) 수직방향 크기 투영

일반적인 얼굴에 대한 크기 투영의 특징은 훈련에 의해 획득된다. 훈련영상집합내의 각 영상의 수평/수직 방향 크기 투영을 구하고 훈련영상집합으로부터 구해진 평균영상 (mean face)의 수평/수직 방향 크기 투영을 구한다. 그리고 이들 간의 상호상관계수 값들을 식 (2)와 같이 구한다.

$$\rho = E \left[\frac{(X - \bar{X})(Y - \bar{Y})}{\sigma_X \sigma_Y} \right] \quad (2)$$

- X : 훈련영상의 수직/수평 방향 크기 투영 값
- \bar{X} : 훈련영상의 수직/수평 방향 크기 투영 값의 평균
- Y : 평균영상의 수직/수평 방향 크기 투영 값
- \bar{Y} : 평균영상의 수직/수평 방향 크기 투영 값의 평균

일반적인 얼굴의 크기 투영을 모델링 하기 위해 본 논문에서는 수평/수직 투영의 상호상관계수 값의 2차원 Gaussian 분포를 가정한다. 즉, 수직 방향 및 수평 방향에 대한 상호상관계수가 2차원 벡터를 형성하게 되고 훈련영상 집합과 입력영상간의 계산된 상호상관계수를 이용하여 일반적인 얼굴의 크기 투영특성을 Gaussian 모델링 하는 것이다. Gaussian 모델링은 2차원 Gaussian 함수를 구성하는 데 필요한 평균 및 공 분산행렬을 구하는 과정으로 얼굴 훈련 영상집합과 비 얼굴 훈련영상집합에 대해 각각 계산되어 얼굴과 비 얼굴의 크기 투영 특성을 모델링 하게 된다.

크기 투영을 위한 얼굴 후보 검출단계에서는 입력 영상에 대해 수평/수직 크기 투영 값을 구하고 평균 영상과의 상호상관 계수 값을 구한다. 상호상관계수 값을 이미 모델링 된 얼굴과 비 얼굴에 대한 Gaussian 함수에 적용시켰을 때 더 높은 값을 가지는 클래스로 분류하게 된다. 이 때 계산을 용이하게 하기 위해 Gaussian 함수로부터 유도된 판별함수를 이용한다. 각 클래스에 대한 판별함수[11]는 식 (3)에 나와 있다.

$$g_i(x) = x^t W_i x + w_i^t x + w_{i,0} \tag{3}$$

where, $W_i = -\frac{1}{2} \Sigma_i^{-1}$, $w_i = \Sigma_i^{-1} \mu_i$,

$$w_{i,0} = -\frac{1}{2} \mu_i^t \Sigma_i^{-1} \mu_i - \frac{1}{2} \ln |\Sigma_i| + \ln P(w_i),$$

$$\Sigma_i = \begin{bmatrix} \sigma_{x_1}^2 & \rho \sigma_{x_1} \sigma_{x_2} \\ \rho \sigma_{x_1} \sigma_{x_2} & \sigma_{x_2}^2 \end{bmatrix},$$

$$\Sigma_i^{-1} = \frac{1}{(1-\rho^2)} \begin{bmatrix} 1/\sigma_{x_1}^2 & -\rho \sigma_{x_1} \sigma_{x_2} \\ -\rho \sigma_{x_1} \sigma_{x_2} & 1/\sigma_{x_2}^2 \end{bmatrix},$$

- $\sigma_{x_1}^2$: 수직방향 상호상관계수 값들의 분산
- $\sigma_{x_2}^2$: 수평방향 상호상관계수 값들의 분산
- ρ : 수직방향과 수평방향의 상호 상관계수 값들의 상호 상관계수 값
- Σ_i : 공 분산 행렬
- μ_i : 수직방향, 수평방향 상호상관계수 값들의 평균값을 성분으로 하는 열 벡터
- x : 판별을 위한 입력 영상과 얼굴 평균영상의 수직방향, 수평방향 상호상관계수 값을 각각 열의 성분으로 하는 열 벡터

위의 판별함수를 적용하여 얼굴이 아닐 것 같은 영상을

미리 제거 하고 얼굴검출의 정확도와 속도 개선을 이루어낸다. 판별함수를 이용한 얼굴 후보에 대한 판단은 다음과 같이 이루어진다.

$$g_{positive}(x) = x^t W_{positive} x + w_{positive}^t x + w_{positive,0} \tag{4}$$

$$g_{negative}(x) = x^t W_{negative} x + w_{negative}^t x + w_{negative,0} \tag{5}$$

$$g_{positive}(x) - g_{negative}(x) > 0 \tag{6}$$

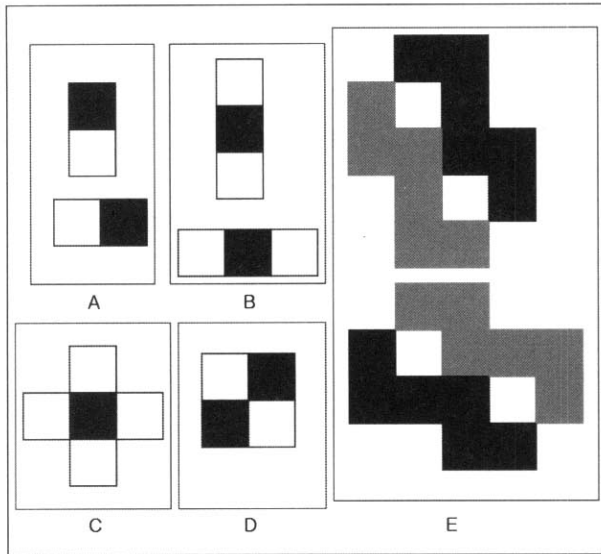
$$g_{positive}(x) - g_{negative}(x) = \text{Confidence} \tag{7}$$

식 (4)와 (5)로부터 입력 영상의 얼굴 클래스와 비 얼굴 클래스에 대한 판별함수 값을 얻고 식 (6)과 같이 얼굴 클래스의 판별함수 값이 비 얼굴 클래스의 판별함수 값보다 크면 얼굴 후보로 판정한다. 두 식의 차의 크기는 결국 얼굴일 신뢰도를 나타내는 것으로 이 값이 크면 클수록 얼굴일 확률이 크고 작으면 얼굴일 확률이 상대적으로 작음을 의미한다.

2.4 AdaBoost를 이용한 얼굴 검출

크기 투영을 이용해 얼굴 후보에 대해서 정확하게 얼굴인지를 판단하기 위한 본 얼굴 검출 과정이 이루어지게 된다. 본 논문에서는 본 얼굴 검출 방법으로서 Viola와 Jones에 의해 제안된 Haar 웨이블릿 특징 기반의 AdaBoost 얼굴 검출 방법을 적용했다[5]. AdaBoost는 Adaptive Boosting의 약어로서, 약한 분류기의 조합으로 강한 분류기를 생성하는 boosting을 적용적으로 수행하는 알고리즘이다. 이것은 얼굴의 세부 요소 분류기를 약한 분류기로 이용할 수 있다는 점에서 유용하며 특히 직사각형 특징인 Haar 웨이블릿 특징은 이러한 요구를 잘 만족시킨다.

Haar 웨이블릿 특징은 인접한 직사각형 영역들의 픽셀 값들의 합의 차로 단순하면서도 계산하기에 용이하다. 본 얼굴 검출 시스템에서 적용하고 있는 Haar 웨이블릿 특징들은 (그림 3)과 같이 Viola와 Jones에 의해 2001년도와 2003년도에 각각 제안된 특징들과 자체적으로 추가한 특징을 포함해 모두 8가지로 구성된다[5, 12]. 이 특징들은 모두 같은 모양, 같은 면적을 가진 수평, 수직으로 인접한 직사각형으로 구성되어있다. 주 검출단계에서는 24x24 픽셀크기의 훈련영상을 사용하는데 24x24 픽셀크기에서 가능한 모든 Haar 웨이블릿 특징의 총 개수는 매우 많으며(A타입: 86,400, B타입: 55,200, C타입: 1,436, D타입: 20,736, E타입: 6,600) 이들 각 특징들을 효율적으로 계산하기 위해 누적 영상 표현법(integral image representation)이 사용된다. 누적 영상은 식 (9)와 같이 픽셀 당 간단한 연산으로 쉽게 구현이 가능하며 누적 영상이 얻어지면 (그림 4)와 같이 직사각형의 픽셀 값들의 합을 세 번의 더하기 연산만으로 구할 수 있어 빠르게 Haar 웨이블릿 특징 값을 계산할 수 있다. 특히 앞서 설명한 크기 투영 역시 픽셀 값들의 합을 구하는 과정이기 때문에 누적 영상으로부터 쉽게 구해 질 수 있어 매우 효율적이다.



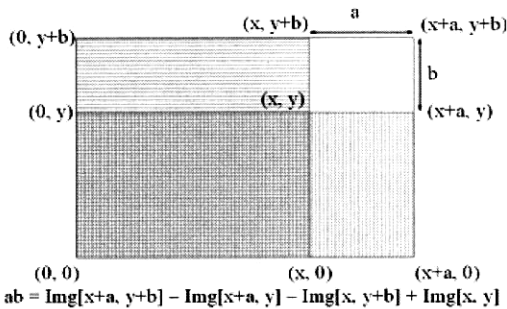
(그림 3) 얼굴 검출을 위한 Haar 웨이블릿 특징

$$ii(x, y) = \sum_{x' < x, y' < y} i(x', y') \quad \text{식 (8)}$$

여기서, $ii(x, y)$ 는 누적영상이고 $i(x, y)$ 는 원 영상이다.

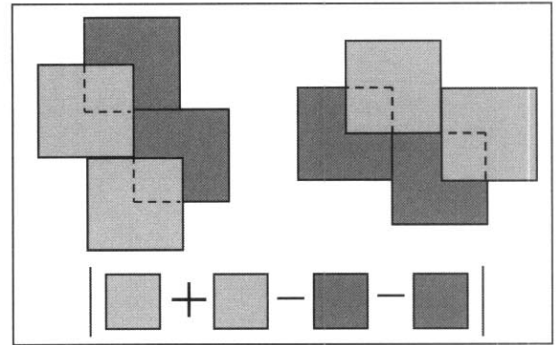
$$\begin{aligned} s(x, y) &= s(x, y-1) + i(x, y) \\ ii(x, y) &= ii(x-1, y) + s(x, y) \end{aligned} \quad \text{식 (9)}$$

여기서, $s(x, y)$ 는 행에 대한 누적 합이며, $s(x, -1) = 0$, $ii(-1, y) = 0$ 로 초기화된다.



(그림 4) 누적 영상으로부터 사각형ab의 픽셀 값들의 합의 계산

특징 값은 A, B, C와 D 타입의 경우 검정색 영역의 픽셀 값들의 합과 흰색 영역의 픽셀 값들의 합의 차로 계산되며 E 타입의 경우 검정색 영역의 픽셀 값들의 합과 회색 영역의 픽셀 값들의 합의 차로 계산된다. 여기서 특징 값이 음의 실수를 가지지 않도록 모든 특징 값은 절대 값을 유지한다. 각 영역의 픽셀 값들의 합은 앞서 설명한 바와 같이 누적영상으로 구해질 수 있는데 A, B, C와 D 타입의 경우 (그림 3)으로부터 계산되어야 할 직사각형의 영역을 쉽게 알 수 있으며 E 타입의 경우 역시 (그림 5)에서와 같이 모두 네 개의 직사각형 영역에 대한 픽셀 값들의 합을 계산함으로써 쉽게 구해질 수 있다.



(그림 5) Haar 웨이블릿 특징 값의 계산(E 타입)

수많은 Haar 웨이블릿 특징 가운데 분류 성능이 좋은 특징만을 선정하기 위해 AdaBoost알고리즘이 적용된다. 훈련 영상집합은 $S_m = \{(x_1, y_1), \dots, (x_n, y_n)\}$ 으로 표현되며 여기서 각 x_i 에 대해 $y_i = 1$ 일 경우에는 얼굴인 영상을, $y_i = 0$ 일 경우에 얼굴이 아닌 영상을 나타낸다. AdaBoost알고리즘은 T라운드 동안 반복적으로 약한 학습기를 호출하여 약한 분류기를 선정하고 분류성능에 맞게 가중치를 부여한다. 약한 학습기는 각 라운드에서 훈련영상집합에 대해 가장 적은 에러를 가지는 약한 분류기를 선정하도록 되어 있다. 약한 분류기는 앞서 설명한 Haar 웨이블릿 특징을 이용하여 다음과 같이 구성 된다.

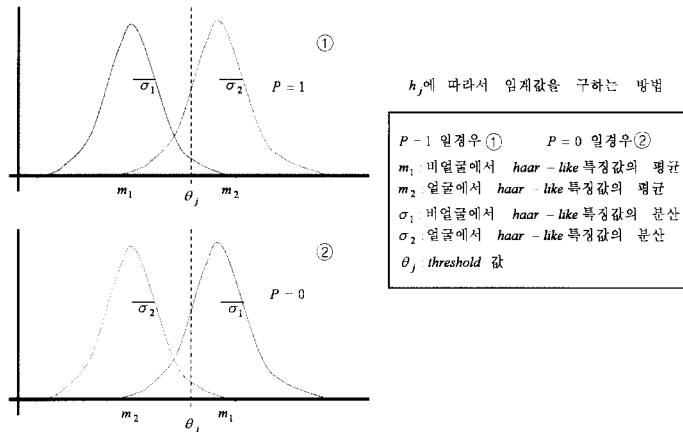
$$h_j(x) = \begin{cases} 1 & \text{if } p_j f_j(x) < p_j \theta_j \\ 0 & \text{otherwise} \end{cases} \quad \text{식 (10)}$$

훈련 영상에 대해 가능한 모든 약한 분류기에 대해 j 번째 분류기를 $h_j(x)$ 라고 하면 임계값 θ_j 와 극값 p_j 를 이용해 Haar 웨이블릿 특징 값인 $f_j(x)$ 가 일정 값보다 크거나 작을 경우 얼굴이라고 판단하게 된다. 본 논문에서는 부호 및 임계값을 훈련영상집합에 대한 Haar 웨이블릿 특징 값의 Gaussian 분포를 가정함으로써 획득한다.

일반적으로 임계값은 훈련영상집합에 대한 Haar 웨이블릿 특징 값의 평균으로 정하게 되는데 (그림 6)과 같이 Gaussian 분포를 가정하여 임계값을 결정할 경우 보다 정확한 결과를 얻을 수 있다. 약한 분류기가 훈련영상집합을 이용해 구성되면 AdaBoost알고리즘[4]을 이용해 강한 분류기를 생성하게 된다.

2.5 단일 얼굴 후보 영역 선정

얼굴의 검출은 앞서 설명한 바와 같이 크기 투영을 이용한 판별 함수 및 주 얼굴 검출 알고리즘을 스캐닝을 통해 입력 받은 24×24 픽셀크기의 서브윈도우에 적용시켜 얼굴인지 아닌지를 분류하게 되는데 가능한 모든 서브윈도우에 대해 이러한 작업을 수행하게 되면 실시간 검출성능을 보장할 수 없게 된다. 그렇기 때문에 본 논문에서는 스캐닝 과정에서 얼굴에서 중복된 검출 결과를 보이는 현상을 막고 불필요한 스캐닝에 따르는 검출시간의 증가를 최대한 줄이기 위



(그림 6) 약한 분류기의 생성 방법

한 단일 얼굴 후보 영역 선정 방법을 제안한다. 단일 얼굴 후보 영역 선정 방법은 스캐닝 과정에서 얼굴이라고 분류되는 지점이 발생하면 적용된다. 얼굴이라고 최초로 판단되는 지점에서 오른쪽으로 7픽셀까지, 아래쪽으로 7픽셀까지 1픽셀크기씩 서브윈도우를 이동하면서 각 서브영상의 신뢰도를 조사한다. 조사하는 동안 얼굴이라고 최초로 판별한 지점보다 더 높은 신뢰도를 가진 지점이 존재 한다면 그 부분을 얼굴로 분류한다. 그렇지 않다면 얼굴이라고 최초로 판별된 부분이 얼굴이 될 것이다. 즉 서브윈도우들 중에서 최고의 신뢰도를 가진 서브윈도우를 찾음으로써 얼굴주변에서 얼굴이라고 분류되는 서브윈도우중 정확히 얼굴에 해당하는 한개의 서브윈도우만을 얼굴로 검출하는 것이다. 이후 얼굴이라고 검출된 서브윈도우 영역에 대해서는 스캐닝을 수행하지 않고 지나친다. 이후 기존의 스캐닝을 입력 영상의 우측 하단까지 계속하면서 또 다른 얼굴이 있는지 없는 지를 검색한다. 만약 얼굴이라고 판별되는 부분이 다시 검출된다면 이전의 과정을 다시 한 번 거쳐서 최대의 신뢰도를 가진 부분을 얼굴로 검출한다. 얼굴 검출 시 단일 얼굴 영역 선정 방법을 적용하면 테스트 영상 내에 실제로 얼굴이 있는 부분에서 과 검출이 일어나는 것을 방지하므로 얼굴검출시스템의

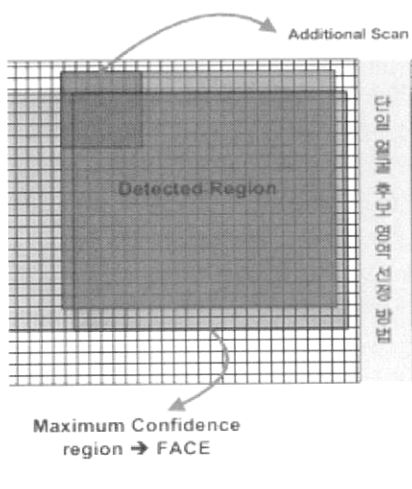
검출 정확도를 향상시키고 얼굴이라고 검출된 영역에 대해서는 스캐닝을 수행하지 않음으로써 모든 픽셀에 걸쳐서 스캐닝을 수행하지 않기 때문에 얼굴의 검출 속도를 향상시킨다. (그림 7)은 단일 후보 영역 선정 방법을 보여 준다.

3. Condensation기반의 얼굴 추적

본 장에서는 얼굴을 추적하는 방법을 설명한다. 본 논문에서 제안하는 얼굴 추적 방법은 Condensation알고리즘[9]에 기반하고 있다. 얼굴만을 효과적으로 추적하기 위해 앞서 얼굴 검출에 이용된 피부색 필터와 Haar 웨이블릿 특징이 이용되는데 얼굴 검출에 이용한 특징을 추적에 동시에 이용함으로써 불필요한 연산을 줄일 수 있으며 강인한 추적 성능을 보장할 수 있다. Condensation알고리즘에 대해 간단하게 설명한 후 본 논문에서 제안하는 추적 방법을 설명하고 실시간 어플리케이션에 적용하기 위한 몇 가지 요구사항과 이를 해결하기 위한 방법을 소개한다.

3.1 Condensation 알고리즘

Condensation(CONDitional dENSity propagaTION) 알고리즘은 1998년에 Isard와 Blake에 의해 제안된 방법으로 표본화 과정을 통해 추적 대상의 상태를 모델링하고 표본의 분포 및 확률을 전파하는 방법으로 시간적 요소를 이용하는 점에서 Kalman 필터[8]와 매우 유사한 방법이다. 그렇지만 Kalman 필터가 Gaussian 분포를 가정하는 것과는 달리 Condensation알고리즘은 임의의 확률 분포를 모델링 할 수 있기 때문에 복잡한 배경이나 다양한 상황에서의 물체 추적이 가능한 장점을 가지고 있다. X_t 가 시간 t 에서 추적하는 물체의 상태를 나타내고 Z_t 가 시간 t 에서 관측된 값이라고 하면 물체의 상태 X_t 가 진행되어 온 과정을 벡터 $X_t = X_1, \dots, X_t$ 로 나타낼 수 있으며 Z_t 가 진행되어 온 과정을 벡터 $Z_t = Z_1, \dots, Z_t$ 로 나타낼 수 있다. 물체의 동적 특성이 Markov chain을 형성한다고 가정하면 시간에 따른 상태 분포의 전파는 다음 식과 같이 나타낼 수 있다.



(그림 7) 단일 얼굴 후보영역 선정방법

$$p(x_t | Z_t) = k_t p(z_t | x_t) p(x_t | Z_{t-1})$$

$$\text{where } p(x_t | Z_{t-1}) = \int_{x_{t-1}} p(x_t | x_{t-1}) p(x_{t-1} | Z_{t-1}) \quad \text{식 (11)}$$

$p(x_t | Z_t)$ 는 일반적으로 Gaussian분포가 아니기 때문에 Condensation 알고리즘은 상태분포를 추정하기 위한 방법으로서 factored sampling을 이용한다. Factored sampling은 시간 t 에서의 관측 값을 기반으로 샘플들에 가중치를 할당함으로써 사후 확률 분포 $p(x_t | Z_t)$ 를 추정하는 방법으로 factored sampling에 의한 Condensation 알고리즘은 다음과 같다.

- 샘플링: 시간 $t-1$ 에서의 표본 집합 $\{s_i^{(n)}, n=1, \dots, N\}$ 으로부터 표본을 선택하는 과정을 N 번 반복한다. 이 과정에서 높은 가중치를 가지는 표본들이 주로 선택되게 된다.
- 예측: 동적 모델 $p(x_t | x_{t-1})$ 을 이용해 샘플을 예측한다.
- 관측: 관측된 값을 기반으로 예측된 샘플들에 대한 가중치를 할당하는 Factored sampling을 적용한다.

3.2 Haar 웨이블릿 특징과 피부색 밀도를 이용한 얼굴 추적
Condensation기반의 추적기는 각 프레임에서 동적 모델과 관측 모델에 의해 다음 프레임에서의 상태를 예측하고 업데이트하는 과정을 반복한다. 얼굴 추적에 Condensation기반의 추적기를 적용하기 위해 본 논문에서는 다음과 같이 얼굴에 대한 파라미터를 정의한다.

(x, y) : 추적되는 얼굴의 좌표

s : 추적되는 얼굴의 크기

이와 함께 각 t 단계에서의 상태를 $S_t = (x_t, y_t, s_t)$ 로서 정의한다. 앞서 언급한 바와 같이 Condensation기반의 추적기는 추적되는 대상에 대한 동적 모델과 관측 모델을 잘 정의하는 것이 요구된다. 본 논문에서는 얼굴의 움직임에 일정 패턴이 없다는 점과 얼굴 검출에 적용된 Haar 웨이블릿 특징이 이미 얼굴의 일반적인 특성을 잘 모델링하고 있다는 점을 이용해 이에 적합한 동적 모델과 관측 모델을 제시한다. 얼굴 추적을 위해 본 논문에서 적용한 동적 모델과 관측 모델은 다음과 같다.

■ 동적 모델

본 연구에서는 동적 모델로서 단지 Gaussian분포에 따른 잡음 요소만을 고려한다. 모델은 간단하게 다음과 같이 표현된다.

$$x_t = x_{t-1} + u_{t-1}$$

$$y_t = y_{t-1} + w_{t-1}$$

$$s_t = s_{t-1} (1.2)^k \quad \text{where } u_{t-1} \in N(\sigma_x), w_{t-1} \in N(\sigma_y), k \in \{-1, 0, 1\}$$

식 (12)

σ_x 와 σ_y 는 영상의 크기나 얼굴의 일반적인 움직임에 맞게 적절하게 설정되어야 한다. 예를 들어 너무 크게 설정되면 얼굴의 빠른 움직임을 더 잘 추적할 수 있지만 표본의 수 역시 많아져야 하기 때문에 실시간 추적을 보장하기 힘들 수가 있다. 반면 너무 작으면 적은 수의 표본만으로 추적이 가능하지만 얼굴의 빠른 움직임을 추적할 수 없게 된다. 본 논문에서는 많은 실험을 통해 가장 적절한 값으로 σ_x 와 σ_y 둘 다 20으로 설정하였다. 본 동적 모델은 단순하지만 특정한 움직임의 패턴이 없는 얼굴에 적합하며 이 후 소개될 관측 모델을 잘 설정함으로써 얼굴의 급작스런 방향 전환 등의 움직임을 잘 추적할 수 있다.

■ 관측 모델

동적 모델에 의해 예측된 각 표본들은 관측 모델 $p(Z_t | S_t)$ 에 의해 업데이트된다. 얼굴 추적을 위한 관측 모델은 앞서 정의된 얼굴에 대한 파라미터에 의해 $p(\text{face} | x, y, s)$ 로 표현된다. 각 표본에 대한 $p(\text{face} | x, y, s)$ 를 계산하기 위해 피부색 밀도 $\Omega(x, y, s)$ 가 추가로 정의된다.

$\Omega(x, y, s)$: t 단계에서 검출 파라미터 x, y, s 에 의해 정의된 영역에서 피부색의 밀도

AdaBoost기반의 얼굴 검출 방법과 피부색 밀도를 이용해 다음과 같이 관측 모델이 정의된다.

if $\Omega(x, y, s) > 0.5$ and $\Delta(x, y, s) > 0.4$,

$$p(\text{face} | x, y, s) = \alpha \Delta(x, y, s) + \beta \Omega(x, y, s)$$

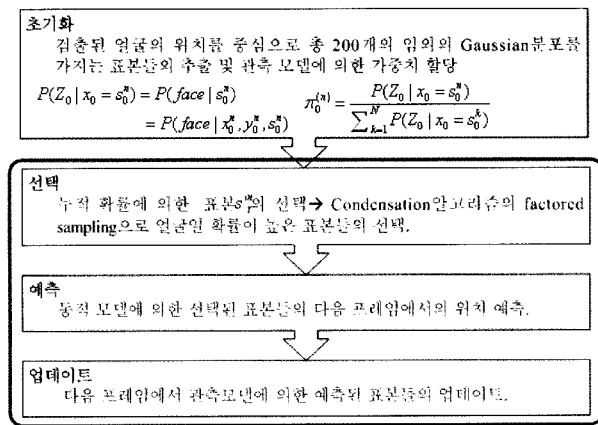
else

$$p(\text{face} | x, y, s) = 0$$

$$\text{where } \Delta(x, y, s) = \frac{\sum_{i=1}^{100} w_i h_i(x, y, s)}{\sum_{i=1}^{100} w_i} \quad \text{식 (13)}$$

얼굴 검출에 이용된 AdaBoost의 경우 200개 이상의 Haar 웨이블릿 특징들을 결합하는 것과 달리 추적에서는 실시간 적용을 위해 단지 100개의 Haar 웨이블릿 특징들만을 결합하였다. Condensation은 각각의 표본에 얼굴일 확률에 따라 가중치를 할당하고 이를 기반으로 표본들의 기대치로서 얼굴의 위치 및 크기를 추적하므로 각 표본들이 반드시 얼굴인지 아닌지를 판단할 필요는 없다. 게다가 약하게 설정된 검출 기준은 정면 얼굴뿐만 아니라 회전된 얼굴이나 일부 차단된 얼굴에 대해서도 얼굴을 추적할 수 있는 가능성을 가지기 때문에 표본화에 의해 추적이 이루어지는 Condensation 알고리즘에 적합하다고 할 수 있다. 실제로 200개의 약한 분류기를 모두 이용한 경우 얼굴의 자세 변화나 가려짐에 민감하게 반응하거나 속도가 느려짐을 알 수 있었다. Haar 웨이블릿 특징과 함께 피부색 밀도를 적용함으로써 얼굴을 움직임을 보다 효율적으로 추적할 수 있도록

하고 있다. 이러한 과정은 Haar 웨이블릿 특징만을 이용함으로써 추적에 실패할 경우를 최소화하기 위한 것으로 일정 피부색 밀도를 가진 표본들만이 Haar 웨이블릿 특징에 의해 평가될 수 있도록 하는 것이다. 식 (13)과 같이 피부색 밀도를 적용하면 단순한 동적 모델이 가지는 문제를 효과적으로 해결할 수 있다. 많은 실험을 통해 α, β 는 각각 0.9, 0.1로 설정 하였으며 이 두 개의 파라미터가 추적 성능에 많은 영향을 미치는 것을 알 수 있었다. 특히 β 를 너무 크게 설정할 경우 피부색의 의존도가 높아 표본의 크기가 점으로 수렴하는 결과를 초래하거나 피부색과 유사한 물체로 인해 추적에 실패할 확률이 높아진다. 이와 같이 정의된 동적 모델과 관측 모델을 이용한 Condensation기반의 추적은 크게 초기화, 선택, 예측, 업데이트의 네 단계로 이루어지며 (그림 8)은 이 과정을 설명하고 있다.



(그림 8) Condensation 기반의 얼굴 추적

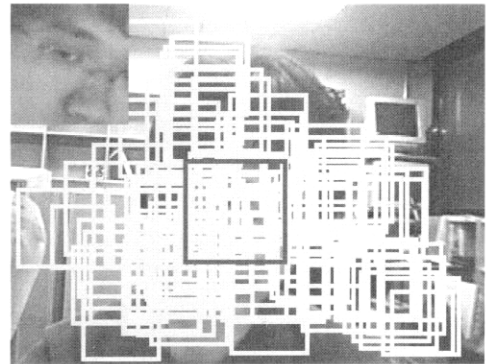
각 프레임에서 추적되는 얼굴의 위치와 크기는 다음과 같이 추정된다.

$$E[S] = \sum_{n=1}^N \pi_i^{(n)} S_i^{(n)} \quad \text{식 (14)}$$

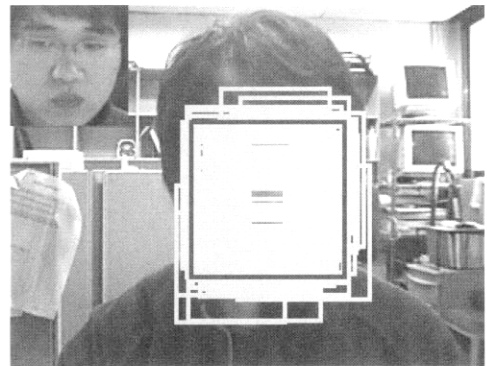
3.3 얼굴 인식이나 표정 인식 등의 실시간 어플리케이션에 적용되기 위한 요구 사항

본 논문에서 제안하는 얼굴 검출 및 추적 방법은 궁극적으로 얼굴 인식 및 표정 인식에 적용되는 것을 목적으로 한다. 사용자 친화적인 방향으로 얼굴 인식 및 표정 인식이 이루어지기 위해서는 무엇보다도 자연스럽게 부드럽게 화면에 나타난 얼굴을 검출하고 추적하는 것이 요구되며 오 검출에 의해 엉뚱한 객체를 추적하는 것을 막아야 한다. 본 절에서는 이러한 요구를 만족시키기 위한 방법을 설명한다.

· 대부분의 추적 방법이 추적할 객체를 설정하는 초기화가 반드시 요구된다. 특히 얼굴 추적의 경우 화면 내에 얼굴이 나타났는지를 자동으로 판단하고 얼굴 추적의 입력으로 전달하는 것이 요구되는데 복잡한 배경에서는 얼굴 검출이 잘 못됨으로 인해 엉뚱한 객체를 추적하게 될 수 있다.



(a) 추적 실패



(b) 추적 성공

(그림 9) 추적의 정확도에 따른 표본의 분산

본 논문에서는 이러한 문제를 자동적으로 해결하기 위해 각 프레임의 추적되는 영역에 대해 피부색 밀도를 계산하고 연속된 5 프레임의 추적되는 영역에 대해 피부색 밀도가 임계값 θ 보다 작으면 얼굴 검출을 다시 수행함으로써 잘못된 영역을 추적하는 것을 막는다. 그렇지만 피부색 밀도만으로는 잘못된 영역을 추적하는 것을 완전히 막기는 힘들다. 이와 함께 Condensation알고리즘의 각 표본들의 중심좌표의 분산을 측정할 수 있다. 얼굴이 아닌 객체에서는 (그림 9) (a)와 같이 표본들의 분산이 커지기 때문에 표본들의 중심좌표의 x축과 y축의 분산이 모두 임계값 τ 보다 크면 얼굴 검출을 다시 수행한다. 본 논문에서 설정한 θ 와 τ 는 각각 0.5와 100이다. 이 값은 많은 실험을 통해 설정된 것으로 본 논문의 실험에 그대로 적용되었다.

$$\sigma_s^2 = \sum_{n=1}^N \pi^{(n)} (S^{(n)})^2 - (E[S])^2 \quad \text{식 (15)}$$

· 얼굴 인식이나 표정 인식은 얼굴의 크기를 정확하게 추적하는 것이 요구된다. 본 논문에서는 동적 모델에 의해 표본의 크기가 1.2의 비율로 확대하거나 축소할 수 있도록 하는데 모든 표본들에 대해 각각 축소나 확대를 수행하게 되면 실시간의 요구사항을 만족시키기 힘들다. 그렇기 때문에 본 논문에서는 현재 표본들의 위치와 크기를 수렴할 수 있는 최대 영역을 구하고 이렇게 구해진 영역에 대해 1.2의 비율로 확대한 영상과 축소한 영상을 구한다. 예측된 표

본의 크기는 이전 크기와 동일하거나 1.2의 비율로 확대되거나 축소된 경우만이 존재하기 때문에 예측된 크기에 맞게 미리 계산된 영상으로부터 표본을 얻을 수 있다. 이러한 방법은 표본의 개수가 늘어날수록 그 효과가 매우 크다.

4. 실험 및 고찰

제안된 방법은 Visual C++ 6.0을 이용해 구현되었으며 펜티엄 4 2.8GHz의 CPU를 가진 워크스테이션에서 320×240의 해상도를 가진 웹 캠을 이용해 테스트되었다. 크기 투영과 AdaBoost에 의한 얼굴 검출기의 훈련을 위해 모두 5000장의 얼굴 훈련 영상과 15000장의 비 얼굴 훈련 영상을 수집하였다. 훈련 영상들은 피부색이나 얼굴 형태, 혹은 얼굴 요소들의 다양성을 잘 표현함으로써 학습 알고리즘이 얼굴의 공통적이고 고유한 특성을 발굴해낼 수 있어야 한다. 이를 만족시키기 위해 얼굴 데이터베이스로서 최근 얼굴 처리 연구에서 많이 이용되고 있는 CMU_PIE 데이터베이스[13], BioID 데이터베이스 그리고 Slovenian CVL 데이터베이스를 이용해 500장의 정면 얼굴 영상을 추출하고 미러링된 영상을 포함해, 총 1000장의 정면 얼굴 영상을 취득했으며 이와 함께 아시아인의 얼굴 특성을 효과적으로 반영하기 위해 한국인을 포함하고 다양한 조명 환경에서 획득된 영상들을 인터넷으로 취득하고 이에 대해 얼굴 영역을 추출하여 1500장의 영상을 취득한 후 Gaussian 잡음 추가와 미러링을 통해 모두 4000장의 얼굴 영상을 취득했다. 이와 함께 비 얼굴 훈련영상집합은 웹 사이트로부터 획득한 얼굴이 전혀 포함되지 않고 사이즈가 상당히 큰 자연 이미지들로부터 추출되었다. 30장의 사이즈가 큰 이미지로부터 10000장의 비 얼굴 영상을 획득하고 Bootstrap과정을 통해 얼굴과 유사한 특성을 가지는 5000장의 영상을 추가로 획득하였다. 얼굴 영상 및 비 얼굴 훈련영상집합을 위해 수집된 영상들은 얼굴의 크기가 모두 다르기 때문에 학습 조건에 맞게 얼굴 훈련영상집합과 비 얼굴 훈련영상집합의 크기는 모두 24×24 픽셀크기로 제한하고 칼라영상일 경우 흑백영상으로 변환시켰으며 파일은 모두 raw파일로 저장되었다. (그림 10)은 얼굴 훈련영상집합의 일부를 보여준다. 구성된 훈련영상집합에 AdaBoost 알고리즘을 적용해 훈련하여 모두 200개의 Haar 웨이블릿 특징을 선택하였다. 크기 투영과 AdaBoost알고리즘 모두 전처리 과정으로써 각 훈련 영상에 히스토그램 평활화를 적용하였다.

훈련된 검출기의 성능을 평가하기 위해 인터넷으로부터 다양한 크기와 자세를 가진 261개의 얼굴을 가진 97장의 영상을 획득하고 테스트한 결과 92%의 검출 정확도를 얻을 수 있었다. 얼굴 검출 방법은 대부분 회전된 얼굴에 대해 검출에 실패하는 것을 볼 수 있었는데 이는 제안된 얼굴 검출 방법이 주로 정면 얼굴을 훈련영상으로 이용했기 때문에 회전된 얼굴에 대해서는 정확한 검출을 보장하기 힘들기 때문이다. 그러나 이러한 문제는 제안된 얼굴 추적 방법을 이용해 해결될 수 있다. 제안된 얼굴 추적 방법의 성능을 보

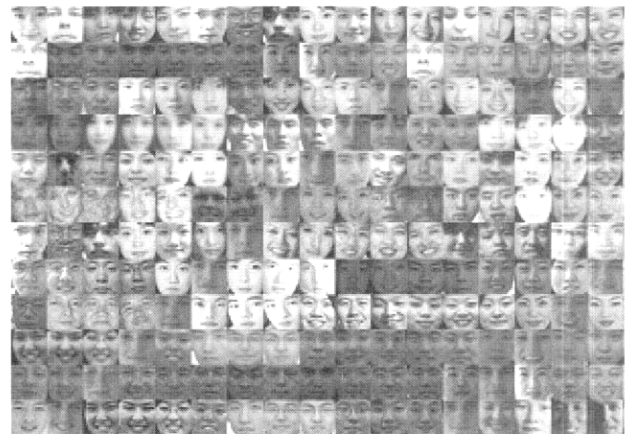
이기 위해 다양한 실험을 수행되었다. 실험은 얼굴의 크기나 자세 변화 및 조명 변화에 대해 이루어졌다. 추적되는 얼굴을 효과적으로 보여주기 위해 추적되는 얼굴 영역의 20% 확장된 영역을 사각형으로 나타내었다.

먼저 제안된 방법의 실시간 성능을 테스트 한 결과 제안된 방법은 초당 약 17~20 프레임의 처리속도를 보여주었다. 제안된 방법은 얼굴의 크기가 커짐에 따라 영상의 확대와 축소에 따른 연산 량이 증가하기 때문에 프레임마다 약간의 처리속도의 차이가 나게 된다.

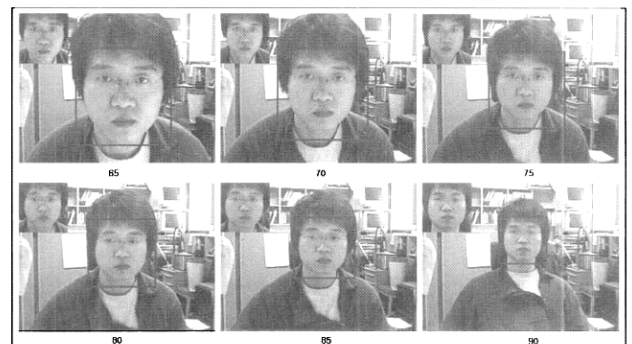
(그림 11)은 얼굴의 크기 변화를 추적하는 과정을 보여준다. 제안된 방법이 빠르게 작아지는 얼굴을 잘 추적하는 것을 볼 수 있다.

특히, 크기를 추적하는 과정이 매우 부드럽게 이루어지는데 이러한 이유는 추적하는 매 프레임마다 모든 표본들에 대한 기대치를 구함으로써 얼굴 영역을 결정하기 때문이다. (그림 11)과 같이 얼굴이 작아지고 있다면 동적 모델에 의해 1.2의 비율로 축소된 표본들이 관측모델에 의해 더 많은 가중치를 할당 받게 될 것이고 표본들의 기대치는 축소된 얼굴 크기로 나타날 것이다.

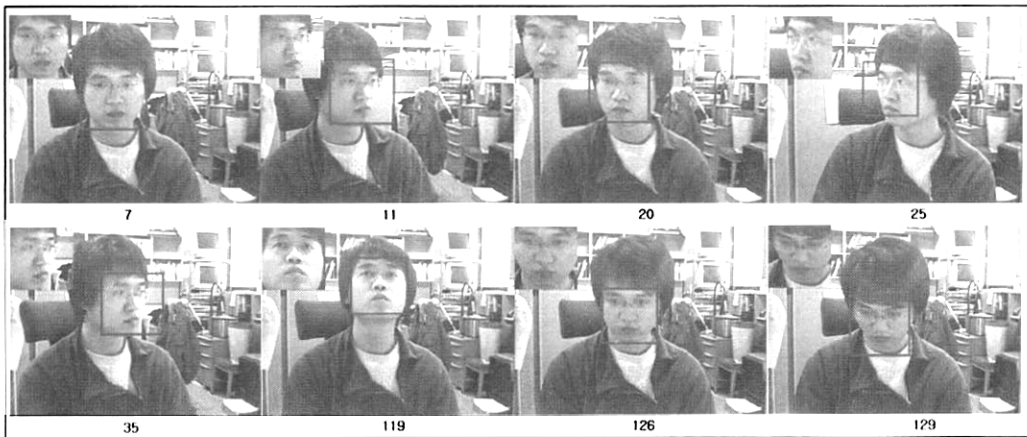
(그림 12)는 얼굴의 자세 변화에 따른 추적 성능을 보여준다. 위 아래 및 좌우로 회전되는 얼굴에 대해 눈, 코, 입과 같은 주요 특징들이 식별 가능하다면 잘 추적하는 것을 볼 수 있다. 이 실험 결과는 앞서 얼굴 검출을 정지영상에 적



(그림 10) 얼굴 훈련영상집합



(그림 11) 크기가 급격하게 변하는 얼굴에 대한 추적 결과



(그림 12) 회전하는 얼굴에 대한 추적 결과



(그림 13) 회전하는 얼굴에 대해 각 프레임마다 검출기만을 적용한 추적 결과

용했을 때 회전된 얼굴을 검출하지 못하는 것과 비교되는 결과이다. 추적 과정에 이용되는 특징은 검출에 이용되는 Haar 웨이블릿 특징과 동일하며 단지 그 개수를 적게 이용할 뿐이다. 그러나 정확하게 얼굴인지를 판단해야 하는 얼굴 검출과는 달리 Condensation기반의 추적방법은 확률 분포를 전파하는 표본들에 얼굴일 가능성을 할당하고 가능성에 따라 표본들에 대한 기대치를 구해 얼굴 영역을 결정하므로 해당 표본들이 얼마나 얼굴에 가까운지를 결정하기만 하면 되는 것이다. (그림 13)은 회전하는 얼굴에 각 프레임마다 얼굴 검출만을 이용했을 때의 결과를 보여준다. 회전된 얼굴에 대해서 검출에 실패하는 것을 볼 수 있다.

피부색을 가지는 물체에 의해 얼굴이 가려지는 경우에 대한 추적 성능 역시 테스트되었다. 단순히 피부색 정보를 이용하는 경우에 손에 의해 얼굴이 가려지면 손을 추적하게 되는 등 추적 성능에 심각한 영향을 초래 할 수 있으나 제안된 방법은 피부색과 함께 Haar 웨이블릿 특징을 이용하기 때문에 이러한 문제를 해결할 수 있음은 물론 일반적인 물체에 의한 가려짐에도 비교적 강인함을 알 수 있다. (그림 14)는 다양한 가려짐이 일어나는 상황에서 얼굴을 추적하는 과정을 보여준다. 특히, 그림에서 볼 수 있는 것처럼 제안된 방법은 얼굴의 일부에 가려짐이 일어나고 동시에 얼굴이 움직일지라도 얼굴을 잘 추적한다.

일반적으로 추적방법이 보다 일반적인 환경에 적용되기 위해서는 비교적 빠른 얼굴의 움직임 역시 잘 추적해야 한다. (그림 15)는 제안된 방법이 비교적 빠른 얼굴의 움직임에 대해 얼굴을 추적하는 것을 보여준다. 빠른 움직임으로 인해 얼굴이 정확하게 나타나지 않는 상황에서도 정확하게

추적하는 것을 볼 수 있다. 그러나 제안된 방법이 빠른 얼굴의 움직임을 항상 잘 추적하지는 못한다. 빠른 얼굴의 움직임이나 복잡한 배경으로 인해 얼굴의 추적에 실패하는 경우가 발생하는데 이러한 경우에는 전 장에서 설명한 바와 같이 얼굴을 다시 검출하여 잘못된 물체를 추적하는 것을 막는다.



(그림 14) 손에 의해 가려지는 얼굴에 대한 추적 결과



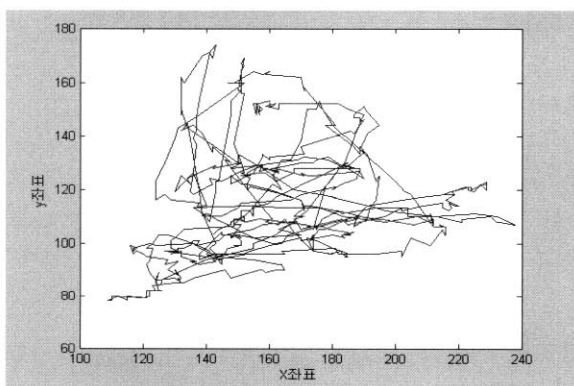
(그림 15) 빠르게 움직이는 얼굴에 대한 추적 결과



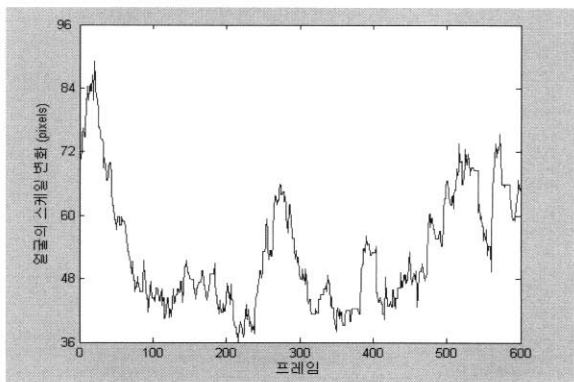
(그림 16) 보다 자연스러운 상황에서의 추적 결과

지금까지 소개한 얼굴의 회전이나 크기변화 및 얼굴의 크기가 보다 작은 경우를 포함하는 자연스러운 상황에서의 영상이 추가로 획득되었다. (그림 16)은 영상의 일부를 보여주며 (그림 17)과 (그림 18)은 추적을 통해 획득된 프레임에 따른 얼굴의 이동 경로 및 크기 변화를 보여준다.

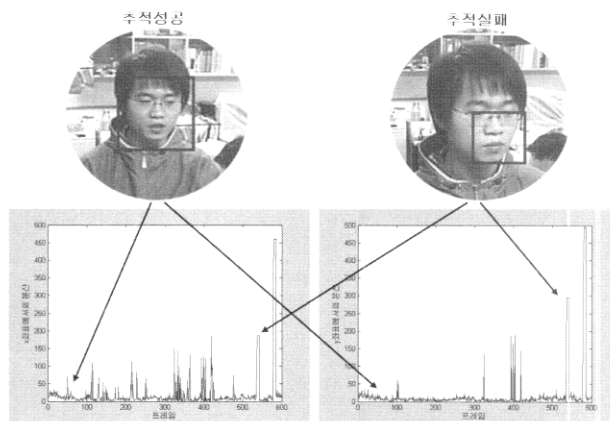
3장에서 설명한 바와 같이 본 논문에서는 표본의 분산 및 피부색 밀도를 이용해 추적이 실패하는 상황을 자동적으로



(그림 17) 추적을 통해 획득된 얼굴의 이동 경로



(그림 18) 추적을 통해 획득된 얼굴의 크기 변화



(그림 19) 표본의 분산에 따른 추적의 성공과 실패

판단하고 재 검출을 수행해 계속적으로 추적이 정확하게 이루어 질 수 있도록 한다. 이를 보다 자세히 보이기 위해 표본의 x좌표에 대한 분산과 y좌표에 대한 분산이 각각 획득되었다. (그림 19)는 프레임에 따른 분산의 변화 및 추적이 성공적으로 이루어지는 상황과 실패하는 상황에 대한 영상을 각각 보여준다.

5. 결 론

본 논문에서는 Haar 웨이블릿 특징과 피부색에 기반을 둔 얼굴 검출 및 추적 방법을 제안하였다. 간단한 피부색 필터와 콤포넌트 레이블링에 의해 얼굴이 존재하는 후보영역을 찾고 각 후보영역에 대해 크기 투영에 의한 분류기와 AdaBoost 알고리즘에 의해 훈련된 Haar 웨이블릿 특징에 의한 분류기를 순차적으로 적용해 얼굴을 검출하였다. 검출된 얼굴을 추적하기 위해서 Condensation 알고리즘을 적용하였는데 이 과정에서 관측모델로서 Haar 웨이블릿 특징과 피부색정보를 이용해 보다 강인한 추적이 가능하게 함은 물론 검출에 이용되는 특징을 추적에 동시에 이용함으로써 효율을 높일 수 있었다. 다양한 상황에 대해 실험을 한 결과 제안된 방법이 얼굴의 자세 변화나 가려짐 등에서도 얼굴을 잘 추적하는 것을 알 수 있었다. 향후 제안된 방법에 대해 개선해 나가고자 하는 문제는 다음과 같다.

- 조명에 보다 더 강인한 성능을 위해 피부색 모델에 의존도를 줄이면서 얼굴을 빠르고 정확하게 추적할 수 있도록 검출 방법을 개선
- 복수의 얼굴에 대한 추적 성능 추가
- 제안된 방법과 얼굴 인식 및 표정 인식과의 융합

참 고 문 헌

[1] C.Liu, "A Bayesian Discriminating Features Method for Face Detection," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.25, No.6, pp.725-740, June, 2003.

[2] H.A.Rowley, S.Baluja and T.Kanade, "Neural Network-Based Face Detection," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.20, No.1, pp.23-38, January, 1998.

[3] E.Osuna, R.Freund and F.Girosi, "Training Support Vector Machines: an Application to Face Detection," Proc. CVPR'97, Puerto Rico, June, 1997.

[4] Yoav Freund and Robert E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," In computational learning Theory : Eurocolt '95, pp.23-37. Springer-Verlag, 1995.

[5] Paul Viola and Michael Jones, "Rapid object detection using a boosted cascade of simple features," Conference on Computer Vision and Pattern Recognition, 2001.

[6] G. Hager and K. Toyama, "X Vision : A Portable Substrate for Real-Time Vision Applications," Computer Vision and Image Understanding, Vol.69, No.1, pp.23-37, 1998.

[7] S. Birchfield, "Elliptical Head Tracking Using Intensity Gradients and Color Histograms," Proc. Conf. Computer Vision and Pattern Recognition, pp.232-237, 1998.

[8] Applied Optimal Estimation, A. Gelb, ed. MIT Press, 1992.

[9] M. Isard and A. Blake, "CONDENSATION-conditional density propagation for visual tracking," International Journal of Computer Vision, Vol.29, pp.5-28, 1998.

[10] C Garcia and G Tziritas, "Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Analysis," IEEE Transactions on Multimedia, Vol.1, pp.264-277, 1999.

[11] Richard O. Duda, Peter E. Hart, David G.Stork, "Pattern Classification," Wiley-Interscience, 2000.

[12] Michael Jones and Paul Viola, "Fast Multi-view Face Detection," Mitsubishi Electric Research Lab TR-20003-96, 2003.

[13] T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression Database," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.25, No.12, pp.1615-1618, 2003.

[14] R. C Verma, C Schmid, K Mikolajczyk, "Face Detection and Tracking in a Video by Propagating Detection Probabilities," IEEE Trans on Pattern Analysis and Machine Intelligence, pp.1215-1226, 2003.

[15] MH Yang, DJ Kriegman, N Ahuja, "Detecting Faces in Images : A Survey," IEEE Trans on Pattern Analysis and Machine Intelligence, pp.34-58, 2002.

박 영 경



e-mail : multipym@skku.edu

2002년 성균관대학교 전자공학과(학사)
2004년 성균관대학교 대학원 전기전자 및
컴퓨터공학부(공학석사)
2004년~현재 성균관대학교 대학원
정보통신공학부 박사과정

관심분야: 컴퓨터 비전, 패턴 인식, 영상신호처리, 데이터 방송

서 해 종



e-mail : rokaf539@skku.edu

2005년 성균관대학교 전자공학과(학사)
2005년~현재 성균관대학교 대학원 정보통신
공학부 석사과정

관심분야: 컴퓨터 비전, 패턴 인식, 영상
신호처리

민 경 원



e-mail : minkw@keti.re.kr

1994년 홍익대학교 전자공학과(학사)
1996년 홍익대학교 전자공학과(공학석사)
1996년~현재 전자부품연구원 책임연구원
관심분야: 컴퓨터 비전, 영상 처리

김 중 규



e-mail : jkkim@skku.edu

1980년 서울대학교 전자공학과(공학사)
1982년 서울대학교 전자공학과(공학석사)
1989년 The University of Michigan,
Electrical Engineering(Ph.D).
1980년~1981년 한국통신기술연구소
(현 ETRI) 위촉연구원

1989년~1990년 The University of Michigan, Post Doctoral
1990년~1992년 삼성전자 특수연구소 TASR 개발팀장
1992년~현재 성균관대학교 정보통신공학부 교수
관심분야: 컴퓨터 비전, 패턴 인식, 의용공학, 적응신호처리, 데
이터 방송