

# 복합형 필터와 CNN 모델을 이용한 효과적인 얼굴 검출 기법

조일국, 박현정, 김호준  
한동대학교 전산전자공학부

cik0225@seed.handong.edu, hjpark79@empal.com, hjkim@handong.edu

## Robust Face Detection Using Hybrid Filters and Convolutional Neural Networks

Il Gook Cho, Hyun Jung Park, Ho Joon Kim

School of Computer Science and Electronic Engineering, Handong Global University

### 요 약

본 논문에서는 수정된 CNN(Convolutional Neural Network) 모델과 다중 필터가 상호 결합된 형태의 얼굴 패턴 검출 기법을 소개 한다. 이는 로봇 시각의 응용문제에서 실내영상의 실시간 인식문제를 대상으로 한다. 검출 과정의 효율성 향상을 위하여 도입된 다중 필터는 후보 영역의 개수와 범위를 줄일 수 있게 한다. 제안된 모델에서 CNN 신경망은 가보변환(Gabor Transform)계층을 두어 검출 과정의 첫 단계에서 영상 내의 기본 특징 지도를 생성 하도록 하였다. 보다 강인한 검출기능을 위하여 조명보정 기법이 시스템의 전처리 단계로 구현 된다. 실제 영상을 통한 실험 결과로부터 제안된 이론의 타당성을 고찰 한다.

### 1. 서론

얼굴 검출 문제는 단일 영상 또는 일련의 연속영상으로부터 얼굴 패턴의 영역과 위치를 찾아내는 작업으로 매우 다양한 응용 분야를 갖는다. 예컨대 보안 시스템 분야나 모델기반 비디오 코딩, 인간과 컴퓨터의 대화형 시스템, 로봇 시각 등의 넓은 분야에서 연구되고 있다.

얼굴 검출에 장애가 되는 다양한 요소들을 극복하고 정확한 얼굴을 검출하기 위한 방법들은 많은 연구들을 통해 발표된 바 있다[1-6]. 최근의 연구된 바를 보면 다음과 같은 특정 기술을 이용함을 알 수 있다 [3]. 신경망 모델, 학습 이론 연구, 정보 이론적 분석, 기하학적 모델링, 움직임 추출과 피부색 분석 등의 방법이 제안된 바 있다. 이러한 방법들은 조명 변화, 얼굴의 크기 등의 얼굴 검출에 장애가 되는 문제들에 강인한 방법을 가지기 위해 복잡한 알고리즘들을 적용 했으나, 계산의 복잡도가 증가함으로 인하여 얼굴 추출 시간이 매우 긴 경우가 대부분이다.

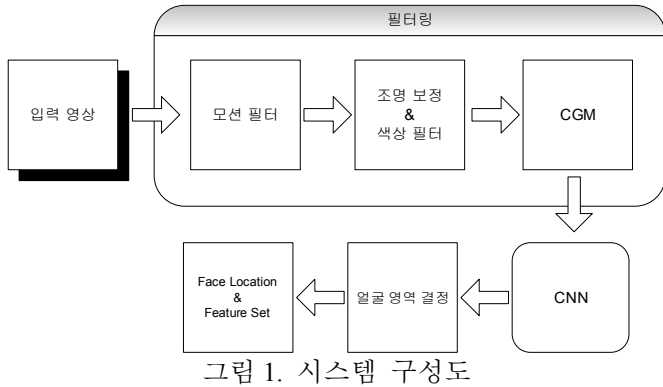
본 연구에서는 실내영상의 실시간 인식문제를 대상으로 얼굴 패턴 검출을 위한 신경망 기법을 소개한다. 이는 CNN 모델과 다중 필터가 상호 결합된 형태를 가지며 전처리 단계로서 조명 보정기법을 고려한다.

검출 과정에서 후보 영역의 개수와 범위를 줄이기 위하여 움직임, 색상, 명암의 특징 요소를 고려한 3 종류의 필터가 도입된다. 본 연구에서는 기존의 CNN 신경망의 C1 계층을 가보변환(Gabor Transform)계층으로 대체 하여 검출 과정의 첫 단계에서 영상 내의 기본 특징 지도를 생성 하도록 하였다. 실제 영상을 통한 실험 결과로부터 제안된 이론의 타당성을 고찰 한다. 본 논문의 2 장에서는 대상 시스템의 구조를 간략히 소개하고, 3 장에서는 다중 필터링 기법에 대하여, 4 장에서는 얼굴검출 모듈의 구조 및 동작 특성을 소개한다. 이어서 5 장에서는 실제 영상을 통한 실험 결과로부터 제안된 이론의 타당성을 고찰 한다.

### 2. 시스템 개요

본 연구는 그림 1 과 같은 구조의 얼굴 검출 시스템을 대상으로 한다. 그림에 보인 바와 같이 시스템은 크게 두 개의 모듈로 이루어 진다. 전 단계인 필터링 과정에서는 입력 영상의 움직임 요소, 색상 정보를 이용하여 후보 영역을 선별한다. 이는 이어서 명암 데이터를 사용하여 학습된 신경망 필터를 거침으로써 이후 검출 과정에 적용될 후보 영역의 수를 최소화 하게 된다. 이어서 두 번째 단계인 검출 모듈에서는

CNN 을 사용하여 후보 영역에 대하여 얼굴 패턴 여부를 판정하게 된다.



### 3. 필터링

주어진 입력 영상에서 사전 전제 없이 모든 영역을 대상으로 하는 영역 추출 과정은 많은 연산 시간이 소요 되고, 조명 변화에 민감할 뿐만 아니라, 복잡한 배경에서의 영역 추출의 정확도를 저하 시킨다. 이에 본 연구에서는 얼굴 검출 모듈의 전 단계로서 다양한 특징 요소를 고려한 필터링 작업을 통해 얼굴 검출을 위한 후보 영역을 결정하여 상기의 문제점들을 개선하고자 한다.

#### 3.1 모션 필터

본 연구는 로봇 비전의 응용으로써 실내 영상의 얼굴 패턴 검출 문제를 대상으로 한다. 따라서 연속 영상을 입력 데이터로 고려 함으로서 움직임 정보를 이용할 수 있다. 입력되는 실내 영상에서 배경은 고정되어있다고 가정할 수 있다. 따라서 영상 안에서 움직임에 관한 정보는 얼굴 패턴 검출의 매우 유용한 요소가 된다. 기존의 연구에서 모션 필터에 관하여 매우 다양한 방법론이 제안된 바 있다. 그러나 본 연구에서는 시스템의 실시간 응답 특성을 고려하기 위하여 단순화 된 필터를 사용하여 실행 시간의 최적화를 시도하였다.

실제 사용된 방법은 연속 영상의 각 픽셀 별로 색상의 변화를 산출하고 이에 임계치를 적용함으로써 움직임 정보를 생성한다.

#### 3.2 조명 보정

입력되는 영상에서 피부색은 조명 변화에 매우 민감하다. 이에 본 연구에서는 시스템의 전처리 단계로서 조명 보정 과정을 수행 한다. 이를 위하여 "Reference White"[2] 기법을 채택하였다. 이는 입력되는 전체 영상에서 luma(비 선형적으로 감마 보정된 밝기값)값의 범위를 구한 후, luma 값의 상위 일정 비율의 픽셀을 특정 그룹으로 정의 한다. 이에 속한 픽셀수가 임계치 보다 많으면 전체 픽셀에 대해 R, G, B 값을 조정하게 되는데, 이 때 Y 값의 범위를 Reference White pixel 의 평균 값으로 나눈 값이 되도록 재조정 한다.

### 3.3 색상 필터

본 연구에서는 실제 적용 환경에서 획득한 다량의 데이터로부터 피부색의 특성을 분석, 학습하도록 하였다. 그림 2 는 실제 영상을 대상으로 분석한 피부색에 대한 특징 분석 결과의 예이다. 그림. 3 은 이를 각 특징의 종류별로 분석한 결과를 보인다. 이들 분석 결과로부터 적용할 특징 요소의 종류를 적용적으로 선별해 내게 되었다. 이러한 분석 결과로부터 색상 필터를 구현 하게 된다. 고려한 특징의 종류는 R, G, B, 명암, Cb, Cr 등 9 가지를 대상으로 분석하고 이들 중 선별적으로 적용하게 된다.

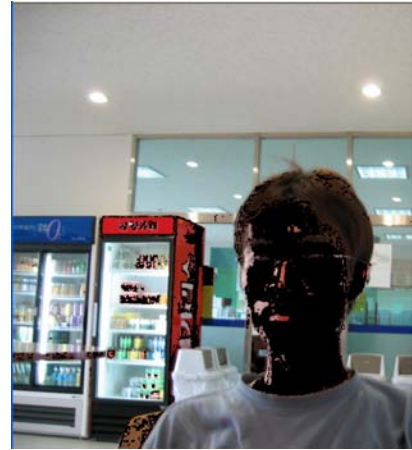


그림 2. 피부색 검출 및 분석 실험의 예

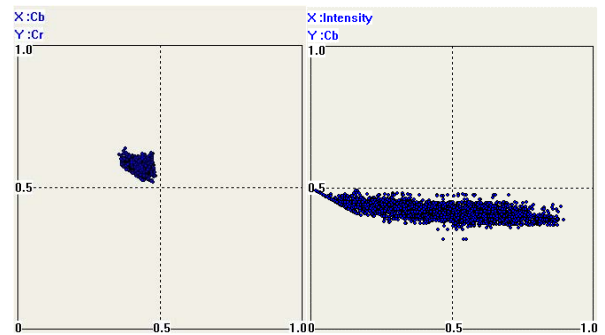


그림 3. 피부색에 대한 특징 분석 결과의 예

#### 3.4 명암요소를 고려한 신경망 필터

본 연구에서는 명암 요소를 고려한 필터를 두 가지 형태의 신경망으로 고려한다. 먼저 다층 퍼셉트론을 가진 전 방향 신경망으로 필터를 구성한다. 그 다음 Feraud 등이 제안한 Constrained Generative Model(CGM)을 이용하여 명암 필터를 구성한다[4]. CGM 은 3 개의 계층을 갖는다. 입력 계층과 출력 계층은 같은 숫자의 노드를 갖고, 각각의 노드는 입력되는 영상의 각각의 픽셀에 대응된다. CGM 신경망은 입력 및 출력 계층에 일련의 얼굴 패턴을 학습시켜 자기 연상 기능(Auto-Association)을 지원 하게 된다. 이는 다음과 같은 목적 함수를 최적화 시키는 과정으로서 이루어 진다.

$$Cw = \sum_i (P_w(x_i) - P_{MN}(x))^2, \quad (1)$$

if  $x \in v$ , then  $P_{NV}(x) = x$ ,

if  $x \notin v$ :  $P_{NV}(x) = P_{knn}(x)$ .

식에서  $W$ 는 신경망의 가중치 벡터 이며,  $P_{NV}$ 은 입력에 대한 투영된 값  $P$ 의 근사치를 의미한다. 따라서 이 식은 입력이 얼굴 패턴 집합에 속하는 경우 원 영상 패턴을 연상하게 되며, 그렇지 않은 경우  $P_{knn}$ 으로 결정됨을 보인다.

#### 4. 얼굴검출 신경망

입력된 영상에서 필터링을 거쳐 결정된 후보 영역을 CNN을 통과 시켜 얼굴 후보 영역중 얼굴 영역을 추출한다. CNN은 다양한 형태로 필기체 인식, 인쇄체 인식, 얼굴 인식 등의 여러 해결하기 어려운 문제들에 성공적으로 적용된 바 있다[1]. CNN은 생물학적 형태를 반영한 계층적 다중 신경망으로써 목표물의 위치 이동, 크기 변화 및 왜곡에 강한 인식 기능을 제공한다.

본 연구에서 제안하는 얼굴검출 신경망은 다음과 같은 구조를 가진다.

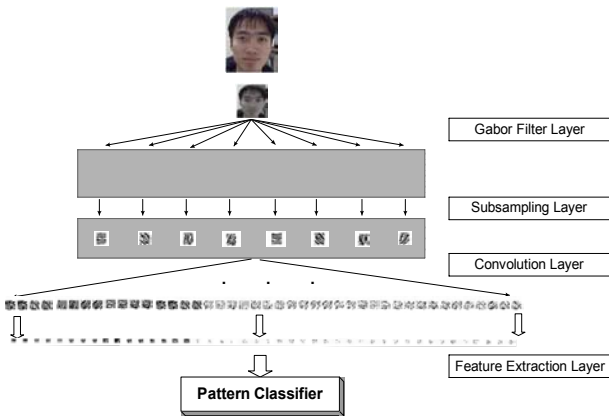


그림 4. 얼굴 검출 신경망의 구조

첫 번째 계층인 가보 필터 계층에서는 입력 영상으로부터 가보 변환을 적용하여 8방향의 기본 특징을 추출한다. 가보 필터의 과정은 다음과 같이 요약될 수 있다. 가보 함수는 식 (2)와 같은 구조를 갖는다.

$$g(x, y) = ae^{-j(\omega_x + \omega_y)} e^{-(x/\lambda_x)^2 - (y/\lambda_y)^2} \quad (2)$$

식에서  $a$ 는 크기를 제어하는 요소로서 이에 대한 변환 함수는 식 (3)과 같이 결정된다.

$$G(\omega_x, \omega_y) = ae^{-j(\lambda_x(\omega_x - \omega_{0x}))^2 - j(\lambda_y(\omega_y - \omega_{0y}))^2} \quad (3)$$

$\lambda_y / \lambda_x$ 의 비율은 가보 필터의 방향을 정의한다. 이로부터 제안된 모델에서는 8방향의 특징 지도를 생성하도록 하였다. 38 X 42의 최초 입력 영상은 가보 필터를 거쳐 28 X 32의 특징 지도로 생성되며, 두 번째 계층에서 2 X 2영역으로 표본 추출하여 14 X 16크기로 조정된다. 8개의 특징 영역 중 2개의 서로 다른 조합을 선택하여 총 44개의 특징 지도를 생성하고 이로부터 최종 패턴 분류를 위한 신경망의 입력이 결정된다.

제안된 시스템에서 패턴 분류기(Pattern classifier)는 수정된 구조의 퍼지 최대최소신경망을 사용하여 구현하였다. 이는 기존의 FMM 신경망[6]에서 가중치 요소를 결합한 형태의 신경망으로 WFMMNN(Weighted Fuzzy Min-Max Neural Network)으로 명명하였다. 신경망의 활성화 특성은 다음과 같이 정의 되는 하이퍼박스의 멤버쉽 함수로써 결정된다.

$$B_j = \{X, U_j, V_j, C_j, F_j, f(X, U_j, V_j, C_j, F_j)\} \quad \forall X \in I^n \quad (4)$$

식에서  $U_j$  및  $V_j$ 는 각각 하이퍼박스  $j$ 의 특징범위에 대한 최대값 및 최소값 벡터를 의미하며  $C_j$  및  $F_j$ 는 특징 집합의 평균값 및 특징의 유형을 의미한다. 이와 같이 정의된 하이퍼 박스는 다음과 같은 식으로 그 활성화 특성이 정의 된다.

$$b_j(A_n) = \frac{1}{\sum_{i=1}^n w_{ji}} \cdot \sum_{i=1}^n w_{ji} [\max(0, 1 - \max(0, \gamma_{jiU} \min(1, a_{ni} - v_{ji}))) + \max(0, 1 - \max(0, \gamma_{jiV} \min(1, u_{ji} - a_{ni}))) - 1.0] \quad (5)$$

$$\begin{cases} \gamma_{jiU} = \frac{\gamma}{R_U} & R_U = \max(s, u_{ji}^{new} - u_{ji}^{old}) \\ \gamma_{jiV} = \frac{\gamma}{R_V} & R_V = \max(s, v_{ji}^{old} - v_{ji}^{new}) \end{cases} \quad (6)$$

식에서  $w_{ji}$ 는 특징  $i$ 와 하이퍼박스  $j$ 과의 가중치 요소를 의미한다. 또한 매개변수  $\gamma_{jiU}$  and  $\gamma_{jiV}$ 는 퍼지 멤버쉽함수에서 특징범위의 가장자리에서 기울기를 제어하는 상수이다. 식에서 보인 바와 같이 기존의 FMM 모델과는 달리 특징과 하이퍼박스 간의 가중치 요소를 포함하게 된다. 이는 특징값의 빈도와 범위를 고려하여 학습과정에서 조정되게 되는데 기존의 FMM 신경망의 학습에서 하이퍼박스 중첩의 해결을 위한 축소과정의 부작용을 개선한다.

#### 5. 실험 결과 및 고찰

본 연구는 실내 영상에서 실시간 얼굴 검출 시스템을 대상으로 한다. 그림 5는 조명 보정 기법과 모션 필터, 색상 필터 및 신경망 필터를 적용한 결과를 보인다.

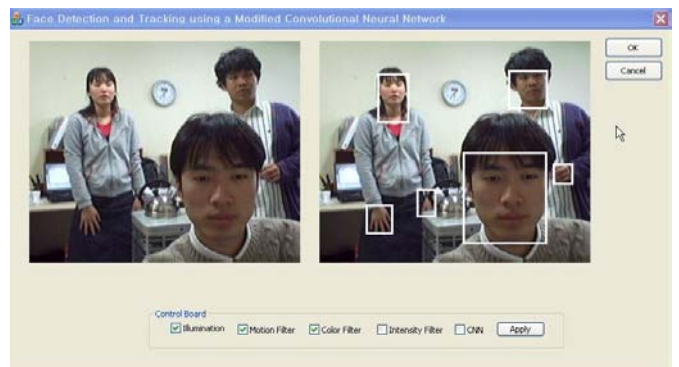


그림 5. 전처리 과정이 끝난 후 얼굴 후보 영역 추출 결과

제안된 시스템에서 전처리의 결과는 얼굴 검출 모듈의 입력이 될 대상 영역을 결정한다. 그림에 보인 바와 같이 필터링 등의 전처리 과정을 통하여 대상 영역의 개수와 규모를 획기적으로 줄일 수 있게 된다. 이는 얼굴 검출 과정에서 소요되는 계산 시간을 단축할 수 있게 함으로서 시스템의 실시간 응답특성을 개선한다.

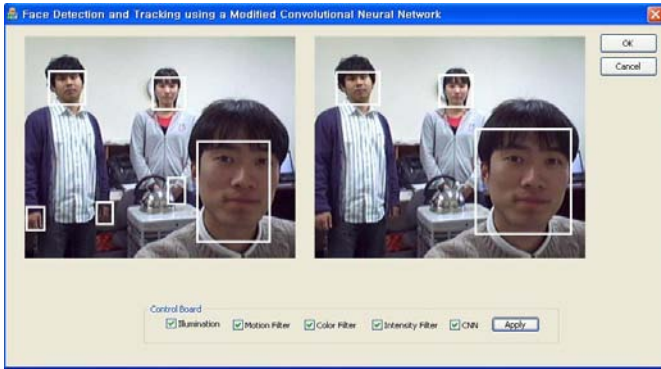


그림 6. 시스템의 얼굴 영역 검출 결과

인식율은 점진적으로 감소하였다.

## 6. 결론

본 연구에서는 얼굴검출을 위한 방법론으로서 다단계 필터와 복합형 신경망을 이용한 모델을 제시하였다. 이는 실내 영상을 대상으로 한 실시간 인식 및 추적시스템을 목표로 한다. 전처리 단계에서 조명변화에 대한 보정기능은 시스템이 보다 강인하고 실용적인 성능을 보일 수 있게 한다. 얼굴 검출 모듈에서 사용한 신경망 모델은 CNN 모델의 변형과 수정된 구조의 FMM 모델을 상호 결합한 형태이다. 보다 미세한 특징의 추출을 위하여 CNN 모델의 첫 번째 계층에서 가보필터를 적용하여 기본 특징지도를 생성하도록 하였으며, 이는 3 개의 계층을 거쳐 분류 신경망의 입력 특징을 생성하게 된다. 패턴 분류 신경망으로 사용된 신경망은 가중치를 갖는 구조의 FMM 신경망 모델이다. 이는 기존의 FMM 모델의 갖는 효율적인 학습기능과 분류기능을 제공하면서, 특징집합에서 특징 값의 빈도와 분포에 따른 변화를 효과적으로 반영함으로써, 비 정규 데이터 및 노이즈에 의한 학습효과의 저하를 개선할 수 있게 한다.

표 1 다양한 얼굴 패턴에 대한 얼굴 인식율과 오인식율.

학습 패턴 수		얼굴 인식율	오인식율
얼굴 패턴	비얼굴 패턴		
40	100	76	18
60	100	88	16
80	100	93	21
100	100	94	17

표 2 반대 예제 패턴에 대한 얼굴 인식율과 오인식율.

학습 패턴 수		얼굴 인식율	오인식율
얼굴 패턴	비얼굴 패턴		
100	40	92	33
100	60	95	24
100	80	92	18
100	100	94	17

그림 6 은 인식 및 추적 시스템의 사용자 인터페이스를 보인다. 실험에서 얼굴 패턴 및 배경 영상 정보를 각각 100 개를 추출하여 학습 패턴으로 사용하였다. 예제 패턴에 의한 학습으로부터 검출 율을 상승시킬 수 있으나 오인식율 (False alarm rate)은 개선이 미미하였다. 이는 반대 예제(Counter example)패턴을 사용한 추가 학습으로 개선할 수 있다. 이에 배경 영상으로부터 비얼굴(Non-face) 패턴을 추출하여 점진적으로 학습시켜 그 효과를 평가하였다. 표 1 은 예제 패턴 수의 변화에 따른 인식률의 변화를 보인다. 표에서 보는 바와 같이 예제 학습 패턴 수가 점진적으로 증가함에 따라 검출율이 증가함을 볼 수 있다. 표 2 는 반대 예제 패턴수의 변화를 주어 오인식율의 변화를 관찰한 것이다. 비얼굴 학습 패턴의 개수가 증가함에 따라 오

\* 이 연구는 과학기술부 뇌과학 연구개발사업으로 수행되었음

## 참고문헌

- [1] Christophe Garcia, Manolis Delakis, "Convolutional Face Finder: A Neural Architecture for Fast and Robust Face Detection," IEEE Transactions on Neural Networks, Vol. 26, No. 11, 2004
- [2] Rein-Lien Hsu, Mohamed Abdel-Mottaleb, Anil K. Jain, "Face Detection in Color Images," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No.5, 2002.
- [3] Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja, "Detecting Faces in Images: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 1, 2002
- [4] Raphael Feraud, Olivier J. Bernier, Jean-Emmanuel, and Michel Collobert "A Fast and Accurate Face Detector Based on Neural Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 1 2001
- [5] Steve Lawrence, C. Lee Giles, Ah Chung Tsoi and Andrew D. Back, "Face Detection: A Convolutional Neural-Network Approach," IEEE Transaction on Neural Networks, Vol.8, No.1, pp.98-113, 1997.
- [6] P. K. Simpson, "Fuzzy Min-Max neural networks Part 1:classification," IEEE Transaction on Neural Networks, Vol.3 No.5, pp.776-786, 1992.