[UCI]I804:24011-200000240080

2010年 8月
博士學位論文

# MLNN을 이용한 실시간 얼굴인식에 관한 연구

朝鮮大學校 大學院

電算統計學科

林 憙 京

# MLNN을 이용한 실시간 얼굴인식에 관한 연구

The study on Real Time Face Recognition using MLNN

2010년 8월 25일

朝鮮大學校 大學院

電算統計學科

林 憙 京

# MLNN을 이용한 실시간 얼굴인식에 관한 연구

指導敎授　丁　采　榮

이 論文을 理學博士學位申請 論文으로 提出함.

2010년　4월　17일

朝鮮大學校 大學院

電算統計學科

林 憙 京

# 林 憙 京의 博士學位論文을 認准함

委員長　　敎授 _____(인)

委　員　　敎授 _____(인)

委　員　　敎授 _____(인)

委　員　　敎授 _____(인)

委　員　　敎授 _____(인)

2010年 6月　　日

朝鮮大學校 大學院

# 목 차

표 목 차 ·····

그림 목차 ·····

ABSTRACT ·····

Ⅰ. 서 론 ····· 1

Ⅱ. 얼굴인식에 관한 연구 및 얼굴 검출 ····· 6

2.1 얼굴 인식의 주요 기술 ····· 6

2.2 기존의 얼굴 영역 추출 ····· 9

2.3 특징 추출 방법 ····· 13

2.4 Haar-like 특징을 이용한 얼굴검출 ····· 21

2.4.1 Haar-like Feature와 인테그랄 이미지 ····· 22

2.4.2 AdaBoost 알고리즘을 이용한 얼굴 후보 영역 검출 ····· 25

Ⅲ. 제안한 얼굴 인식 시스템 ····· 35

3.1 제안한 얼굴 인식 순서도 ····· 35

3.2 얼굴 영상의 정규화 과정 ····· 37

3.2.1 얼굴 영상의 크기 정규화 ····· 38

3.2.2 얼굴 영상의 밝기 정규화 ····· 41

3.2.3 얼굴 영상의 noise 제거 ····· 47

3.3 정규화 영상을 이용한 얼굴 영상 공간 구성 ····· 49

3.3.1 얼굴 영상 구성 ····· 50

3.3.2 주성분 분석을 이용한 얼굴 고유 공간 구성 ····· 52

3.4 인공 신경회로망 ····· 58
3.4.1 인공 신경회로망 모델 ····· 60
3.2.2 신경망 학습 규칙 ····· 63
3.4.3 오류역전파 알고리즘 ····· 65

**Ⅳ. 실험 결과 및 고찰** ····· 73
4.1 실험 환경 ····· 73
4.2 실험 얼굴 영상 ····· 74
4.3 얼굴 인식 시스템의 성능분석 ····· 77
4.3.1 MLNN을 이용한 인식 결과 ····· 77
4.3.2 학습률에 따른 인식률 ····· 78
4.3.3 실시간 얼굴 인식 실험 결과 ····· 81

**Ⅴ. 결 론** ····· 86

**참 고 문 헌** ····· 88

# 표 목 차

[표 2-1] 얼굴 특징점 ····· 18
[표 2-2] AdaBoost 알고리즘 ····· 27
[표 3-1] 그레이 레벨 빈도수의 예 ····· 43
[표 3-2] 평활화된 그레이 레벨의 예 ····· 43
[표 3-3] 신경망의 학습 절차 ····· 63
[표 4-1] 얼굴 인증 시스템의 시스템 실험 환경 ····· 73
[표 4-2] 매칭 방법에 따른 매칭 성공률 ····· 78
[표 4-3] 학습에 사용된 얼굴 영상 개수에 따른 실험 환경 ····· 78
[표 4-4] 실험환경 1의 학습률에 따른 인식률 ····· 79
[표 4-5] 실험환경 2의 학습률에 따른 인식률 ····· 80

## 그림 목차

[그림 2-1] 여러 방향의 조명에 대한 얼굴 영상 ····· 6
[그림 2-2] 얼굴 인식에 사용되는 기하학적 특징 ····· 7
[그림 2-3] 얼굴 인식에 사용되는 고립밀도지도 특징 ····· 8
[그림 2-4] 평균 형판을 이용한 템플릿 매칭 ····· 10
[그림 2-5] 밝기 정보를 이용하여 얼굴 영역을 추출 ····· 11
[그림 2-6] 피부색 구간 설정을 통한 얼굴 영역 추출 ····· 12
[그림 2-7] 고유 공간(Eigenspace) 상에서 얼굴 특징점들의 분포 ····· 14
[그림 2-8] 특징점 추출에 의해 얻어진 얼굴 영역 ····· 16
[그림 2-9] 얼굴 영역의 기하학적 정보에 의한 특징점 분포 영역 ····· 18
[그림 2-10] 특징점 후보 영역으로부터 눈썹과 눈, 입 영역 검출 결과 ····· 19
[그림 2-11] 눈 후보 영역으로부터 좌·우측 눈 영역 획득 ····· 19
[그림 2-12] Eli Saber[41]의 얼굴 모델 ····· 20
[그림 2-13] 얼굴 영역 내의 특징 예 ····· 22
[그림 2-14] Haar-like Feaute의 프로토타입 ····· 23
[그림 2-15] 사각형의 인테그랄 이미지 ····· 24
[그림 2-16] 45° 회전 사각형의 인테그랄 이미지 ····· 25
[그림 2-17] Boost 알고리즘 ····· 26
[그림 2-18] AdaBoost의 알고리즘의 가중치 부여 ····· 28
[그림 2-19] AdaBoost 알고리즘의 단계별 그룹화 개요 ····· 29
[그림 2-20] 얼굴 검출을 위한 Cascade 구조 ····· 29
[그림 2-21] 단계별 사용된 프로토타입의 개수 ····· 30
[그림 2-22] AdaBoost의 학습에 의해 선택된 특징 ····· 31
[그림 2-23] 얼굴 후보 영역 검출을 위한 학습과정 ····· 31
[그림 2-24] 얼굴 검출을 위한 검색 윈도우 ····· 32

[그림 2-25] 다양한 조건하에서 실시간 얼굴 검출 ····· 33
[그림 2-26] 다중 얼굴 검출 ····· 34
[그림 3-1] 제안한 알고리즘 순서도 ····· 35
[그림 3-2] 전처리 전체 개략도 ····· 37
[그림 3-3] Haar-like Feature를 통해 검출된 얼굴 영상 집합 ····· 38
[그림 3-4] 선형 보간법에 의한 영상 축소 방법 ····· 39
[그림 3-5] 원 영상과 축소된 영상 ····· 40
[그림 3-6] 히스토그램 데이터를 평활화 ····· 43
[그림 3-7] 히스토그램 평활화 ····· 44
[그림 3-8] 일반 영상에 히스토그램 평활화를 적용한 영상 ····· 45
[그림 3-9] 조명의 세기가 다른 각각의 영상을 히스토그램 평활화 후의 영상 ····· 46
[그림 3-10] 마스크의 크기가 3인 1차원 메디언 필터의 적용 예 ····· 48
[그림 3-11] 임펄스 잡음이 포함된 영상에 메디언 필터를 적용한 영상 ····· 48
[그림 3-12] 얼굴 영상 공간을 이용한 전체 매칭 과정 ····· 49
[그림 3-13] 동양 사람의 방향이 틀어지는 경우 영상 ····· 50
[그림 3-14] 회전 영상에서 Haar-like Feature를 이용한 얼굴 영역 검출 영상 ····· 51
[그림 3-15] 입력 영상들과 평균 영상의 거리 관계 ····· 52
[그림 3-16] 일부 얼굴 영상 집합 ····· 53
[그림 3-17] 전체 얼굴 영상의 평균 영상 ····· 54
[그림 3-18] $V$행렬의 직교 행렬 데이터 ····· 56
[그림 3-19] $U$행렬의 직교 행렬 데이터 ····· 56
[그림 3-20] 트레이닝 얼굴 이미지에 대한 고유 영상의 예 ····· 57
[그림 3-21] 뉴런의 기능 모델 ····· 60
[그림 3-22] 일반적인 신경회로망 모델 ····· 61
[그림 3-23] 다층 신경회로망의 구조 ····· 66
[그림 3-24] BP 알고리즘의 순서도 ····· 72

[그림 4-1] ORL 얼굴 DB에서의 일부 얼굴 영상 ····· 74
[그림 4-2] 남자 5의 회전 영상의 모델영상 ····· 75
[그림 4-3] Haar-like Feature 알고리즘을 이용하여 [그림 4-2]의 얼굴 영역을 추출한 영상 ·· 76
[그림 4-4] [그림 4-3]에서 얼굴영역만 검출한 얼굴 영상 집합 ····· 76
[그림 4-5] [그림 4-5] 실험환경에 따른 인식률 변화 ····· 80
[그림 4-6] 얼굴 인식 시스템에 얼굴영상 입력 및 얼굴 영역 검출 ····· 81
[그림 4-7] 얼굴 영역 추출 및 밝기 정규화 ····· 82
[그림 4-8] 모델 얼굴 영상 초기화 ····· 82
[그림 4-9] 평균 영상 생성 및 데이터베이스 구축 시간 ····· 83
[그림 4-10] 오류역전파 알고리즘을 이용한 얼굴인식 ····· 83
[그림 4-11] 서양 남자의 회전 영상 입력 ····· 84
[그림 4-12] 회전 이미지의 대한 얼굴 검출 및 얼굴인식 ····· 85

# ABSTRACT

## The study on Real Time Face Recognition using MLNN

Advisor : Prof. Jung Chae-Young
Department of Computer Science and Statistics,
Graduate School of Chosun University

This study proposed stable and fast face detection and identification system. First, for reliable and stable face detection, Haar-like Feature and AdaBoost were proposed. AdaBoost can calculate simple features using integral image fast. Its cascade structure can calculate them fast by removing much of the background data in initial stage with different number of features in each stage. As it needs less cost for calculation than existing approaches, its real-time detection performance is remarkably improved.

The face recognition system proposed in this study obtains Eigenvector from face image set and analyses Principal Component using Eigenspace only consisting of useful vectors. It has been demonstrated through tests that it is simpler and more stable to process in comparison with approaches using geometric data or stereo image. However, the Principal Component tends to have lower recognition rate when there is change in illumination of images. In particular, when there are differences in illumination of learning images, as they may work as wrong feature points, the recognition rate

may significantly lower. Therefore, this study equalized histogram to recognize who is in database even when there is change in illumination of input images, resulting in decreased change of lighting. It is not affected by changes in lighting and has maintained consistent recognition rate in comparison with PCA approach and approach using Intensity Normalization. In matching input image with model image, matching failure may occur even though real face images are successfully matched in face space where many faces are projected.

To solve the problem, this study used Error Back Propagation learning algorithm, which is Multiple Perceptron learning. Using feature vectors obtained through Principal Component Analysis as input data for neutral network, this study presents suggestions for improvement of recognition performance. As a result, recognition rate of the proposed approach was 95.3%. A mean of 6.5% was improved in comparison with the existing approaches. In particular, matching failures occur frequently when there is hardly change in feature data such as rotated image of face. However, in this approach, only 2% of failure rate was found.

This study obtained the optimal learning value by testing changes in recognition rates under a variety of environments. Finally, the proposed face detection system showed the best performance in real-time application, but changes in face expression and face detection in complex backgrounds should be further considered. If stereo vision is applied and face recognition is added for it to improve security and recognition rate, it can be effectively used for diverse systems.

# I. 서론

최근 지능형 자율 이동 로봇에 대한 연구가 본격적으로 진행되고 있다. 지능형 이동 로봇 개발에서 중요한 것은 기구부, 제어, 전력 공급 장치와 같은 H/W 기술이 있겠지만, 로봇이 진정한 지능을 갖기 위해서는 인간과 같이 독립된 개체로서 주어진 환경을 이해하고 자율적으로 특정한 일을 수행할 수 있는 능력이 확보되어야 한다. 인간의 시각은 인간 생활의 90% 이상을 차지할 정도로 중요하며, 특히 인간의 시각을 통한 물체인식 능력은 다른 어떤 시스템보다 탁월하다.

컴퓨터 비전에서 최종 목표로 하는 것은 바로 이와 같은 인간의 시각 인식특성을 컴퓨터로 모사하는 것이다[1]. 최근 인간의 시각을 이용한 인식 구조를 이해하기 위해 심리학, 신경생리학, 인지과학, 기하학, 광학, 인공지능, 패턴인식 분야에서 방대한 연구를 해왔지만 아직까지 완전하게 정립된 인식 메커니즘이 없다. 한편, 컴퓨터비전 분야에서는 알려진 인간의 시각 인식특성을 이용하여 컴퓨터로 구현하고자 다양한 시도를 하였다. 대표적으로 모델 기반 인식에서는 인간 시각의 상향식 및 하향식 처리 과정을 가설-생성-검증과정으로 모델링한 바 있다[2][3].

인간이 살아가는 과정은 외부로부터의 많은 데이터에 대한 의식적인 혹은 무의식적인 반응이라 할 수 있다. 이에 따라 입력으로 들어온 수많은 데이터를 분석하고 표현함과 동시에 그것들로부터 정보를 끌어내는 일, 즉 입력을 이해하는 일이 필요하게 된다. 인간이 어떠한 방식으로 감각 기관으로부터 정보를 습득하고, 저장하고, 활용하는가에 관한 탐구는 인간의 인지능력에 관한 연구에 있어서 가장 중요한 주제 중 하나이다. 컴퓨터비전은 인간의 여러 감각들 중에서도 가장 중요한 역할을 담당한다고 할 수 있는 시각적 인지능력에 대한 연구를 통해 궁극적으로 기계에게도 인간과 유사한 인지능력을 심어주려는 과학적인 목표를 갖고 이러한 문제들에 접근하고 있는 학문분야이다. 이러한 컴퓨터비전의 여러 분야 중에서도 얼굴인식은 컴퓨터비전 제반 분야의 방법론을 한데 아울러 종합적인 결과를 산출해 내는 중추적인 역할을 담당하고 있으며, 그 중에서 얼굴 인식 기술은 국내·외적으로 활발하게 연구가 수행되어 왔고 현재도 많은 연구가 진행 중이다.

얼굴인식 기술은 지문인식 등 다른 생체인식기술과 다르게 자신의 신체일부를 인식 장치에 직접 접촉시키지 않아도 되는[4] 등 생체정보 획득방법에서 강제성이 적어 다른 생체인식기술이 사용자에게 줄 수 있는 거부감과 불편함이 존재하지 않

는다는 특성이 있다. 즉 가장 인간친화적인 기술이라는 것이다. 이러한 비접촉성. 비강제성, 편리성의 특징으로 인해 얼굴인식 기술은 신원확인분야 외에도 신분확인 대상자가 모르는 사이에 자연스럽게 정보를 획득해야 하는 지능형 무인감시 등의 분야에도 효과적으로 응용할 수 있다.

생체인식은 지문이나 손 모양, 음성, 홍채, 망막, 혈관, 서명에 이르기까지 개인에 따라 그 특징이 명확하게 다른 신체 부위나 행동 특성으로 사람을 식별하고 인증하는 것을 말하며, 이러한 생체 특징들은 분실 또는 망각의 우려가 없고 나아가서는 복제나 위조가 불가능하다는 점에서 최근 열쇠나 카드식 출입통제 시스템 등 전통적인 보안시스템을 대신할 차세대 보안 기술로 각광받고 있다.

얼굴인식 기술은 광범위한 응용 가능성을 지닌 반면에 일상생활의 각 분야에 실제로 적용되기 위해서는 반드시 해결되어야 할 몇 가지 제약도 지니고 있다. 그 중 대표적인 것이 인식 환경이나 인식 대상의 변화에 민감하다는 것이다. 즉 얼굴인식 시스템은 안경, 머리모양, 표정 등에 의해 수시로 바뀌는 얼굴 모양과 주변 조명 및 배경의 다양한 변화를 고려하여 설계되어야 한다. 현재 많은 응용 시스템들이 얼굴 영상을 등록하고 인증하는 과정에서 가능한 배경 및 조명을 일정하게 고정시키거나 사용자로 하여금 무표정한 상태를 유지하도록 지시함으로써 얼굴 모양의 변화를 통제하고 있다. 그러나 안정적인 얼굴인식 성능을 위해서는 이와 같은 문제점들을 시스템의 인식 단계에서 자체적으로 해결할 수 있어야 하고, 동시에 인식 속도도 응용분야에 적합한 수준이 보장되어야 한다[5].

얼굴 인식에 대한 연구는 크게 인식기에 중점을 둔 연구와 특징추출에 중점을 둔 연구로 나눌 수 있다. 인식기에 중점을 둔 연구는 에러함수에 바탕을 둔 통계학적 인식기, 신경망을 이용한 인식기 등으로 분류될 수 있다. 이들 인식기에 대한 연구는 보다 인식률을 높이려는 노력과 인간과 유사한 인식 과정을 재현하려는 목적으로 이루어져 왔으며, 각 인식기에 적합한 특징들을 입력으로 사용한다[7]. 에러함수는 단순한 LMSE(Least Mean Square Error)나 에너지 함수 등을 사용하고, Bayer의 결정함수를 사용하여 인식한다[6-9]. 이 밖에 fuzzy set이나 moment 등을 이용한 변형된 형태의 인식에 이용되고 있다[10][11]. 신경망 회로는 정교한 조정이 가능한 반면, 다수의 테스트 이미지로 미리 학습을 해야 한다는 제약이 있다. 이런 신경망 회로는 입력 차수가 작아야 유리하므로 원래 이미지를 변환하여 차수를 줄이는 방법인 PAC(principal component analysis)등과 결합하여 쓰이고 있다

[12][13].

얼굴 인식을 위한 특징들은 기하학적인 특징, 고립밀도지도(isodensity map), 질감 특징 등으로 분류될 수 있다. 기하학적인 특징들은 가장 쉽게 생각할 수 있는 특징으로 초기에 많이 사용되었으며, 고립밀도지도 특징은 조명이 고정된 환경에서만 이동되었고[10][14], 질감 특징은 신경망의 입력으로 많이 사용되고 있다[12]. 기하학적인 특징들은 조명에 거의 영향을 받지 않으나 원래 이미지에 대한 정보의 손실이 많아 저해상도 이미지에서는 인식률이 떨어진다. 고립영역지도는 조명의 변화에 치명적이지만 고정된 조명에서는 높은 신뢰도를 나타낸다. 질감 특징은 정보의 손실은 적으나 입력 정보의 양이 많아 별도의 변환이 필요하다는 단점이 있다. 이러한 단점을 극복하기 위해 수학적 모델을 통해 조명의 방향을 예측하거나 얼굴을 여러 각도에서 촬영한 충분한 이미지를 학습시키는 방법을 사용한다[15][16].

얼굴 표정의 동적 변화를 실시간으로 분석하기 위해서는 얼굴 동작의 시간적 변이를 효과적으로 반영할 수 있는 최적의 표정 정보의 추출과 실시간 추적이 필요하며 특정 표정 사이의 변화를 능동적으로 설명할 수 있는 표정변화 모델에 기반을 둔 얼굴 표정의 해석 방법이 필요하다. 그러나 실시간 분석을 위한 얼굴표정 변화의 추적은 조명 및 배경의 변화에 따라 매우 어려운 작업으로 인식되고 있다.

얼굴 표정 인식은 1971년 EKman과 Friesen[17]에 의해 행복, 슬픔, 공포, 혐오, 놀람, 화남의 주요 감성표현 방법이 제안된 이후 얼굴표정 인식에 관한 다양한 연구가 진행되고 있다[18-20].

일반적으로 얼굴표정 인식 시스템은 얼굴 특징벡터의 측정단계와 얼굴표정의 인식의 두 단계로 구성된다. 얼굴 표정은 눈썹, 눈, 코, 입, 피부와 같은 얼굴 요소들의 변형이 일어날 때 발생하는 얼굴 근육의 수축에 의하여 나타나며, 얼굴 표정의 강도는 얼굴 특징의 기하학적 변화 또는 근육 표현의 밀도에 따라서 결정된다. 이러한 얼굴의 특징 추출 방법으로는 크게 변형기반 방법과 모션기반 방법이 있다.

변형기반 방법은 영상기반과 모델기반 방법이 있으며, 영상기반 방법은 얼굴 영상의 전체적인 변형을 분석하거나, 얼굴의 주요 특성 부분으로부터의 변화를 추출하는 방법으로 정확한 추출을 위하여 얼굴과 배경을 분리하는 처리가 필요하며, 빛의 변화에 민감한 특성이 있으므로 이러한 빛의 영향을 최소화하는 처리과정이 필요하다. 모델기반의 얼굴 특징 추출 방법은 얼굴의 형태 또는 텍스쳐 정보를 이용하여 얼굴의 변형을 추출하는 방법이다. 대표적으로 Active Shape Model(ASM)은

추출하고자 하는 객체에 대하여 정점으로 구성된 점 분포모델(Point Distribution Model)을 이용하여 얼굴 변화를 분석하며, Active Appearance Models(AAM)는 얼굴의 형태와 그에 대응되는 텍스쳐 정보를 이용하여 얼굴 영상을 해석하는 방법이다[21].

반면 모션기반 방법으로는 밀집광류, 특징점추적, 차영상방법, 그리고 마커추적등이 있으며, 공간적 분석법과 시공간적 분석법으로 분류할 수 있다. 공간적 분석법은 주성분 분석, 독립요소 분석, 가버 웨이블릿 필터와 같은 방법을 이용하여 영상을 분류한다. 시공간적 분석법은 시간의 변화에 따른 얼굴 표정을 해석하는 방법으로 신경망, 은닉마코프 모델 등의 방법이 있다[22][23]. 시공간적 모션 에너지 템플릿 방법은 얼굴의 움직임을 2차원의 움직임 공간으로 표현하여 두 템플릿 사이의 유클리디언 거리를 계산하여 얼굴 표정을 추출한다[24]. 이렇듯 많은 분야와 많은 연구자들에 의해 컴퓨터 비전을 기반한 물체인식과 생체인식 등과 같은 시스템이 활발히 개발되어지고 있으며, 그 중에서도 얼굴 인식 분야가 가장 각광받고 있는 인식 분야이다.

최근 들어 컴퓨터 비전은 다양한 분야에서 활발하게 연구가 진행되어 왔다. 특히, 얼굴 인식은 신분 증명, 보안 시스템, 인사 자료, 회원 관리, 비디오 검색과 같은 실생활에 이용하기 위한 연구가 두드러지게 나타나고 있다. 얼굴 인식 시스템은 일반 환경에서 사람의 얼굴을 찾아내고 인식하는 시스템으로 사용자는 카메라만 응시하면 되므로 생체정보의 획득 방법에서 강제성이 적어 다른 생체 인식 기술에 비해 사용자에게 줄 수 있는 거부감과 불편함이 존재하지 않는 특징을 가지고 있다. 얼굴 인식 기술의 응용 분야를 살펴보면 출입통제, 근태관리 시스템, 전자상거래, 금융거래, 전자결재 시스템, 지능형 영상 감시 장치, 주요 고객관리, 공항검색시스템, 웹 사용자 인증 시스템, ATM, 인터넷 뱅킹 사용자 인증 시스템, 컴퓨터 보안, 정보보안 시스템 등 다양한 분야에 활용되고 있다.

사람의 얼굴 영상을 일반 환경에서 인식할 경우 설비된 현장 영상이 실시간으로 변화되는 영상들을 획득하게 되는데, 접근해 오는 얼굴 영상의 흔들림과 얼굴의 반사, 주변의 조도 등 환경요소의 변화에 따른 얼굴 영상을 인증할 수 있어야 한다. 직관적으로 볼 때 얼굴은 얼굴영역과 비 영역으로 구분하고 구성 요소의 특징을 이루는 눈, 눈썹, 코, 입 등의 위치가 변하지 않는다. 이들만의 요소문제를 실시간으로 받아들이는 영상에서 개인 얼굴이 지니고 있는 불변의 특징값을 추출하여 학

습된 인식알고리즘을 통하여 인증 여부를 확인하게 된다. 인간의 시각을 통한 인식 방법은 학습된 기존 지식을 이용하거나, 경험 및 습득한 자료를 이용해서 대상 얼굴을 찾아내기도 하고 다른 사람과 구분하기도 한다. 하지만 컴퓨터를 통한 얼굴인식은 매우 복잡한 패턴 인식 절차를 따르게 된다.

본 연구에서는 일반 환경에서 얼굴의 동적인 시각정보를 실시간으로 인증하는 문제에 접근하고자 한다. 이 방법은 매우 편리한 인터페이스를 제공할 수 있고 보안효과가 우수하여 단독으로, 또는 다른 방법과 혼합하여 인식률을 높일 수 있다.

본 논문의 구성은 다음과 같다.

제 Ⅰ 장에서는 얼굴인식에 관한 연구 배경과 얼굴인식 연구 동향 및 논문의 필요성을 기술 하였다.

제 Ⅱ 장에서는 얼굴인식에 관한 다양한 연구 활동 및 얼굴 검출에서는 기존의 얼굴영역 추출 및 특징 추출 방법, Haar-like Feature와 AdaBoost를 이용하여 얼굴영역을 검출하는 방법을 기술하였다.

제 Ⅲ 장에서는 본 논문에서 제안하는 얼굴 인증 시스템의 알고리즘을 기술하였고, 전처리 과정에서는 얼굴 영상의 크기, 영상의 밝기 정규화 방법에 대하여 기술하였다. 인식 분야에서는 주성분 분석과 오류역전파를 이용한 얼굴 인식 방법을 기술하였다.

제 Ⅳ 장에서는 본 논문에서 제안한 얼굴 인식 시스템에 대한 실험 분석 및 평가 결과를 기술하였다.

마지막으로 제 Ⅴ 장에서는 연구 결론과 향후 추가적인 연구 과제를 제시하였다.

# II. 얼굴인식에 관한 연구 및 얼굴 검출

## 2.1 얼굴 인식의 주요 기술

얼굴 인식에 사용되는 특징들을 비교, 분석하기 위해서는 조명의 밝기와 위치, 얼굴의 각도 등에 대한 영향 정도, 원래 이미지에 대한 정보 손실 정도, 처리하기 위한 계산량, 신뢰도 등을 다각적으로 검토해 볼 필요가 있다. [그림 2-1]은 조명이 비치는 방향에 따른 얼굴 이미지들을 보여주고 있다.

(a) 측면 조명 (b) 정면 조명 (c) 위쪽 조명

[그림 2-1] 여러 방향의 조명에 대한 얼굴 영상

얼굴 인식에 관한 기존의 연구 중 주요 기술을 살펴보면 다음과 같다.

- **얼굴의 기하학적인 특징을 이용한 인식 기술**
- **고립밀도지도(Isodensity map) 특징을 이용한 방법**
- **에러함수를 이용한 통계학적 방법**
- **형판 기반(Template Matching)을 이용한 방법**

이 외에도 무수히 많은 연구가 진행되었으며 현재도 인식률을 높이기 위한, 조명이나 외적 요소로부터 강인한 알고리즘을 찾기 위해 많은 연구가 진행되고 있다.

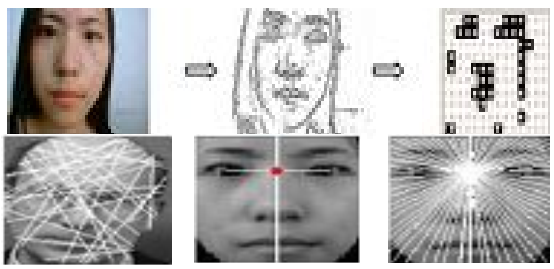
### 2.1.1 얼굴의 기하학적인 특징을 이용한 인식 기술

얼굴의 기하학적인 특징을 이용한 인식 기술은 얼굴 인식에서 가장 보편적으로 이용하는 방법론으로 눈, 코, 입과 같은 얼굴의 특징점들의 위치나 크기 또는 이들 간의 거리와 같은 기하학적인 인자들만으로도 각 개개인의 얼굴을 인식할 수 있다는 사실에 착안한 것이다. 이들 기하학적 특징들은 입력 화상의 해상도를 낮추었을 때 최종적으로 남는 요소들이다.

Santini는 눈, 코, 입 등의 길이와 얼굴 영역의 너비 등을 특징으로 사용하였고[6], Brunelli는 턱의 형태를 포함시키기 위해 입에서 턱까지의 11등분한 각도에서의 거리를 첨가했다[7]. Cox는 이들 특징들을 독립적으로 사용하지 않고, 적정한 가중치로 이들 특징들을 결합하여 혼합 거리를 특징으로 사용하기도 하였다[8].

기하학적 특징의 가장 큰 장점은 조명에 무관하다는 것이다. 조명의 방향에 관계없이 이들 길이나 형태는 변하지 않는다. 단지 얼굴의 회전각에 의해서만 길이가 변형될 뿐이다. 그러나 조명이 충분치 않다면 그림자 부분이나 어두운 부분의 영향으로 기하학적 특징 자체를 추출하기 어려울 수 있다. [그림 2-1]의 얼굴 윤곽이 좋은 예이다. 조명의 위치에 따라 가장자리가 그림자로 가려 전혀 다른 모양을 나타낸다.

기하학적인 특징이 비교적 단순하고 비교하기 쉬운 장점이 있는 반면 얼굴 이미지가 원래 포함하고 있는 많은 정보를 버리게 된다. [그림 2-2]에서 기하학적인 특징의 예를 보여주고 있는데, 여기서도 알 수 있듯이 기하학적인 특징을 이용한다는 것은 그림과 같은 단순화된 얼굴 도형만을 가지고 사람을 인식하는 것과 같다.



[그림 2-2] 얼굴 인식에 사용되는 기하학적 특징

정보의 손실을 줄이기 위해 최근에는 스케치 형태의 얼굴 윤곽을 이용하거나 보다 많은 특징점을 선정하여 사용하기도 하고, affine transform 등을 적용하여 얼굴 각도의 변화를 보정하려는 연구도 시도되었다[9].

### 2.1.2 고립밀도지도(Isodensity map) 특징을 이용한 방법

고립밀도지도란 얼굴에서 일정 범위의 밝기를 가지는 영역의 분포나 모양을 말한다. 각 밝기 영역은 일정 간격의 밝기 범위를 두고 결정되며 일종의 등고선 형태로 분포하게 된다. 고립밀도지도의 가장 큰 단점은 조명의 밝기나 각도에 매우 민감하다. [그림 2-3]에서 고립밀도지도의 한 예를 보여주고 있다.

고립밀도지도가 조명에 민감한 반면 기하학적인 특징보다는 많은 정보를 포함하고 있고, 얼굴의 3차원적인 형태정보를 일부 포함하고 있음으로 조명을 고정시킨 환경에서는 높은 신뢰도를 보이고 있다[10][15].

[그림 2-3] 얼굴 인식에 사용되는 고립밀도지도 특징

### 2.1.3 에러함수를 이용한 통계학적 방법

가장 널리 알려진 에러함수는 기하학적 거리일 것이다. n차원의 참조 벡터 A와 일렬 벡터 B가 주어졌을 때 기하학적 거리의 공식의 한 예가 (식 2-1)과 같다.

$$d(A,B) = \left[\sum_{i=1}(a_i - b_i)^2\right]^{\frac{1}{2}} \qquad \text{(식 2-1)}$$

$$A = (a_1, a_2, a_3, \ldots, a_n), B = (b_1, b_2, b_3, \ldots, b_n)$$

인간의 물체 인식과정은 선형적 대수관계를 따르지는 않는다. 인간을 대상으로 한 실험에서도 인간의 유사도 척도가 선형적이지 않다는 여러 증거가 밝혀졌다[6].

벡터의 각 원소들이 서로 상관관계가 크지 않다면, 기하학적인 거리가 큰 의미를 부여하지는 못한다. 사각형의 모양과 크기를 변경시키며 행한 인식 실험에서 Atteave는 실험치에 의한 설득력 있는 에러 함수를 유도하였다[6]. 이를 구획거리 모델(city block distance model)이라고 하는데 이에 대한 공식은 (식 2-2)와 같다.

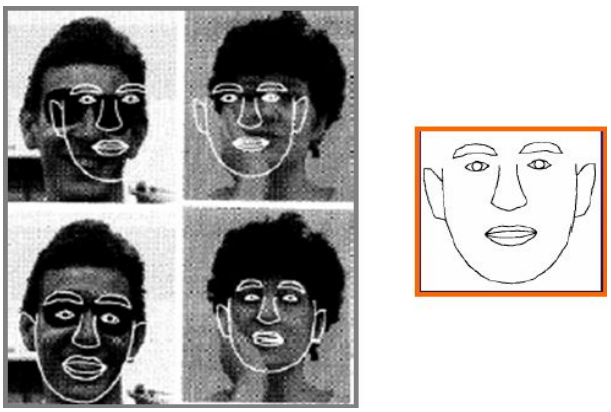
$$d(A,B) = \sum_{i=1}^{n}|a_i - b_i| \qquad \text{(식 2-2)}$$

이외에도 유사한 거리함수들로 Hamming distance[25], Minkowski distance[6] 등의 모델이 있으며 각 경우에 적합한 용도로 선택될 수 있다. 예를 들면 Hamming distance는 이진 데이터에 적용하기 편리하고, Minkowski distance는 fuzzy set 등에 응용될 수 있다. 이들 에러함수를 이용하면 통계학적인 인식기의 구현이 가능하다. Bayer 인식기[6]와 fuzzy set을 응용한 인식기[11], 은닉 마코프 모델을 이용한 인식기[26] 등이 여기에 속한다. 대표적인 인식기로 nearest neighbour method[27]를 살펴보면 하나의 입력 특징 벡터에 대해 저장되어 있는 모든 참조 특징 벡터들과의 각각의 에러 함수를 구한 뒤, 그 중 최소값을 가지는 참조 클래스에 입력이 포함된다고 결정하게 된다.

### 2.1.4 형판 기반(Template-based)방법

형판 정합 방법이란 사전에 적당한 템플릿을 이용하여 찾고자 하는 영상과 유사한 위치를 알아내는 방법으로서, 얼굴의 전체 또는 일부분을 템플릿으로 미리 등록

하고 입력영상과 템플릿 매칭을 통하여 얼굴을 찾는 방법이다. 하지만, 이 방법은 검색 영역이 넓은 경우에는 처리해야 하는 양이 커지므로 연산 속도 면에서 저조한 결과를 가져오며 템플릿의 형태가 찾고자 하는 형태와 얼마만큼 동일하게 제작되었는가에 따라 유사도의 차이를 나타낼 수 있다. 즉, 사람마다 얼굴 특징 요소의 크기나 모양이 다르기 때문에 일반화된 템플릿의 설계가 쉽지 않다는 것과 얼굴의 중요한 특징 요소 부분이 가려지거나 얼굴이 기울어지고 변형되면 효과적인 템플릿 매칭이 이루어지지 않는다. 따라서 최근에는 이러한 단점을 보완하고자 변형된 템플릿을 이용한 탄성 템플릿 매칭이 연구 중이다. [그림 2-4]는 평균 형판을 이용한 매칭을 보여주고 있다.



[그림 2-4] 평균 형판을 이용한 템플릿 매칭

형판 기반의 가장 두드러지는 특징은 얼굴요소를 찾는 과정이 필요 없고 복잡한 배경에서도 수행이 가능한 장점을 지니고 있지만 개개인의 얼굴마다 차이가 크기 때문에 표준 형판의 생성에 따라 영향을 많이 받는 단점을 가지고 있다.

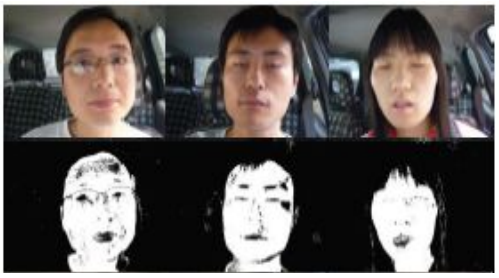
## 2.2 기존의 얼굴 영역 추출

얼굴 영역을 추출하는 문제는 얼굴 인식 뿐 아니라 과거 영역 기반 영상 압축 분야의 객체 분할과정에서 많이 연구가 되어 왔으며, 근본적으로 사람이 포함된 입력 영상일 경우 얼굴 영역을 추출하는 시도는 동일하다고 볼 수 있다. 이에 밝기

정보를 이용한 기법, 움직임 정보를 이용한 기법, 밝기 및 움직임을 동시에 이용한 기법 및 칼라정보를 이용한 얼굴 영역 추출 기법에 대해서 간략하게 기술하면 다음과 같다.

### 2.2.1 밝기 및 움직임 정보를 이용한 기법

밝기 정보를 이용하여 얼굴 영역을 추출하는 기법은 입력 영상에서 배경 영역과 얼굴 영역의 밝기가 다르다고 가정하고 영상 분할을 통하여 얼굴과 배경을 분할함으로써 얼굴 영역을 추출한다. 배경이 단순하여 얼굴 영역과 배경 영역의 경계 부근 화소들의 밝기값의 차가 큰 경우에는 얼굴 영역을 정확히 추출할 수 있으며 아래 [그림 2-5]는 이에 대한 예를 보여준다. 그러나 경계부근 화소들의 밝기값이 유사할 경우 또는 복잡한 배경에서는 얼굴 영역과 배경 영역이 같은 영역으로 추출될 수 있어 얼굴 영역을 정확히 추출할 수 없는 단점이 있다.



[그림 2-5] 밝기 정보를 이용하여 얼굴 영역을 추출

### 2.2.2 움직임 정보를 이용한 기법

얼굴 영역과 배경 영역의 경계 부근 화소들의 밝기값이 유사하더라도 얼굴 영역에 움직임이 존재한다면 인접 프레임간의 움직임 정보를 이용하여 얼굴 영역을 정확히 추출할 수 있다. 움직임 정보를 이용한 얼굴 영역 추출 기법은 크게 인접 프레임간의 차영상을 이용하는 경우, 움직임 추정을 통한 움직임 벡터(Motion Vector)를 이용하는 경우로 구분된다. 움직임 정보를 이용하는 기법들은 동영상에서 사람의 얼굴은 움직임이 있는 객체라고 가정하고, 인접 프레임간의 차 또는 움

직임 추정을 통해 유사한 움직임 벡터값을 나타내는 화소들을 그룹화 함으로써 얼굴 영역 추출하였다.

또한 움직임 추정을 통한 얼굴 영역 추출 기법들은 얼굴 영역의 움직임이 분명하고, 배경 영역의 움직임이 없을 경우 얼굴 영역을 정확히 추출할 수 있다. 그러나 배경 영역에서의 움직임 추정 정보가 얼굴 영역에서의 움직임 추정 정보와 유사할 경우 움직임 정보만을 이용하여 얼굴 영역을 정확히 추출할 수 없다.

### 2.2.3 색상 정보를 이용한 기법

피부색은 얼굴 영역 추출 시 유용한 정보가 될 수 있어 널리 이용되어 왔다. 이는 크게 HSI, YIQ 및 YCbCr 등과 같은 칼라 모델에 기반을 두어 피부색 구간을 설정하고 피부색 화소들을 추출함으로써 얼굴 영역을 추출하는 기법들과, 피부 영역의 색분포를 모델링 한 후 입력 화소들에 대한 피부색 및 비피부색 화소 여부를 결정하여 얼굴 영역을 추출하는 기법들로 구분된다.

피부색 구간 설정을 통한 기법으로는 HSI 및 YCbCr 칼라 모델에 기반을 둔 기법들이 많이 연구 되었다. HSI 칼라 모델에 기반한 기법은 색상(Hue) 정보에서 피부색 구간 범위를 정한 후 입력영상 화소의 색상값이 피부색 구간 범위에 속하면 얼굴 화소로 결정한다. YCbCr 칼라 모델을 이용할 경우 Cb 및 Cr의 피부색 구간 범위를 각각 정한 후 이에 속하는 화소들을 추출함으로써 얼굴 영역을 추출하였다. 배경 영역의 색상이 단순하고 또한 이의색상값이 피부색 구간 범위내의 값과 차이를 나타내는 경우 얼굴 영역을 정확히 추출할 수 있다.

[그림 2-6]은 배경 영역에 피부색과 유사한 색상을 나타내는 화소들이 존재하고, 배경 영역이 복잡한 경우 정확한 얼굴 영역을 추출하기 어려움을 보여준다.

[그림 2-6] 피부색 구간 설정을 통한 얼굴 영역 추출
(복잡하고 피부색과 유사한 배경)

## 2.3 특징 추출 방법

특징 추출에 앞서 카메라의 줌인 및 줌아웃으로 인한 얼굴 크기 변화, 회전 변화 및 주위 환경의 영향으로 인한 밝기 변화 등을 고려하기 위하여 정규화 과정을 수행한다. 이는 보통 패턴 인식의 정규 과정에서 널리 이용되는 어파인(affine) 변환을 사용하여 이동, 크기 및 회전 등을 정규화 한다. 그리고 히스토그램 평활화를 이용하여 밝기 정규화를 수행한다.

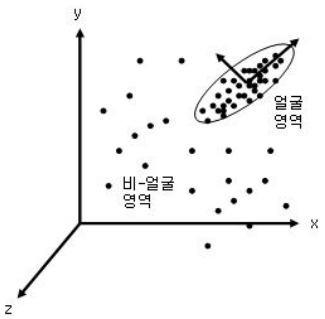
얼굴 영역 추출 후 각 개인을 구분하는 척도가 되는 특징 추출에 관한 연구는 KL(Karhunen Loeve) 변환을 이용하여 고유 벡터를 추출하고 이를 특징으로 하는 기법, 영상에서 방향 정보와 기울기의 크기 정보를 추출해 낼 수 있는 허프(Hough) 변환을 이용하는 기법, 눈, 코 및 입 간의 거리를 특징으로 이용하는 기법, 주성분 분석(PCA, Principal Component Analysis)을 이용하는 기법 및 선형 변별 해석(LDA, Linear Discriminant Analysis)등 많은 연구가 수행되어 왔다.

Pentland 등[29]은 주성분 분석으로 물체간의 상관관계가 없는 특징들을 추출하여 최 근접 이웃(nearest neighbor) 알고리즘으로 분류하는 방법을 제안하고 이를 얼굴 인식에 적용하였다. 주성분 분석은 입력 벡터를 공분산의 고유치와 고유벡터로 정의되는 부공간(subspace) 상에 수직 투영시킨다. 이 때 입력 자료의 투영값을 나타내는 주성분들은 평균 자승 오차 측면에서 기존의 방법 보다 입력 자료를 잘 표현한다.

Swet 등[30]은 주성분의 변별 성능을 높이기 위하여 주성분을 분석한 후 선형 변별하는 해석 방법을 제안하고 이를 적용하였다. 한편 Chatterjee 등[31]은 Swet 등의 선형 변별 해석 방법을 개선하여 학습 과정에서 클래스 평균을 추정하여 주성분 분석과 선형 변별 해석을 함께 할 수 있는 방안을 제시하였다. 이들 방법에는 클래스내 산포도(within-class scatter)에 대한 클래스간 산포도(between-class scatter)의 비를 변별 기준으로 이용하기 때문에 단봉형태(uni-modal)로 분포된 자료 또는 여러 개의 가우스 분포를 갖는 군집으로 구성된 자료들에 대해서 변별력이 우수하다고 알려져 있다.

### 2.3.1 Eigenface Algorithm

Eigenface는 256*256 이미지를 65536 차원공간으로 만드는데 이것은 noise과 빛의 조건에 민감하여 원래의 이미지를 가지고 얼굴인식을 하기에는 너무 복잡하고 시간 또한 오래 걸린다. 그러므로 가장 중요한 몇 개의 고유 벡터(Eigen-Vector)만으로 얼굴을 표현하는데 사용할 수 있다. Eigenface 기본 개념은 고차원으로 표현된 이미지를 저차원으로 표현해서 얼굴을 쉽게 인식하게 하는 것이다. 이렇게 차원을 줄이면 이미지의 모든 픽셀을 비교하지 않고도 Testing Face를 Training face들과 비교할 수 있다. [그림 2-7]는 실제 얼굴 영상을 대상으로 특징점을 추출하였을 때 그 특징점들의 분포가 고유 공간(Eigenspace) 상에서 어떻게 분포되는 지를 보여주고 있다.




[그림 2-7] 고유 공간(Eigenspace) 상에서 얼굴 특징점들의 분포

### 가. Eigenface Training Method

얼굴 이미지는 $N^2 * 1$ vector로 표현되며, X1,X2,....,XM의 M장의 이미지가 있다. (식 2-3)과 같이 M장의 이미지에 대해서 평균을 구한다.

$$\overline{X} = \frac{1}{M}\sum_{i=1}^{M} X_i \qquad \text{(식 2-3)}$$

$$\Phi_i = X_i - \overline{X} \qquad \text{(식 2-4)}$$

(식 2-4)과 같이 계산된 평균 이미지에서 각각 이미지의 차를 구하여 편차인 $\Phi_i$를 구한다음, 이 과정을 통해서 (식 2-5) vector는 normalize 된다.

$$C = \frac{1}{M}\sum_{n=1}^{M}\Phi_n\Phi_n^T = AA^T \quad (N^2 * N^2 \text{ matrix}) \qquad \text{(식 2-5)}$$

$$\text{where} \quad A = [\Phi 1\ \Phi 2 \ldots . \Phi M] \quad (\text{N} * \text{M matrix})$$

이 이미지들의 편차로 이루어진 matrix $A$와 $A^T$ 의 곱으로 공분산 행렬(covariance matrix)을 구할 수 있다. 이 공분산 행렬을 이용하여 Eigen-value와 Eigen-vector를 구하여 이미지를 선형결합(linear combination)의 형태로 만들 수 있다. (식 2-6)은 Eigen-value와 Eigen-vector를 구하는 과정이다.

$$A^T A v_i = \mu_i v_i \qquad \text{(식 2-6)}$$

공분산 행렬은 (식 2-7)과 같다

$$C = A \triangle A^T \qquad \text{(식 2-7)}$$

$$\text{where } A = [\Phi 1\ \Phi 2 \ldots . \Phi M]$$

이미지 vector의 공분산 행렬 크기가 $N^2 * N^2$이기 때문에 만약 256*256의 이미지라면 65536*65536의 행렬이고 여기서 Eigen-value를 구한다는 것은 엄청난 비용을 초래하게 된다.

$A^T A$는 M * M size로 공분산 행렬보다는 작다. $A^T A$를 이용하여 Eigen-value를 구하면 (식 2-8)과 (식 2-9)과 같다.

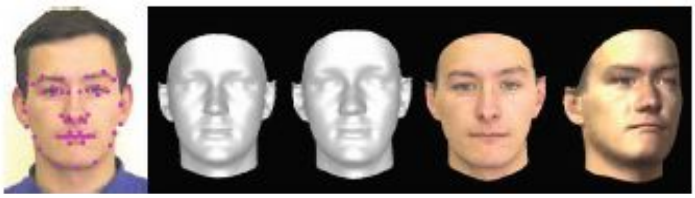
$$A^T A v_i = \mu_i v_i \Rightarrow A A^T A v_i = A \mu_i v_i \Rightarrow C A v_i = A \mu_i v_i \quad \text{(식 2-8)}$$

$$C A v_i = A \mu_i v_i \text{ or } A^T C A v_i = \mu_i A v_i \Rightarrow C u_i = \mu_i u_i (u_i = A v_i) \quad \text{(식 2-9)}$$

위위 식을 보면 $A^T A$와 $A A^T$를 이용하여 각각 Eigen-value를 구할 수 있다. 그 두 Eigen-value를 비교하면 C를 이용하여 구한 상위 M개의 Eigen-value와 $A^T A$를 이용하여 구한 Eigen-value가 같다는 것을 알 수 있다. 이 Eigen-value를 이용하여 Eigen-vector를 구하면 $A^T A$는 $v_i$가 되며, $A A^T$는 $u_i$가 된다. 즉, $A A^T$로 계산을 하면 계산량을 줄이면서 같은 Eigen-value를 얻을 수 있으며 이에 따른 Eigen-vector는 $u_i$를 사용한다. 상위 몇 개의 Eigen-value만으로도 본래의 이미지를 표현할 수 있다. 상위 몇 개의 Eigen-value가 이미지의 주요 성분을 거의 포함하고 있으므로 예를 들어 100개의 Eigen-value 중 10% 정도의 큰 값을 제외하고는 나머지는 0에 가까운 값을 갖는다. 0에 가까운 Eigen-value는 원래 이미지를 복원하는데 많은 영향을 끼치지 못하기 때문에 이 정도의 오류는 감안하고 처리 시간을 줄이는 것이 더 효과적이다.

$$\hat{X} = b_1 u_1 + b_2 u_2 + .... + b_{N'} u_{N'} = \sum_{j=1}^{N} b_i u_i (b_i = u_i^T X) \quad \text{(식 2-10)}$$

각 이미지의 계수(Coefficient)를 (식 2-10)로 구해서 Training image set의 각 계수들을 데이터베이스에 저장하는 방법으로 얼굴 데이터베이스가 만들어진다. [그림 2-8]은 위의 특징점 추출 과정을 거쳐서 얻어진 얼굴 영역이다.



[그림 2-8] 특징점 추출에 의해 얻어진 얼굴 영역

### 나. Eigenface Recognition Method

Eigenface 인식 방법은 (식 2-11)에 의해서 인식되어지며, 그 인식 과정을 간단히 살펴보면 4단계로 이루어진다.

$$w_i = Au_j \ (A = \overline{X_{\text{평}}} - X_{input}) \qquad \text{(식 2-11)}$$

1) $\Gamma = \overline{X} - X$ 번째 이미지가 들어오면 이미지를 normalize 한다.
2) Training 과정을 거쳐 입력된 이미지 M개의 계수를 구한다.
3) 데이터베이스에 저장된 계수와 거리가 가장 작은 이미지를 찾는다.
4) 일정 거리 안에 들어오면 같은 사람의 이미지를 판단한다.

위와 같은 방법으로 Eigenface를 인식할 수 있으나 얼굴영상을 Eigenface에 적용하기 위해서는 통계학적 특성을 바탕으로 한 주성분분석법(PCA: Principal Component Analysis)을 적용함으로써 인식률을 좀 더 높일 수 있다.

## 2.3.2 기하학적 정보를 이용한 특징점 추출

얼굴의 특징점은 MPGE 4에 정의된 얼굴의 주요 특징점 가운데 얼굴 표정 인식에 필요한 18개의 특징점을 추출하여 사용하며 그 내용은 [표 2-1]과 같다.

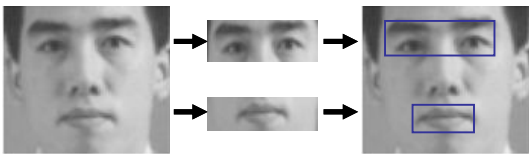
[그림 2-9]는 정확한 얼굴의 특징점을 검출하기 위하여 눈썹, 눈, 입의 얼굴에 대하여 어떤 비율로 분포하는지를 나타낸다. 얼굴 영역을 3등분 했을 때 눈썹과 눈은 2/3 영역에 분포하고, 입은 나머지 1/3 영역에 분포하게 된다. 이와 같이 주어진 특징점의 위치 정보를 이용하여 각 특징점을 검출하기 위한 후보 영역을 설정한다. 각 후보 영역에 대하여 눈썹과 눈이 포함된 템플릿과 입이 포함된 템플릿을 이용하여 템플릿 매칭을 수행한다.

[표 2-1] 얼굴 특징점

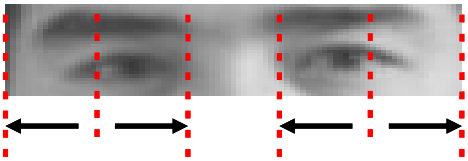
| 구분 | 개수 | 위치(x,y) |
|---|---|---|
| 눈썹 | 3개*2 | $(eyebrow\,area_L, eyebrow\,area_H/2)$ |
| | | $((eyerow\,area_L + eyebrow\,area_H)/2, (eyebrow\,area_H)/2)$ |
| | | $(eyebrow\,area_R, eyebrow\,area_H/2)$ |
| 눈 | 4개*2 | $(eye\,area_L, eye\,area_H/2)$ |
| | | $((eye\,area_L + eye\,area_R)/2, eye\,area_T/2)$ |
| | | $(eye\,area_R, eye\,area_H/2)$ |
| | | $((eye\,area_L + eye\,area_R)/2, eye\,area_B/2)$ |
| 입 | 4개 | $(mouth\,area_L, mouth\,area_H/2)$ |
| | | $((mouth\,area_L + mouth\,area_B)/2, mouth\,area_T/2)$ |
| | | $(mouth\,area_R, mouth\,area_H/2)$ |
| | | $((mouth\,area_L + mouth\,area_B)/2, mouth\,area_B/2)$ |
| 총 | 18개 | |

[그림 2-9] 얼굴 영역의 기하학적 정보에 의한 특징점 분포영역

검출된 후보 영역에 대하여 템플릿 매칭을 수행한 결과는 [그림 2-10]와 같다. 이 때, 검출된 눈썹과 눈 영역 대해서 좌·우측에 존재하는 눈썹과 눈 영역으로 나누어야 한다. 검출된 눈 후보 영역에 대하여 최종 눈 영역을 획득하기 위하여 [그림 2-11]과 같이 눈 후보 영역을 나눈다.



[그림 2-10] 특징점 후보 영역으로부터 눈썹과 눈, 입 영역 검출 결과



(좌측 눈 영역) (우측 눈 영역)

[그림 2-11] 눈 후보 영역으로부터 좌.우측 눈 영역 회득

눈동자 영역에 포함되는 픽셀의 밝기값이 가장 낮기 때문에 (식 2-12)를 이용하여 가장 어두운 화소값을 임계치로 하여 이진 영상으로 변환한다.

$$basis\,\pi xel = MIN(\pi xel \in eye\,area)$$
$$eye\,\pi xel \leq basis\,\pi xel + \sigma \qquad \text{(식 2-12)}$$

또한 사용자가 안경을 사용할 경우 정확한 좌·우측 눈 영역의 검출이 어렵기 때문에 마스크 영상을 이용하여 각 검출된 후보 영역에 대한 AND 연산을 적용해 안경 정보를 제거할 수 있다.

정확한 입 영역의 획득은 캐니에지(Canny edge) 알고리즘을 이용하여 검출된 에지영상의 히스토그램 분석을 수행하며, 입의 후보 영역에 대하여 에지를 검출하게 되면 배경 및 턱 선과 같은 에지 정보도 함께 획득된다. 눈 영역의 검출 과정과 마찬가지로 마스크 영상을 이용하여 입 영역을 제외한 나머지 부분의 에지 정보를 제거하여 입 영역을 획득한다.

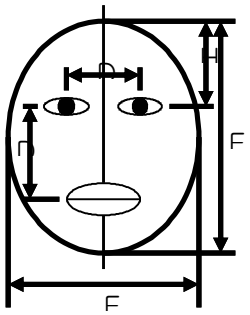
### 2.3.3 평가함수를 이용한 얼굴 특징 추출

평가함수를 이용하는 기존의 연구들은 대부분 얼굴 내에서 눈의 위치와 두 눈의 상대적 위치 및 대칭성에 기반하여 얼굴 특징들을 인식한다[37][38]. Eli Saber는

이러한 정보뿐만 아니라 얼굴 내에서 눈과 입의 상대적 위치관계 및 얼굴에서 눈의 위치를 세밀하게 반영하는 평가함수를 도입하였다. 평가함수는 다음과 같은 얼굴 특징들 간의 관계를 반영한다.

1. 눈과 입 사이의 거리(D1)는 두 눈 사이의 거리(D)의 약 1.2배이다.
2. 입은 중심을 지나며 입술에 수직인 직선에 대해 두 눈은 대칭이다.
3. 얼굴 영역의 높이($F_H$)는 얼굴 영역 폭($F_W$)의 약 1.2배이다.
4. 두 눈간의 거리(D)는 이마 부분의 길이(H)와 비슷하다.

위의 관계를 그림으로 표현하면 [그림 2-12]와 같이 표현 된다.




[그림 2-12] Eli Saber[37]의 얼굴 모델

첫 번째 특성에 대한 평가함수는 (식 2-13)와 같이 표현된다. (식 2-13)에서 D는 두 눈 사이의 거리이며 $W_{LEFT}$(이하 $W_L$)는 왼쪽 눈의 중심점 y 좌표와 입의 중심점 y 좌표간의 차이, $W_{RIGHT}$(이하 $W_R$)는 오른쪽 눈의 중심점 y 좌표와 입의 중심점 y 좌표간의 차이다. 두 번째 특성은 (식 2-14)과 같이 표현된다. (식 2-14)에서 $N_{LEFT}$($N_L$)는 왼쪽 눈의 중심점 x 좌표와 입의 중심점 x 좌표간의 거리이고, $N_{RIGHT}$($N_R$)는 오른쪽 눈의 중심점 x 좌표와 입의 중심점 x 좌표간의 거리이다. 세 번째 특성은 (식 2-15)과 같이 나타내며, $F_W$는 얼굴 영역의 폭이고 $F_H$는 얼굴 영역의 길이이다. 네 번째 특성은 (식 2-16)와 같이 나타내며, H는 얼굴 영역의 가장 큰 y 좌표와 눈의 중심점 y 좌표간의 차 D는 두 눈 사이의 거리이다.

이들을 모두 반영하는 전체 평가함수는 (식 2-17)과 같고 $\omega_i$는 가중치이다. Eli Saber[41]는 입력 영상 내에 한 사람만이 존재한다고 가정하기 때문에 얼굴 영역과

두 눈, 입으로 구성되는 쌍들 모두에 평가함수를 적용하고 가장 작은 평가함수의 값을 갖는 한 쌍의 눈, 입, 얼굴을 최종 인식 결과로 설정되어진다.

$$C_k^1 = Abs[1 - \max(\frac{W_L}{1.2D}, \frac{W_R}{1.2D})] \quad \text{(식 2-13)}$$

$$C_k^2 = Abs[1 - \frac{\min(N_L, N_R)}{\max(N_L, N_R)}] \quad \text{(식 2-14)}$$

$$C_k^3 = Abs[1 - \frac{1.2F_W}{F_H}] \quad \text{(식 2-15)}$$

$$C_k^4 = Abs[1 - \frac{H}{D}] \quad \text{(식 2-16)}$$

$$C_k = \sum_{i=1}^{4} \omega_i C_k^i \quad where, \sum_{i=1}^{4} \omega_i = 1.0 \quad \text{(식 2-17)}$$

## 2.4 Haar-like 특징을 이용한 얼굴검출

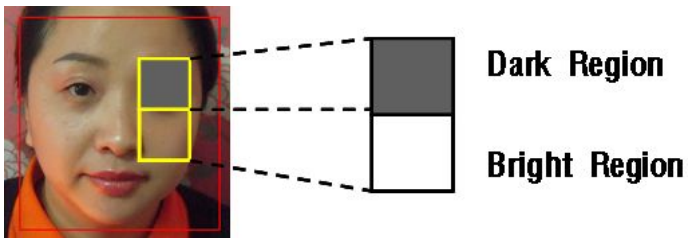
기존의 얼굴 검출 시스템은 상당한 계산 시간을 요구하므로 카메라로부터 획득한 영상 이미지를 실시간 시스템에 적용하기에는 부적합하다. 따라서 본 논문에서는 빠른 연산 능력을 가진 Haar-like Feature를 이용하여 얼굴특징을 생성하고 캐스케이드 알고리즘을 이용하여 얼굴 후보 영역을 검출하였다.

Haar-like Feature는 Viola[27]가 얼굴 검출에 적용한 것으로 단순 합 이미지를 이용하여 특징값을 표현하는 방법이다. 이는 위치, 모양 크기에 따라서 수많은 형태를 가질 수 있다. 이를 AdaBoost 학습 알고리즘을 사용하여 약한 인식자들을 합쳐서 강인한 인식자의 그룹으로 만든다. 이때 생성된 특징값은 얼굴의 특징을 잘

포용하는 장점을 가진다. 계산 방식이 단순 합이기 때문에 동영상에서의 적용이 매우 용이하며 평균 0.5초안에 얼굴 이미지 데이터를 추출하는 것이 가능하다.

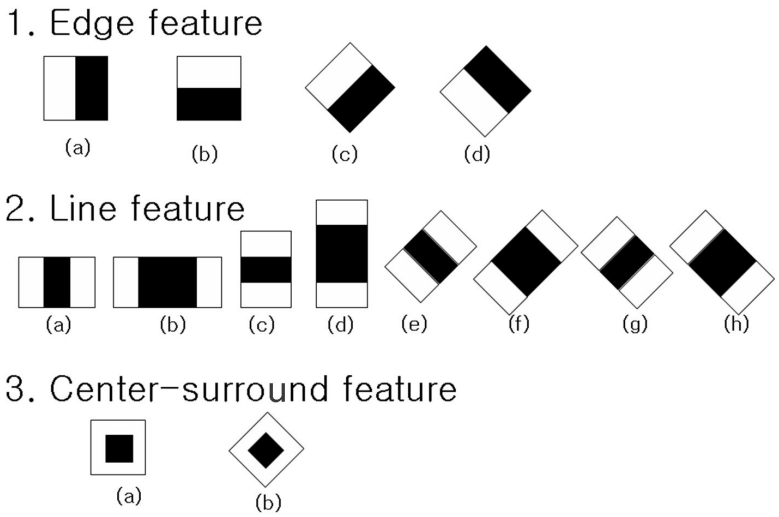
### 2.4.1 Haar-like Feature와 인테그랄 이미지

얼굴 검출에서 사용한 Haar-like Feature는 가장 간단하면서도 효율적인 인식자로 쓰인다. 이를 얼굴 검출에서 사용하면 얼굴을 배경으로 판단할 확률이 1%보다 작아진다. 뿐만 아니라 특징값을 구하는데 요구되는 계산량은 단순히 합 연산밖에 없기 때문에 빠른 얼굴 검출을 요하는 응용분야에서 사용될 수 있다. 그러므로 카메라로부터 입력되는 얼굴 영상은 다양한 외부환경 요소에 영향을 많이 받는다. 이와 같은 환경 요소에 강인하면서 빠르게 연산하여 얼굴을 검출해야한다. 이를 위해 [그림 2-13]과 같이 얼굴 영역 내에서 간단하고 객체간의 구별 능력이 있는 특징을 사용하여 해결할 수 있다.




[그림 2-13] 얼굴 영역 내의 특징 예

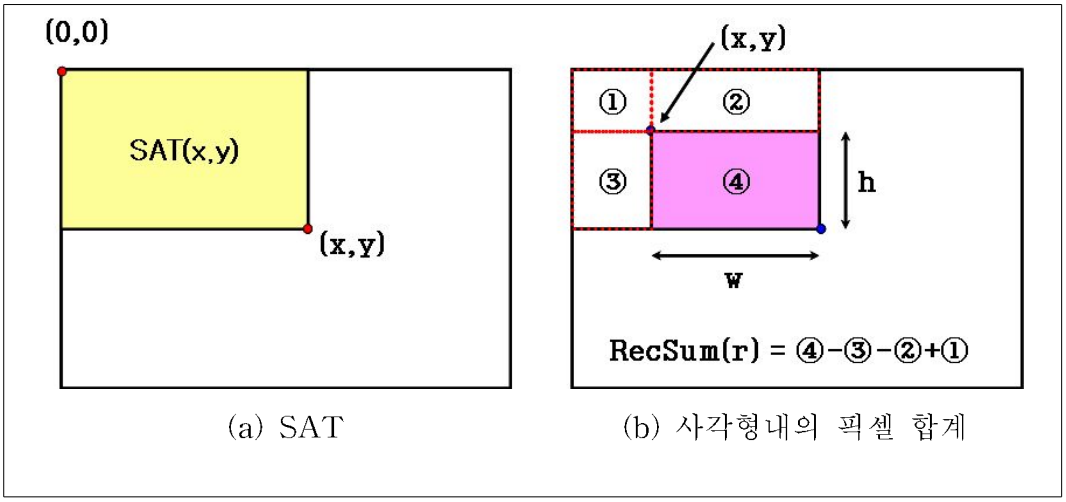
본 논문에서는 픽셀 자체를 이용하는 것보다 영상이 가지고 있는 특징을 더 잘 압축할 수 있는 Haar-like Feature를 얼굴을 인식하는 인식자로 사용하였다. 이는 학습 알고리즘에서 학습 입력 정보를 줄여주고 반면에 학습 결과의 다양성을 증가시켜주면서 더욱 쉬운 분류를 가능하게 한다. 또한 픽셀 자체를 가지고 연산을 수행하는 것보다 특징값을 가지고 수행하는 것이 더 효율적이다.




[그림 2-14] Haar-like Feaute의 프로토타입

이는 개별 영역 안에 있는 픽셀들의 값을 더하여 영역 합을 구하고 그 값들에 가중치(Weight)를 곱하여 합을 계산함으로써 하나의 특징값을 구할 수 있는 Haar-like Feature의 특징에 기인한다. 이러한 Haar-like Feature와 인테그랄 이미지의 사용은 복잡한 연산을 줄이고, 높은 검출 성능이 요구되는 실시간 객체 검출에 효과적으로 적용된다. [그림2-14]는 Haar-like Feature의 프로토타입의 예를 나타낸다. [그림 2-14]의 Haar-like Feature 프로토타입들의 이미지내의 얼굴을 검출하기 위한 효과적인 특징으로 사용된다. Haar-like 특징 계수를 이용해서 얼굴 영역을 검출하면, 유동적인 얼굴의 특성에도 강인한 검출 성능을 보여준다. 이미지내에서 Haar-like 사각형의 픽셀의 합 계산을 빠르게 수행하기 위해 인테그랄 이미지를 사용하였다. 인테그랄 이미지는 영상처리에서 사용하는 영역 합 테이블(Summed Area Table : SAT)[6]과 유사하다. SAT는 중복 연산을 최소화하여 연산 속도를 빠르게 한다.

[그림 2-15]의 (a)의 SAT는 입력영상의 원점(0,0)에서부터 좌표 영역 $(x,y)$를 포함하는 사각형 영역의 모든 픽셀값의 합이며, (식 3-1),(식 3-2)와 같이 계산된다. [그림 2-15]의 (b)에서 좌측 상단의 좌표가 $(x,y)$이고 너비와 높이가 각각 $w$와 $h$이고 회전각도가 $\alpha$=0인 사각형 영역 $r=(x,y,w,h,0)$에 대한 RecSum$(r)$의 값은 [그림 2-15] (b)와 같이 (식 3-3)에 의해 계산된다.



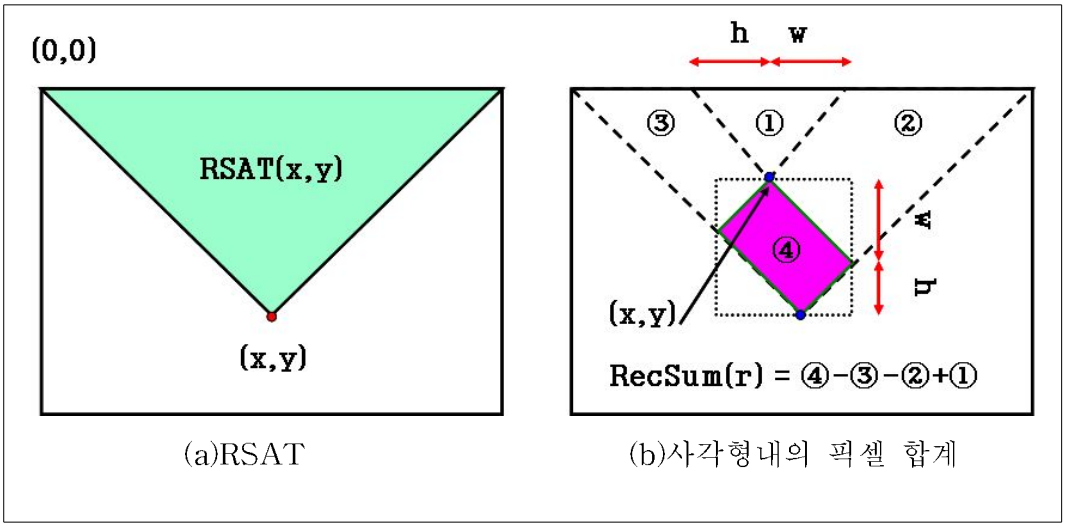

[그림 2-15] 사각형의 인테그랄 이미지

$$SAT(x,y) = \sum_{x' < x, y' < y} I(x',y') \qquad \text{(식 2-17)}$$

$$SAT(x,y) = SAT(x,y-1) + SAT(x-1,y) + I(x,y) - SAT(x-1,y-1) \qquad \text{(식 2-18)}$$

$$RecSum(r) = SAT(x-1,y-1) + SAT(x+w-1,y+h-1) - SAT(x-1,y+h-1) - SAT(x+w-1,y-1) \qquad \text{(식 2-19)}$$

이미지 내에서의 45° 회전한 사각형의 RSAT는 사각형 영역의 모든 픽셀값의 합이며, (식 2-20),(식2-21)과 같이 계산되고, [그림 2-16]의 (a)의 RSAT를 이용하여 특정 위치에 존재하는 사각형 영역 $r=(x,y,w,h,45)$의 합계 RecSum$(r)$은 [그림 2-16]의 (b)와 (식 2-22)에 의해 계산된다.




(a)RSAT (b)사각형내의 픽셀 합계

[그림 2-16] 45° 회전 사각형의 인테그랄 이미지

$$RSAT(x,y) = \sum_{y' < y, y' \le y - |x - x'|} I(x', y') \quad \text{(식 2-20)}$$

$$RSAT(x,y) = RSAT(x-1,y-1) + RSAT(x+1,y-1) - RSAT(x,y-2) + I(x,y) + I(x,y-1) \quad \text{(식 2-21)}$$

$$RecSum(r) = RSAT(x-h+w,y+w+h-1) + RSAT(x,y-1) - RSAT(x-h,y+h-1) - RSAT(x+w,y+w-1) \quad \text{(식 2-22)}$$

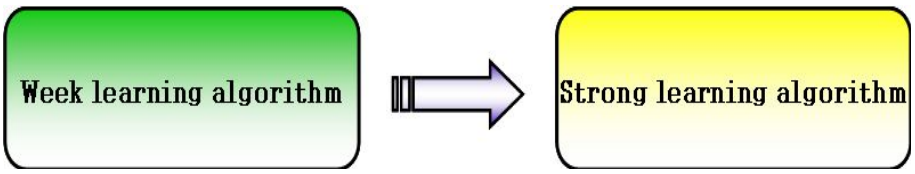
## 2.4.2 AdaBoost 알고리즘을 이용한 얼굴 후보 영역 검출

Haar-like Feature 추출 단계에서 생성된 얼굴 특징 계수를 이용하여 얼굴 후보 영역을 분류하기 위해서 본 논문에서는 Adaptive Boosting(AdaBoost)알고리즘과 캐스케이드(Cascade) 구조를 이용하였다.

### 가. AdaBoost 학습 알고리즘

Boost[29]란 [그림 2-17]에서 보이는 것과 같이 데이터들로부터 결과를 도출할 경우 약한 선택 기준들을 합쳐서 정확도가 높은 선택 기준을 만들어 주는 알고리

즘이다.




[그림 2-17] Boost 알고리즘

AdaBoost 알고리즘은 t 개의 약한 분류기(weak classifier)의 선형적인 결합을 통하여 최종적으로 높은 검출 성능을 가지는 강한 분류기(strong classifier)를 생성하여 분류하는 기법이다. Freund와 Schapire[31]에 의해 소개되었으며 에러율은 약한 분류기의 수가 증가할수록 영점에 지속적으로 근접함이 증명되었다[28-30].

AdaBoost 알고리즘은 기본적으로 데이터의 확률 분포를 가지고 학습이 진행된다. 현재 약한 분류기를 학습 시킨 후 올바르게 분류된 데이터에 대해서는 확률 분포를 낮추고 올바르지 않게 분류된 데이터에 대해서는 확률 분포를 높인다. 그러면 다음번 약한 분류기를 만들어 낼 때는 잘못 분류된 데이터에 집중해서 학습이 된 약한 분류기가 탄생하게 된다.

[표 2-2]의 AdaBoost 알고리즘을 이용한 얼굴 검출 방법은 얼굴과 비얼굴 패턴으로 구성된 훈련영상 집합에서 추출된 사각형 특징 집합을 이용하여 학습된다.

[표 2-2] AdaBoost 알고리즘

▷ 입력 : $(x_1,y_1),\ldots,(x_N,y_N), x_i$ : 입력 영상, $y_i=0,1$ 0: 비얼굴, 1: 얼굴

▷ 초기값 : 가중치 $w_{1,i}=\frac{1}{2m},\frac{1}{2l}$ for $y_i=0,1$ $m$ : 비 얼굴 수 $l$ : 얼굴 수

▷ 루프반복 : $t=1,\ldots,T$ ($T$ : 약한 분류기 수)

1. 가중치 정규화

$$w_{t,i}=\frac{w_{t,i}}{\sum_{j=1}^{N} w_{t,j}} \quad (w_{t,i}: \text{확률분포})$$

2. 약한 분류기 $h_j$의 에러 평가

( $h_j(x)$: 1개의 Haar-like 특징 )

$$\epsilon_j=\sum_i w_i|h_j(x_i)-y_i|$$

3. 가장 작은 에러 $\epsilon_t$을 갖는 분류기 $h_t$선택

4. 가중치 갱신

$$w_{t+1,i}=w_{t,i}(\frac{\epsilon_t}{1-\epsilon_t})^{1-e_i}$$

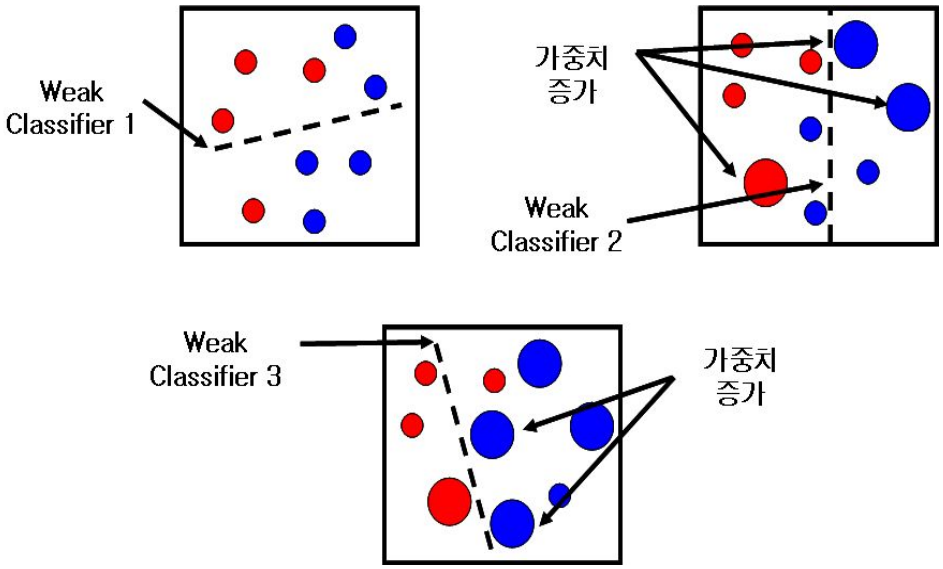
( $x_i$가 정확하게 분류되면 $\epsilon_i=0$이고 나머지는 $\epsilon_i=1$ )

▷ 최종 강한 분류기

$$h(x)=\begin{cases}1, & \sum_{t=1}^{T}\alpha_t h_t(x)\geq \frac{1}{2}\sum_{t=1}^{T}\alpha_t \qquad (\alpha_t=\log\frac{1-\epsilon_t}{\epsilon_t})\\ 0, & \text{나머지}\end{cases}$$

잘못 분류된 훈련 패턴은 가중치 $w_{t,i}$을 증가시키고, 바르게 분류된 훈련 패턴은 $w_{t,i}$을 감소시키는 과정을 반복하면서 최소의 에러율을 나타나는 Haar-like Feature 특징만을 선택한다. 약한 분류기는 1개의 사각형 특징으로 분류함을 의미하며 강한 분류기는 $t$개의 사각형 특징을 선형적 결합을 통해 분류함을 의미한다. 즉, 분류기의 가중치를 수정해 가면서 단계적으로 반복하여 각 단계마다 특징의 선형 집합

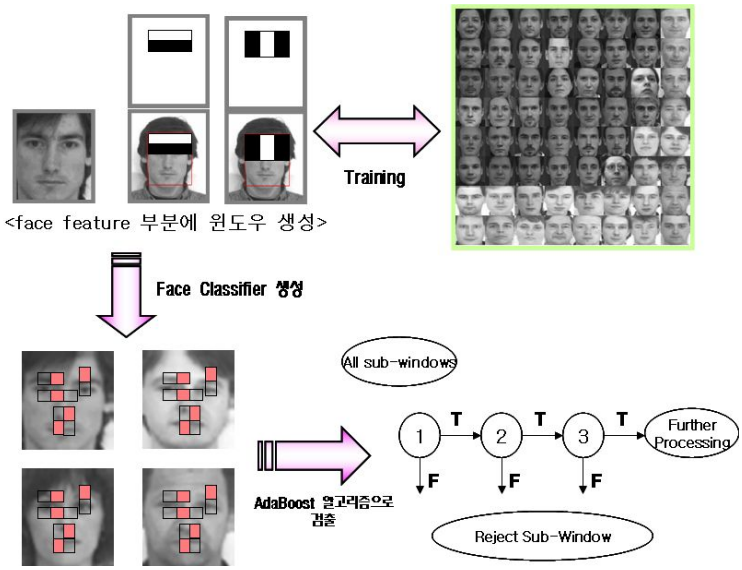
중에서 얼굴을 검출하는데 결정적인 역할을 하는 특징만을 남기고 나머지는 제거하는 방식이다. [그림 2-18]은 가중치 부여의 예를 보여주고 있다.




[그림 2-18] AdaBoost의 알고리즘의 가중치 부여

AdaBoost 학습 알고리즘을 사용하면 학습을 거듭할수록 얼굴을 잘 표현하는 특징값들을 많이 뽑아 낼 수 있어 강력한 검출 알고리즘을 만들 수가 있다.AdaBoost 학습 알고리즘을 통해 뽑은 특징값들은 [그림 2-19]와 같이 단계(Stage)별로 그룹화한다.

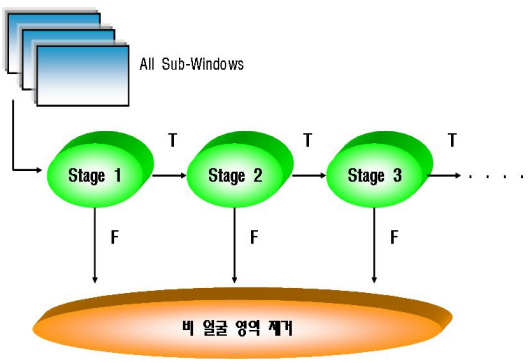
그룹화를 통하면 더 강력한 검출 알고리즘을 구현할 수 있다. 첫 단계에서는 적은 수의 특징값들의 값들을 그룹화하고 다음 단계에서는 첫 단계에서 보다 더 많은 수의 특징값들을 가지고 전 단계와 비슷한 판별력을 가지는 그룹을 만든다. 결국 이렇게 트레이닝된 값을 이용하여 영상에서 얼굴 영역을 찾아내는 것이다.




[그림 2-19] AdaBoost 알고리즘의 단계별 그룹화 개요

## 나. Cascade를 적용한 얼굴 검출

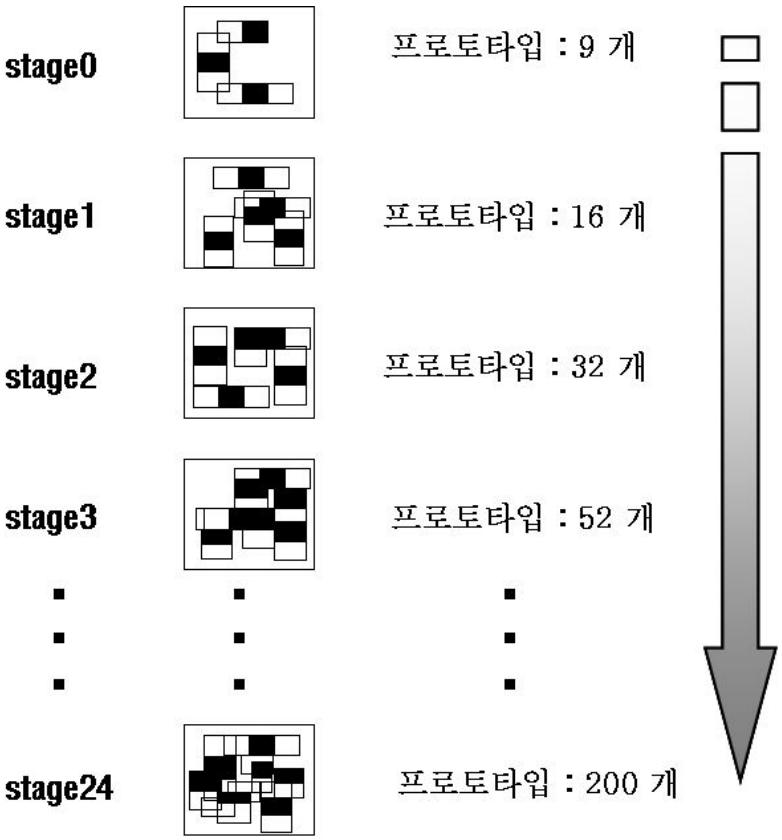
본 논문에서는 얼굴 검출의 연산 속도를 향상시키기 위해 [그림 2-20]과 같이 AdaBoost 분류를 Cascade 구조로 만들어 적용하였다. Cascade 구조는 여러 개의 스테이지를 나누고 스테이지별로 약한 분류기의 수를 다르게 하여 수행하는 방법이다. 처음 단계에서 대부분의 배경(비 얼굴)이 제거되는 기능을 가지고 있어 단일 단계를 적용하는 방법보다 훌륭한 성능을 보이고 있다.




[그림 2-20] 얼굴 검출을 위한 Cascade 구조

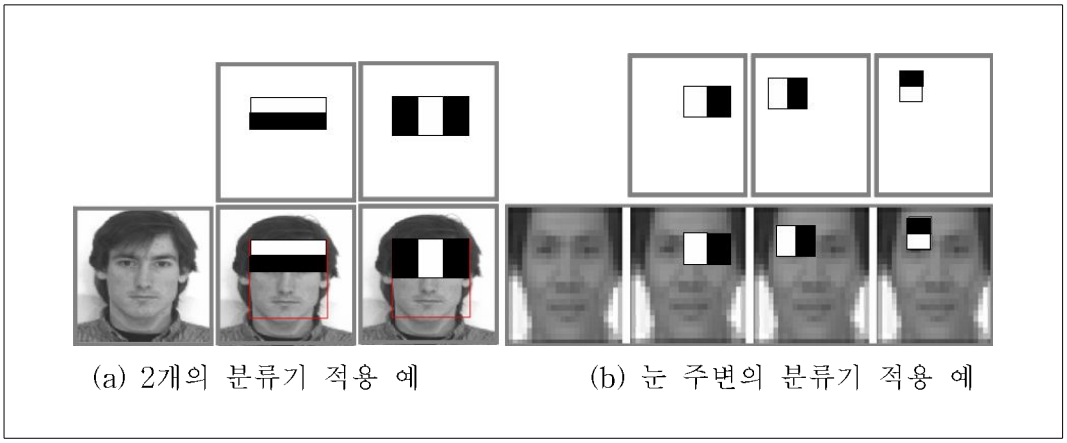
AdaBoost 학습을 위해 총 25 단계를 이용하였다. 서브 윈도우의 크기는 24*24로 정해졌고, [그림 2-21]와 같이 높은 단계로 올라갈수록 Haar-like Feature 개수도 증가하게 되고 각각의 프로토타입들도 세밀한 부분까지 위치하게 된다[23].

프로토타입의 개수의 설정은 실험에 의한 최적의 수를 사용하였다. 즉, 단계가 높을수록 세밀한 얼굴 영역 마스크를 생성하게 된다.




[그림 2-21] 단계별 사용된 프로토타입의 개수

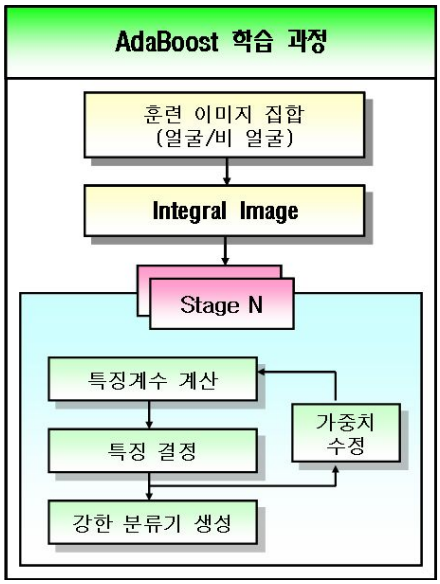
[그림 2-22]의 (a)는 첫 번째 특징(약한 분류기)과 두 번째 특징(약한 분류기)의 예를 표시했다.




[그림 2-22] AdaBoost의 학습에 의해 선택된 특징

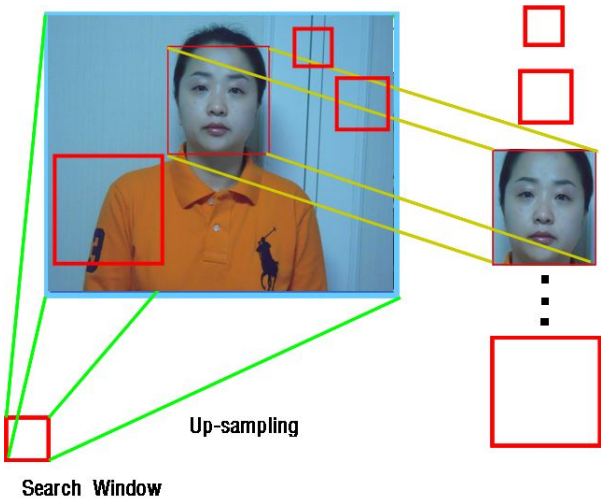
첫 번째 특징은 눈 영역과 뺨 영역의 다른 명도가 측정되어 선택되었고, 두 번째 특징은 두 눈 사이의 다른 명도로 측정되어 선택되었다. [그림 2-22]의 (b)는 실질적인 픽셀을 확대하여 눈 주변의 명도에 따른 특징을 나타내었다.

[그림 2-23]은 얼굴 후보 영역 검출을 위한 Cascade 구조를 적용한 AdaBoost 학습과정을 나타내고 있다.



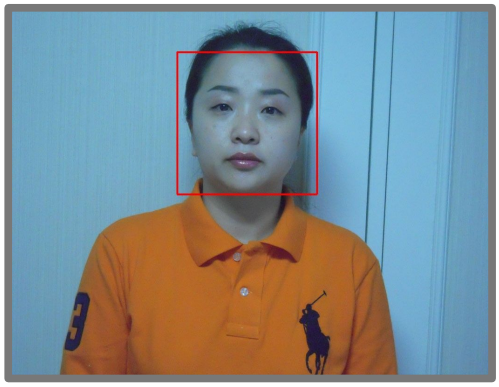

[그림 2-23] 얼굴 후보 영역 검출을 위한 학습과정

얼굴 검출 단계에서 입력 이미지에 대하여 얼굴 검색 윈도우를 일정 크기로 멀티스케일 하였고, 일정 간격으로 입력 영상을 시프트하면서 스캔을 수행하였다. 모든 서브 윈도우에 대해 스테이지 분류기를 통해 얼굴 영역을 분류하였다.
초기 스테이지에서 많은 수의 배경 영상이 제거되고, 다음 스테이지는 이전 스테이지에서 제거되지 않은 배경영상을 제거하여 고려할 영상의 개수를 감소시킨다. 각 스테이지에서 배경영상이 제거되므로 연산량이 대폭 감소되고, [그림 2-24]와 같이 입력 이미지에 대하여 얼굴 영역을 검출하기 위해 검색 윈도우를 멀티스케일하고 시프트하여 모든 서브 윈도우를 스테이지 분류기를 통해 얼굴 영역을 분류하였다.

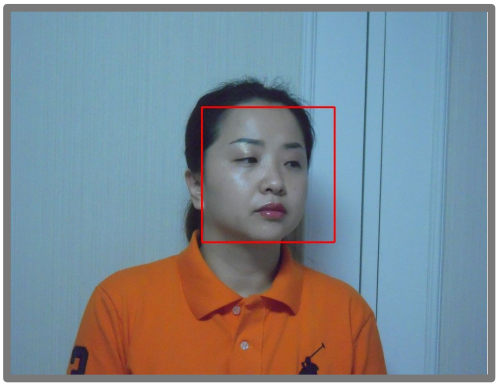



[그림 2-24] 얼굴 검출을 위한 검색 윈도우

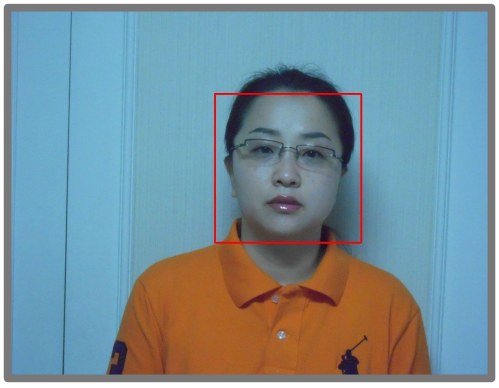
[그림 2-25]은 Web Camera를 통한 실시간 영상을 다양한 조건하에서 얼굴을 검출 하였다. [그림 2-25]의 (a)는 정면 얼굴을 검출한 결과이고, (b)는 동일한 인물이 우측으로 회전한 영상의 검출결과이고, (c),(d)는 동일한 인물에 안경을 추가하여 정면, 회전을 검출한 결과이다. [그림 2-25]의 (e),(f)는 얼굴 영상의 대표적인 착용품인 안경과 모자를 착용 후 검출결과이다. 본 논문에서 제안한 얼굴 검출 알고리즘은 좌우 회전은 45°까지 검출 가능하며 상하 이동은 30°까지 검출이 가능하였다. 또한 얼굴 영상의 모자 및 안경의 착용에도 강인한 얼굴 검출효과를 실험을 통하여 확인하였다. [그림 2-26]는 한 영상에 다중 사람이 입력되었을 때 검출한 결과이다.



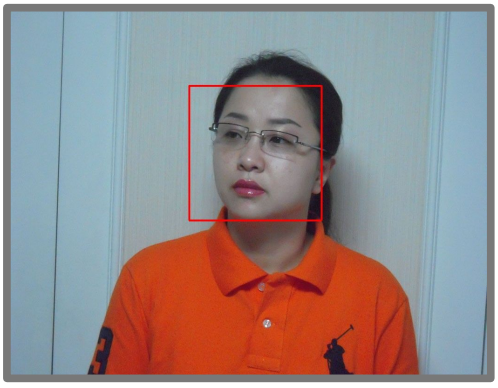
(a) 정면 얼굴 검출



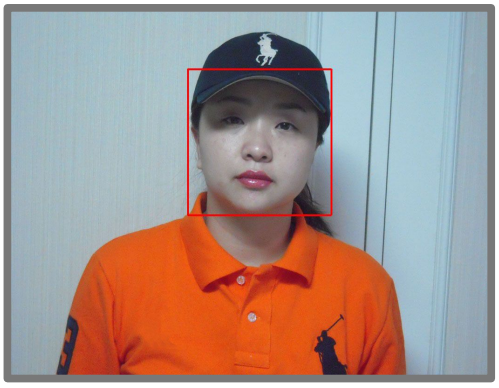
(b) 우측으로 회전 후 얼굴 검출



(c) 안경 착용 후 얼굴검출



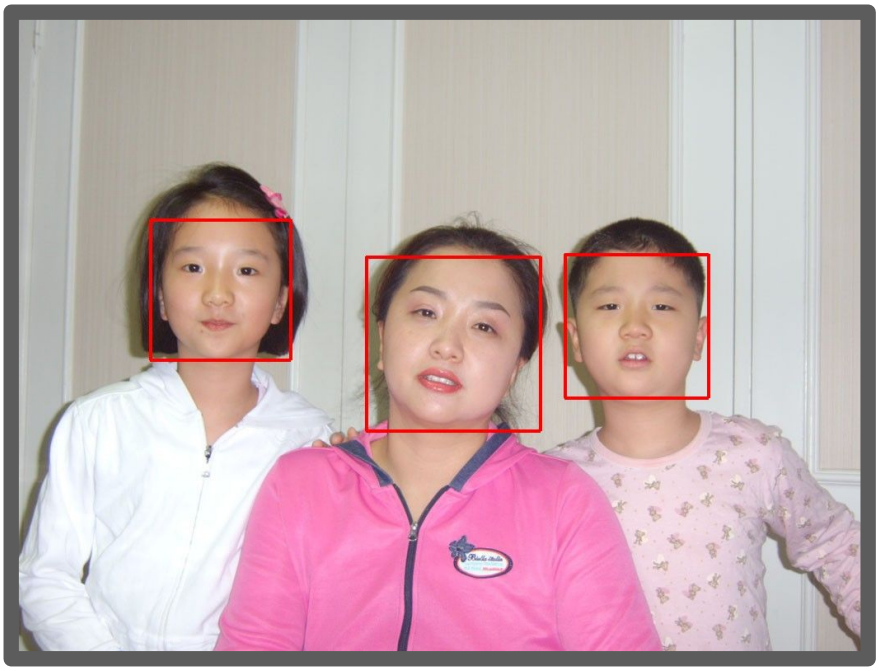
(d) 안경 착용 후 좌측으로 회전 후 얼굴 검출



(e) 안경, 모자 착용 후 얼굴검출

(f) 안경, 모자 착용 후 회전 후 얼굴 검출

[그림 2-25] 다양한 조건하에서 실시간 얼굴 검출

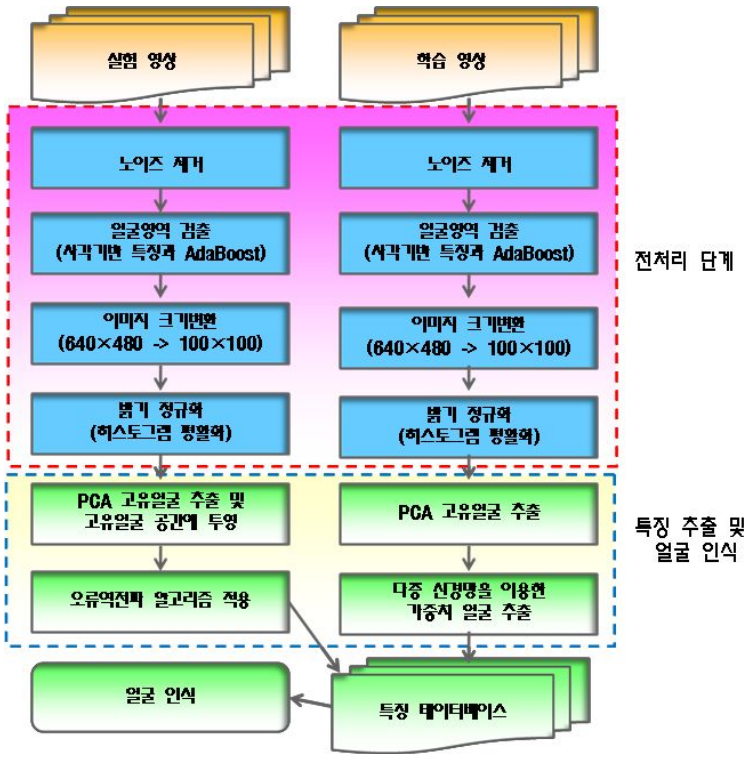


[그림 2-26] 다중 얼굴 검출

# Ⅲ. 제안한 얼굴 인식 시스템

## 3.1 제안한 얼굴 인식 순서도

본 논문에서 제안하는 얼굴 인식 시스템의 알고리즘은 카메라를 통해 얻어진 배경 영상에서 사각형 특징 기법을 이용하여 얼굴만을 추출하고 영상의 크기 정규화 및 노이즈를 제거한 후 PCA를 통해 고유벡터를 산출한 후 다층 신경망을 적용하여 얼굴 인식을 수행한다. [그림 3-1]은 본 논문에서 제안하는 알고리즘의 순서도를 나타낸다.




[그림 3-1] 제안한 알고리즘 순서도

전처리에 해당하는 부분은 실시간 입력 영상에서 AdaBoost를 이용하여 얼굴영역만을 검출한 후 조명의 변화 속에서도 얼굴을 인식하기 위해 영상에서 크기 정규

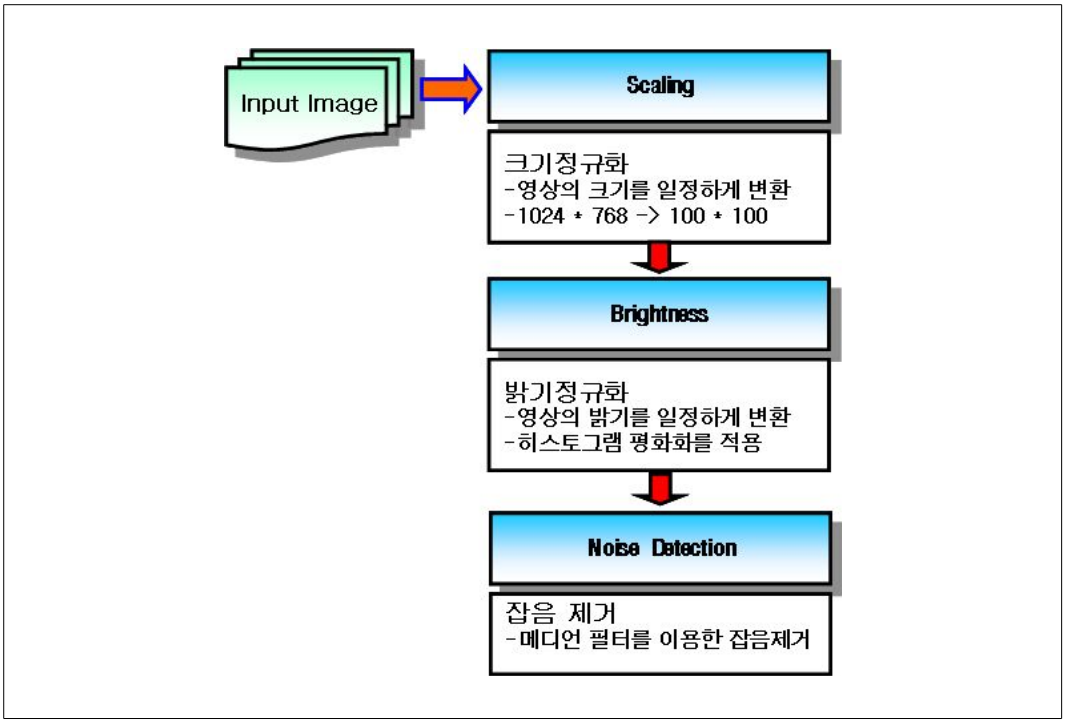
화를 수행 한 후 히스토그램 평활화를 이용하여 조명의 밝기를 정규화해 준다. 그러나 밝기 정규화를 수행했다 하더라도 영상에서는 많은 노이즈들이 얼굴의 형상에 영향을 주기 때문에 메디언 필터를 이용하여 노이즈를 제거한다. 제거된 얼굴영상을 인식하기 위해 주성분 분석을 통하여 고유공간을 구성하는데 많은 시간이 소요된다. 그래서 본 논문에서는 특징벡터를 줄이기 위해 고차 극소 자동 상관계수 방법을 이용하여 얼굴의 특징을 급격하게 축소시켰다. 또한 급격하게 줄여든 특징만을 인식을 수행했을 경우 잘못된 인식이나 인식이 실패한 입력영상의 경우 다시 그레이 영상을 재전송하여 인식률의 안전성을 추구하였다.

인식 부분에서는 추출된 얼굴을 주성분 분석을 이용하여 얼굴의 전체적인 외관 특징을 표현할 수 있는 저차원 벡터 공간, 즉 파라메트릭 얼굴 공간을 생성하여 입력 영상의 얼굴을 얼굴 공간에 투영시켜 인식을 수행한다. 또한 기존의 거리 분석 방법인 유클리디안(Point to Point) 방식이 아닌 개선된 오류역전파 알고리즘을 이용하여 인식이 성공했을지라도 다른 얼굴로 인식하는 경우와 매칭 실패의 오류를 개선하였다.

## 3.2 얼굴 영상의 정규화 과정

얼굴 영상 집합의 정규화는 카메라로부터 연속적으로 얻어진 영상들을 모델구성에 필요한 전처리 과정으로 시각 기반의 얼굴 인식에 필수적인 처리 과정이다. 특히 일반 환경에서 연속적 영상들은 각각 서로 다른 많은 noise들을 포함하기 때문에 정규화 처리에 의해 인식률이 크게 좌우될 수 있다.

정규화 과정은 크게 크기와 밝기 정규화와 영상의 특징을 줄이기 위한 2진화 과정으로 나누어진다. 크기와 밝기 정규화를 거친 얼굴 영상들은 많은 noise을 가지고 있기 때문에 noise을 제거한 후 인식 속도를 개선하기 위한 방법인 고차 국소 자동 상관계수를 이용하여 특징벡터를 생성한다. [그림 3-2]는 전처리에 대한 전체 흐름도이다.



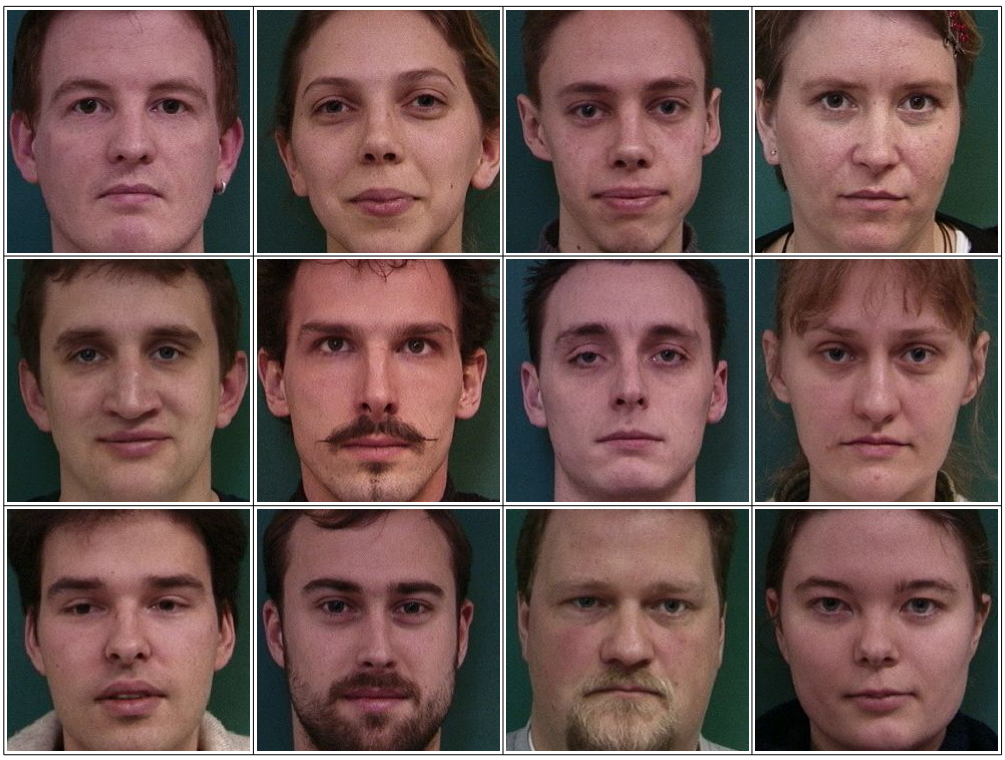

[그림 3-2] 전처리 전체 개략도

### 3.2.1 얼굴 영상의 크기 정규화

이 절에서는 입력 영상의 크기 정규화에 대하여 설명한다. 먼저 영상 카메라로부터 입력영상이 들어오면 목차 2.4단계에서 Haar-like feature와 AdaBoost 알고리즘을 이용하여 얼굴영역만이 검출되어진 [그림 3-3]과 같은 영상을 얻는다. 이러한 영상의 집합을 $\hat{X}$라 하고 (식 3-1)과 같이 표현한다.

$$\hat{X} = [\hat{X}_1, \hat{X}_2, ....., \hat{X}_N]^T \quad (3\text{-}1)$$

여기서 $N$ 은 얼굴 영상의 개수이다.

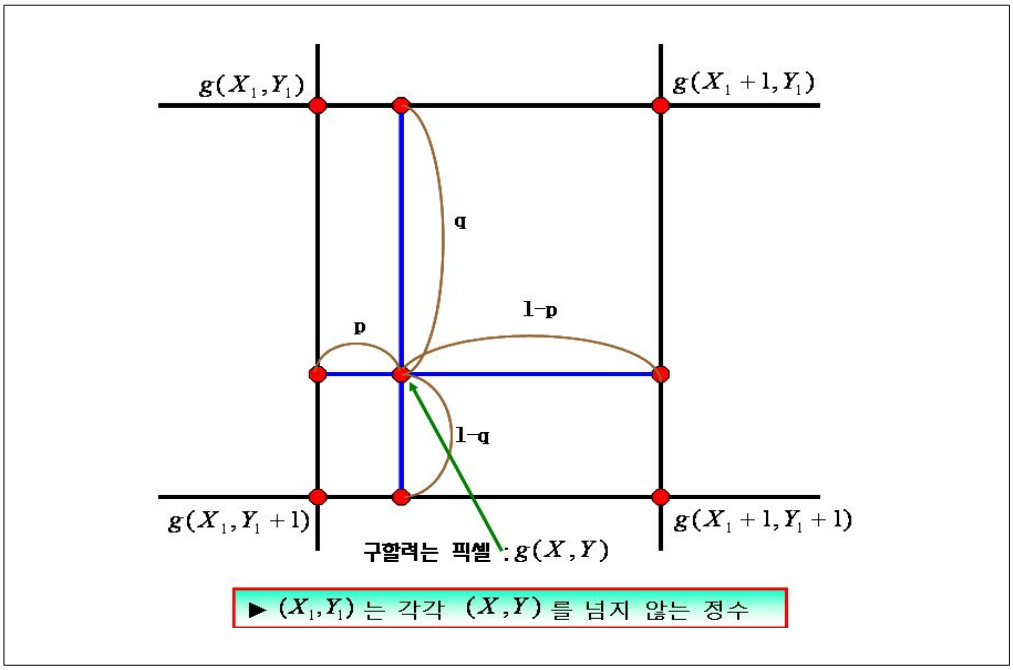


[그림 3-3] Haar-like Feature를 통해 검출된 얼굴 영상 집합

본 논문에서는 1024 * 768의 입력영상으로부터 얼굴영역만 검출된 영상은 항상

일정치가 않기 때문에 영상을 일정한 크기로 정규화하고 또한 계산량을 줄이기 위해 영상의 크기를 100 * 100으로 정규화시켰다.

영상의 크기를 변환하는 방법은 여러 가지가 있으나 여기에서는 확대/축소에 많이 사용하는 선형 보간법을 이용하여 영상을 축소하였다. 이 방법은 영상을 확대하거나 축소하는 방법으로 얼굴의 모양은 변경하지 않고 다만 크기만이 변경되는 닮은 꼴 변환이 된다. 즉 $(x,y)$위치의 화소가 원점을 중심으로 가로 방향으로 $a$배만큼, 세로 방향으로 $b$만큼 확대/축소된다면 새로운 좌표 $(X,Y)$는 $\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} ax \\ by \end{bmatrix}$로 표현된다. 여기서 $a$와 $b$가 1보다 크면 확대가 되고 1보다 작으면 축소가 된다. 만약 확대나 축소로 인한 새로운 좌표 $(X,Y)$가 정수이면 다행이나 그렇지 않다면 2차원 선형 보간법을 사용한다.



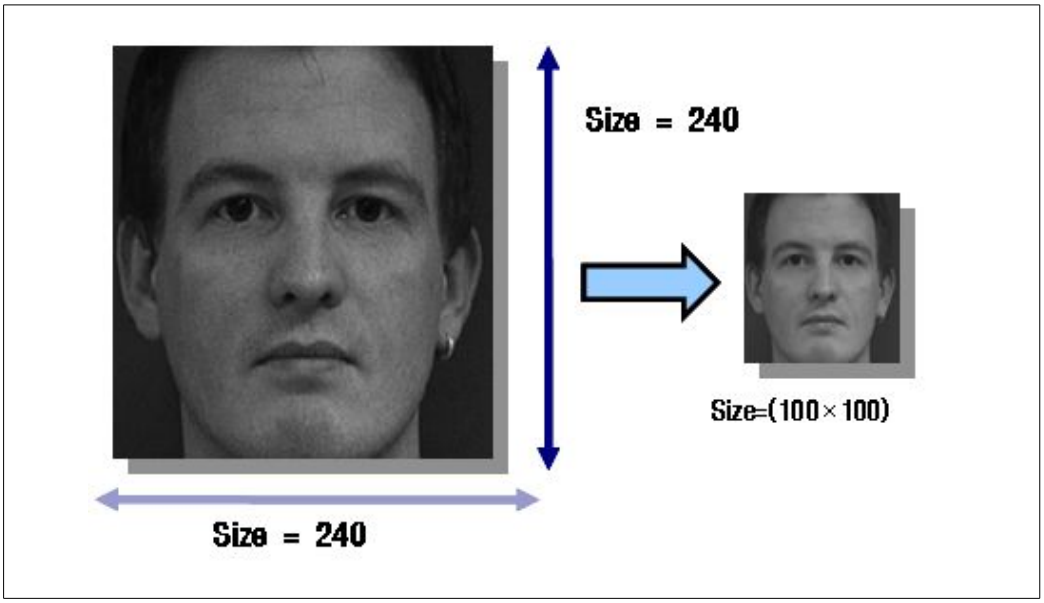

[그림 3-4] 선형 보간법에 의한 영상 축소 방법

이 방법은 [그림 3-4]에서 나타낸 것처럼 구하려는 어드레스가 격자점과 일치하지 않는 경우, 근방 4개의 격자점으로부터 비율을 구하고, 이 비율로 근방 4픽셀의 농도값으로 보간하는 방법이다. 이 농도값의 계산을 (식 3-2)로 나타내었다. [그림

3-5]는 영상의 크기 변환 방법을 나타낸 것이다. [그림 3-5]는 선형 보간법에 의해 축소된 영상을 나타내었다.

$$g(X,Y)=pqg(X_1,Y_1)+p(1-q)g(X_1,Y_1+1)+(1-p)qg(X_1+1,Y_1) \\ +(1-p)(1-q)g(X_1+1,Y_1+1) \quad \text{(식 3-2)}$$

여기서 $X_1$, $Y_1$는 각각 $(X,Y)$의 왼쪽 및 위 화소의 좌표를 의미하며, $p$와 $q$는 그 화소와 $(X,Y)$화소와의 거리를 의미한다.




[그림 3-5] 원 영상과 축소된 영상

### 3.2.2 얼굴 영상의 밝기 정규화

특별한 조명을 사용하지 않았기 때문에 각 입력 영상은 조명의 밝기에 따라 인식에 큰 차이를 나타낸다. 이를 해결하기 위한 방법으로 본 논문에서는 히스토그램 평활화를 이용하여 밝기 변화에 강한 얼굴 인식을 수행한다.

#### 가. 히스토그램 평활화

휘도치의 분포가 잘 펼쳐지지 않은 영상에서 히스토그램 평활화 처리는 영상을 분석하는데 많은 도움을 준다. 결과 영상은 새롭고 균등한 히스토그램 분포를 가지게 된다.

히스토그램 평활화는 히스토그램을 평평하게 하는 것이 아니고 휘도치 분포를 새롭게 하는 것이다. 만약 히스토그램에 극대점과 극소점이 있다면 이것을 실행한 후에도 여전히 그 극대점과 극소점을 유지시킨다. 단지 그 극대점과 극소점은 이동될 뿐이다. 즉 히스토그램 평활화는 출력 영상의 히스토그램 분포가 균등하게 되도록 처리하는 것으로, 영상의 대비를 향상시켜 영상을 개선시킨다. 먼저 수식을 예로 들어 간단히 설명하면, 구간 [0,1]에서 연속적인 그레이 레벨 $x$에 대한 변환 $y = T(x)$를 정의하고 변환 $y$가 균일한 분포를 갖도록 하는 것이 히스토그램 평활화 기법이다. 즉, $y$에 대한 확률 밀도 함수가 $p_y(y) = c$($c$는 상수)를 성립되도록 히스토그램을 처리하는 것이다.

변환함수 $T(x)$에 대해서는 다음과 같은 조건을 만족한다고 가정한다.

(1) $T(x)$은 구간 [0,1] 내에서 단조증가하며 화소는 $x$에 대해, 변환된 $y$는 일대일로 대응한다.
(2) $0 \leq x \leq 1$에 대해 $0 \leq T(x) \leq 1$ 이다.

여기서 $x$에 대한 확률 밀도 함수인 $p_x(x)$와 $T(x)$가 정의되고 역변환 함수인 $T^{-1}(y)$가 $y$에 대한 위의 조건 (1)을 만족한다면 $y$에 대한 확률 밀도 함수인

$p_y(y)$는 (식 3-3)과 같이 표현된다.

$$p_y(y) = [p_x(x)\frac{dx}{dy}]_{x = T^{-1}(y)} \quad \text{(식 3-3)}$$

히스토그램 평활화를 위해서 $y$에 대한 확률 밀도함수가 구간 [0,1]사이에서 균등한 확률 밀도함수 $p_y(y) = 1$을 갖는다면 (식 3-4)로부터 (식 3-5)와 같은 누적 분포함수(Cumulative Distribution Function, CDF)를 변환함수로 갖게 된다.

$$y = T(x) = \int_0^x p_x(t)dt \quad \text{(식 3-4)}$$

여기서 t는 적분을 위한 가변수(dummy variable)이다.

(식 3-4)를 디지털 영상처리에 사용하기 위해 이산 함수(discrete function)형태로 바꾸면 다음 (식 3-5)와 같다.

$$y_k = T(x_k) = \sum_{j=0}^{k} p_k(x_j) = \sum_{j=0}^{k} \frac{n_j}{n} \quad \text{(식 3-5)}$$

**[예]** 히스토그램 평활화 기법을 처리하는 과정을 단계별로 설명한다.

그레이 레벨[0 , 7]을 갖는 8 * 8 영상에서 전체 64개의 화소 중 각각의 그레이 레벨의 빈도수가 [표 3-1]과 같다고 가정하자.

단계 1 : 먼저 $x$에 대한 히스토그램의 값을 누적시킨다. 누적된 결과는 각각 10, 10+8=18, 18+9=27, 27+1=28, 28+12=40, 40+16=56, 56+4=60, 60+4=64이다.

단계 2 : 전체 화소의 개수로 누적된 결과를 나누어 $x$에 대한 정규화된 누적 확

률분포를 계산한다. 정규화된 값은 10/64, 16/64, 27/64, 28/64, 40/64, 56/64, 60/64, 64/64이다.

단계 3 : 최대 그레이 레벨 7을 단계 2에서 계산된 값과 곱하고 그 결과를 반올림한다. 수행된 결과로부터 1, 2, 3, 3, 4, 6, 7, 7을 구한다.

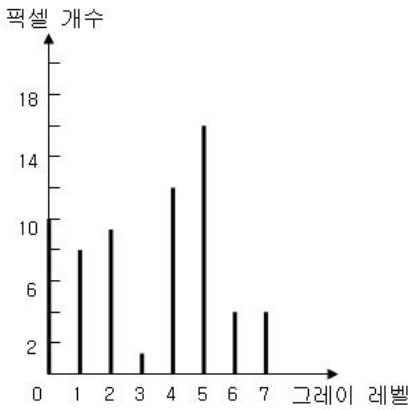
단계 4 : 단계 3에서 계산된 값을 원영상의 그레이 레벨에 대한 변환값으로 1:1 대응시킨다.

[표 3-1] 그레이 레벨 빈도수의 예

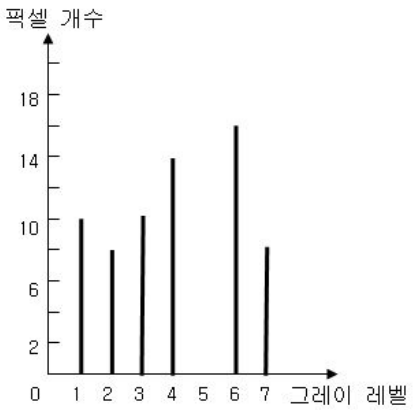
| 그레이 레벨 | 화소의 개수 (히스토그램의 값) |
|---|---|
| 0 | 10 |
| 1 | 8 |
| 2 | 9 |
| 3 | 1 |
| 4 | 12 |
| 5 | 16 |
| 6 | 4 |
| 7 | 4 |

[표 3-2] 평활화된 그레이 레벨의 예

| 그레이 레벨 | 평활화된 히스토그램의 화소값 |
|---|---|
| 0 | 1 |
| 1 | 2 |
| 2 | 3 |
| 3 | 3 |
| 4 | 4 |
| 5 | 6 |
| 6 | 7 |
| 7 | 7 |

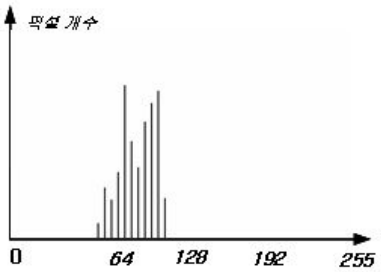



(a) 원영상의 히스토그램

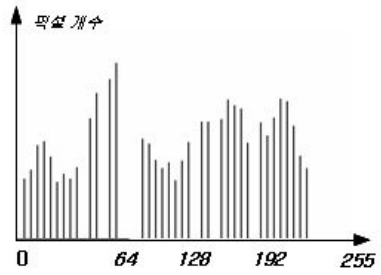



(b) 균등화된 영상의 히스토그램

[그림 3-6] 히스토그램 데이터를 평활화




(a) 저 대비 영상 (b) 저 대비 영상의 히스토그램



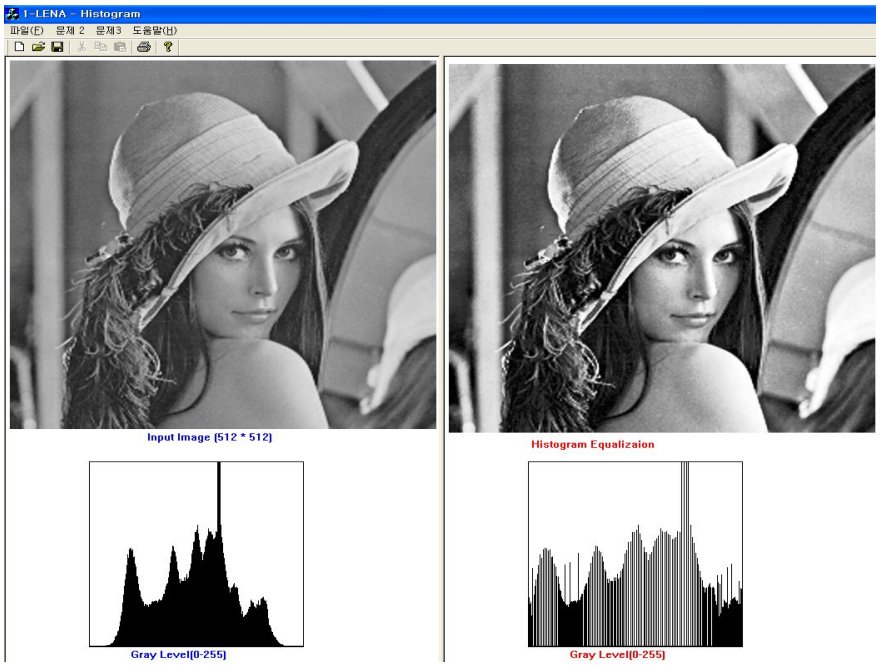

(c) 평활화된 영상 (d) 평활화된 영상의 히스토그램

[그림 3-7] 히스토그램 평활화

히스토그램 평활화를 실제 영상에 적용해 보면 먼저 [그림 3-7]의 (a)를 보면 전

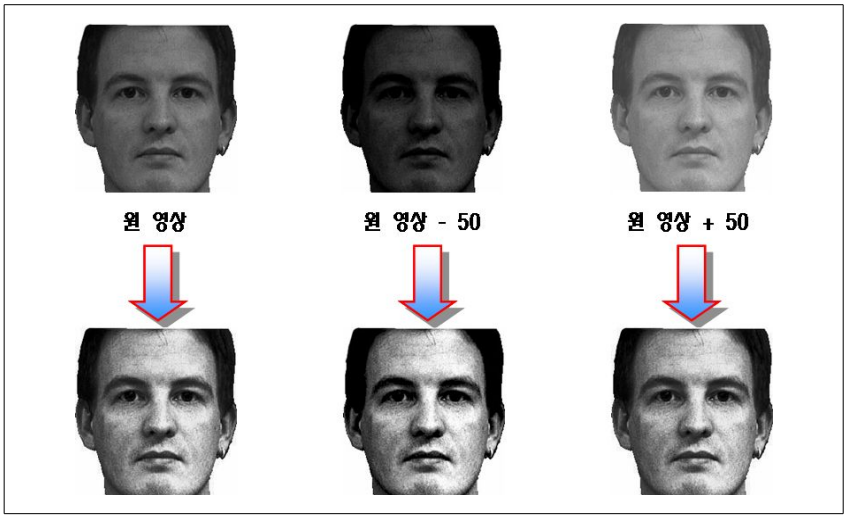
체적으로 어두운 저 대비 영상임을 알 수 있다. [그림 3-7]의 (b)는 (a) 영상의 히스토그램으로 상대적으로 좁고 왼쪽 어두운 방향으로 치우쳐 있음을 알 수 있다. [그림 3-7]의 (c)는 히스토그램 평활화의 결과를 나타내며, 원 영상과 비교해 상당히 개선되어져 있으며, [그림 3-7]의 (d)는 [그림 3-7]의 (c) 영상의 히스토그램으로 평활화된 모양을 나타낸다. 이 그림에서 알 수 있는 것처럼 히스토그램 평활화 처리는 그레이 레벨의 분포를 고르게 하여 결과적으로 영상의 대비를 증가시킨다.

[그림 3-8]은 일반 영상에서 평활화를 적용한 결과 영상이 전체적으로 데이터가 펼쳐짐으로서 영상의 화질이 개선됨을 확인할 수 있다.




[그림 3-8] 일반 영상에 히스토그램 평활화를 적용한 영상

[그림 3-9]는 본 논문에서 제안한 얼굴 인식 방법에 적용하여 밝기의 변화에서도 일정한 인식을 행할 수 있다는 것을 보여준다.




[그림 3-9] 조명의 세기가 다른 각각의 영상을 히스토그램 평활화 후의 영상

얼굴 인식에서 히스토그램 평활화를 사용하는 이유는 이것을 실행하면 영상의 히스토그램이 그냥 펼침뿐만 아니라 영상의 대조(contrast)를 높여준다. 그래서 영상의 조명이 밝거나 어두우면 영상은 영상의 휘도치 분포는 한곳에 모일 것이다. 그 경우 히스토그램 평활화를 사용함으로써 얼굴의 특징점 찾기에 도움을 주게 된다.

### 3.2.3 얼굴 영상의 noise 제거

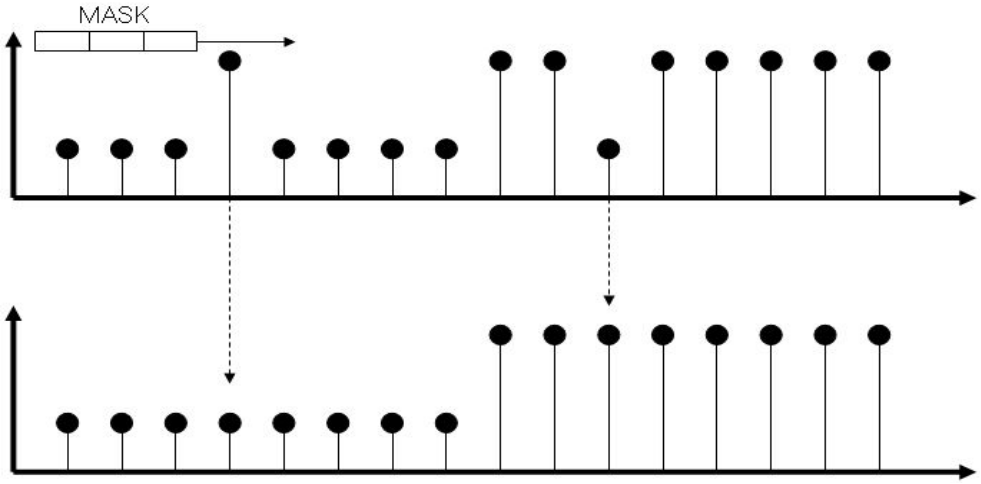
이 절에서는 밝기가 정규화된 영상에서도 noise가 발생하기 때문에 noise를 감소하기 위해 메디언 필터 사용 방법 대하여 기술을 한다.

#### 가. 메디언 필터(Median filter)

메디언 필터는 윈도우에 의해 씌워진 화소값들을 크기 순서로 정렬한 후 중간 위치에 해당하는 값 즉 메디언값을 선택하는 필터이다. 예를 들어, 먼저 윈도우 내에 $n$개의 화소값이 존재하고 이를 크기 순서로 정렬한 값을 $x_i$, $i=1,....n$로 하자. 즉 $x_1$이 $n$개의 화소값 중 최대의 값이고 $x_n$은 최소의 값이다. 이 $n$개의 화소들의 메디언값은 (식 3-6)으로 표현될 수 있다.

$$med(x_i) = \begin{cases} x_{v+1}, & n = 2v+1 \\ \frac{1}{2}(x_v + x_{v+1}), & n = 2v \end{cases} \qquad \text{(식 3-6)}$$

메디언 필터는 평균필터와 달리 영상의 경계(edge)부분은 잘 보존하면서 임펄스 잡음(impulse noise)을 제거하는데 효과적이다. 왜냐하면 임펄스 잡음은 주변 값들과 차이가 많이 나기 때문에 수열의 중간값으로서 선택될 확률이 거의 없기 때문이다. 이런 성질 때문에 메디언 필터는 영상에서 주변 값들과 상관 관련이 적은 값들을 제거하는데 효과적으로 사용된다. [그림 3-10]은 윈도우의 크기가 3인 일차원 메디언 필터를 사용한 예로 영상의 경계를 보존하면서 임펄스 잡음을 제거함을 알 수 있다. [그림 3-11]은 임펄스 잡음이 포함된 영상에 메디안 필터를 적용하여 잡음을 제거한 영상을 보여준다.



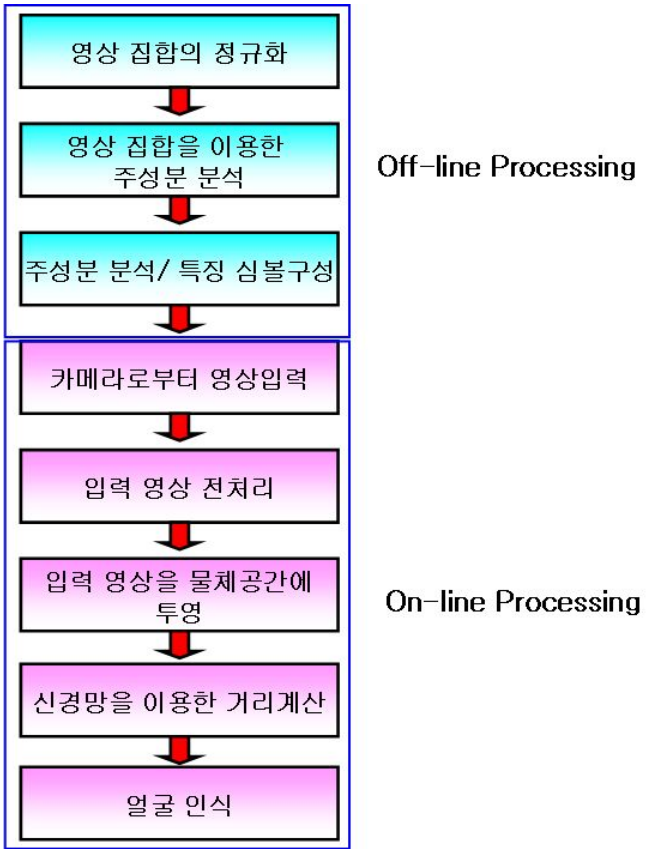

[그림 3-10] 마스크의 크기가 3인 1차원 메디언 필터의 적용 예

[그림 3-11] 임펄스 잡음이 포함된 영상에 메디언 필터를 적용한 영상

본 논문에서 히스토그램 평활화를 수행한 후에 약간의 임펄스 잡음이 포함되어 있기에 이를 해결하기 위해 [그림 3-11]과 같이 실험에 가장 효과적인 방법으로 밝혀진 메디언 필터를 이용하여 noise을 제거하였다.

## 3.3 정규화 영상을 이용한 얼굴 영상 공간 구성

앞 절에서 설명한 방법을 통하여 얻어진 정규화된 영상들과 특징벡터를 이용하여 얼굴의 전체적인 외관 특징을 표현할 수 있는 저차원 벡터 공간인 파라메트릭 얼굴 공간을 생성한다. 이 공간은 주성분 분석법이라는 통계적 방법에 의해 만들어진다. [그림 3-12]는 얼굴 영상 공간에서 전체 매칭과정을 나타낸 것이다.




[그림 3-12] 얼굴 영상 공간을 이용한 전체 매칭 과정

### 3.3.1 얼굴 영상 구성

Mono 카메라를 이용하여 얼굴의 방향이 틀어지는 경우를 같은 형상으로 인식하기에는 많은 어려움이 있다.

[그림 3-13] 동양 사람의 방향이 틀어지는 경우 영상

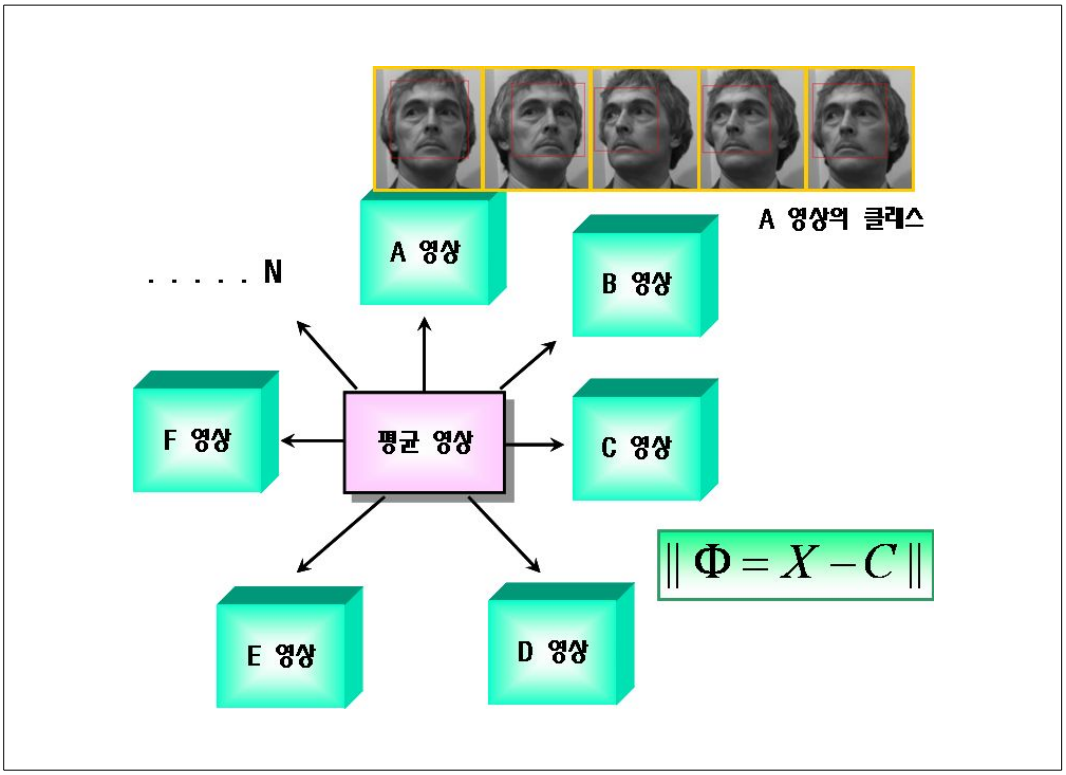
따라서 본 논문에서는 [그림 3-13]과 같이 얼굴의 방향성이 틀어지는 경우 같은 얼굴 영상으로 인식하기 위해 주성분 분석을 이용한 얼굴 방향성 문제의 해결 방법을 제안한다. 이 문제를 해결하기 위해서는 한 사람의 영상을 5°씩 회전을 해서 얼굴 영상을 습득하여 데이터베이스를 구축해야 한다. 그러나 주성분 분석에서는 단지 영상의 외관기반으로 인식을 수행하기 때문에 사람의 머리 모양의 변화만으로도 같은 사람으로 인식하기가 어렵다. 그래서 본 논문에서는 앞의 2.4장에서 설명한 Haar-like Feature을 이용하여 얼굴 영역이 추출 가능한 영상만을 모델 영상으로 사용하였다. [그림 3-14]는 [그림 3-13]에서 얼굴 영역이 추출 가능한 영상만을 나타내었다. 그러나 Haar-like Feature을 이용하여 얼굴 영역을 추출하는 방법은 얼굴의 상하 좌우의 큰 회전에 대해서는 추출이 불가능하며 좌우 회전 30°까지만 얼굴 추출이 가능하였다.

[그림 3-14] 회전 영상에서 Haar-like Feature를 이용한 얼굴 영역 검출 영상

### 3.3.2 주성분 분석을 이용한 얼굴 고유 공간 구성

주성분 분석법은 앞에서 간단히 설명했듯이 각 얼굴 영상 집합들에 대한 픽셀 값들의 공간적 위치값들이 주로 각 얼굴 영상에서 어디에 분포하는가를 계산하여 확률 빈도가 높은 벡터값들을 고유치에 비례하여 재구성하는 방법이다.

따라서 이 방법은 고유벡터(eigenvector)와 고유치(eigenvalue)를 계산하여 얼굴 영상의 평균 모델을 구하여 이를 이용한다. 고유 벡터를 계산하기 위해서는 먼저 [그림 3-15]에서처럼 하나의 얼굴 영상은 5° 돌린 영상들의 집합이라고 할 수 있는 모든 얼굴 영상들과의 차를 구한다. [그림 3-17]에서 나타낸 평균 영상 $C$ 와 새로운 영상 집합 $X$를 (식 3-23)과 (식 3-24)와 같이 나타낸다.




[그림 3-15] 입력 영상들과 평균 영상의 거리 관계

(식 3-7)을 이용하여 [그림 3-16]의 얼굴 영상을 [그림 3-17]의 평균 영상 $C$ 를 구

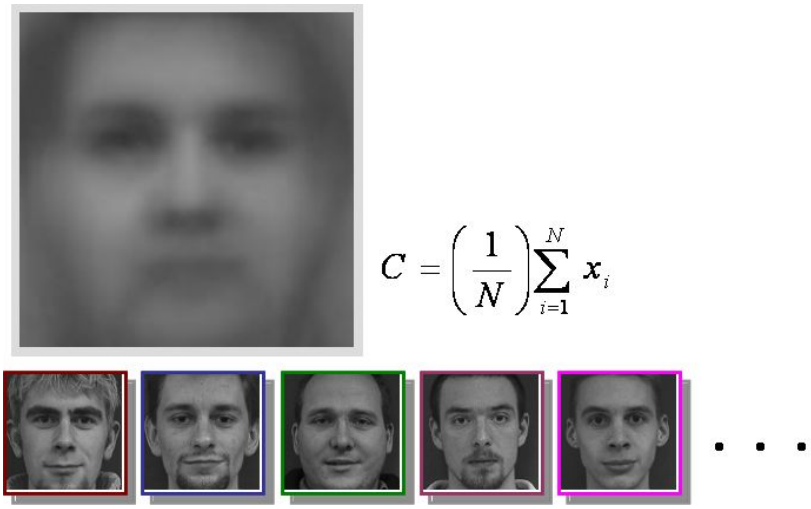
성하였다.

$$C = (1-N)\sum_{i=1}^{N} x_i \qquad \text{(식 3-7)}$$

$$X_{=}^{\triangle}\left\{x_1^{(1)}-c, x_2^{(1)}-c, \cdots\cdots\ x_R^{(1)}-c, \cdots\cdots\ x_R^{(p)}-c\right\} \qquad \text{(식 3-8)}$$

여기서 $N$은 얼굴 영상의 개수이며 $x_i$는 입력 영상이다.

[그림 3-16] 일부 얼굴 영상 집합




[그림 3-17] 전체 얼굴 영상의 평균 영상

고유 공간을 구하기 위해서는 M × N의 크기를 지닌 영상 집합 $X$를 (식 3-9)과 같이 계산하고 (식 3-10)을 만족하는 고유벡터를 구하면 된다.

즉, 공분산 행렬 $Q$에 대한 고유치 λ와 고유벡터 $e$를 구한다.

$$Q \underset{=}{\triangledown} XX^T \qquad \text{(식 3-9)}$$

$$\lambda_i e_i = Q e_i \qquad \text{(식 3-10)}$$

여기서 M은 한 영상의 픽셀수이고 N은 전체 영상의 개수를 나타내는 정수이다. 본 논문에서는 특이치 분해(Singular Value Decomposition)를 이용하여 고유 공간을 구하였다. 특이치 분해를 이용하여 (식 3-11)과 같이 행렬이 분해되는데, 영상 집합 $X$의 공분산 행렬에 대한 고유 벡터를 쉽게 얻을 수 있다.

$$\begin{bmatrix} X \end{bmatrix} = \begin{bmatrix} V \end{bmatrix} \cdot \begin{bmatrix} \alpha^1 & & \\ & \alpha^2 & \\ & & \alpha^3 \end{bmatrix} \cdot \begin{bmatrix} U^T \end{bmatrix} \qquad \text{(식 3-11)}$$

행렬은 (식 3-11)과 같이 분해가 가능하며 (식 3-12), (식 3-13), (식 3-14), (식 3-15)의 표현을 이용하면 원 식은 (식 3-16)과 같이 나타낼 수 있다.

$$X = \sum_{i=1}^{r} \sigma_i V_i U_i^T \qquad \text{(식 3-12)}$$

$$V = [v_i, ....., v_r] \qquad \text{(식 3-13)}$$

$$U = [u_i, ....., u_r] \qquad \text{(식 3-14)}$$

$$\sum = diag(\sigma_1, ....., \sigma_r) \qquad \text{(식 3-15)}$$

여기서 행렬 $X$의 특이치(singular value)는 $\sigma_i = \sqrt{\lambda_i}$ , $i = 1, ...$로 정의 된다.

$$X = V \sum U^T \qquad \text{(식 3-16)}$$

$V$행렬과 $U$행렬은 각각 직교 행렬이기 때문에 (식 3-17)에서 나타낸 것처럼 표현된다.

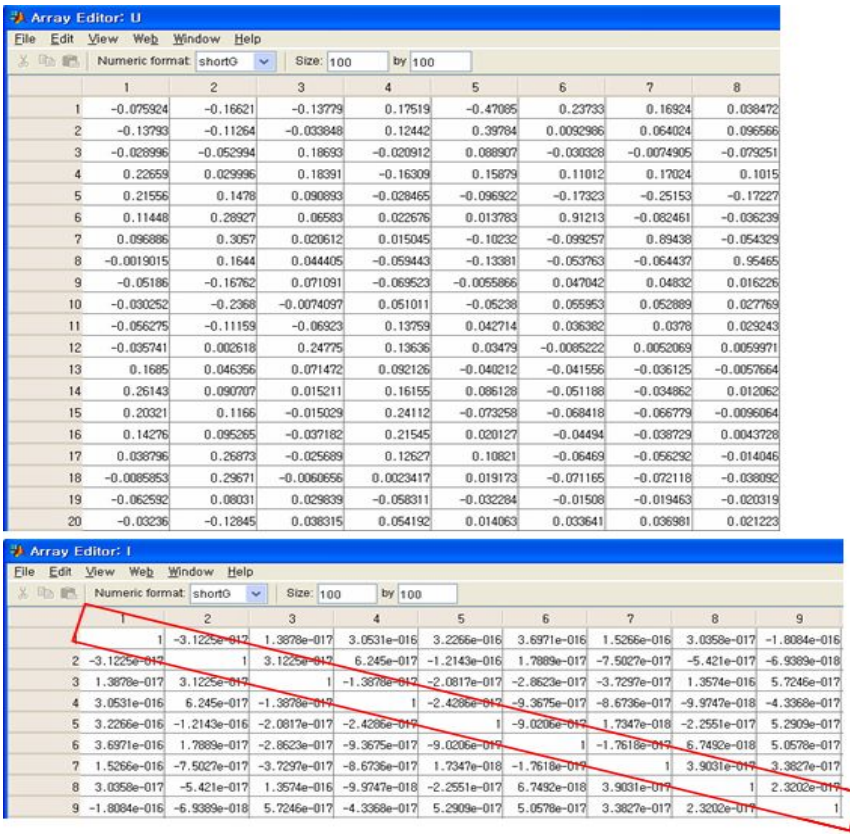
$$[V^T] \bullet [V] = [U^T] \bullet [U] = [I] \qquad \text{(식 3-17)}$$

[그림 3-18]과 [그림 3-19]는 특이치 분해 후에 $V$행렬과 $U$행렬이 직교 행렬이라는 사실을 증명한 그림이다.

| | | | | |
|---|---|---|---|---|
| 0.64297 | 0.50463 | 0.32472 | -0.16279 | 0.44721 |
| -0.18519 | -0.068809 | 0.27306 | 0.8285 | 0.44721 |
| -0.56844 | -0.12077 | 0.43559 | -0.52206 | 0.44721 |
| -0.27862 | 0.42403 | -0.73614 | -0.025841 | 0.44721 |
| 0.38929 | -0.73907 | -0.29723 | -0.1178 | 0.44721 |

| | | | | |
|---|---|---|---|---|
| 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 0.0000 | 1.0000 | -0.0000 | 0 | 0 |
| 0.0000 | -0.0000 | 1.0000 | -0.0000 | 0.0000 |
| 0.0000 | 0 | -0.0000 | 1.0000 | -0.0000 |
| 0.0000 | 0 | 0.0000 | -0.0000 | 1.0000 |

[그림 3-18] $V$행렬의 직교 행렬 데이터




[그림 3-19] $U$행렬의 직교 행렬 데이터

이런 과정으로 얻어진 행렬 중 공분산의 고유 벡터로 이용되는 행렬은 공분산 행렬 $X$와 크기가 일치하는 $U$행렬이다. 특이치 분해 과정에서 나온 고유벡터를 고유치가 큰 순서대로 재구성한다. 각 고유벡터가 지닌 고유치의 크기는 그 고유벡터의 중요도를 의미하므로, 그 고유 공간을 규정하는 중요 고유벡터를 (식 3-18)을 이용하여 선택한다. 따라서 모든 고유벡터를 고유 공간 구성에 이용하지 않고 많은 영상을 대표할 수 있는 주성분의 벡터만을 이용할 수 있다.

$$\frac{\sum_{i=1}^{k}\lambda_i}{\sum_{i=1}^{N}\lambda_i} \geq T_1 \qquad \text{(식 3-18)}$$

여기서 $T_1$은 고유벡터의 개수를 조정하는 임계치이며, 인식과 얼굴 영상 평가 시 이용되는 고유벡터$\{e_i \mid i = 1, 2, 3, \ldots\ldots\ldots, k\}$는 저차원 공간을 구성하기 위해 $(k << N)$을 만족하며 실험에서는 $k$ = 5를 이용했다. 이렇게 구성된 공간을 앞으론 얼굴 영상 공간으로 표현한다.

아래 [그림 3-20]은 얼굴의 트레이닝 이미지에 대해서 고유 벡터의 개수에 따른 고유 영상(eigen-images)을 나타내고 있다. 각 트레이닝 이미지는 고유 영상의 선형 조합(linear combination)으로 나타내질 수 있다.

[그림 3-20] 트레이닝 얼굴 이미지에 대한 고유 영상의 예

## 3.4 인공 신경회로망

인간 두뇌의 신경회로망 조직을 모델링한 것이 신경회로망 모델이며, 엄밀한 의미에서는 신경회로망 구조를 이용하여 만든 컴퓨터가 신경회로망이지만 일반적으로는 뉴로 컴퓨팅, 신경회로망, 신경망 등의 용어로써 혼용되고 있다.

디지털 컴퓨터는 프로그램에 의해 순차적으로 업무를 처리하는 반면 신경 회로망은 인간과 마찬가지로 어떤 일을 수행하기 위해서는 미리 학습이 필요하며, 병렬로 업무를 처리한다. 이미 알려진 바와 같이 신경회로망을 구성하는 뉴런들 각각은 처리속도가 아주 느리지만 상호간의 병렬처리에 의해 현존하는 어떠한 컴퓨터보다도 빠르게 연산이나 패턴 인식이 가능하다.

정보 저장 및 기억에 있어서도 상당한 차이가 있다. 디지털 컴퓨터는 어드레스 지정에 의하여 정보를 저장하거나 검색하는 반면 신경회로망은 뉴런간의 연결강도에 의해 정보를 저장하며, 원하는 정보를 찾고자 하는 경우에도 내용에 의해 정보를 검색한다.

신경회로망의 정보 저장에 있어서 특정한 응용 목적에 따라 뉴런간의 연결강도(connection weight)를 변화시키는 과정을 학습이라 한다. 동일한 정보를 기억하는데 있어서도 사람의 IQ 또는 학습 방법에 따라 학습 시간이 다르듯이 신경회로망의 학습에 있어서도 어떠한 신경회로망 모델 또는 학습 방법을 선택하느냐에 따라 학습 시간이 빨라지기도 하고 느려지기도 하며, 심지어는 영원히 학습이 되지 않을 수도 있다. 신경회로망 모델은 이미 상당히 오래 전부터 연구되기 시작하였으나 디지털 컴퓨터의 급속한 발전으로 인해 관심을 끌지 못하다가 최근에 와서야 디지털 컴퓨터의 약점을 보완하기 위하여 다시 새로운 연구 분야로서 주목받고 있다.

신경회로망에 대한 연구는 인간의 두뇌와 동일한 형태의 독자적인 시스템으로 제작하려는 목표를 가지고 출발하였지만, 현재의 반도체 기술로는 도저히 실현 불가능할 정도로 복잡하기 때문에 요즈음은 특정 응용분야에 국한하여 보조 처리장치로써 활용하는 것이 타당하다는 인식이 확산되고 있다.

신경회로망의 응용에 있어서 가장 먼저 고려해야 할 사항은 다양한 신경회로망 모델 중에서 응용 목적에 적합한 신경회로망 모델을 선택하는 것이다. 신경회로망 모델의 선택이 잘못되면 원하는 결과를 얻지 못할 수도 있기 때문이다.

신경회로망 모델을 선택한 다음에는 그 구조를 설계하여야 한다. 구조설계라 함

은 뉴런 수는 몇 개로 하며, 뉴런들을 어떤 형태로 배치하고 어떻게 연결시킬 것인지, 입력과 출력은 어떻게 할 것인지를 결정하는 것이다.

신경회로망의 구성이 완료되면 원하는 일을 할 수 있도록 신경회로망을 학습시켜야 한다. 사람이 새로운 학문을 습득하는 방법에는 학교에 가서 선생님으로부터 배우거나 자습서를 이용하여 혼자 독학하는 방법이 있듯이 신경회로망의 학습 방법에도 지도 학습(supervised learning)과 자율 학습(unsupervised learning) 방법이 있다. 이들 학습방법은 신경회로망 모델에 따라 적합한 방법을 선택하여야 한다.

학습 방법이 결정되면 신경회로망을 학습시킬 학습 패턴을 선정하여야 한다. 여기서 학습 패턴이라 함은 지도 학습의 경우에는 입력 패턴과 원하는 출력 패턴의 짝을 의미하며, 자율 학습의 경우에는 신경회로망이 자율적으로 결과를 출력하므로 입력 패턴만을 의미한다.

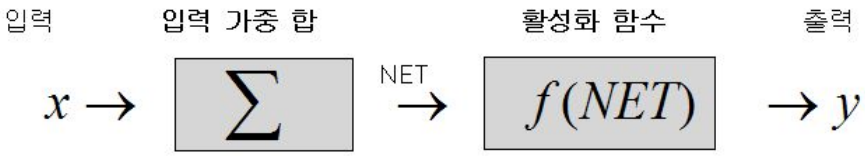
학습 패턴의 특징 추출이란 패턴 중 특히 응용 목적에 크게 영향을 미치는 중요한 부분만을 학습 패턴으로 사용함으로써 정보량을 줄이는 것이다.

최종적으로는 실제 신경회로망을 어떻게 구현할 것인가를 결정하여야 한다. 신경회로망을 구현하는 방법에는 VLSI로 구현하는 방법, 광학적으로 구현하는 방법, 디지털 컴퓨터로 시뮬레이션하는 방법 등이 있다.

### 3.4.1 인공 신경회로망 모델

인간의 두뇌는 계산, 연산, 추론, 기억 등 상당히 복잡한 일들을 수행하지만 디지털 컴퓨터처럼 CPU, 기억장치와 같은 하드웨어적인 소자들이 존재하는 것이 아니라 이러한 모든 작업들이 뉴런들의 결합체인 신경회로망의 적용에 의해서 이루어진다.

신경회로망의 기본 소자는 뉴런이며, 뉴런은 입력된 외부 자극을 합하여 그 결과에 따라 반응하는 단순한 기능만을 할 뿐이므로 뉴런의 기능을 [그림 3-21]과 같이 모델링할 수 있다.




[그림 3-21] 뉴런의 기능 모델

[그림 3-21]에서 NET와 OUT은 각각 입력된 외부 자극의 가중의 합과 뉴런의 출력을 의미하며, (식 3-19)와 같이 표현할 수 있다.

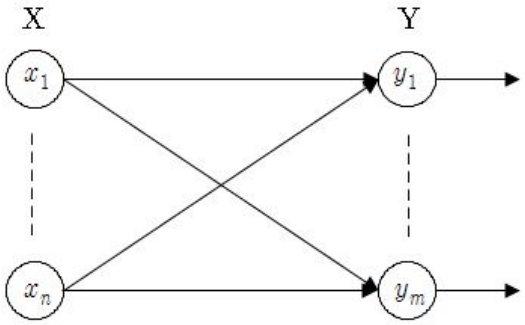
$$NET = \sum 입력$$
$$OUT = f(NET) \qquad \text{(식 3-19)}$$

f(x)는 뉴런의 반응 여부를 결정하는 역할을 하며, 활성화 함수라고 한다. 임계치가 T인 계단 함수라고 가정하면 뉴런의 출력은 다음 (식 3-20)과 같다.

$$OUT = \begin{cases} 1 ; NET \geq T \\ 0 ; NET < T \end{cases} \qquad \text{(식 3-20)}$$

생물학적 신경회로망에 있어서 뉴런 단독으로 어떤 기능을 수행하기 보다는 여러 뉴런들이 거미줄처럼 복잡하게 연결되어 서로 상호 작용을 하고 있으므로 신경

회로망은 방향성 그래프를 이용하여 모델링할 수 있으며, 이를 인공 신경회로망 모델 또는 신경회로망 모델이라 한다. 신경회로망 모델의 방향성 그래프에 있어서 노드는 뉴런을 나타내며, 링크는 뉴런의 시냅스 연결강도를 나타낸다.




[그림 3-22] 일반적인 신경회로망 모델

일반적인 신경회로망 모델은 [그림 3-22]와 같으며, 수식적으로 전개하면 다음과 같다. 입력 $X$와 연결강도 $W$는 (식 3-21)과 같이 벡터 형태로 표시할 수 있다.

$$X = [X_1 X_2 \cdots X_n]$$
$$W = [W_1 W_2 \cdots W_n] \quad \text{(식 3-21)}$$

출력 뉴런 y의 NET 값은 입력 가중 합이므로 (식 3-22)와 같이 구할 수 있다.

$$NET = \sum_{i=1}^{n} X_i W_i$$
$$= X_1 W_1 + X_2 W_2 + \cdots + X_n W_n \quad \text{(식 3-22)}$$

NET를 벡터 형태로 표현하면 (식 3-23)과 같다.

$$NET = XW^T \quad \text{(식 3-23)}$$

여기서, $W^T$는 W 의 치환 벡터이며, 뉴런 y의 출력 OUT은 (식 3-24)와 같다.

$$\begin{aligned} OUT &= f(NET) \\ &= f(XW^T) \end{aligned} \qquad \text{(식 3-24)}$$

따라서, 계단함수를 활성화 함수로 사용하면 출력 OUT은 (식 3-25)와 같다.

$$OUT = \begin{cases} 1 \; ; \; NET \geq T \\ 0 \; ; \; NET < T \end{cases} \qquad \text{(식 3-25)}$$

결국, 입력 가중 합 NET가 임계치 T보다 크면 뉴런이 활성화되지만, 그렇지 않으면 활성화되지 않음을 알 수 있다.

### 3.4.2 신경망 학습 규칙

인간이 어떤 일을 하기 위해서는 반복 학습이 이루어져야 하듯이 신경회로망도 역시 응용 분야에 활용하기 위해서는 신경회로망의 학습이 선행되어야 한다.

신경회로망에서의 학습이라 함은 특정한 응용 목적에 적합하도록 뉴런 간의 연결강도를 적응시키는 과정이다. 신경회로망의 학습 방법은 지도 학습(supervised learning), 자율 학습(unsupervised learning), 경쟁식 학습(competitive learning)과 같이 크게 3가지로 구분할 수 있다.

지도 학습 방법은 신경회로망을 학습시키는데 있어서 반드시 입력 x와 원하는 목표치 d의 쌍(x, d)가 필요하며, 이를 학습 패턴 쌍(training pattern pair)이라고 한다. 일반적인 학습 절차는 [표 3-3]과 같이 8단계로 이루어져 있다.

[표 3-3] 신경망의 학습 절차

| | |
|---|---|
| 1단계 | 응용 목적에 적합한 신경회로망 구조를 설계한다. |
| 2단계 | 연결강도를 초기화한다. |
| 3단계 | 학습 패턴 쌍 (x, d)를 입력하여 신경회로망의 출력 y를 구한다. |
| 4단계 | 출력 y와 목표치 d를 비교하여 그 오차를 산출한다. |
| 5단계 | 오차를 학습 신호 발생기에 입력하여 연결강도의 변화량 Δw를 계산한다. |
| 6단계 | 연결강도를 Δw 만큼 변경한다. |
| 7단계 | 변경된 연결강도 (w + Δw)에 대하여 3단계~6단계를 반복한다. |
| 8단계 | 더 이상 연결강도가 변하지 않으면 학습을 종료한다. |

신경회로망의 학습이라 함은 특정한 응용 목적에 적합하도록 뉴런 간의 연결 강도를 적응시키는 과정이다. 우선, 학습 입력 패턴을 신경회로망에 입력을 하면 입력층에서 은닉층으로 은닉층에서 출력층으로 각 층을 통과하면서 가중된 합계를 구하고, 활성화 함수를 통해 계산되어 최종 출력값을 얻게 된다. 다음으로 출력값

과 목표값과의 차에 해당하는 오차를 구하고, 마지막으로 오차값을 감소시키는 방향으로 역전파하면서 각 층 사이의 연결 강도를 변경한다. 학습은 오차의 최소 임계치보다 작아질 때까지 진행하게 된다.

학습에 있어서 중요한 문제는 각 층의 초기 연결강도를 적절히 선택하는 것이다. 초기 연결강도를 잘 못 설정하면 응용 목적에 적합한 학습이 이루어지지 않으며, 오차의 국부 최소점에 빠져 학습을 진행할 수 없게 된다. 또한 은닉층의 뉴런 개수선택과 학습율도 학습의 진행 속도에 관련하게 된다. 따라서 초기 연결 강도와 은닉층 뉴런 개수, 학습률에 대한 최적의 선택이 학습능률을 향상시키는 중요한 성분이 된다.

### 3.4.3 오류역전파 알고리즘

BP 알고리즘이라고도 부르는 오류 역전파 학습 알고리즘은 [그림 3-23]와 같이 다층 신경회로망으로 구성되어 있다. 다층 신경회로망(Multiple Neural Network: MNN)의 다층 퍼셉트론을 확장한 형태로서 단층 퍼셉트론의 한계를 극복 할 수 있는 우수한 특성을 가지고 있다.

이와 같은 특성은 신경소자의 활성함수가 가지고 있는 비선형성에 기인한다. 그러나 이러한 비선형성으로 인하여 단층 퍼셉트론에 비해 학습하기 어려운 단점이 있다. 다층 퍼셉트론을 학습하는 가장 대표적인 학습규칙에는 역전파 알고리즘이 있고, 이것은 미분 가능한 비선형 활성함수를 가진 다층 퍼셉트론에 최소 평균자승(Least mean of square)법을 확장하여 적용한 것이다. 이것을 일반화된 델타규칙(Generalized delta rule)이라고도 한다.

BP 알고리즘을 이용한 다층 신경회로망의 세부적인 학습 절차는 다음과 같이 진행된다.

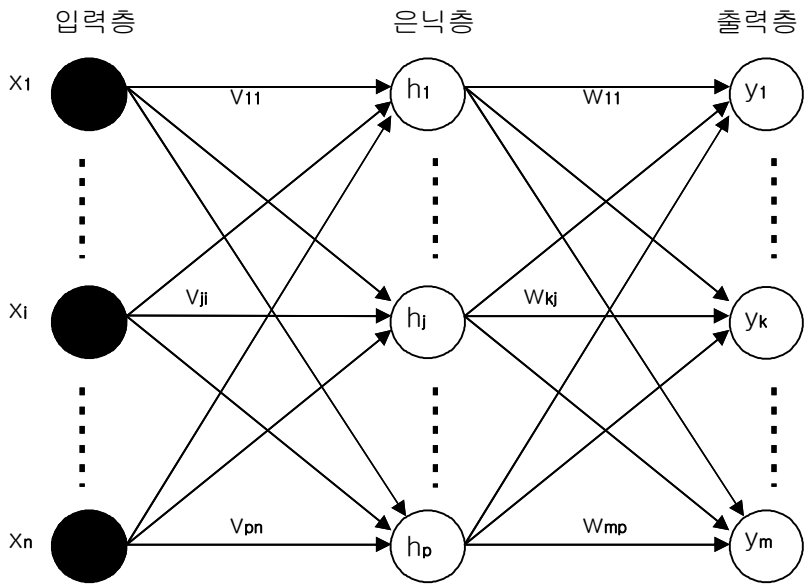
입력층 패턴 벡터 $X$, 은닉층 출력 벡터 $H$, 출력층 출력 벡터 $Y$는 (식 3-25)와 같이 표현한다. [그림 3-23]은 다층 신경회로망의 구조이다.

$$X=[x_1 x_2 \cdot \cdot \cdot x_n] \qquad \text{(식 3-25a)}$$

$$H=[h_1 h_2 \cdot \cdot \cdot h_p] \qquad \text{(식 3-25b)}$$

$$Y=[y_1 y_2 \cdot \cdot \cdot y_m] \qquad \text{(식 3-25c)}$$

입력층과 은닉층간의 연결 강도 $V(p\times n)$, 은닉층과 출력층간의 연결강도 $W(m\times p)$ 매트릭스 형태로 다음 (식 3-26)과 같이 표현한다. 여기서 연결강도 $v, w$는 -0.5에서 +0.5 사이의 작은 값으로 초기화시킨다.




[그림 3-23] 다층 신경회로망의 구조

$$V=\begin{bmatrix} v_{11}\, v_{12} & \cdots & v_{1n} \\ v_{21}\, v_{22} & \cdots & v_{2n} \\ & \vdots & \\ v_{p1}\, v_{p2} & \cdots & v_{pm} \end{bmatrix} \qquad \text{(식 3-26a)}$$

$$W=\begin{bmatrix} w_{11}\, w_{12} & \cdots & w_{1p} \\ w_{21}\, w_{22} & \cdots & w_{2p} \\ & \vdots & \\ w_{m1}\, w_{m2} & \cdots & w_{mp} \end{bmatrix} \qquad \text{(식 3-26b)}$$

은닉층의 가중 함 $NET_h$와 출력 $H$를 구하면, (식 3-27)과 같다. 여기서 $f()$는 활성화 함수이며, 단조 증가 함수, 연속 함수, 미분 가능 함수의 조건을 만족하여야 한다.

$$NET_h = XV^T$$
$$H = f(NET_h) \qquad \text{(식 3-27)}$$

주로 다층 신경회로망 학습에서는 활성화 함수로 시그모이드 함수가 사용되며, 단극성 시그모이드 함수와 양극성 시그모이드 함수가 있다. (식 3-28)은 각각 단극성 시그모이드 함수와 양극성 시그모이드 함수를 나타낸다.

$$f(NET) = \frac{1}{1+\exp(-NET_h)} \quad unipolar\,sigmoid$$
$$= \frac{1-\exp(-NET_h)}{1+\exp(-NET_h)} \quad bipolar\,sigmoid \qquad \text{(식 3-28)}$$

출력층의 가중 합 $NET_y$와 출력 $Y$를 구하면 (식 3-29)와 같다.

$$NET_y = ZW^T$$
$$Y = f(NET_y) \qquad \text{(식 3-29)}$$

원하는 목표값과 신경회로망의 출력을 뺀 값의 평균 제곱 오차 $E$를 구하면 (식 3-30)과 같다.

$$E = \frac{1}{2}\sum_{i=1}^{m}(d_i - y_i)^2 \qquad \text{(식 3-30)}$$

평균 제곱 오차 $E$를 최소화하는 방향으로 연결 강도 $w_{kj}$와 $v_{ji}$를 변경시켜야 한다. 연결 강도 $w_{kj}$의 변화량인 $\Delta w_{kj}$는 (식 3-31)과 같다.

$$\begin{aligned} \Delta w_{kj} &= -\alpha \nabla E \\ &= -\alpha \frac{\partial E}{\partial w_{kj}} \end{aligned} \qquad \text{(식 3-31)}$$

여기서, $\alpha$는 학습률이라 하며, 양수의 값을 가진다. 또한, 출력층 뉴런 $k$에서 발생한 오차 신호 $\delta_{y_k}$는 (식 3-32)와 같이 정의한다.

$$\delta_{y_k} = -\frac{\partial E}{\partial (NET_k)} \qquad \text{(식 3-32)}$$

출력층 뉴런 $k$의 출력 오차는 $w_{kj}$에만 관련되므로, 정리하면 (식 3-33)과 같다.

$$\begin{aligned} \frac{\partial E}{\partial w_{kj}} &= \frac{\partial E}{\partial (NET_k)} \cdot \frac{\partial (NET_k)}{\partial w_{kj}} \\ &= \frac{\partial E}{\partial (NET_k)} \cdot \frac{\partial}{\partial w_{kj}} [h_1 w_{k1} + h_2 w_{k2} + \cdots h_m w_{km}] \\ &= -\delta_{y_k} h_j \end{aligned} \qquad \text{(식 3-33)}$$

(식 3-33)의 $\delta_{y_k}$를 $y_k$에 관련된 관계식으로 나타내면 (식 3-34)와 같다.

$$\begin{aligned} \delta_{y_k} &= -\frac{\partial E}{\partial y_k} \cdot \frac{\partial y_k}{\partial (NET_k)} \\ &= -\frac{\partial E}{\partial y_k} \cdot \frac{\partial [f(NET_k)]}{\partial (NET_k)} \\ &= -\frac{\partial E}{\partial y_k} f'(NET_k) \end{aligned} \qquad \text{(식 3-34)}$$

(식 3-34)의 평균 제곱 오차 $E$를 신경회로망의 출력으로 편미분한 부분을 계산하면 (식 3-35)와 같이 나타낸다.

$$\frac{\partial E}{\partial y_k} = -(d_k - y_k) \qquad \text{(식 3-35)}$$

(식 3-33), (식 3-34), (식 3-35)과 (식 3-36)의 활성화 함수 $f(NET)$을 신경회로망의 출력 $Y$로 치환 적용하여 (식 3-27)을 정리하면 (식 3-36)과 같다.

$$\begin{aligned} \triangle w_{kj} &= \alpha \delta_{y_k} h_j \\ &= \alpha \cdot (d_k - y_k) \cdot y_k \cdot (1 - y_k) \cdot h_j \quad unipolar\, sigmoid \\ &= \alpha \cdot \frac{1}{2} \cdot (d_k - y_k) \cdot (1 - y_k^2) \cdot h_j \quad bipolar\, sigmoid \end{aligned} \qquad \text{(식 3-36)}$$

연결 강도 $v_{ji}$의 변화량인 $\triangle v_{ji}$는 (식 3-37)과 같다.

$$\triangle v_{ji} = -\alpha \frac{\partial E}{\partial v_{ji}} \qquad \text{(식 3-37)}$$

은닉층 뉴런 $j$의 출력 오차는 $v_{ji}$에만 관련되므로 정리하면 (식 3-38)과 같다.

$$\begin{aligned} \frac{\partial E}{\partial v_{ji}} &= \frac{\partial E}{\partial (NET_{h_j})} \cdot \frac{\partial (NET_{h_j})}{\partial v_{ji}} \\ &= \frac{\partial E}{\partial (NET_{h_j})} \cdot \frac{\partial}{\partial v_{ji}} [x_1 v_{j1} + x_2 v_{j2} + \cdot\cdot\cdot x_m v_{jm}] \\ &= -\delta_{h_j} x_i \end{aligned} \qquad \text{(식 3-38)}$$

(식 3-38)에서의 $\delta_{h_j}$는 은닉층의 오차 신호이며, (식 3-39)와 같이 전개된다.

$$\begin{aligned}\delta_{h_j} &= -\frac{\partial E}{\partial h_j} \cdot \frac{\partial h_j}{\partial (NET_{h_j})} \\ &= -\frac{\partial E}{\partial h_j} \cdot \frac{\partial [f(NET_{h_j})]}{\partial (NET_{h_j})} \\ &= -\frac{\partial E}{\partial h_j} f'(NET_{h_j})\end{aligned} \qquad \text{(식 3-39)}$$

(식 3-39)에서 평균 제곱 오차 $E$를 은닉층 출력값 $H$로 편미분한 부분을 계산하면 (식 3-40)과 같다.

$$\begin{aligned}\frac{\partial E}{\partial h_j} &= \frac{\partial}{\partial h_j}\left\{\frac{1}{2}\sum_{k=1}^{m}[(d_k - y_k)^2]\right\} \\ &= -\sum_{k=1}^{m}(d_k - y_k)\frac{\partial}{\partial h_j}[f(NET_{y_k})] \\ &= -\sum_{k=1}^{m}(d_k - y_k) f'(NET_{y_k})\frac{\partial (NET_{y_j})}{\partial h_i} \\ &= -\sum_{k=1}^{m}(d_k - y_k) f'(NET_{y_k}) w_{kj}\end{aligned} \qquad \text{(식 3-40)}$$

(식 3-36)을 (식 3-40)에 대입하면 (식 3-41)과 같다.

$$\frac{\partial E}{\partial h_j} = -\sum_{k=1}^{m}\delta_{y_k} w_{kj} \qquad \text{(식 3-41)}$$

따라서, 은닉층 뉴런 $j$에서의 오차 신호 $\delta_{h_j}$는 (식 3-39), (식 3-40)을 대입하면 (식 3-42)와 같이 나타낸다.

$$\begin{aligned}\delta_{h_j} &= \sum_{k=1}^{m}\delta_{y_k}w_{kj} \cdot \frac{\partial h_j}{\partial(NET_{h_j})} \\ &= f'(NET_{h_j})\sum_{k=1}^{m}\delta_{y_k}w_{kj}\end{aligned} \qquad \text{(식 3-42)}$$

입력층과 은닉층 간의 연결 강도의 변화량 $\triangle v_{ij}$는 (식 3-40), (식 3-41)을 대입하여 정리하면 (식 3-43)과 같다.

$$\begin{aligned}\triangle v_{ji} &= \alpha\delta_{h_j}x_i \\ &= \alpha f'(NET_{h_j})x_i\sum_{k=1}^{m}\delta_{y_k}w_{kj} \\ &= \alpha h(1-h)x_i\sum_{k=1}^{m}\delta_{y_k}w_{kj} \quad unipolar\,sigmoid \\ &= \alpha\frac{1}{2}(1-h^2)x_i\sum_{k=1}^{m}\delta_{y_k}w_{kj} \quad bipolar\,sigmoid\end{aligned} \qquad \text{(식 3-43)}$$

$n$번째 학습 단계에서의 은닉층과 출력층 간의 연결 강도 변화량 $\triangle w^n$ 및 입력층과 은닉층 간의 연결 강도 변화량 $\triangle v^n$을 정리하면 (식 3-44)와 같다.

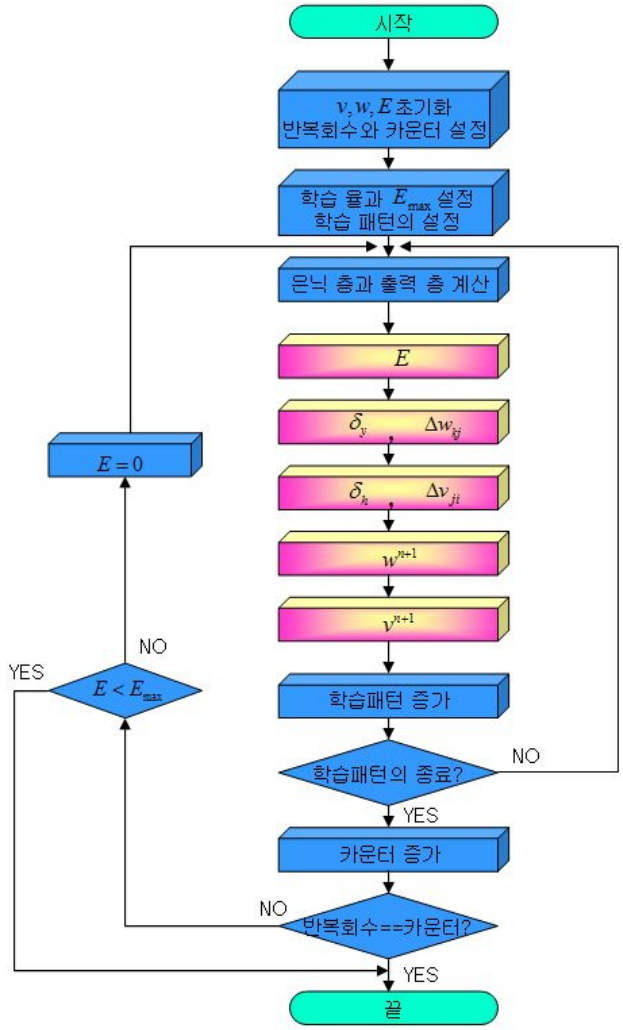
$$\begin{aligned}\triangle w^n &= \alpha\delta_y h \\ \triangle v^n &= \alpha\delta_h x\end{aligned} \qquad \text{(식 3-44)}$$

$n+1$ 단계에서의 은닉층과 출력층 간의 연결 강도 $w^{n+1}$, 입력층과 은닉층 간의 연결 강도 $v^{n+1}$을 구하면 (식 3-45), (식 3-46)과 같이 나타낸다.

$$\begin{aligned}w^{n+1} &= w^n + \triangle w^n \\ &= w^n + \alpha\delta_y h^n\end{aligned} \qquad \text{(식 3-45)}$$

$$\begin{aligned}v^{n+1} &= v^n + \triangle v^n \\ &= v^n + \alpha\delta_h x^n\end{aligned} \qquad \text{(식 3-46)}$$

학습 패턴 쌍을 반복 입력하여 연결 강도를 변경하며, 오차 $E$가 특정 범위 $E_{\max}$ 보다 작아지면 학습을 종료한다. [그림 3-24]는 BP 알고리즘의 순서도를 나타낸다.




[그림 3-24] BP 알고리즘의 순서도

# Ⅳ. 실험 결과 및 고찰

본 연구에서는 실시간으로 웹 카메라를 통하여 입력 영상을 얻으면 AdaBoost 학습 알고리즘을 사용하여 최적화된 얼굴 검출을 위한 프로토타입들의 집합들을 Haar-like Feature에 사용하여 얼굴 영역이 있는지 없는지를 판단하게 되고 얼굴 영역이 존재하면 그 영역만을 검출하고 검출된 이미지를 주성분 분석과 오류 역전파를 사용하여 인식을 수행한다. 또한 얼굴의 회전에 대한 인식을 수행하기 위해 영상을 5°씩 회전하여 모델 영상을 구성하였고, Haar-like Feature를 통해서 검출할 수 있는 회전 영역만을 모델 영상으로 사용하였다.

그리고 본 논문에서 제안한 얼굴 인식 시스템은 신경망 학습률이 얼굴 인식률에 미치는 관계를 관찰하여 최적의 값을 도출하였다. 본 장에서는 실험환경 및 실험에 사용된 조건들에 대하여 살펴보고 각 단계별로 실험 결과 및 성능을 고찰한다.

## 4.1 실험 환경

본 논문에서 제안하는 얼굴 인증 시스템 실험 환경은 [표 4-1]과 같다.

[표 4-1] 얼굴 인증 시스템의 시스템 실험 환경

| 구분 | 환 경 | 비 고 |
|---|---|---|
| 운영체제 | Windows XP Professional | |
| CPU/RAM | Intel Core2 1.86Ghz / RAM : 1GB | |
| 프로그램 언어 | Visual C++ 6.0 | |
| 입력 장치 | 웹카메라 2종 | 팬-틸트 기능(1개) |
| 데이터베이스 | MIT Media Labs, Yale, M2VTS, Chosun-DB | |

## 4.2 실험 얼굴 영상

본 논문에서 사용한 실험 얼굴 영상은 ORL 얼굴 DB와 MIT 대학의 얼굴 DB를 이용하여 얼굴 인식 시스템의 성능을 평가하였다.

ORL 얼굴 영상은 서로 다른 환경에서 40명으로부터 각 개인에 대하여 10장의 얼굴로 구성되어 있으며, 이 영상들은 위치, 회전, 스케일, 얼굴 감정에 변화를 주고 있다. 회전에서 변화는 최대 30도 회전하였고, 스케일에서 변화는 사람과 비디오카메라 사이에 거리를 변화시켰다. 또한 어떤 사람에 대해서는 안경 착용에 변화를 주었고, 각 영상들은 디지털화되고 0에서 255까지의 그레이값을 갖는 영상 집합이다. [그림 4-1]은 ORL 얼굴 데이터베이스의 일부를 나타낸다.

MIT 얼굴 영상은 15명에 대한 650개의 흑백이미지로 구성되어 있으며, 각 사람마다 좌우 회전 이미지를 담고 있다. 각 영상들은 디지털화되고 0에서 255까지의 그레이값을 갖는 영상 집합이다.

그리고 회전 얼굴 영상의 인식 실험을 하기 위해 이용된 얼굴 영상은 원 영상[그림 4-2]]을 Haar-like Feature를 이용하여 얼굴영역만 검출한 영상을 이용하여 실험하였다. [그림 4-3]는 Haar-like Feature와 AdaBoost 알고리즘을 이용하여 얼굴영역을 추출한 영상이고, [그림 4-4]는 얼굴영역만 검출하여 실제 실험에 사용한 얼굴 영상집합이다.

[그림 4-1] ORL 얼굴 DB에서의 일부 얼굴 영상

[그림 4-2] 남자 5의 회전 영상의 모델영상

[그림 4-3] Haar-like Feature 알고리즘을 이용하여 [그림 4-2]의 얼굴 영역을 추출한 영상

[그림 4-4] [그림 4-3]에서 얼굴영역만 검출한 얼굴 영상 집합

## 4.3 얼굴 인식 시스템의 성능분석

### 4.3.1 MLNN을 이용한 인식 결과

본 얼굴 인식 시스템에서 사용된 신경망의 학습 방법은 앞서 기술한 오류 역전파 학습 알고리즘을 사용하였다.

실험을 위하여 사용된 영상은 ORL DB와 MIT 얼굴영상 데이터베이스를 사용하여 실험을 하였다. 그러나 사용된 얼굴 영상은 사각형 특징 기반 분류기를 통하여 검출된 얼굴 영역만을 사용하였다. 신경망 학습을 위해 사용한 이미지는 ORL DB 200개와 MIT DB 300개의 얼굴 영상을 PCA에 적용하여 특징벡터를 산출하였다. 구해진 특징 벡터는 구축된 신경망의 입력 노드값으로 사용되어 진다.

검증을 하기위해서는 학습에 사용되지 않은 나머지 영상을 사용하였다. 제안한 방법의 우수성을 검증하기 위해여 PCA 결과 가중치를 신경망에 입력하여 신경망의 가중치를 학습한 후에 인식 성능을 기존의 Euclidean 거리척도 방법, Mahananobis 거리척도, MLNN을 적용한 결과를 [표 4-2]와 같다.

제안한 신경망은 입력노드 수는 주성분 분석을 통해 고유치의 95% 이상의 값을 가지고 있는 특징벡터 8개를 사용하였고 은닉 층의 개수는 훈련 수에 따라 조절하여 사용하였다. 초기 가중치는 0.01, 오차허용 한계는 0.05로 설정하였다. 오류 역전파 학습 방법은 반복 학습 횟수는 1,000회를 지정하여 학습오차값(0.05) 이하로 감소하는 경우 학습이 완료된 것으로 간주하여 학습을 종료한다.

인식률 실험 결과는 [표 4-2]와 같이 제안한 방법을 이용했을 때 기존의 제안한 방법인 Euclidean 과 Mahananobis의 방식보다 매칭 실패율은 상당히 좋은 양상을 보였고, 잘못된 매칭에서도 기존의 방법보다는 우수성을 나타내었다.

이렇게 인식률이 향상된 이유는 기존의 Euclidean 방식과 Mahananobis 방식은 가장 근접한 데이터를 인식의 기준으로 평가했다. 이런 방식의 경우 얼굴의 회전이미지와 같은 특징 데이터의 변화가 거의 없을 경우 잘못된 매칭이 빈번히 발생하게 된다. 본 논문에서 제안한 방법은 학습된 값에 가장 근사치로 가기 위해 계속 학습함으로써 인식을 수행하기 때문에 매칭 실패나 잘못된 매칭에서 평균 6.5%나 높게 인식률을 향상 시킬 수 있었다.

[표 4-2] 매칭 방법에 따른 매칭 성공률

| 매칭 방법 | 매칭 실패 | 잘못된 매칭 | 매칭 성공 |
|---|---|---|---|
| Euclidean | 3.5 % | 6.0 % | 90.5 % |
| Mahananobis | 7.5 % | 5.5 % | 87.0 % |
| Propesed MLNN | 2.7 % | 2.0 % | 95.3 % |

### 4.3.2 학습률에 따른 인식률

본 논문에서는 최적의 학습률에 따른 인식률의 차이를 알기 위해 [표 4-3]과 같이 실험을 진행하였다. [표 4-3]의 실험환경 1은 얼굴 영상 학습 모델 200개를 가지고 얼굴 영상 100개의 영상에 대해 인식을 수행하는 실험을 하였다. 실험환경 2는 실험환경 1과 동일하게 얼굴 영상 학습 모델을 획득하였으나 실험환경 1과는 다르게 학습 모델을 획득하지 않은 50개의 영상에서 인식하는 실험을 하였다. 실험에서 출력 오차의 임계값은 0.025로 사용하였다.

[표 4-3] 학습에 사용된 얼굴 영상 개수에 따른 실험 환경

| 영상개수 / 실험환경 | 학습에 사용된 얼굴 영상 수 | 인식에 사용된 얼굴 영상 개수 |
|---|---|---|
| 실험환경 1 | 200 | 100 |
| 실험환경 2 | 200 | 50 |

[표 4-4], [표 4-5]는 [표 4-3]에 대한 각각의 실험에 대한 결과이다. 각각의 실험은 모두 학습률이 증가함에 따라 학습 횟수가 감소되는 것을 알 수 있었다. [표 4-4]의 학습률에 따른 인식률에서는 학습률을 0.2로 하여 학습하였을 때 인식률이

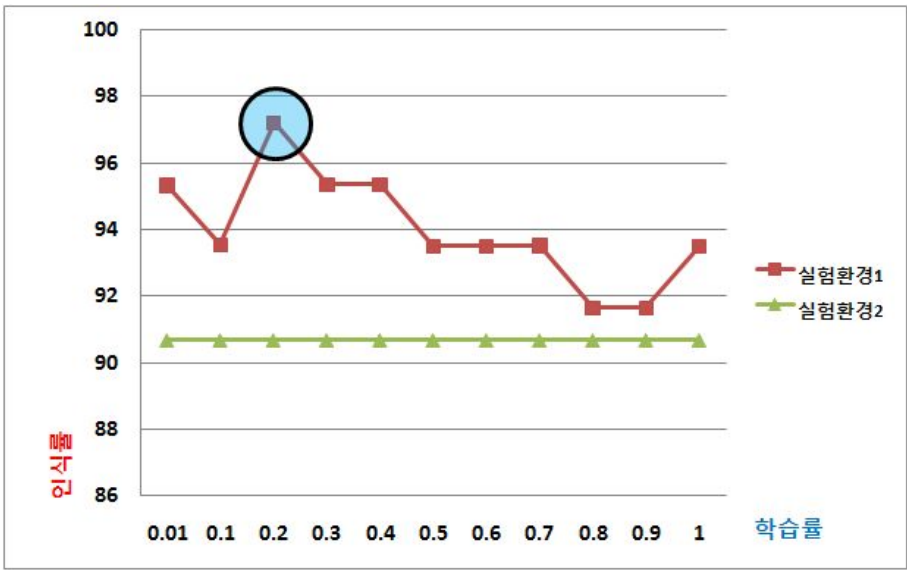
가장 좋음을 알 수 있었고, [표 4-4]는 학습률에 따른 인식률이 크게 차이가 없음을 알 수 있었다. 또한 학습에 사용한 얼굴 영상을 입력 영상에 포함 시킬 때 인식률이 좋음을 알 수 있었다. [그림 4-5]는 실험환경에 따른 인식률에 변화를 나타낸 그림이다. [그림 4-5]에서 알 수 있듯이 실험 환경 1에서 가장 높은 인식률을 얻을 수 있었고, 그 중에서도 0.2일 때 가장 높은 인식률이 나타났다.

[표 4-4] 실험환경 1의 학습률에 따른 인식률

| 학습률 | 학습 횟수 | 인식률(%) |
|---|---|---|
| 0.01 | 3360 | 95.32 |
| 0.1 | 372 | 93.54 |
| 0.2 | 149 | 97.20 |
| 0.3 | 101 | 95.35 |
| 0.4 | 68 | 95.36 |
| 0.5 | 70 | 93.50 |
| 0.6 | 47 | 93.50 |
| 0.7 | 59 | 93.52 |
| 0.8 | 43 | 91.65 |
| 0.9 | 27 | 91.65 |
| 1.0 | 28 | 93.50 |

[표 4-5] 실험환경 2의 학습률에 따른 인식률

| 학습률 | 학습 횟수 | 인식률(%) |
|---|---|---|
| 0.01 | 3360 | 90.68 |
| 0.1 | 372 | 90.68 |
| 0.2 | 149 | 90.68 |
| 0.3 | 101 | 90.68 |
| 0.4 | 68 | 90.68 |
| 0.5 | 70 | 90.68 |
| 0.6 | 47 | 90.68 |
| 0.7 | 59 | 90.68 |
| 0.8 | 43 | 90.68 |
| 0.9 | 27 | 90.68 |
| 1.0 | 28 | 90.68 |

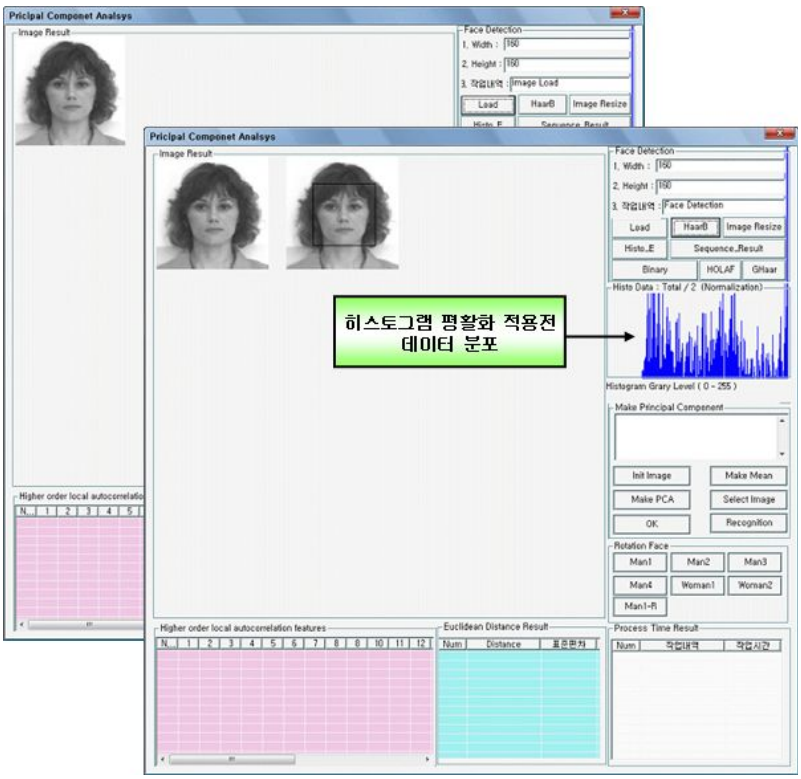



[그림 4-5] 실험환경에 따른 인식률 변화

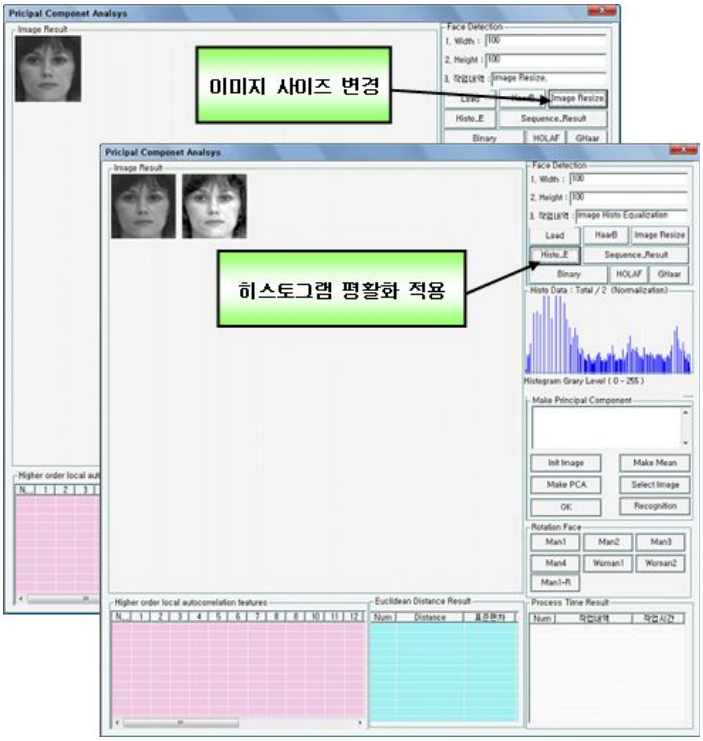
### 4.3.3 실시간 얼굴 인식 실험 결과

실시간 얼굴 인식은 Visual C++로 구현되었으며 이미지의 크기는 정규화 과정을 거쳐 사용하였다. 한 사람이 가지는 이미지의 개수는 회전 그룹의 이미지를 가지고 있으며 총 모델 영상의 개수 500개를 이용하여 고유 공간을 구성하는데 걸리는 시간은 대략 10분 39초가 걸렸다. 새로운 이미지가 입력되어 오류역전파 알고리즘을 통해 인식하여 복원하는데 걸리는 시간은 평균 19.245 밀리 초가 걸렸다.

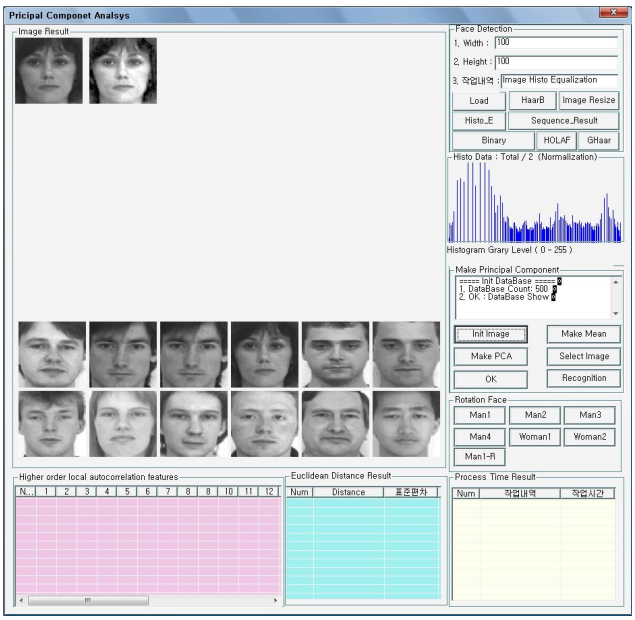
[그림 4-6]은 본 논문의 구현 시스템에서 얼굴 영상이 들어왔을 경우 얼굴영상에서 Haar-like Feature 와 AdaBoost 알고리즘을 이용하여 얼굴 영역만을 검출한 장면이다. [그림 4-7]은 전체적으로 얼굴 영상의 어둡기 때문에 히스토그램 평활화를 이용하여 밝기 정규화를 수행한 장면이다.



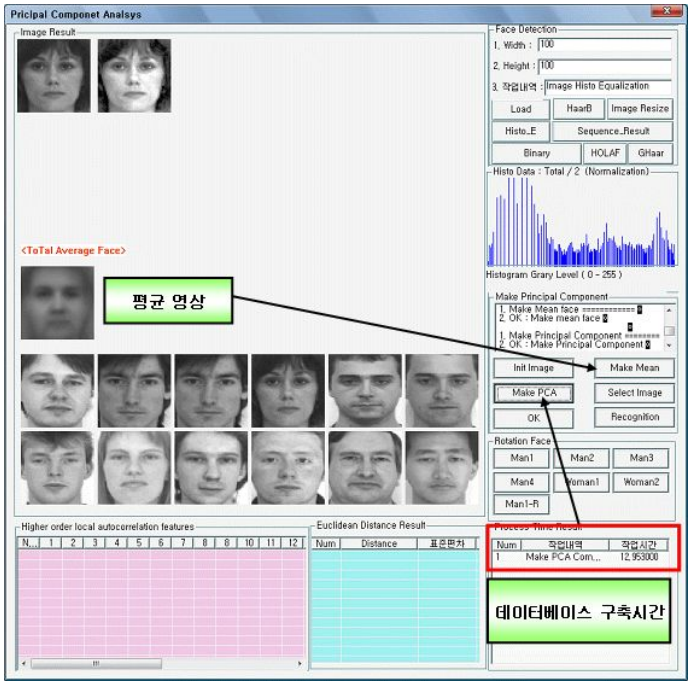

[그림 4-6] 얼굴 인식 시스템에 얼굴영상 입력 및 얼굴 영역 검출



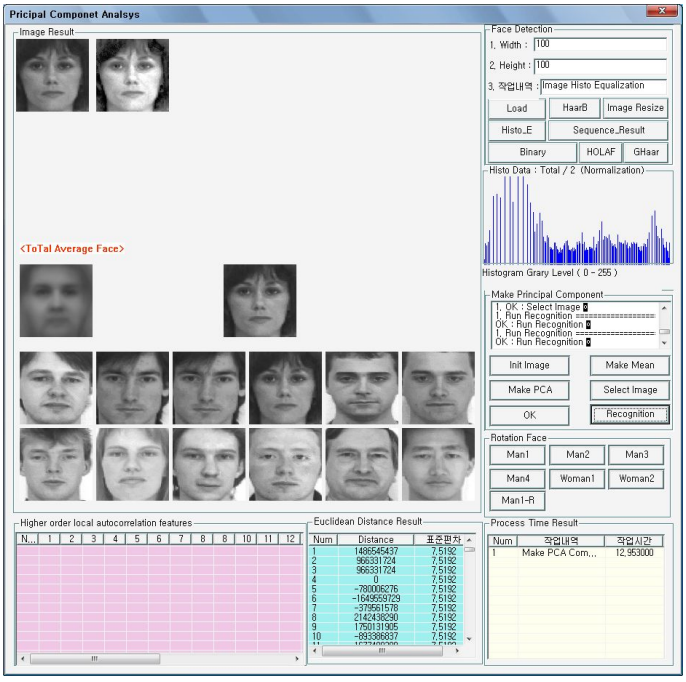

[그림 4-7] 얼굴 영역 추출 및 밝기 정규화




[그림 4-8] 모델 얼굴 영상 초기화



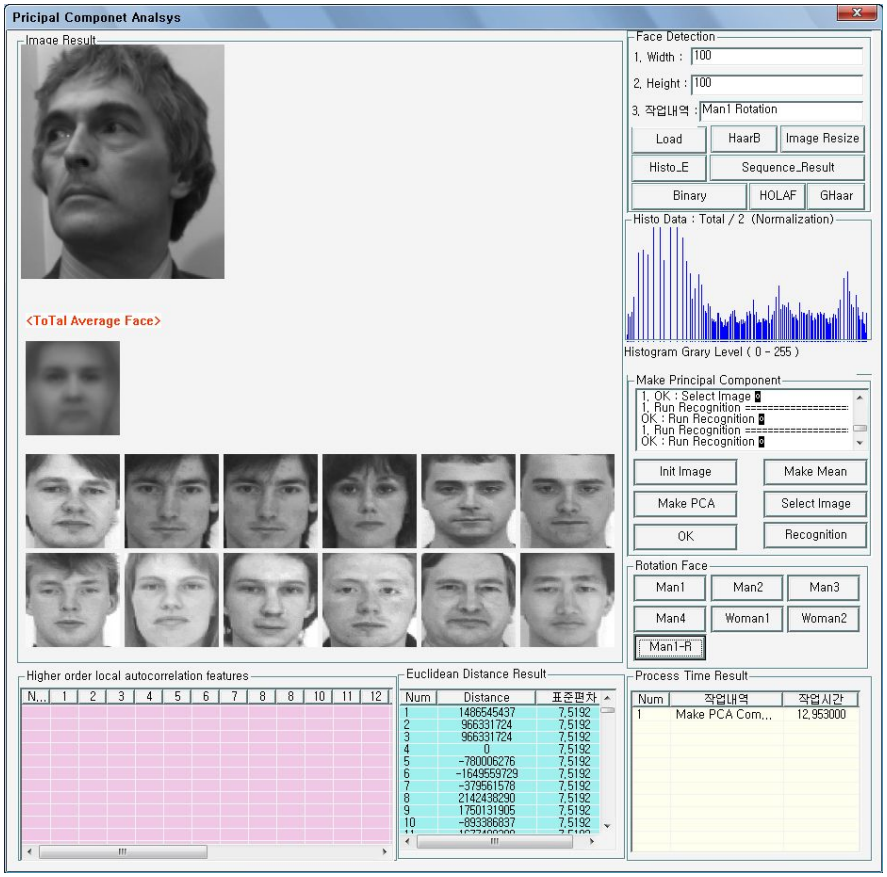

[그림 4-9] 평균 영상 생성 및 데이터베이스 구축 시간



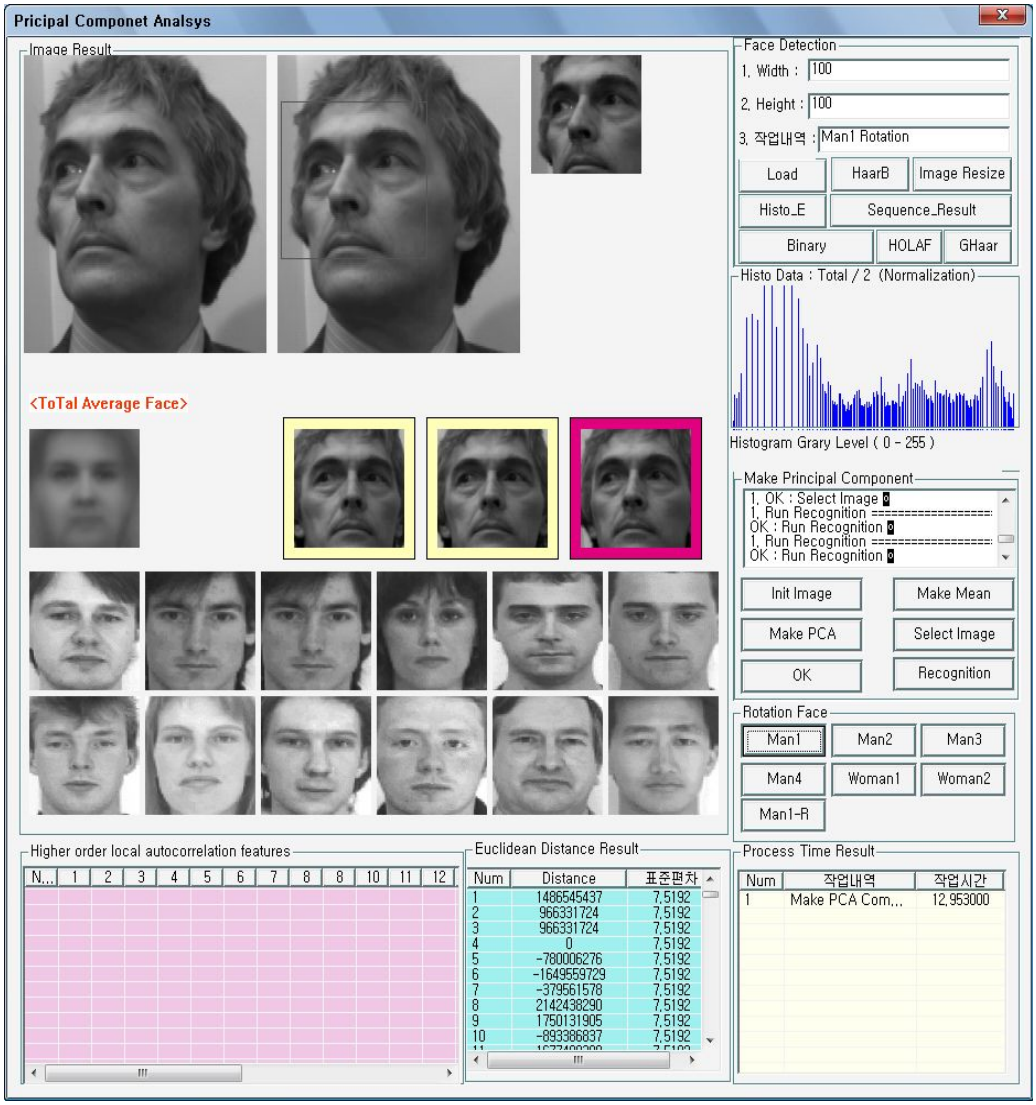

[그림 4-10] 오류역전파 알고리즘을 이용한 얼굴인식

[그림 4-8]은 전체 모델 영상을 초기화 하는 장면이고 [그림 4-9]은 주성분 분석을 하기 위해 평균 영상을 구한 다음 전체 모델영상의 고유 공간을 구성하는 시간을 나타낸 화면이다. [그림 4-10]은 오류역전파 알고리즘을 통해서 매칭을 성공해서 얼굴인식을 수행한 화면이다.

[그림 4-11]과 [그림 4-12]는 한 사람의 회전 이미지의 인식을 알아보기 위해 수행한 화면이다. 가장 유사한 값을 가지는 모델영상을 인식하여 그중 가장 최적의 값을 찾아가는 작업을 보여 주는 화면이다.




[그림 4-11] 서양 남자의 회전 영상 입력




[그림 4-12] 회전 이미지의 대한 얼굴 검출 및 얼굴인식

# V. 결 론

본 논문에서는 기존의 개인 식별 방법의 한계를 해결하는 대안으로 떠오르는 있는 생체 인식 기술 중 하나인 얼굴인식은 사람이 시각을 통해 인식하는 가장 직접적인 식별자로서 가장 보편적으로 인정받는 식별자이고 친숙성과 변별성이 탁월하여 활용이 매우 편하다는 장점으로 주목 받고 있지만 여타 생체 인식 기술과 비교해 보면 불확실한 성능과 기술적인 어려움으로 인하여 기술적 보급이 늦어지는 상태이다. 이러한 문제점들을 해결하기 위해 본 논문에서는 안정적이고 빠른 얼굴 검출방법과 안정적인 얼굴 인식 방법에 대하여 제안하였다.

첫째, 신뢰도 있고 안정적인 얼굴영역 검출을 위해 Haar-like Feature 와 AdaBoost 알고리즘을 제안하게 되었고, AdaBoost는 간단한 특징들을 인테그랄 이미지 이용하여 계산을 빠르게 하였고, 캐스케이드 구조는 스테이지별로 특징 수를 다르게 하여 초기 스테이지에서 많은 배경을 제거하여 계산 속도를 빠르게 하였다. 기존 얼굴 검출 방법보다 계산 비용을 줄였기 때문에 실시간 얼굴 검출 성능이 기존의 방법에 비해 탁월하게 향상되었다.

또한, 본 논문에서 제안한 얼굴 인증 시스템은 얼굴 영상 집합으로부터 고유벡터를 구하고 기여도가 큰 벡터만으로 구성된 고유공간을 이용하여 얼굴을 인식하는 방법인 주성분 분석을 제안하였다. 제안된 방법은 얼굴 영상의 구분뿐만 아니라 구체적인 정보인 얼굴 영상의 크기 정보나 회전 정보를 얻어내는데 유용하다는 것을 알 수 있었다. 기존의 기하학적 정보를 이용하거나 스테레오 영상을 이용하는 방법에 비해 처리가 간단하고 안정된 결과를 얻을 수 있음이 실험을 통하여 증명되었다.

그러나 주성분 분석은 영상의 조명변화가 있을 때 인식률이 떨어지는 경향이 있다. 특히 학습영상에서 각각의 사람마다 조명의 차이가 있을 경우 이것이 하나의 잘못된 특징점으로 작용할 우려가 있어서 인식률이 크게 감소하게 된다. 그래서 본 논문에서는 입력영상이 학습영상에 대해 조명의 차이가 있는 경우에도 데이터베이스 안의 누구인지를 인식하기 위해 히스토그램 평활화라는 방법을 사용하여 히스토그램을 펼침으로서 조명변화에 영향을 감소시키는 결과를 나았고, 기본적인 PCA 방법과 휘도치 정규화를 이용한 방법 등과 비교해본 결과 거의 조명변화에 영향없이 일정한 인식률을 유지하였다. 또한, 인식 과정에서 투영된 입력 영상이 모델

영상과의 매칭의 경우, 여러 얼굴들이 투영되어 있는 얼굴 영상 공간에서 실제 얼굴 영상이 매칭에 성공했을지라도 다른 얼굴 영상으로 인식하는 잘못된 매칭이나 매칭실패가 발생할 수 있다.

이러한 문제를 해결하기 위해 본 논문은 다층 퍼센트론 학습인 오류 역전파 학습 알고리즘을 이용하였다. 주성분 분석을 통해서 얻은 특징 벡터를 산출하여 이를 신경망에 입력 데이터로 활용함으로써 차원을 낮추면서도 인식 성능을 높일 수 있는 방안을 제시하였다. 실험 결과 기존의 방식들과 비교하여 제안한 방법이 95.3%로 기존의 방법과 비교하여 평균 6.5%의 인식률을 향상시켰다. 특히 얼굴의 회전이미지와 같은 특징 데이터의 변화가 거의 없을 경우 잘못된 매칭이 빈번히 발생하게 되는데 이런 부분에서 본 논문이 제안한 방법은 2%의 낮은 실패를 나타내었다. 또한 다양한 환경에서의 학습률에 따른 인식률에 변화를 실험하여 가장 최적의 학습률의 값을 도출하였다.

향후 연구과제는 얼굴인식 기술을 위한  다양한 환경 변화에 강인한 얼굴 인식률을 높이는 방법이다. 다양한 환경 변화는 조명, 촬영 각도, 객체와의 겹침, 다양한 표정 변화 및 노화, 인종, 헤어스타일등의 내외적인 요인에 따라 얼굴인식률의 변화가 심하게 나타나기 때문에 다양한 환경에서도 신뢰도를 높이기 위한 연구가 수행되어 져야 할 것이다. 이러한 목표가 성취된다면 얼굴인식은 진정한 생체인식의 판단기준에 대한 기본 바탕을 제공할 수 있을 것이다. 나아가 컴팩트하고 효율적이며 낮은 가격의 구현을 유지하면서 발전이 이루어진다면 얼굴인식은 광범위하게 사용되어 질 것으로 사료된다.

## 참 고 문 헌


[1] M. Bennamoun, G. J. Mamic, Object Recognition, Springer, 2002, pp.3-28

[2] Y. Kuno, et al., "Model-based Vision by Cooperative Proessing of Evidence and Hypotheses using Configuration Spaces", SPIE Digital and Optical Shape Representation and Pattern Recognition, Vol. 938, 1988.

[3] S. C. Zhu, et al., "Integrating Bottom-up/Top-down for Object Recognition by Data Driven Markov Chain Monte Carlo", Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2000.

[4] 이성환, 이미숙, "얼굴 영상 인식 기술의 연구 현황." 대한전자공학회지, 제23권 제6호, 1996년 6월, pp.688-702

[5] 이성환, "Face Recognition: Theory and Practice," 한국정보과학회 컴퓨터비전 및 패턴인식 연구회 주관 생체인식 기술 워크샵 자료집, 2001년 2월, pp.59-86

[6] Simone Santini, Ramesh Jain, "Similarity Matching," Second Asian Conference on Computer Vision, pp.544-548, Singapore, 1995.

[7] R. Brunelli, T. Poggio, "Face Recognition through Geometrical Features," Proceedings of European Conference on Computer Visioin, pp.792-800, 1992.

[8] Ingemar J. Cox, Joumana Ghosn, Peter N, Yianilos, "Feature-Based Face Recognition Using Mixture Distance," Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.209-216, 1996.

[9] Thomas Maurer, Christoph von der Malsburg, "Single-View Based Recognition of Faces Rotated in Depth, International Workshop on Automatic Face and Gesture Recognition, pp.248-253, Zurich, 1995.

[10] S. S. Ipson, A. A. Somaie, W. Booth, "Face Identification Using 2-D Isodensity Moments," Second Asian Conference on Computer Vision, pp.484-488, Singapore, 1995.

[11] S. Y. Lee, Y. K. Ham, R. H. Park, "Recognition for Human Front Face using Knowlege-Based Feature Extraction and Neuro-Fuzzy Algorithm," Pattern Recognition Vol.29, No.11,pp.1863-1876,1996.

[12] Ying Dai, Yasuaki Nakano, Hidetoshi Miyao, "A Study of Face Recognition

with the Low Quality Image ," The Third International Conference on Automation Robotics and Computer Vision, pp.1442-1446, Singapore, 1994.

[13] Alex Pentland, Baback Moghaddam, Thad Stamer, "View-Based and Modular Eigenspaces for Face Recognition," M.I.T. Media laboratory Perceptual Computing Section Technical Report No.245, 1994.

[14] Osamu Nakamura, Shailendra Mathur, Toshi Minami, "Identification Of Human Face Based On Isodensity Maps," Pattern Recognition Vol.24, No.3, pp.263-272, 1991.

[15] Martin Bichsel, "Human Face Recognition : From Views to Models-From Models to Views," International Workshop on Automatic Face and Gesture Recognition, pp.59-64, Zurich, 1995.

[16] A.S. Georghiades, D.J.Kriegman,P.N.Belhumeur,"Illumination Cones for Recognition Under Variable Lighting : Faces," Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.52-58, 1998.

[17] P. Ekman and W.V.Priesen: Facial Action Coding System(FADS), Consulting Psychologist Press, Inc.,1978.

[18] C.C Chien, Y.J Chang and Y.C Chen, "Facial Expression Analysis Under Various Head Poses" Proceedings of 3rd IEEE Pacific-Rim Conf. on Multimedia, pp.16-18. 2002.

[19 J.C Chun, O.R Kwon, P.Park: A Robust 3D Face Pose Estimation and Facial Expression Control for Vision-Based Animation, LNCS Vol. 4351, pp.700-708, 2007.

[20] P. Michel, R. 따 Kaliouby : Real Time Facial Expression Recognition in Video using Support Vector Machines, ICMI, pp.258-264,2003

[21] M. Evgeny, P. Edmond C, "Tracking Facial Features with Occlusions", Journal of Zhejiang University SCIENCE A, vol. 7, No. 7, pp.1282-1288, 2006

[22] Parcdas and Bonafonte, M. Pardas, A. Donafonte, "Facial Animation Parameters Extraction and Expression Recognition Using Hidden Markov

Models", Signal Processing : Image Communication 17, p.675-688. 2002.

[23] Y. Zhu, L. C. De Silva, C. C. Ko, "Using Monent Invariants and HMM in Facial Expression Recognition", Pattern Recognition letters Archive Volume 23, 83-91, 2002.

[24] Y.Zhu, L. C. DE. Silva, C.C. KO, "A Solution for Facial Expression Representation, and Recognition", Signal Processing : Image Communication 17, 57-673, 2002.

[25] Yoh-Han Pao, "Adaptive Pattern Recognition and Neural Networks," pp. 172-178, Addison-Wesley, 1989.

[26] Tatsumi Sakaguchi, Shigeo Morishima, Jun Ohya, Fumio Kishino, "Facial Expression Recognition from Image Sequences Using Hidden Markov Model," Second Asian Conference on Computer Vision, pp.713-717, Singapore, 1995.

[27] J. T. Tou, R. C. Gonzales, "Pattern Recognition Principles," pp.76-86, Addison-Wesley, 1974.

[28] K. P. Min, J.C Chum, G.R Park, " A Nonparametric Skin Color Model for Face Detection from Color Images", LNCS Vol, 3320, pp. 115-119, 2004

[29] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modula eigenspace for face recognition," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.84-91, Seattle, Washington, June 1994.

[30] D.L Swet and J.J Weng, "SHOSLIF-O: SHOSLIF for object recognition and image retrieval(phase II)," Technical report 첸95-39, Department of Computer Science, Michigan State University, Oct. 1995.

[31] C. Chatterjee and V. P. Roychowdhury, "On self-organizing algorithms and networks for class=separability feature," IEEE Trans. on Neural Networks, vol. 8, no. 3, May 1997.

[32] M. Black, "Robust incremental optical flow," PhD thesis, Yale University, 1992

[33] J.L. Barron, and N.A Thacker, "Tutorial: Computing 2D and 3D Optical

Flow", Tina Memo No. 2004-012, 2005

[34] P. Viola And M. Jones, "Rapid Object Detection using Boosted Cascade of Simple Features," Proceeding of IEEE International conference on Computer Vision and Pattern Recognition, vol.1, pp. I-511-I-518, 2001

[35] M. H, Yang, D. J. Kreiegman "Detecting Faces in Image: A Survey," IEEE Trensactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 1, pp. 34-58, 2002

[36] Y. Wang, Video Processing and Communication, Prentice Hall, 2002.

[37] Eli Saber, A. Murat Tekalp, "Frontal-view Face Detection and Facial Feature Extraction Using Color, Shape and Symmnetry Based Cost Function," Pattern Recognition Letters 19, pp.669-680, 1998.

[38] Eli Saber, A. Murat Tekalp, "Face Detection and Facial Feature Extraction Using Color, Shape and Symmenery-Based Cost Function," Proceedings of the INternational Conference on Pattern REcognition, pp.654-658,1996.

[39] Shen, K. J. and Fu, H. C. and Xu, Y. Y. and Hsu, F. R. and Chang, H. T. and Meng, W. Y., 1996, A principal component based probabilistic DBNN for face recognition, Image Processing, 1996. Proceedings., International Conference on Volume: 3, pp. 499-502

[40] P. Viola and M. Jones, "Robust Real-time Object Detection", IEEE ICCV Workshop on Statistical and Computational Theories of Vision, Vancouver, 2001

[41] Andrew D. Wilson, Aaron F. Bobick, "Recognition and Interpretation of Parametric Gesture", ICCV, 1998

[42] J.Weng, N.Ahuja, and T.S.Huang, "Learning recognition and segmentation of 3-D object from 2-D images." Proc. of Fourth Int'l Conf. on Computer Vision, pp. 121-128, Belin, May 1993.

[43] David F. Ferraiolo, Janet A. Cugini, D.Richard Kuhn, "Role-Based Access Control(RBAC): Features and Motivations", 「Proceedings of the 11th Annual Computer Security Applications Conferences」, December 1995, pp. 241-248

[44] S.Z. Li, L. Zhu, Z.Q. Zhang, and H.J. Zhang, "Statistical Learning of

Multi-View Face Detection", In *Proc. 7th European Conference on Computer Vision,* Copenhagen, Denmark. May 2002.

[45] Geof H. Givens J. Ross Beveridge, Bruce A. Draper and David Bolme, "Using A Generalized Linear Mixed Model to Study the Configuration Space of a PCA+LDA Human Face Recognition Algorithm", Technical Report, Computer Science, 2003.

[46] Wilson, H. R., Loffler, G., & Wilkinson, F. (2002). Synthetic faces, face Cubes, and the geometry of Face Space. *Vision Research, 42,* 2909-2923.

[47] Bartlett, M. S. B., Movellan, J. R. & Sejinowski, T. J. (2002). Face recognition by independent component analysis. *IEEE Transactions on Neural Networks, 13,* 1450-1464.

[48] Xiao-Peng, D.,Hua-wei, W., Huang, H. : Novel Method for Background Modeling and Update. Application Research of Computers, pp141-239, May, 2005

[49] Jong-Min Kim, and Hwan-Seok Yang, "A Study on Object Recognition Technology using PCA in Variable Illumination", Lecture Notes in Computer Science LNAI 4093 pp911-918, 2006

[50] Liang Xiao, Zhihui Wei, and Huizhong Wu, "Robust Orientation Diffusion Via PCA Method and Application to Image Super-Resolution Reconstruction",Lecture Notes in Computer Science LNCS 4688 pp726-735, 2007