

다층신경망을 이용한 임의의 크기를 가진 얼굴인식에 관한 연구

A Study on Face Awareness with Free size using Multi-layer Neural Network

송홍복 · 설지환

Hong-Bok Song · Ji-Hwan Seol

동의대학교 전자공학과

요 약

본 논문에서는 실시간 폐쇄회로 화면으로 받은 컬러 이미지에서 얼굴영상을 추출하고 이미 지정된 특정인의 얼굴영상과 비교를 통해 지하철이나 은행 등 공공장소에서의 수배자 등 어떤 특정인을 검출하는 방법을 제안하고자 한다. 감시카메라의 특성상 화면속의 얼굴정보가 임의의 크기로 가변하고 영상 내에서 다수의 얼굴정보를 포함하고 있음을 가정할 때, 얼굴영역을 얼마나 정확하게 검색 할 수 있는가에 초점을 맞추었다. 이를 해결하기 위하여 F.Rosenblatt가 제안한 퍼셉트론 신경망 모델을 기초로 임의의 얼굴영상에 대한 20x20 픽셀로 서브샘플링을 사용한 규준화 작업을 통해서 전면얼굴에서와 같은 인식기법의 효과를 사용하고, 획득한 얼굴후보 영역에 대하여 조명이나 빛에 의한 외부환경의 간섭을 최소화하기 위하여 최적선형필터와 히스토그램 평활화 기법을 이용하였다. 그리고 불필요한 학습을 최소화하기 위하여 달걀형 마스크의 덧셈연산을 전 처리 과정에 추가하였다. 전 처리 과정을 마친 이미지는 각각 세 개의 수용필드로 쪼개어져 특정 위치에 존재하는 눈, 코, 입 등의 정보를 신경망 학습을 통해 최종 결정된다. 또한 각각 다른 초기값을 가지는 3개의 단일셋 네트워크 시스템을 병렬형태로 구성하여 결과의 정확도를 높여 구현하였다.

Abstract

This paper suggest a way to detect a specific wanted figure in public places such as subway stations and banks by comparing color face images extracted from the real time CCTV with the face images of designated specific figures. Assuming that the characteristic of the surveillance camera allows the face information in screens to change arbitrarily and to contain information on numerous faces, the accurate detection of the face area was focused. To solve this problem, the normalization work using subsampling with 20×20 pixels on arbitrary face images, which is based on the Perceptron Neural Network model suggested by R. Rosenblatt, created the effect of recognizing the whole face. The optimal linear filter and the histogram shaper technique were employed to minimize the outside interference such as lightings and light. The addition operation of the egg-shaped masks was added to the pre-treatment process to minimize unnecessary work. The images finished with the pre-treatment process were divided into three reception fields and the information on the specific location of eyes, nose, and mouths was determined through the neural network. Furthermore, the precision of results was improved by constructing the three single-set network system with different initial values in a row.

Key Words : Surveillance camera, Perceptron Neural Network

1. 서 론

21C를 시작하는 오늘날의 세계는 컴퓨터의 성능과 인공지능 기법처리에 대한 기술의 급속한 발전과 더불어 영상, 음성 등 다양한 멀티미디어 정보를 실시간 혹은 보다 빠르게 검출할 수 있는 연구가 활발히 이루어지고 있다. 이러한

다양한 멀티미디어 정보 중에서 개인의 신원 확인, 보안 시스템 등의 영역에서 컴퓨터를 이용한 자동화된 시스템에 대한 요구가 급증하게 되면서 지문, 홍채, 서명, 얼굴, 음성 등의 생체 정보를 이용한 인증 및 보안, 감시 시스템이 상용화되어 보급되고 있다. 그 중 얼굴인식을 통한 생체학적 인증은 지문, 홍채, 서명 등과 달리 사용자 하여금 자신의 신체 일부를 인식장치에 직접 접촉시키지 않아도 되고 인식시 사용자의 특정행위를 요구하지 않는 비강제적이고 거부감 없이 인증 및 감시를 수행할 수 있는 장점으로 인하여 출입통제 시스템, 보안 시스템, 특정인 검색 시스템 등의 많은 응용분야를 가지고 있다. 또한 이러한 기술들은 초고속 통신망의 확산과 더불어 범죄자를 식별할 수 있는 효과적인 수사 지원

접수일자 : 2004년 7월 15일
완료일자 : 2005년 3월 11일
본 논문은 2004학년도 동의대학교 자체 학술연구 조성비의 지원을 받아 작성하였음

시스템의 주요 기술로도 활용될 수 있다.

따라서 인증시스템에서는 보다 효율적인 영상 감시 시스템을 위한 알고리즘들의 필요성은 점차 커지고 있다. 이러한 영상 감시 시스템이 사용되는 환경은 일반적으로 실험실에서 처럼 잘 조정된 환경에서 만들기 어렵고 예외 상황 발생 시 즉각적인 대응이 요구되는 경우가 많으므로, 영상 감시 시스템에 사용되는 알고리즘은 환경의 변화에 대처하는 능력을 갖는 것이 요구되며 실시간으로 동작하는 것이 바람직하다.

얼굴 인식의 단계는 크게 3단계로 분류할 수 있다. 첫 번째 단계가 얼굴의 탐색 및 추출(Face detection)이며, 두 번째 단계는 얼굴을 인식에 적합하도록 정보를 감소함으로써 밀집된 데이터 형태로 만드는 것(Face encoding)이다. 마지막 단계는 얻어진 얼굴 정보 데이터를 이용하여 적절한 패턴 인식 기법을 통해 얼굴을 인식하는 단계이다. 즉 얼굴 영역을 먼저 찾고 눈, 코, 입과 같은 특징 점 구성 요소들을 추출 및 압축[1-3]하여 최종단계인 인식과 분류가 이루어지게 된다.

폐쇄회로 감시 카메라 등에서 움직이는 인간을 검출하고 얼굴인식에 적합한 좋은 품질의 얼굴 영상을 취득하는 일은 동영상기반 얼굴인식 응용분야의 전 단계로서 매우 중요한 부분이다. 하지만 여기에는 크게 두 가지 풀어야 할 문제를 가진다. 첫째는 다수의 사람들의 움직임이 있는 복잡한 임의의 영상에서 인식 대상만을 추출하는 일이다. 둘째는 얼굴 검출기의 성능에 관한 문제다. 또 한가지 실제적인 문제가 있다면 위의 사항들을 실시간으로 처리할 수 있어야 한다는 것이다. 위 문제들을 풀기 위한 노력은 각기 독립적으로 이루어져 왔다. 우선 움직이는 물체를 추출하기 위해서 크게 차영상[1]이나 광류를 이용[2]하거나 형판정합[3]을 사용하려는 시도가 있었으며 빠른 얼굴 검출을 위해서는 색상정보[4]를 이용하려는 시도가 주류를 이루었다.

본 논문에서는 실제 응용 가능한 분야에 초점을 맞추어 시스템환경을 제한함으로써 일반적으로는 접근하기 어려운 위의 문제들을 해결하려고 시도하였다. 이를 구현하기 위해서 F.Rosenblatt가 제안한 퍼셉트론(Perceptron)의 신경망 모델을 통하여 영상 내에 임의의 크기로 존재하는 얼굴위치 추적을 구현해 본다.

본 논문의 구성은 다음과 같다. 서론에 이어 제2장 얼굴인식기술에서는 얼굴인식을 이용한 개인식별 및 인증방법에 대하여 고찰하고 3장에서 현재까지 진행되어온 얼굴인식 기술과 제안하고자 하는 시스템의 관련연구의 이론적 배경을 서술하였고, 4장에서는 본 논문에서 제안하고자 하는 다층구조의 신경망 학습에 대한 대략적인 이론과 배경을 설명한다. 제 5장은 구현된 시스템의 다층 신경망을 이용한 얼굴인식의 전 처리 과정설명과 구현된 안면인식에 적용되는 알고리즘을 단계별로 설명하고, 본 논문에서 제안하는 신경망을 이용한 얼굴위치추적 시스템의 유용성에 대한 응용프로그램의 결과를 확인한 후 6장에서 결론을 맺는다.

2. 얼굴인식에 관한 연구

2.1 영상처리와 얼굴인식 기술

일반적으로 생체인식기술 중 얼굴인식기술의 장점은 다음과 같이 요약된다.

첫째는 보편성과 용이성이다. 일반적으로 사람들이 상대방을 인식할 때, 그 사람의 지문이나, 정맥이나 눈동자를 보고 자신이 아는 사람이라고 생각하지는 않는다. 사람이 시각적

으로 많은 사람을 구별해 내기 위해 사용하는 정보는 바로 얼굴이며, 가장 쉽게 확인 가능하고, 사람들에게 있어 가장 보편화 된 정보이다.

둘째는 유일성이다. 만일 누군가가 얼굴을 가리고 있다면 그가 누구인지 알기 어려우며, 범죄자 등 특정인의 얼굴사진에서 양쪽 눈만을 가리는 처리를 해도, 누구인지를 알기 어렵다는 것은 얼굴영역, 특히 두 눈을 포함한 영역이 사람의 유일성을 최대로 보유하고 있다는 예측을 가능케 해준다. 반면 사람의 얼굴은 한번만 봐도 다시 만났을 때 기억할 확률이 높는데, 이는 얼굴의 특징적 패턴이 갖는 유일성이 식별 능력을 향상시켜준 결과이다. 이는 얼굴인식 알고리즘에도 영향을 미친다.

셋째는 비 접촉성, 비 강제성, 감시성이다. 얼굴인식의 입력정보가 카메라를 통하여 얻게 되므로 근접식 또는 접촉식 기술이 요구되는 타 생체와 비교할 때, 기계에 어떤 접촉도 없으며 별도의 정보획득 물질이 필요 없이 단순히 카메라로 얼굴을 받아들여 검사한다는 데에 있어 편리하고 위험부담도 없다. 또한 감시 역할을 할 경우는 폐쇄회로 카메라 등으로 비교적 원거리에서 한층 고급의 감시 기능을 수행할 수 있다.

넷째는 정보 획득의 용이성이다. 예를 들어 범죄 사진 속에서 지문이나 음성이나 망막을 인식해 내어서 검사하는 것은 아주 힘들고 거의 불가능하다. 하지만 이러한 사진을 얼굴인식에 이용함으로써 적어도 유사한 범죄 후보자들을 추출해 낼 수 있고, 보안에서의 인력과 시간을 최대한 절약할 수 있다. 이는 몽타주 정도의 자료를 가지고도 기존 범죄 수사에 들이던 노력을 상당히 절약할 수 있으면서 큰 효과를 볼 수 있을 것으로 예상된다.

그밖에도 인식의 속도가 비교적 빠른 알고리즘 구현이 가능하다는지, 다수의 얼굴정보를 동영상 상태에서 동시에 인식 가능하다는지, 가정에 보급되는 PC기반의 고화질 카메라로 인해 인식장비의 비용부담을 덜 수 있다.

2.2 인간의 영상 인식과정

사람이 영상을 인식하는 과정을 살펴보면, 먼저 빛이 사람 안구의 망막 상에 맺히고, 망막에 있는 다수의 시신경 세포가 입력된 패턴에 의하여 자극 받아 전기 신호를 발생하고 이것이 신경에 의해 두뇌에 전달된다. 전달된 신호를 바탕으로 두뇌는 영상을 처리하여 인식한다. 이런 일반적인 사항은 알려져 있고 신경 회로망, 생명 공학 등에서 많은 연구가 진행되어 왔지만, 아직도 확실한 영상 인식 메커니즘은 모르는 상태이다. 인간의 영상 인식 과정을 지금까지 알려진 사실을 바탕으로 모델링 한 것이 그림1 이다.

전 처리 과정에서 입력된 패턴에 따라 잡음 제거, 확대, 축소, 회전등을 하여 패턴을 정규화가 이루어진다. 전 처리 후의 패턴에 대해서 특징을 추출하는 처리를 한다. 어떤 특징을 추출하느냐는 대상 패턴에 의존한다. 추출한 특징과 현재까지 기존의 패턴에 대해서 기억하고 있던 특징을 비교하여 가장 근접한 패턴을 찾아내어 입력 패턴이 어디에 속해 있는가를 찾아낸다.

2.3 카메라 기반 영상인식 시스템

일반적인 얼굴인식 영상처리의 기본과정은 인간의 영상인식과정처럼 입력영상→전처리→분할→표현 및 설명→인식 및 해석→결과로 진행된다.

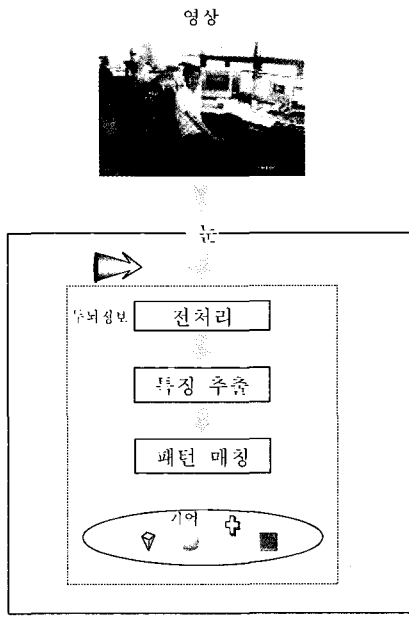


그림1. 인간의 영상인식 처리 과정

Fig. 1. Human's reflex awareness processing process

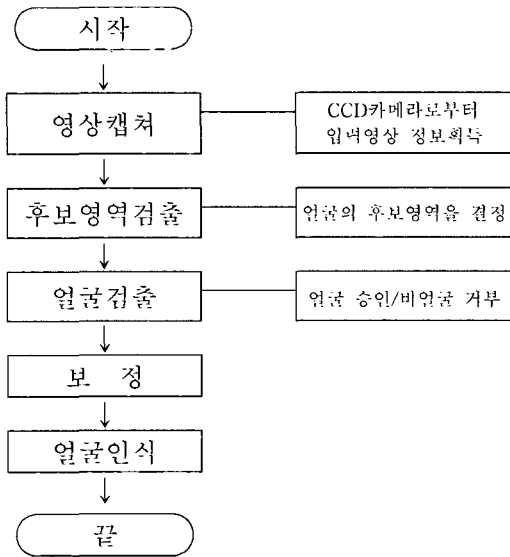


그림 2. 얼굴인식 시스템에서의 인식과정

Fig. 2. Deconvolution process in face awareness system

영상처리에서 현재 많은 얼굴인식 기법들은 비슷한 크기들의 전면 얼굴들(frontal faces)의 이용을 가정한다. 그런데, 이러한 방법은 얼굴외양과 환경조건들의 변화 때문에 충분히 만족되지 않는다. 그림 3에 있는 사진들을 고려하자. 이들 사진들은 얼굴 분류(face classification) 연구에서 사용되는 일반적인 테스트 이미지들이다. 이들 이미지들에서 배경의 배제(exclusion)는 신뢰할 수 있는 얼굴분류 기법들을 위하여 필요하다. 이러한 방법에는 2차 통계(second order statistics) 특징을 이용하여 이진영상에서 정합영역을 추출한 후 정합 하는 방법[5][6]과 얼굴을 구성하는 각 구성성분들의 원형을 정합 하는 방법[7][8] 및 위치관계[10 12]를 이용하는 방법 등이 있다. 2차 통계 특징을 이용한 정합방법의

경우, 정합 영역 추출 시 조명에 의한 특징영역의 크기 및 위치에 대한 변화에 민감하다는 단점이 있고, 원형정합 방법은 입력영상의 크기가 변할 경우 고정된 원형에 의한 인식이 어렵다는 단점이 있다.

그림 4에 있는 예제는 현실적인 응용 시나리오들에서 얼굴은 복잡한 배경에서 그리고 많은 다른 위치들에서 나타날 수 있다. 표준적인 얼굴 이미지들에 기초하는 인식 시스템들은 배경의 어떤 영역들을 얼굴로서 잘못 인식할 수 있다. 이러한 문제를 해결하기 위해서 배경으로부터 얼굴영역을 극대화하고 추출할 필요가 있다. 본 논문에서는 임의의 영상 내에 존재하는 임의의 얼굴크기와 다중 얼굴을 검출하는데 초점을 맞추었다.



그림 3. 얼굴인식을 위해 사용되는 전형적인 이미지들.

Fig. 3. Typical images which is used for face awareness



그림 4. 실제적인 얼굴 검출 시나리오

Fig. 4. Actual face detection scenario

3. 얼굴인식을 위한 모델

얼굴 검출 방법은 여러 가지 분야로 나뉘질 수 있고 목표하는 바에 따라 적절한 방법을 사용한다. 얼굴을 검출하는 방법은 대개 2가지 범주로 분류될 수 있다. 첫 번째는 얼굴 지식을 이용하여 저 수준 특징들 예를 들면 (skin color)과 얼굴형태(face geometry)와 같은 얼굴의 외양 특성들(apparent properties)을 이용하여 검출하는 것이다. 일반적으로 이들 기법들에서 얼굴검출 작업들은 장면으로부터 얻어진 시각적 특징들의 거리, 각도를 그리고 면적 측도 등을 다룸으로써 수행된다. 특징들은 주요한 구성요소들이기 때문에 이들 기법들은 특징-기반의 방법(feature-based approach)이라 불린다. 이들 방법들은 1970년대 초에 시작한 얼굴검출 연구로서 주로 이런 형태가 주를 이루었다.

두 번째 방법은 현재 주로 연구되는 패턴인식을 이용한 얼굴검출이다. 예를 들어 2차원 배열의 얼굴 이미지들은 특징도출(feature derivation) 및 분석 없이 훈련 알고리즘들을 사용하여 얼굴그룹으로 직접 분류된다. 특징기반의 방법과 다르게 이러한 상대적으로 새로운 기법들은 매핑(mapping)과 훈련체계(training schemes)를 통해서 얼굴 지식을 인식하게 된다.

3.1 특징 기반 방법

3.1.1 저 수준 해석

특징 기반 방법의 발전은 세 가지 분야들로 더 나누어질 수 있다. 혼잡한 장면 속에서 얼굴의 위치를 결정하는 전형적인 얼굴검출 문제가 주어지면 먼저 저 수준 해석이 그레이-스케일 및 컬러와 같은 픽셀 특성들을 사용하여 시각적 특징들(visual features)의 분할을 다룬다. 저 수준 본질 때문에 이 분석으로부터 생성되는 특징들은 분명치 않다. 특징해석(feature analysis)에서 시각적 특징들은 얼굴형태(face geometry)의 정보를 사용하여 얼굴 및 안면 특징들의 더 전역적인 개념(global concept)으로 조직된다. 특징해석을 통해서 특징의 모호함이 감소되고 얼굴 및 안면 특징들의 위치들이 결정된다.



그림 5. 색상변환에 의한 눈 위치 강조결과 영상
Fig. 5. Eye position Gangjogyeol and reflex by color conversion

3.1.2 특징 해석(Feature Analysis)

저 수준 해석으로부터 생성되는 특징들은 조금 애매하다. 예를 들어서 피부컬러 모델을 사용하여 안면 영역들의 위치를 결정하는 것에서 유사한 컬러의 배경 객체들 또한 검출될 수 있다. 많은 얼굴검출 기법들에서 얼굴형태의 지식은 그러한 애매한 상태에서부터 다양한 특징들을 특성화하고 이용하기 위해서 사용된다. 과거 몇몇 논문들에서는 얼굴형태의 응용과 관련한 두 가지 방법들이 존재한다. 첫 번째 방법은 그림 6에서처럼 개인의 안면 특징들의 상대적인 위치결정(positioning)에 기초하는 순차적 특징 탐색(feature searching) 하는 방법을 이용하는 것이다. 특징 존재성(feature existence)의 신뢰도(confidence)는 근처에 있는 특징들의 검출에 의해서 향상된다. 두 번째 방법은 다양한 얼굴모델들을 사용하는 유연한 배열들(flexible constellations)을 특징으로 한다.



그림 6. 기하학적 특징 각을 이용한 얼굴인식
Fig. 6. Face awareness which use geometrical characteristic horn

3.1.3 능동형상 모델들(Active Shape Models)

능동형상 모델들은 특징들의 실제적인 물리적 및 더 높은 수준의 외관을 묘사한다. 일단 특징에 대해 가까운 근접(proximity) 내에서 방출되던 활성형상 모델(active shape model)은 국소 이미지 특징들(모서리들, 밝기)과 상호작용하고 점차적으로 특징의 형상을 얻기 위해서 변형된다. 현대의

안면 추출연구에는 일반적으로 세 가지 종류의 활성형상 모델들이 존재한다. 첫 번째 종류는 Kass et al.에 의해서 1987년에 처음 소개된 스넥스(snakes)라 불리는 일반적인 능동윤곽(generic active contour)을 사용한다. 그 뒤에 안면 특징들의 선형성(a priori)을 고려하고 스넥스의 성능을 높이기 위해서 변형 가능한 템플릿들(deformable templates)이 Yuille et al.[7]에 의해서 소개되었다. 그 후에 Cootes et al.는 사람 얼굴의 효율적인 해석을 제공하기 위해서 그들이 스마트 스넥스(smart snakes) 및 PDM으로 명명한 새로운 일반적이고 유연한 모델의 사용을 제안하였다. Cootes et al.의 모델은 훈련절차에 따라서 오직 어떤 형상들로만 변하는 것이 허용되는 라벨이 붙은 점들의 집합에 기초한다.

3.2 이미지 기반 방법

3.2.1 선형 부분공간 방법들(Linear Subspace Methods)

특징해석 얼굴 검출 방법들의 단점은 안면 특징들의 명시적인 모델 링에 의한 얼굴검출이 얼굴의양과 환경조건들의 비 예측성에 의해서 곤란함을 겪는다는 것이다. 비록 최근에 몇몇 특징-기반 시도들이 비 예측성(unpredictability)을 다룰 수 있는 능력을 향상시켰지만 대부분은 여전히 머리와 어깨 및 준-정면(quasi-frontal) 얼굴들(또는 조합된 시스템에서 기법들의 하나로서 포함되는)로 제한된다. 클러터-밀집된 배경들(clutter-intensive backgrounds)을 갖는 다중 얼굴들을 검출하는 것과 같은 더 어려운 시나리오들에서 수행할 수 있는 기법들을 위한 필요성이 여전히 존재한다. 이러한 요구는 얼굴검출을 패턴인식 문제로서 다루는 새로운 연구 분야를 고무하였다. 문제를 예제들로부터 얼굴 패턴을 인식하기 위해 학습하는 것으로 공식화함으로써 얼굴지식의 특정한 적용이 회피된다. 이것은 불완전하거나 부정확한 얼굴지식에 기인하는 잠재적인 모델링 에러를 제거한다. 얼굴패턴들을 인식하는 기본적인 방법은 예제들을 얼굴 및 비-얼굴 프로토타입 클래스들로 분류하는 훈련 과정을 거치는 것이다. 이들 클래스들과 입력 이미지로부터 추출된 2차원 세기배열(intensity array)(이것이 이미지-기반이라는 이름이 붙은 이유이다) 사이의 비교는 얼굴존재의 결정이 이루어지는 것을 허용한다. 가장 간단한 이미지 기반의 방법은 템플릿 정합(template matching)에 의존하지만 그러나 이러한 방법들은 다음절들에서 제시되는 더 복잡한 기법들만큼 잘 수행하지 않는다.

이미지 기반 방법들 대부분은 얼굴들을 검출하기 위하여 윈도우 스캐닝 기법(window scanning technique)을 적용한다. 윈도우 스캐닝 알고리즘은 본질적으로 모든 스케일들에서 가능한 얼굴 위치들을 위한 입력 이미지의 철두철미한 탐색이지만 거의 모든 이미지 기반 시스템들을 위한 이 알고리즘의 구현에서의 변형들이 존재한다. 전형적으로 스캐닝 윈도우의 크기, 하위샘플링 율(subsampling rate), 스텝 크기(step size), 그리고 반복횟수들은 제안되는 방법과 계산적으로 효율적인 시스템을 위한 필요성에 의존하여 변한다. 이미지-기반 방법들을 대략적으로 선형 부분공간 방법들(linear subspace methods), 신경망(neural networks), 그리고 통계학적 접근법들로 나눈다.

3.2.2 신경망 (Neural Network)

신경망은 얼굴검출을 포함하여 패턴인식 문제들을 위해서 널리 사용되는 기법이 되었다. 얼굴 검출에 쓰이는 신경망은 매우 복잡하다. 모듈방식 아키텍처들, 복잡한 학습 알고리즘

들 및 압축 네트워크들 그리고 유전 알고리즘들(genetic algorithms)을 가지고 진화되거나, 가지치기(prune)되는 네트워크들 모두가 패턴인식에서 널리 사용되는 신경망들의 예들이다. 이것은 신경망이 얼굴인식을 위한 시스템의 모든 부분들을 위하여 적용될 수 있다는 것을 함축하며, 이 사실을 몇몇 논문에서 게재하고 있다. 얼굴인식을 위한 다소 기초적인 신경망 방법들에 대한 소개는 Viennet[12] 및 Fougelman Soulie[13]에서 찾아볼 수 있다.

얼굴검출을 위한 최초의 뉴럴 기법들은 다층 퍼셉트론(MLPs)에 기초하였는데 가능성 있는 결과들이 매우 간단한 데이터집합들 상에서 보고되었다. 크고 어려운 데이터집합 상에서 결과들이 보고된 첫 번째의 고급 뉴럴 기법은 Rowley et al.[12]에 의해서 수행된 연구였다. 이러한 시스템은 연결된 신경망에 얼굴지식을 결합한다. 신경망은 20x20 픽셀들(즉, 400개의 입력유닛들) 크기의 윈도우들을 탐색하도록 설계된다. 그 곳에는 26개의 유닛들을 갖는 하나의 은닉 층이 있는데 그 중에서 4개의 유닛들은 10x10 픽셀 부분 영역들을 탐색하고 16개의 유닛들은 5x5 크기의 부분 영역들을 탐색하며 6개의 유닛들은 수평 줄무늬들(horizontal stripes)이 겹치는 20x5 픽셀들을 탐색한다. 입력 윈도우는 조명 교정(lighting correction)에 의해 최적 선형함수가 뿔셈되고 히스토그램 균등화(histogram equalization)를 통해서 전 처리된다. 여기서 윈도우 스캐닝 기법들을 사용할 때 발생하는 문제는 중복검출들(overlapping detections)이다.

3.2.3 통계적 방법들 (Statistical Approaches)

선형 부분 공간 법들 및 신경망들과는 별도로 얼굴검출을 위한 몇 가지 다른 통계적 방법들이 존재한다. 정보이론(information theory), Bayes의 결정규칙(decision rule)에 기초하는 시스템들이 이 절에서 제시된다.

Osuna et al.에서 SVM(support vector machine)이 얼굴검출을 위하여 적용되었다. 제안된 시스템은 이전의 절에서 설명된 Sung[13] 및 Poggio[15]에 의해서 개발된 것과 동일한 틀(frame-work)에 따른다(19x19 크기의 윈도우를 가지고 입력 이미지들을 스캐닝). 커널함수로서 2계 다항식(2nd-degree polynomial)을 갖는 SVM은 전역적 최적성을 보장하는 분해 알고리즘(decomposition algorithm)을 가지고 훈련된다. 훈련은 부트-스트랩 학습 알고리즘을 가지고 수행되며 이미지들은 몇 가지 절차를 가지고 전 처리된다. Kumar 및 Poggio는 최근에 얼굴들의 실시간 추적과 분석을 위하여 Osuna et al.의 SVM 알고리즘을 시스템에 결합하였다. 그들은 과도한 스캐닝(exhaustive scanning)을 피하기 위해서 입력 이미지들에서 분할된 피부영역들에 SVM 알고리즘을 적용한다. SVMs은 또한 다른 부분들을 위하여 독립된 SVMs을 생성함으로써 다중 뷰 얼굴검출을 위하여 사용되었다. Terrillon et al.에서 SVMs은 더 초기에 사용된 다층 퍼셉트론에 비교할 때 얼굴검출기의 성능을 향상시켰다.

4. 인공 신경망

인간의 뇌는 계산, 연산, 추론 등 매우 복잡한 일들을 수행하지만 컴퓨터처럼 CPU나 메모리가 아닌 신경 세포인 뉴런들의 연결 관계에 의하여 이루어진다. 어떠한 특정 기능을 수행하는 부위가 한 세포에 집중이 된 것이 아니라 여러 개의 신경세포의 조합이 그 기능을 수행한다. 예를 들어 컴퓨터는 수많은 부품 중에서 단 한 개만이라도 손상되면 시스템

기능에 큰 영향을 주어 전체 시스템 작동에 커다란 영향을 줄 수 있는 반면에 인간 뇌는 세포 한 두 개가 손상된다 하더라도 한꺼번에 모든 기능을 잃어버리는 일은 거의 없다. 손상된다 하더라도 기능 저하가 서서히 일어나거나 그 손상된 정도가 미미하다. 이러한 증상을 볼 때, 특정 기능을 담당하는 뇌의 부위는 한 곳이 아닌 매우 많은 신경 세포들의 조합들이 한 가지 혹은, 여러 가지의 기능을 담당 한다고 볼 수 있다. 이러한 인간의 신경 세포를 수학적으로 모델링 한 것이 바로 Neural network 이다.

축색돌기(Axon), 시냅스(Synapse), 체세포(Soma), 수상돌기(Dendrite), 세포핵으로 구성된 신경 세포는 다음과 같은 특징을 갖고 있다. 신경 세포는 일정 크기 이상의 자극이 주어질 때만, 그에 대한 반응을 나타내고 신경 세포와 신경 세포간의 자극은 synapse에서 분비되는 신경 호르몬의 정도에 따라서 다른 세포로 전달된다. 이러한 특징을 고려할 때, 다음과 같은 수학적 모델을 세울 수 있다.

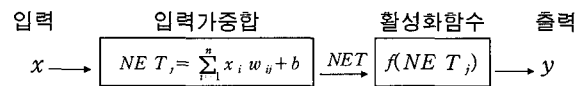


그림 7. 뉴런의 기능 모델
Fig. 7. Function model of neuron

위의 수식은 일정 크기 이상의 자극이 주어져야 반응하고 그 입력된 신호가 출력되고 반응하는 것을 모델링 한 것이다. 한 개의 뉴런에 수많은 뉴런이 상호 연결되어 있음을 가중합으로 표현하였고, 그 자극들의 합이 반응하는 정도를 활성화 함수를 통하여 표현하였다.

뉴런에 입력되는 외부 자극의 가중합 NET와 뉴런의 출력 y는 다음과 같이 표현할 수 있다.

$$NET_j = \sum_{i=1}^n x_i w_{ij} + b \tag{1}$$

위의 수식(1)은 일정 크기 이상의 자극이 주어져야 반응하고 그 입력된 신호가 출력되고 반응하는 것을 모델링 한 것이다. 한 개의 뉴런에 수많은 뉴런이 상호 연결되어 있음을 가중합으로 표현하였고, 그 자극들의 합이 반응하는 정도를 활성화 함수를 통하여 표현하였다. 뉴런에 입력되는 외부 자극의 가중합 NET와 뉴런의 출력 Y는 다음과 같이 표현할 수 있다.

$$Y = f(NET_j) \tag{2}$$

활성화 함수로는 사인함수(sign function), 스텝함수(step function), 시간별 스텝함수(time step function)등을 사용할 수 있다.

단층 퍼셉트론은 선형 분리문제는 해결 할 수 있지만 그 보다 복잡한 조각선출(piece-wise)문제, XOR 문제는 해결할 수 없었다. 이에 이 문제를 해결 할 수 없었던 동안 신경망의 발전 사에 암흑기가 존재하였다. 하지만, 이 문제는 퍼셉트론의 층을 한층 이상으로 설계함으로써 해결 할 수 있었다. 다음 그림8과 같이 입력단(input layer)과 출력단(output layer) 사이에 은닉층(hidden layer)을 둔 다층 퍼셉트론(multi-layer perceptron)은 XOR 문제뿐만 아니라, 선형(linear) 함수만으로는 해결 할 수 없는 비 선형(non-linear) 함수까지도 해결함으로써 다양한 분야에 적용되었다.

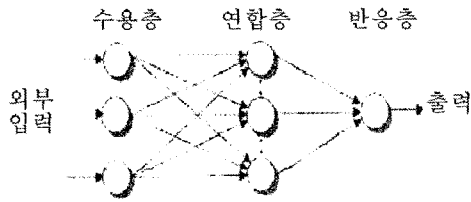


그림 8. Rosenblatt 의 Perceptron
Fig. 8. Rosenblatt's Perceptron

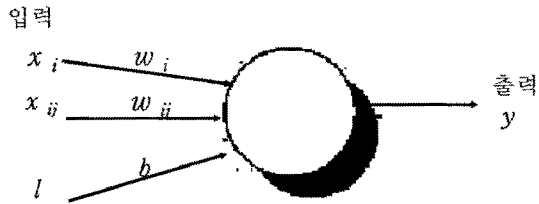


그림 9. Perceptron의 구조
Fig. 9. Perceptron's structure

4.1 단층퍼셉트론(Single Layer Perceptron)의 한계

1969년 Marvin Minsky와 Seymour Papert는 《Perceptrons》 라는 책을 발표하였는데, 이 책에서 퍼셉트론 모델이 선형분리 기능밖에 없고 많은 실제 문제를 해결하지 못한다는 것을 증명하였다. 한 예로, 2 개의 2 진 패턴을 2 개의 종류로 분류하는 문제 중 2 개가 같은 것을 한 종류 (B)로, 다른 것을 2 번째 종류 (A)로 분류하는 XOR 문제는 2 차원 공간에서 최소 2 개의 직선 경계가 필요하고 한 개의 직선으로 경계를 만들 수 없으므로 퍼셉트론 문제로 해결하지 못한다.

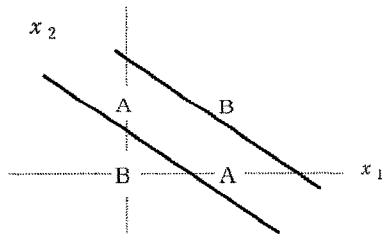
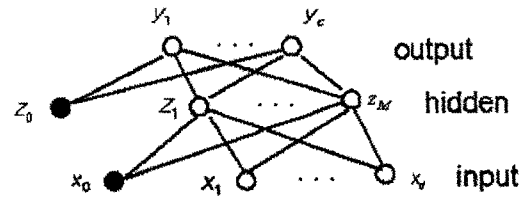


그림 10. 단층 perceptron의 한계
Fig. 10. Single-layer perceptron's limit

4.2 다층 퍼셉트론

선형 분리 문제만을 해결할 수밖에 없는 퍼셉트론을 다시 부흥기로 접어들게 했던 것은 오류 역전파 (error backpropagation) 에 의한 다층구조 퍼셉트론(multi-layer perceptron) 의 학습방법 발견이라 할 수 있다. 즉, Minsky 와 Papert에 의해 퍼셉트론이 할 수 없다고 증명된 일반적인 패턴의 분류를 다층구조 퍼셉트론이 할 수 있음이 증명되고 오차역전파 학습법칙에 의해 보여 줄 수 있게 되었다. 사실 1960년대 Minsky 와 Papert 에 의해 퍼셉트론의 가능성은 알고 있었으나, 이를 훈련시킬 수 있는 적절한 학습법칙이 알려지지 못했다.



$$y_k = \tilde{g} \left(\sum_{j=0}^M w_{kj}^{(2)} g \left(\sum_{i=0}^d w_{ji}^{(1)} x_i \right) \right) \quad (3)$$

그림 11. 다층 퍼셉트론
Fig. 11. Multi-Layer Perceptron

이처럼 그림 12에서 보듯이 다층 퍼셉트론의 응용은 XOR 문제뿐만 아니라 조각선출 문제의 형태도 해결 할 수 있다.

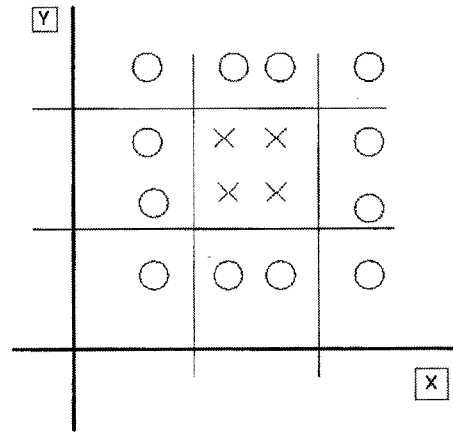


그림 12. Piece-wise 문제
Fig. 12. Piece-wise problem

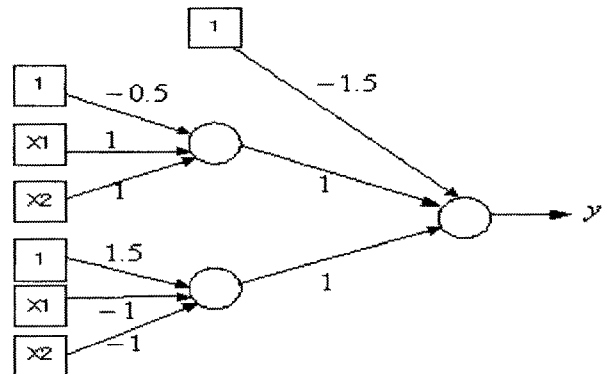


그림 13. 일반적인 Perceptron 구조
Fig. 13. General Perceptron rescue

즉, 그림 13의 퍼셉트론형태는 다음의 그림14와 같은 형태와 같이 한 개의 은닉 층을 가진 다층 퍼셉트론으로 바뀌어 나타낼 수 있다.

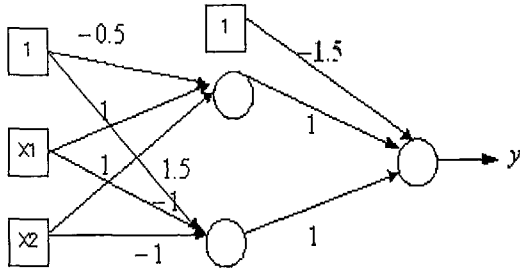


그림 14. 다층 Perceptron 구조
Fig. 14. Multi-layer perceptron

구조	결정영역	XOR 문제	맞물린 형태	일반적인 형태
단층 	반 평면			
2층 	불록다각형			
3층 	임의의 형태			

그림 15. 단층 퍼셉트론 및 다층 퍼셉트론 구조에 의해 형성되는 결정 경계

Fig. 15. The decision border that is formed by fault Peosepteuron and Dacheung Peosepteuron structure

그림 15는 단층 퍼셉트론의 한계를 다층 퍼셉트론으로 극복할 수 있는 방법을 보여준다.

4.3 Back Propagation Algorithm

다층 퍼셉트론에서 연결된 강도 가중치(weight)들의 값을 찾는 방법으로 오류 역전파 알고리즘 (Backpropagation: BP)을 사용한다. 이 방법은 델타-룰에 의해서 일반화되었으며 GDR(generated-delta-rule)이라고도 불린다.

BP 신경망은 지도학습(supervised learning)패턴의 대표적인 신경망 모델로서, 이는 오차 역전파라고 하는 기법을 프로세싱 유닛에 적용한 것이다. 출력 층의 각 프로세싱 유닛들은 목표로 하는 출력 값에 따른 오차신호를 단거리로, 각 프로세싱 유닛간의 연결강도(weights)를 수정함으로써 다음 학습시 목표 값에 더욱 접근된 출력 값을 갖게 한다. 출력 값이 목표 값과 유사하게 될 때까지 학습을 반복하게 되며, 학습이 끝나면 학습한대로 출력을 하게된다. 따라서 BP신경망의 학습에 상당한 시간이 소요되지만 일단 학습이 끝나면 이 용단계에서는 매우 빠른 결과로 출력된다.

오류 역전파 알고리즘은 다음과 같은 특징, 조건을 지닌 다.

- 미리 정해진 입출력의 한 쌍이 한 세트를 이룬다.
- 사이클이 바뀔 때 2상 증식.
- 입력패턴은 각 항이 출력 층까지 네트워크층이 만들어 지게 한다.

- 이 아웃풋 패턴은 각 출력 유닛을 오류신호를 위해 비교한다.
- 그때 오류신호가 바로 출력에 기여하고 있는 중간 수준의 계층의 각 노드 뒤로 출력계층으로부터 전하여 진다.
- 중간 수준의 계층의 각 노드는 유닛이 한 상대적인 원래의 출력에의 기여에 기초를 둔 전체적인 오류 신호의 일부만을 받는다.
- 오류 신호 상에서 그때 받은 연결강도(가중치)가 업데이트 된다.
- 훈련 후 잡음이 섞였거나 불완전한 임의의 입력패턴을 맞이했을 때 네트의 은닉 층들에 있는 유닛들은 새로운 입력이 개개의 유닛들이 학습하는 동안에 인식되도록 학습된 특징을 담은 패턴을 포함하면 액티브출력을 낼 것이다. 반대가 되면 출력을 억제(inhibit)시킬 것이다.
- 입력 벡터를 네트워크에 적용하면 대응한 아웃풋 값을 계산한다. 실제 아웃풋이 적당한 아웃풋과 비교하면 어느 정도의 오류를 결정한다. 모든 벡터를 위해 놓여진 트레이닝에 오류가 받아들일 만한 값(value)으로 만들어 질 때까지 각 연결강도를 바꾼다. 이것을 모든 트레이닝 벡터를 가지 고 수정을 연결강도 반복 아이টে에 적용하는 것을 결정한다.

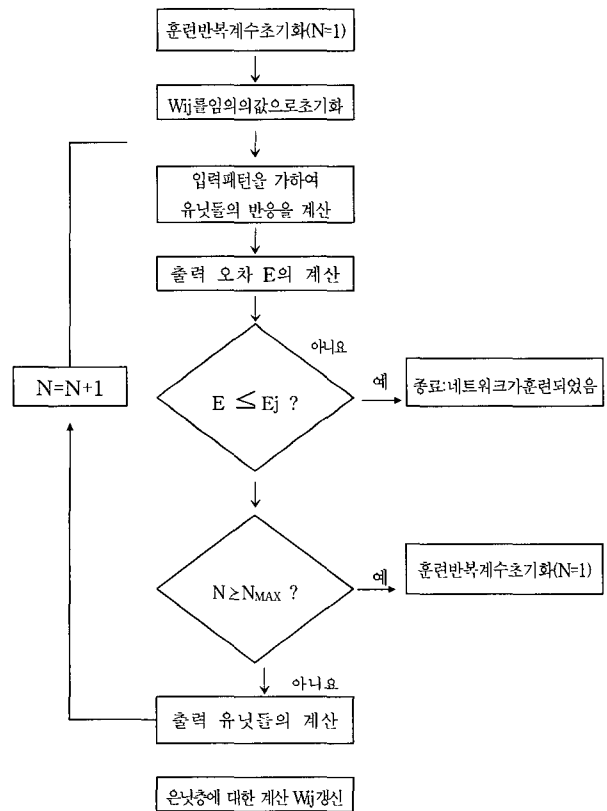


그림 16. 역전파 알고리즘의 순서도

Fig. 16. Flowchart of hard fight welsh onion algorithm

4.4 서브 샘플링(Sub-sampling) 기법

서브 샘플링이라 하면 주어진 정보의 일부분만을 사용하는 것을 말한다. 입력받은 영상신호가 디지털화 할 경우 픽셀은 RGB의 3가지 성분으로 표시된다. 그러나 같은 색을 표시하는 경우 RGB 대신 YUV로 표시할 수 있다. YUV에서 Y는 밝기를 나타내고 U와 V는 색상 정보만을 포함하고 있

다. 사람의 눈은 밝기에는 민감하나 색상 정보에는 상대적으로 둔감하다. 따라서 Y정보를 4번 사용할 때 U나 V 정보를 2번씩 사용하면(Y:U:V=4:2:2) 전체적으로 정보량은 원래 정보의 2/3으로 줄어들게 된다. 이 경우 4:2:2 서브 샘플링이라 하며 가장 널리 사용된다.

서브 샘플링은 다른 형태로도 사용된다. RGB를 각각 8비트씩 총 3바이트를 사용하는 대신 5비트씩 사용하여 2바이트만을 사용하면 역시 2/3으로 압축할 수 있다. 이러한 방법을 사용하면 빨간 색조를 256단계로 보여주는 대신 32단계로 보여주는 것으로 자세히 관찰하면 차이가 나타나 일반 목적으로 사용할 경우 별로 압축하였다는 느낌을 주지 않는다.

본 논문에서는 후자의 기법을 이용하여 영상 내에 임의의 크기로 검색되는 얼굴영상에 대하여 20×20 픽셀로 표준화 작업을 통한 전면얼굴에서와 같은 효과를 낼 수 있다.

5. 다층신경망 기반 얼굴인식 시스템의 구현

5.1 신경망 기반 얼굴인식의 구현

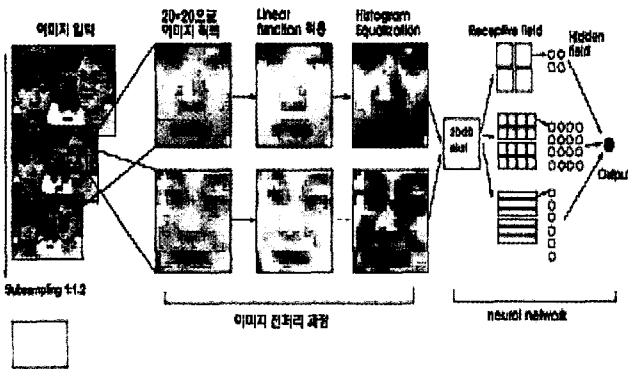


그림 17. 한 영상내의 임의의 크기의 실시간 얼굴 검출
Fig. 17. Real time face detection of free size in premier in video

본 논문에서 제안하는 임의의 검색영상내의 얼굴을 검출하는 방법은 위의 그림17이 개략적으로 잘 설명해 주고 있다. 검색하고자 하는 영상은 일정 크기 비율 1.2로 서브 샘플링 하여 피라미드 구조 형태로 만든다. 그리고 피라미드 구조의 각 영상을 20×20 픽셀 크기만큼 차례로 스캔하여 전처리 과정과 학습된 신경망을 거쳐 얼굴, 또는 비-얼굴을 판단하게 된다.

5.2 전처리 단계

5.2.1 최적 선형 필터 시스템

어떠한 영상에서 임의의 위치에서 20×20 만큼의 서브-이미지가 신경망 최적 선형 필터를 통과한 이미지가 얼굴 영상이라면, 임의의 위치에서 들어올 빛의 영향에 의한 얼굴 그림자, 좌우 빛의 밝기에 의한 차이를 보상에 줄 필요가 있다. 즉, 20×20 서브-이미지에 가장 적합하게 맞는 선형 함수를 구한 후 그 값을 뺄셈(subtract) 연산을 통해 좌우, 상하 균형을 맞출 수가 있다.

최적 선형함수를 구하는 방법은 2차원 공간상에서 질서 없이 산재되어 있는 데이터들 가운데 가장 적합한 1차원함수

를 구하는 방법을 3차원공간으로 확장하는 방법으로 구할 수 있다.

다음은 2차원 공간에서 최적 선형 함수를 구하는 방법을 설명한다.

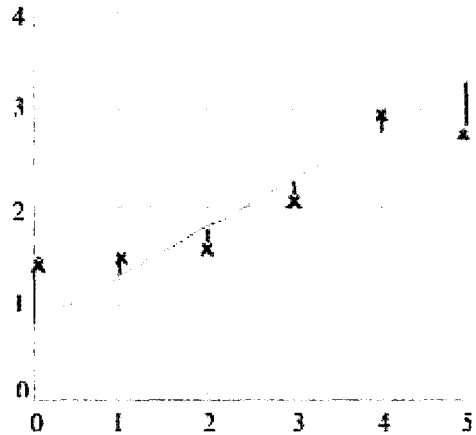


그림 18. 2차원 공간에서의 최적 선형 함수를 구하는 방법
Fig. 18. Method to save most suitable fan shape function from 2 dimensions space

즉, 그림 18과 같이 (1, 1.5), (2, 1.5), (3, 2), (4, 3), (5, 2.7)의 데이터가 2차원 공간에 산재되어 있을 때 이 점들에 가장 적합한 직선 $y = mx + b$ 를 구하는 방법은 다음과 같다. 최적의 직선을 $y = mx + b$ 라고 가정할 때 x 가 각각 $x = 0, 1, 2, 3, 4, 5$ 일 때의 y 값을 $y_0, y_1, y_2, y_3, y_4, y_5$ 라고 하고 linear function과 산재되어 있는 데이터 값들과의 거리를 각각 $d_0, d_1, d_2, d_3, d_4, d_5$ 라 할 때 $D = d_0, d_1, d_2, d_3, d_4, d_5$ 가 최소가 되는 m, b 값을 편미분을 통해서 구할 수가 있다.

즉, 2차원 공간에서 best fit 1-차원 선형 함수는 다음과 같은 공식으로 표현할 수 있다

The best fit line associated with the-points $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ - as the form

$$y = mx + b$$

where

$$\text{slope} = m = \frac{m(\sum xy) - (\sum x)(\sum y)}{(\sum x^2) - (\sum x)^2}$$

$$\text{Intercept} = b = \frac{\sum y - n(\sum x)}{n}$$

Here, \sum means "the sum of", Thus

$$\sum xy = \text{sum of products} = x_1 y_1 + x_2 y_2 + \dots + x_n y_n$$

$$\sum x = \text{sum of } x\text{-values} = x_1 + x_2 + \dots + x_n$$

$$\sum y = \text{sum of } y\text{-values} = y_1 + y_2 + \dots + y_n$$

$$\sum x^2 = \text{sum of squares of } x\text{-values} = x_1^2 + x_2^2 + \dots + x_n^2$$

그림 19. 2차원 공간에서 best fit 1-차원 선형 함수
Fig. 19. Best fit 1 - dimension fan shape function from 2 dimensions space

이처럼 최적 선형 함수는 같은 방식으로 3차원 공간에서 존재하는 데이터들 $\{(x_1, y_1, z_1) (x_2, y_2, z_2) (x_3, y_3, z_3) \dots (x_n, y_n, z_n)\}$ 중에서 최적의 평면을 구하는 문제로 확장 할 수 있다. 즉, 최적평면을 $z = ax + by + c$ 로 가정 할 때, 이 중에서 미지의 계수 a, b, c 를 구하는 문제와 같다. a, b, c 를 구하는 방법은 위에서 구했던 1차원 직선을 구하는 방법과 같은 방법으로 구할 수 있다. 즉, $\{(x_1, y_1) (x_2, y_2)(x_3, y_3) \dots (x_n, y_n)\}$ 에서 최적 평면 위의 값을 각각 $k_1, k_2, k_3, \dots, k_n$ 이라고 한다면 $\{(x_1, y_1, z_1) (x_2, y_2, z_2)(x_3, y_3, z_3) \dots (x_n, y_n, z_n)\}$ 의 주어진 데이터와 $\{(x_1, y_1, k_1) (x_2, y_2, k_2)(x_3, y_3, k_3) \dots (x_n, y_n, k_n)\}$ 의 거리의 합이 최소화 될 때 a, b, c 의 계수를 찾을 수 있다. 편미분을 통해 푼 결과는 다음과 같다.

$$a = \frac{(\sum y^2 - \frac{(\sum y)^2}{N})(\sum xz - \frac{\sum x \sum z}{N}) - (\sum xy - \frac{\sum x \sum y}{N})(\sum yz - \frac{\sum y \sum z}{N})}{(\sum x^2 - \frac{(\sum x)^2}{N})(\sum y^2 - \frac{(\sum y)^2}{N}) - (\sum xy - \frac{\sum x \sum y}{N})^2}$$

$$b = \frac{(\sum xy - \frac{\sum x \sum y}{N})(\sum xz - \frac{\sum x \sum z}{N}) - (\sum yz - \frac{\sum y \sum z}{N})(\sum x^2 - \frac{(\sum x)^2}{N})}{(\sum xy - \frac{\sum x \sum y}{N})^2 - (\sum y^2 - \frac{(\sum y)^2}{N})(\sum x^2 - \frac{(\sum x)^2}{N})}$$

$$c = \frac{-\sum(ax + by - z)}{N}$$

위에서 $\sum x, \sum y, \sum z, \sum x^2, \sum y^2, \sum z^2$ 값은 각 격자에서 주어진 데이터들의 합, 제곱의 합을 말한다. 3차원 상의 최적 선형 함수를 적용한 결과는 다음과 같다.

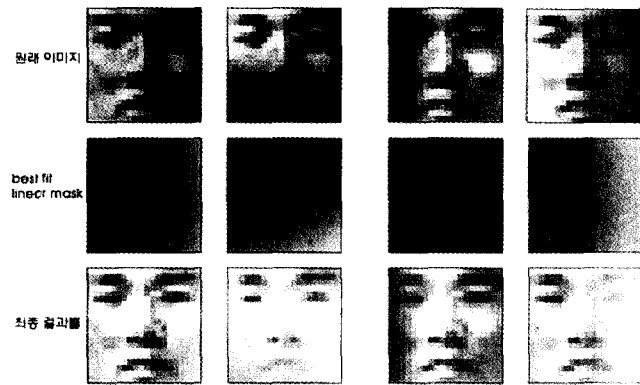


그림 20. best fit linear equation을 사용함으로써 얻은 효과
Fig. 20 Effect that get using best fit linear equation

즉 위의 결과는 같은 얼굴이라 할지라도 주위의 빛의 방향에 따라 얼굴의 이미지라 다를 수 있는 차이를 최적 선형 함수를 통해 보정한 것이다. 이것은 처리를 한 후의 이미지는 얼굴 영상 내에서 코를 기준으로 한 좌우의 영상의 밝기가 균형이 거의 일정하게 변화했음을 처리후의 영상에서 확인할 수 있다. 최적 선형 함수의 기능은 신경망 기반 필터에 통과하는 영상 데이터의 값이 빛에 의한 그림자의 영향을 최소화하기 위함이다. 따라서, 신경망에 통과되는 값은 같은 다른 각도에서 들어간 빛이 얼굴 영상을 왜곡시키더라도 최적 선형 함수를 거치면 같은 얼굴 영상 데이터가 입력된다고 할 수 있다.

5.2.2 히스토그램 평활화

어떤 영상의 명암값 분포가 빈약할 때 히스토그램 평활화 기법에 의해 향상시킬 수가 있다. 최적 선형 평활을 통해 좌우의 빛의 균형이 잡힌 영상에서 눈, 코, 입 부분의 특징들을 더 명확하게 표현하기 위해 영상에 히스토그램 평활 화를 적용하였다.

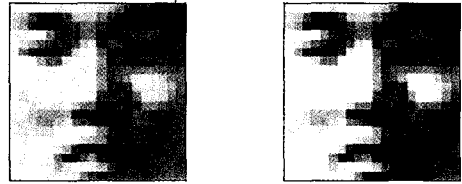


그림 21. Histogram equalization을 적용
Fig. 21. Histogram equalization application

결과에서 알 수 있듯이 눈, 코, 입 부분의 명암도가 다른 곳 보다 차이가 남을 알 수 있다.

5.2.3 달걀형 마스크(Oval mask) 알고리즘

보통 사람의 얼굴은 달걀 모양의 형태를 띄고 있다. 그리고 얼굴의 외형은 다른 부분에 비해 두 눈과 코, 입 부분이 두드러지게 나타남을 알 수 있다. 이러한 얼굴의 특징들을 신경망을 통해 학습하기 위함으므로 눈, 코, 입 부분을 제외한 부분은 필요 없는 부분이라 할 수 있다. 신경망을 통과하기 전처리 과정에서 이러한 부분을 학습 과정에 포함하지 않기 위하여, 달걀 모양의 마스크를 씌워서 신경망 학습 데이터에 영향을 주지 못하도록 하였다.

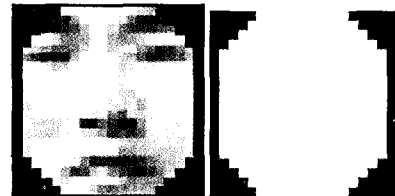


그림 22. Oval mask의 형태와 Oval mask를 씌운 모습
Fig. 22. Oval mask's form and Oval mask Ssuinmoseup

5.3 구현된 신경망 기반 필터

본 논문에서 제안된 신경망 기반 필터의 기본 구조인 단일 셋 네트워크(simple set network)의 형태는 위의 그림 23과 같이 구성되었다. 400 입력 노드층, 1 개의 출력 노드로 구성된 단일셋 구조는 얼굴 부위의 특징을 강조하기 위하여 리셉티브 필드(receptive field)를 두었다. 눈, 코, 입, 부위가 20 by 20 영상의 특정 부위에 입력되는 것을 가정할 때, 얼굴 부위는 네 부분으로 나누었을 때 다른 부분에 비해서 두 눈의 위치, 코, 입이 두드러지게 나타나고, 얼굴 부위를 가로 측면으로 프로젝션한 결과는 받아들인 특정 눈, 코, 입 부위가 특정 지역에 몰려 나타날 것을 예측 할 수 있다. 따라서 이 부분을 강조하고 다른 부위와의 차별성을 강조하기 위하여 입력 층과 은닉 층 사이의 연결을 풀리(fully connect)하지 않고 리셉티브(receptive filed) 형식으로 구성하였다. 은닉 층의 세 개의 리셉티브는 각각 10 by 10으로 4 개 5 by 5로 16개 20 by 5로 오버랩(overlapping)해서 6개의 영상으로 쪼갬다.

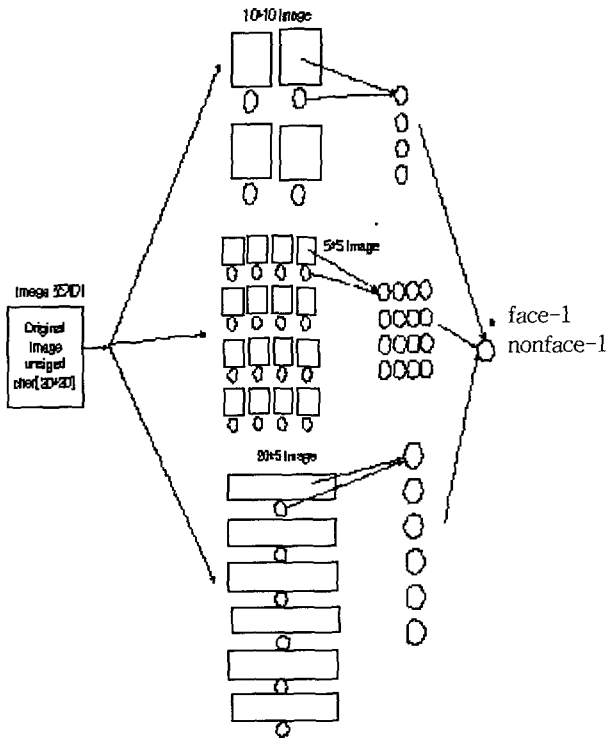


그림 23. simple Network 기본구성도
Fig. 23. Simple Network basis composition

이렇게 만들어진 은닉 층이 하나의 셋이 된다. 이 은닉 층을 병렬로 한 개 또는 두 개 이상을 배치하여 2set 또는 3set의 네트워크를 구성하였다. 2set 네트워크의 은닉노드 수는 58개 연결 수는 2905이고, 3set 네트워크의 은닉 노드 수는 78개 연결 수는 4357개이다.

5.4 부트스트랩 학습 알고리즘

임의의 영상에 존재하는 얼굴을 검색 할 시에 얼굴 영역 비해 비얼굴의 영역은 참으로 광범위하여 그 데이터를 어떤 방법으로 선정하는 것이 매우 중요하다. 본 논문에서는 수작업으로 추출한 얼굴 데이터 1500개와 컴퓨터로 생성한 랜덤 데이터 1500개를 비 얼굴 데이터로 두고 학습을 시작하였다. 하지만 이렇게 학습 된 신경망은 얼굴 영상을 비록 검출할 지라도 얼굴이 아님에도 불구하고 얼굴로 오인식 하는 경우가 발생 할 수가 있다. 따라서 다음과 같은 학습 알고리즘방법으로 비 얼굴 데이터를 업데이트하고 재학습하였다.

- step 1.
 - 초기 비얼굴이미지 1000개를 무작위로 생성해 낸다. 각 이미지에 프로세싱 과정을 거친다.
- step 2.
 - 얼굴은 출력이 1로 비얼굴은 출력이 -1로 목표치를 두고 신경망을 학습한다.
 - 학습알고리즘은 오류 역전파방법으로 학습한다.
 - 최초 반복층(first iteration)에 신경망의 연결강도는 랜덤 하게 초기화한다.
 - 최초 반복층 이후 수정된 연결강도값을 바탕으로 연결 강도 가중치에서 에러를 최소화하는 방향으로 연결강도를

업데이트 한 뒤 수정한다.

- step 3.
 - 얼굴이 전혀 포함되지 않는 배경이미지를 학습된 신경망으로 검색한다. 얼굴 부위 영역 보다 비얼굴 영역이 훨씬 광범위하기 때문에 얼굴이 전혀 포함이 되지 않은 영상 안에서 얼굴로 오인식 하는 부위가 나타나기 마련이다. 따라서 이러한 데이터들은 다시 재수정을 해주어야 한다. 오인식 하는 데이터를 추출 한 뒤 값을 -1로 하고 학습 데이터에 추가한 뒤 다시 학습한다.

- step 4.
 - 위의 과정을 얼굴 데이터 : 비얼굴 데이터 비율이 1 : 8 정도 이상이 되도록 계속 반복하여 데이터를 업데이트, 재 학습한다. 여기서 주의할 점은 한꺼번에 많은 데이터를 업데이트 해서 학습을 할 수 없다는 점이다. 왜냐 하면 얼굴 영상이 포함되지 않은 풍경 이미지에서 같은 형태의 비얼굴 데이터가 중복 될 수 있기 때문이다. 따라서 본 논문에서는 1회에 업데이트 할 수 있는 데이터의 개수를 300 ~ 400개로 한정하였다.

다음 그림 24는 얼굴 데이터가 전혀 없는 풍경 이미지에서 얼굴로 오인식한 데이터를 추출하여 그 중에서 임의로 선택한 것이다.



그림 24. 풍경이미지에서 비얼굴 데이터를 추출하는 장면
Fig. 24. Scene that draw rain face data in scenery image

5.5 통계적 추출 영상 병합

단일 셋 네트워크의 같은 쌍이 두 개, 세 개의 병렬형태로 구성되어 있는 2set, 3set 네트워크가 각각 검출한 결과를 하나의 결과로 정합, 결정하는 단계이다. 이처럼 복잡한 네트워크를 구현하는 것은 임의의 영상 내에 존재하는 얼굴 영역 보다, 얼굴이 아닌 영상 영역이 훨씬 다양하고, 넓기 때문에 부트 스트랩 알고리즘을 통하여 아무리 많은 얼굴 영상을 보상으로 학습시킨다 하더라도, 얼굴이 아님에도 불구하고, 신경망의 결과가 얼굴로 판단하는 결과가 나오기 마련이다, 따라서, 결과의 정확도를 높이기 위하여 단일 셋층을 병렬형태로 구성시킨 2set, 3set 형태의 네트워크를 사용하면 이 문제를 해결 할 수 있다. 각각 다른 초기값, 다른 데이터로 학습된 여러 개의 네트워크로 똑같은 영상을 검색하면 다음 그림 25와 같이 각기 다른 결과를 보인다.

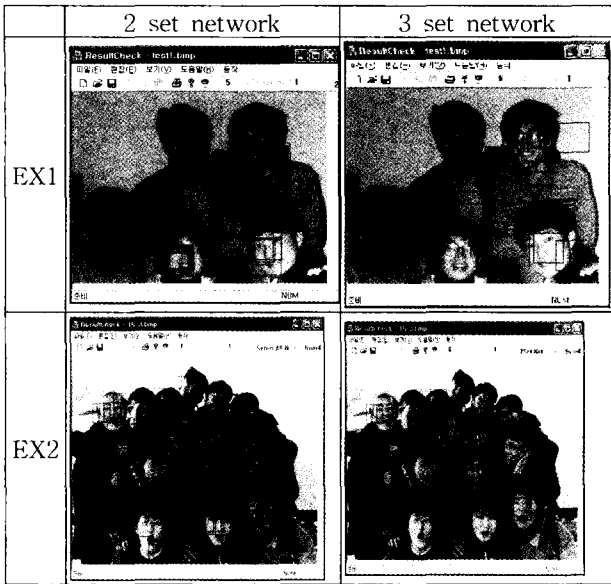


그림 25. boot strap 알고리즘을 이용한 실험결과 화면
Fig. 25. Experiment wave and screen that use boot strap algorithm

이와 같이 각각 다른 set의 신경망이 보이는 결과를 정합하는 방법 3가지를 다음절에 소개한다.

5.5.1 1단계 : MOD(Merging Overlapping Detection)

이 방법은 한 종류의 set 층으로 나타낸 결과를 바탕으로 검출 범을 결정하는 방법이다.

아래 그림 26과 27은 2set 레이어 신경망 한가지로 얼굴을 검색한 결과를 보여 주고 있다. 얼굴 영역을 차례로 검색해 나갈 때, 연속된 서브 샘플링 단계에서 나타낸 결과가 2개 이상의 서브 샘플 단계에서 넘어갈 때, 검출 결과로 처리한다.

sub sampled	image size	detection
	source image	
	0.8 sub sampling	
	0.8 * 0.8 sub sampling	

그림 26. overlapping 결과 화면 (1)
Fig. 26. Overlapping wave and result screen (1)

deting position	result
	detecting position이 overlapping 된 부분 중 하나로 결과 처리한 결과 화면

그림 27. overlapping 결과 화면 (2)
Fig. 27. Overlapping wave and result screen (2)

5.5.2 2 단계 : ANDing 방법

ANDing 방법은 1단계 MOD 방법을 두 개의 set 층으로 확장한 방법이다. 즉, 그림 28에서와 같이 한 개의 set 층에서 오버랩 방법으로 나타낸 결과가 같은 방법으로 다른 set 층의 신경망으로 검출한 결과를 비교해 보았을 때, 같은 단층, 같은 위치에서 결과를 나타낸 것들을 최종 검출한 것으로 처리한다.

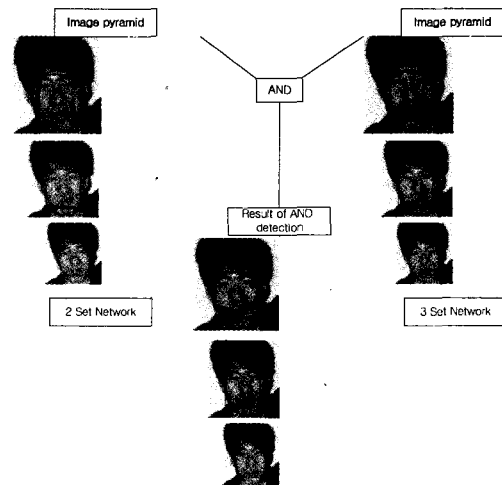


그림 28. ANDing 방법에 의한 결과
Fig.28 ANDing method of overlapped

5.5.3 3단계 : AMN 방법(Arbitration among Multiple Networks.)

세 번째 방법은 각각 다른 초기 값과, 다른 학습 데이터를 통하여 학습 된 3개 이상의 신경망의 결과를 정합, 결정하는 것이다. 그림 29에서와 같이 3x3의 관심위치 주위영역을 설정한다. 이 영역은 1단계의 오버랩 방법을 토대로 결정된 검출결과를 다시 정합하는 방법이다. 한 개의 영상을 3개의 네트워크를 통하여 검색하였을 때, 결과를 보이는 위치는 한 영상 내에서 몇 픽셀 간격으로 조금씩 다를 수 있다. 이 방법은 이 차이를 극복하고 하나로 정합해주는 과정을 또 다른 신경망(arbitration network)을 두어서 결정, 정합

하는 것이다. 즉, 아래 그림 29와 같이 영상내의 같은 얼굴을 검출한다 하더라도, 포인트 위치 주위에서 크게 벗어나지 않음을 알 수 있다. 9개의 입력 노드, 1개의 출력 노드, 은닉노드 5개를 둔 신경망을 꾸밀 수 있다. 학습 과정 또한 BP를 통한 학습을 시킬 수 있다. 이 신경망을 학습하고, 데이터를 입력받을 동안에는 얼굴을 검색하는 각각의 신경망(2set, 3set)의 연결강도 값을 변경시키면 안 된다.

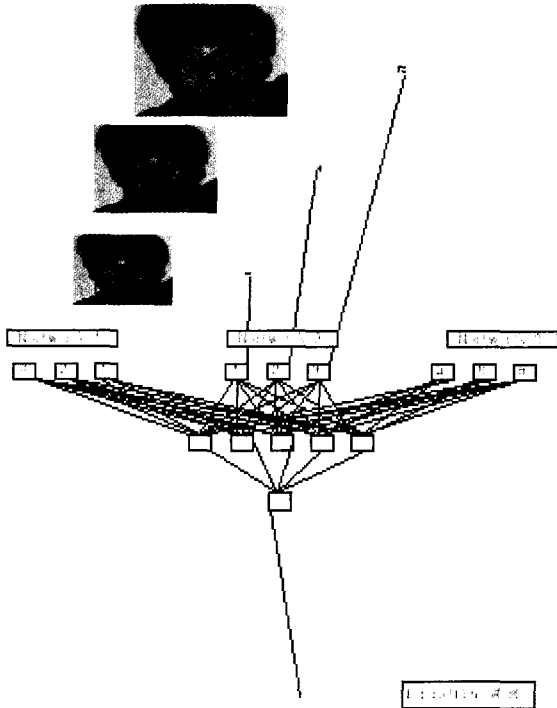


그림 29. Arbitration 방법에 의한 결과
Fig. 29. Result by Arbitration method

5.6 응용프로그램의 구현

현재 존재하는 영상감시 시스템은 폐쇄회로 카메라를 이용하여 모든 영상을 각종 미디어에 저장하고 있다. 그리고 특정 사건이 발생하였을 경우, 방대한 자료 안에서 수작업으로 장면 하나하나를 검색하고 있는 실정이다. 예를 들어 대형 은행에 출입하는 사람들 중에서 특정 인물을 검색 하고자 할 경우 사건 발생 전후의 모든 기록을 사람의 수작업을 통해서 검색하고 있어서 많은 인력과 시간을 필요로 한다. 그런데 영상감시 시스템에 저장되는 기록에 등장하는 인물의 얼굴만 정확하게 추적하여 저장한다면 특정 시간, 특정 장소에 출입하는 사람을 검색하는데 걸리는 시간을 몇 배는 단축할 수 있을 것이다. 그리고 신경망 기반 필터의 얼굴 검색 시스템과 얼굴 인식 시스템을 연동하면 그 보다 훨씬 정확하고 빠르게 검색하는 것은 물론 침입 탐지 시스템등 여러 분야에 적용할 수 있다.

그림 30과 그림 31은 영상감시 시스템에서 어떤 특정인을 검출할 수 있는 방법을 보여준다. 다음의 그림 32와 그림 33은 현실적인 시나리오와 유사한 조건에서의 응용프로그램에서 구현된 결과화면이다.

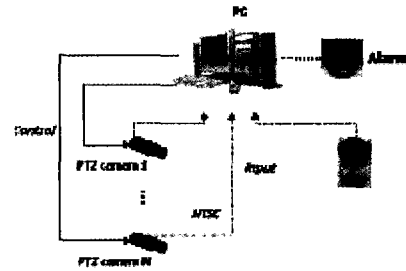


그림 30. 실시간 특정한 검출 감시시스템
Fig. 30. Detection watch system that is real time specification

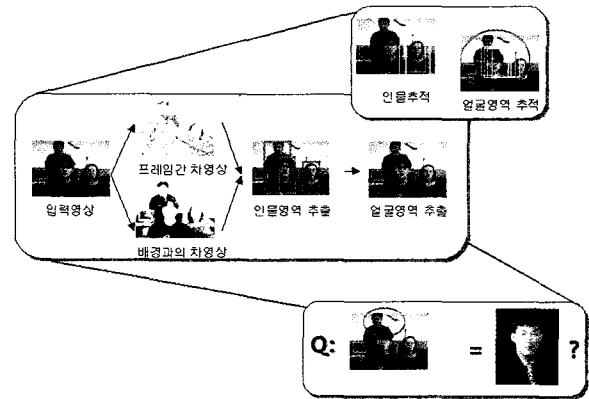


그림 31. 특정한 검출의 예
Fig. 31 Example of detection that is specification



그림 32. 특정한 얼굴 특징점 정보
Fig. 32 Specification person's face characteristic point information

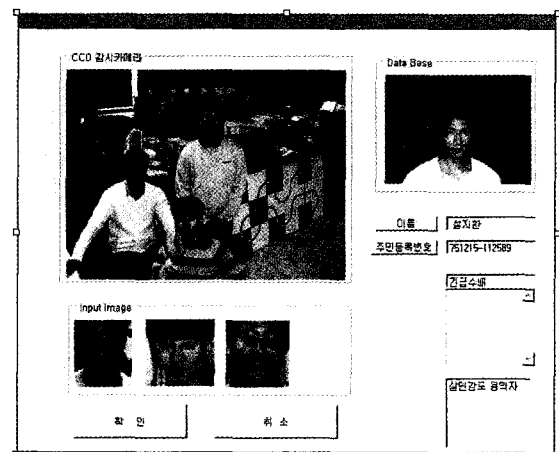


그림 33 다중 얼굴영역에서 특정한 검출
Fig. 33 Specification person detection in multiplex face sacred ground

먼저 특정한 안면특징점, 즉 눈, 코, 입, 눈썹 등의 정보

를 데이터베이스에 등록하고, 영상으로부터 받은 이미지에서 검출된 얼굴영역을 토대로 각 네트워크에서 학습된 결과를 비교하여 특징인을 검출한다. 특징인의 안면정보는 인간의 표정변화에 변동이 적은 눈썹의 양 포인트 점과, 눈동자, 코의 가장자리 등의 특정 점을 각 셋의 조건으로 두었다. 실제 폐쇄회로와 같이 2인 이상의 이미지 영상정보로부터 특징인을 검출하는데 매우 정확한 결과를 보였다. 하지만 1.6GHz PC 환경에서 600x400의 영상 하나를 검색하는 데 걸리는 시간이 무려 1분 이상이 소요됨으로써 검색 속도를 얼마나 올릴 수 있느냐가 매우 중요하다. 만약 전처리 단계에서 다음과 같은 제한 조건을 준다면 충분히 속도를 1/10까지 줄일 수 있다.

1) 샘플링 단계 조정

입력되는 얼굴 영상의 크기가 일정한 크기로 입력된다면 서브 샘플링 단계를 한정 할 수 있다. 얼굴 크기를 벗어나거나, 아주 작은 단계에서부터 검색하지 않는다면 검색할 얼굴 영역이 없음에도 불구하고 검색하는 연산 시간의 소요를 줄일 수 있다. 예를 들어 은행, 공공장소에서 작동하는 폐쇄회로에 입력되는 얼굴 부위 크기는 임의로 변하지 않고, 거의 같은 크기로 입력이 된다. 따라서 서브 샘플링 단계를 3-4단계로 한정할 수 있다.

2) 색깔 영역을 고려

사람 색깔 부분이 존재하는 영역만 검색하는 방법을 사용할 수 있다. 동양인의 피부 색깔은 밝은 장소에서 Hue 채널 130~150 레벨에 주로 분포하는 것을 통계자료를 통하여 확인 할 수 있다.

3) 모션벡터(motion vector)를 이용하는 방법

또 다른 방법의 하나로 모션벡터를 응용 할 수 있다. 실시간으로 입력되는 영상에서 움직이는 물체, 즉 사람의 동작을 추적하는 방법과 신경망 기반 필터를 병행하면 정확도와 속도를 한층 끌어올릴 수 있다.

6. 결 론

본 논문에서는 복잡한 배경을 갖는 실시간 폐쇄회로 화면으로 받은 컬러 이미지에서 얼굴영상을 추출하고 이미 저장된 특징인의 얼굴영상과 비교를 통해 지하철이나 은행 등 공공장소에서의 수배자 등 어떤 특징인을 검출하는 방법을 제안하였다. 특히, 감시카메라의 특성상 얼굴정보가 임의의 크기로 가변하고 영상 내에서 다수의 얼굴정보를 포함하고 있음을 가정할 때, 얼굴 영역을 얼마나 정확하게 검색 할 수 있느냐에 초점을 맞추었다. F.Rosenblatt가 제안한 퍼셉트론 신경망 모델을 기초로 임의의 얼굴영상에 대하여 20×20 픽셀로 서브샘플링을 사용하여 규준화 작업을 통한 전면얼굴에서와 같은 인식기법의 효과를 사용하였다. 획득한 얼굴후보 영역에 대하여 조명이나 빛에 의한 외부환경의 간섭을 최소화하기 위하여 최적선형필터와 히스토그램 평활화기법을 이용하고 불필요한 학습을 최소화하기 위하여 달걀형 마스크의 덧셈연산을 전 처리 과정에 추가하였다. 전 처리 과정을 마친 이미지는 각각 세 개의 수용필드로 쪼개어져 특정 위치에 존재하는 눈, 코, 입 등의 정보를 신경망 학습을 통해 최종 결정된다. 오인식률을 최소화하기 위하여 얼굴영역이 없는 복잡한 배경이미지를 부트스트랩 알고리즘을 이용하여 학습시

켰다.

또한 각각 다른 초기 값을 가지는 3개의 단일넷 네트워크 시스템을 병렬형태로 구성하여 결과의 정확도를 높였다.

제안된 시스템은 기존의 논문에서 제시한 방법들에 비하여 복잡한 알고리즘의 병합으로 처리속도가 상당히 느리지 만, 매우 정확한 결과를 낼 수 있다.

또한, 인간의 색상정보를 이용한 추적이나 차영상을 이용한 방법이 아닌 실시간 다중검색의 가능성을 가지므로 비단 얼굴검출뿐만 아니라 화재나, 강도 사건 등의 특정정보에 대한 보다 빠른 인간의 대처능력을 향상시키는 등 많은 부분에서 응용 가능할 것으로 사료된다.

참 고 문 헌

- [1] R. Jain, "Extraction of motion information from peripheral processes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 3, pp.489-503, 1981.
- [2] A. Shio and J. Sklansky, "Segmentation of people in motion," *IEEE Workshop on Visual Motion*, pp 325-332, 1991.
- [3] M. Bichsel and A. Pentland, "Human face recognition and the face image set's topology," *Computer Vision, Graphics, and Image Processing : Image Understanding*, vol. 59, pp254-261, 1994.
- [4] Haiyuan Wu, et al., "Detecting Human Face in Color Image", *IEEE International Conf. on Systems, Man and Cybernetics*, pp. 2232-2237, 1996.
- [5] K. Matsuda, T. Kageyama and T. Aibara, "Application of the second order statistics for the recognition of human faces," in *Proc. Int. Conf. Image Processing, Singapore*, pp. 543-546, Sept. 1992.
- [6] R. Brunli and T. Poggio, "Face Recognition: Feature versus Templates," *IEEE Trans. Pattern analysis and machine intelligent*, Vol. 15, no. 10, 1993.
- [7] A. L. Yuille, P. W. Hallinan and D. S. Cohen, "Feature extraction from faces using deformable templates," *Int. Journ. Computer Vision*, Vol. 8, no. 2, pp. 99-112, Aug. 1992.
- [8] C.L. Huang and C. W. Chen, "Human facial feature extraction for face interpretation and recognition," *Pattern Recognition*, Vol. 25, no. 12, pp. 1435-1444, Dec. 1992.
- [9] 지은미, 최형일, 이철희, "문체영역에 관한 지식을 이용한 얼굴인식," *정보과학회 논문집 제 17권, 제6호*, pp. 666-675, 1990년 11월.
- [10] K. M. Lim and K. W. Oh, "A face recognition system using fuzzy logic," in *Proc. First Korea Japan Joint Conf. Computer Vision*, pp. 506-509, Seoul, Korea, Oct. 1991.
- [11] Olivier Faugeras, *Three-Dimensional Computer Vision - A Geometric Viewpoint*, The MIT

- Press, 1989.
- [12] Rowley, Baluja, and Kanade: Neural Network-Based Face Detection (*PAMI*, January 1998)
- [13] Kah-Kay Sung. Learning and Example Selection for Object and Pattern Detection. *PhD thesis, MIT AI Lab, January 1996. Available as AI Technical Report 1572.*
- [14] Pawan Sinha. Object recognition via image invariants: A case study. *Investigative Ophthalmology and Visual Science*, 35(4), march 1994.
- [15] Robert Brunelli and Tomaso Poggio, "Face Recognition : Feature versus Templates," *PAMI*, Vol, 15, no. 10, pp.1042-1052, Oct, 1993
- [16] 김영일, 이웅주, "얼굴 특징벡터 및 안면각 정보를 이용한 실시간 자동얼굴검출 및 인증시스템", 정보처리학회논문지, 제9권, pp. 491-500, 2002.
- [17] J. R. Parker, " Algorithms for Image Processing and Computer Vision, " *Wiley Computer Publishing*, 1997.
- [18] 박용성, 유태웅, "복잡한 배경에서 얼굴영역 추출", 군산대학교 논문집 제 23권 pp. 355-365, 1997. 2
- [19] 전인자, 이재훈, 이정훈, 이필규, "오류-역전과 신경망 기반의 얼굴검출 및 포즈추정", 정보처리학회논문지 제9권, pp. 853-862, 2002.
- [20] Face recognition for smart environments, Pentland, A., Choudhury, T. *IEEE Computer*, Vol. 33, Issue 2, Feb 2000.
- [21] Young & Yamane, Sparse Population Coding of Faces in the Inferotemporal Cortex, *Science*, Vol. 256, 1992.
- [22] H.J.Park, H.S.Yang, Invariant Object detection based on evidence accumulation and Gabor Features, *Pattern Recognition Letters*, Vol. 22, Issue.8, June 2001.
- [23] 임춘환, 고낙용, 박종안, "차영상과 DCT 기반 특징 추출과 신경망을 이용한 얼굴인식에 관한 연구", 한국통신학회논문지 '99-8 vol. 24 No.8B pp 1549-1557, 1999. 8.
- [24] 유태웅, 오일석, "정면 칼라 영상에서 얼굴영역 및 특징 추출", 전북대학교 논문집 제 40권 자연과학편 pp. 237-242, 1995
- [25] 김일정, 차의영, "시점 기반 고유공간을 이용한 얼굴인식", 한국정보과학회 가을 학술발표논문집 Vol. 25 No.2, pp. 458-460, 1998.
- [26] 조동욱, 한길성, 김지영, "AVI 시스템 구축을 위한 정면 얼굴의 인식", 서원대학교 응용과학연구 제9권 제1호, pp103-120, 1998.2.

저 자 소 개

송홍복(Hong Bok Song)

제 12권 제2호(2002년 4월호) 참조

설지환(Ji Hwan Seol)

제 12권 제2호(2002년 4월호) 참조