

# 대용량 훈련 데이터의 점진적 학습에 기반한 얼굴 검출 방법

(Face Detection Based on Incremental Learning from Very Large Size Training Data)

박 지 영<sup>†</sup> 이 준 호<sup>\*\*</sup>

(Ji-Young Park) (June-Ho Yi)

**요약** 본 연구는 대용량 훈련 데이터를 사용하는 얼굴 검출 분류기의 학습과정에서 새로운 데이터의 추가 학습이 가능한 새로운 방법을 제안한다. 추가되는 데이터로부터 새로운 정보를 학습하여 이미 습득된 기존의 지식을 갱신하는 것이 점진적 학습의 목표이다. 이러한 학습 기법에 기반한 분류기의 설계에서는 최종 분류기가 전체 훈련 데이터 집합의 특성을 반영하는 것이 매우 중요한 문제이다. 제안하는 알고리즘은 최적화된 최종 분류기 획득을 위하여 훈련 집합의 전역적인 특성을 대표하는 검증집합을 생성하고, 이 집단 내에서의 분류성능을 기준으로 중간단계 분류기들의 가중치를 결정한다. 각 중간단계 분류기는 개별 데이터 집합의 학습 결과로써 가중치 기반 결합 방식에 의해 최종 분류기로 구성된다. 반복적인 실험을 통해, 제안한 알고리즘을 사용하여 학습한 얼굴 검출 분류기의 성능이 AdaBoost 및 Learn++ 기반의 분류기보다 우수한 검출 성능을 보임을 확인하였다.

**키워드** : 점진적 학습법, 얼굴 검출, 검증집합, 가중치 기반 결합

**Abstract** Face detection using a boosting based algorithm requires a very large size of face and nonface data. In addition, the fact that there always occurs a need for adding additional training data for better detection rates demands an efficient incremental learning algorithm. In the design of incremental learning based classifiers, the final classifier should represent the characteristics of the entire training dataset. Conventional methods have a critical problem in combining intermediate classifiers that weight updates depend solely on the performance of individual dataset. In this paper, for the purpose of application to face detection, we present a new method to combine an intermediate classifier with previously acquired ones in an optimal manner. Our algorithm creates a validation set by incrementally adding sampled instances from each dataset to represent the entire training data. The weight of each classifier is determined based on its performance on the validation set. This approach guarantees that the resulting final classifier is learned by the entire training dataset. Experimental results show that the classifier trained by the proposed algorithm performs better than by AdaBoost which operates in batch mode, as well as by Learn++.

**Key words** : Incremental Learning Method, Face Detection, Validation Set, Weighted Majority Voting

## 1. 서 론

얼굴 검출은 영상을 얼굴과 얼굴이 아닌 클래스로 구분하는 이진 분류의 문제이다. 문제 해결을 위해 다양한 패턴 분류 알고리즘의 적용이 시도되어 왔으나 여전히

많은 문제점들을 안고 있다. 얼굴은 개인별로 컬러, 크기, 모양이 다를 뿐 아니라 다양한 표정과 머리 모양, 안경 착용 및 화장에 따라 매우 다양한 변화를 포함하는 클래스이다. 뿐만 아니라 얼굴이 아닌 클래스는 그 범위를 제한하는 것조차 불가능하다. 이와 같이 클래스 내의 변화가 매우 심한 경우에는 클래스의 분류를 위해 복잡한 형태의 결정 경계(decision boundary)가 필요하다. 이러한 상황에 적합한 분류기 학습 알고리즘으로 AdaBoost[1]를 들 수 있다. 이는 여러 약한 분류기

<sup>†</sup> 비 회 원 : 성균관대학교 정보통신공학부  
jiyp@ece.skku.ac.kr

<sup>\*\*</sup> 중 심 회 원 : 성균관대학교 정보통신공학부 교수  
jhyi@ece.skku.ac.kr

논문접수 : 2004년 1월 29일  
심사완료 : 2004년 4월 30일

(weak classifier)들의 조합을 통해 성능이 우수한 강한 분류기(strong classifier)를 생성해내는 방식이며 얼굴 검출을 위한 분류기 학습에 적합함을 입증한 연구가 발표된 바 있다[2].

얼굴 검출을 위한 분류기 학습에 AdaBoost 알고리즘을 적용한 경우, 학습의 결과로 얻게 되는 강한 분류기의 성능은 학습 데이터 집단의 구성에 크게 영향을 받는다. 학습을 위해 요구되는 이상적인 데이터는 얼굴과 얼굴이 아닌 두 클래스의 다양한 변화를 포함하는 방대한 양의 데이터나 이러한 훈련 예제들을 모두 수집하여 학습한다는 것은 매우 어려운 일이다. 학습 시간 또한 매우 길어 일반적으로 몇 주 단위의 시간이 소요된다. 각 클래스의 특성을 대표하는 이상적인 데이터 집단을 한번에 모두 확보하기는 어려우며, 따라서 새롭게 추가되는 데이터의 학습이 요구되는 상황이 빈번히 발생한다. 새로운 데이터가 추가 될 때마다 전체 데이터에 대하여 재학습을 수행하는 방법은 매우 비효율적이다. 따라서 이미 학습된 지식에 새로운 데이터의 학습 결과를 추가할 수 있는 점진적인 학습(incremental learning)방식이 필요하다. 이러한 문제 해결을 위한 연구로서 Learn++[3]라는 점진적인 학습 방법이 최근 소개되었다. Learn++는 AdaBoost 알고리즘의 개선을 통해 새롭게 추가된 데이터 집합으로부터 생성한 분류기를 이미 학습된 분류기에 결합함으로써 AdaBoost의 비효율적인 일괄 학습 방법을 극복하였다. 그러나 Learn++는 추가되는 소규모의 데이터 집합 각각에 독립적으로 최적화된 분류기들을 단순 결합함으로써 최종 분류기에 전체 데이터 집합의 특성을 반영할 수 없는 문제점이 있다. 새롭게 추가될 데이터의 성질을 미리 예측할 수 없는 점진적 학습의 특성상 분류기가 개별 데이터 집합 고유의 성질에 영향을 받지 않도록 하는 것은 매우 중요한 문제이다.

본 연구에서는 얼굴 검출을 위한 분류기 학습에 점진적 학습 알고리즘을 적용하여 해당 데이터 집합 내에서만 최적화된 형태의 중간단계분류기를 효과적으로 조합함으로써 최종 분류기의 성능을 획기적으로 개선하였다. 제안한 방법은 중간 분류기의 효율적인 가중치 결정을 위해 전체 훈련 집합을 대표하는 검증집합(validation set)을 생성하는 방법을 사용하여 Learn++ 알고리즘에 적용하였다. 검증집합에서의 분류 성능을 기준으로 각 분류기의 가중치(weight)를 결정함으로써 전역적으로 최적화된 최종 분류기를 생성함으로써 우수한 검출 성능을 얻을 수 있었다. 반복적인 실험을 통해 제안한 알고리즘이 AdaBoost, Learn++ 보다 우수한 검출 성능을 보임을 확인하였다.

## 2. 관련연구

영상으로부터 효과적으로 얼굴을 검출하기 위한 다양한 방법들이 연구되어 왔다. 얼굴 검출은 얼굴 인식, 표정 인식, 3차원 모델링 등의 분야에 필수적인 전처리 단계이다. 얼굴 검출에 관한 연구는 크게 두 가지 부류로 나누어 볼 수 있는데 하나는 얼굴의 특징(Feature based)에 기반을 둔 방법이며 다른 하나는 얼굴의 전체 형상(Image based)을 이용하는 방법이다. 표 1 에는 대표적인 얼굴 검출 방식들이 요약되어 있다. 본 연구는 영상 기반 방식의 연구로서 여기서는 영상기반 방식들에 대하여 주로 기술하기로 한다.

특징 기반 방법들은 다양한 얼굴 변화와 복잡한 주변 환경 조건을 충분히 고려하지 않음으로써 실제 환경에서의 얼굴 검출에 많은 어려움이 있다. 또한 복잡한 배경 속에서 하나 이상의 얼굴을 찾아야 하는 경우라면 문제는 더욱 심각하다. 이러한 문제점들을 해결 하고자, 얼굴 검출을 패턴 인식의 문제로 다루는 연구가 시작

표 1 대표적인 얼굴검출 관련 연구의 분류

| 분류                   | 저자                 | 알고리즘   | 장점                            | 단점                      |
|----------------------|--------------------|--|-------------------------------|-------------------------|
| 특징 기반<br>얼굴 검출<br>방식 | Graf[11]           | 흑백영상의 명암 향상 후 임계 값으로 특징을 분류                      | 얼굴의 특징을 명암 값으로 간단하게 분류        | 조명의 영향을 많이 받음           |
|                      | Cai[12]            | CIE Lab 칼라 공간상의 피부색을 기반으로 얼굴 영역 검출               | 칼라 정보를 이용하여 분별력을 높임           | 조명 빛의 스펙트럼에 영향을 많이 받음   |
|                      | Craw[13]           | 얼굴의 특징들을 사전에 템플릿으로 정의하고 유사도 계산                   | 구현이 용이 하며 빠른 검출 가능            | 크기, 회전, 포즈의 변화를 다루지 못함  |
| 영상 기반<br>얼굴 검출<br>방식 | Sung and Poggio[9] | 고차원 공간상의 점들로 표현되는 훈련 영상의 분포를 기반으로 얼굴과 배경 구분      | 각 클래스의 대표적 데이터 확보 시 고성능 보장 가능 | 배경을 대표하는 샘플 획득이 어려움     |
|                      | Rowley[7]          | 다중네트워킹과 멀티스케일 기반의 방식으로 검출결과 중첩의 문제를 중재자의 도입으로 해결 | 복잡한 얼굴 패턴 모델링을 위한 시스템 훈련에 적합  | 연산이 많아 속도가 느림           |
|                      | Viola and Jones[2] | AdaBoost 기반의 학습과 Cascade 구조의 적용으로 효율적인 분류기 구성    | 실시간 얼굴검출 가능                   | 학습 데이터 추가 시 전체 데이터의 재학습 |

되었다. 이러한 접근방식은 예제들로부터 얼굴 패턴을 학습함으로써 얼굴의 특징을 미리 명확히 설정하여 인위적으로 모델링할 필요가 없으므로 얼굴 특징을 모델링 하는 과정에서 발생할 수 있는 다양한 에러를 줄일 수 있다.

대표적인 이미지 기반 방식으로는 신경망 기반 방법, 분포 기반 방법 그리고 AdaBoost를 이용한 얼굴검출 기법이 있다. 신경망 기반 방법에는 계층적 신경망[4], 자기 결합 신경망(autoassociative neural network)[5], 확률적 판단에 근거한 신경망방식(probabilistic decision-based neural network)(PDBNN)[6]등을 적용한 얼굴 검출 기법이 있다. 신경망을 이용한 최근의 얼굴 검출 기법 중 가장 대표적인 연구는 Rowley의 시스템이다 [7,8]. 그들이 제안한 시스템의 특징은 다중 신경망과 중재자 개념의 도입이다. 얼굴 패턴의 검출을 위해 사용된 다중 신경망의 결과는 AND, OR 그리고 Voting과 같은 중재자를 통해 효과적으로 조합될 수 있음을 보였다. Sung과 Poggio는 훈련 데이터의 분포 학습에 기반을 둔 얼굴 검출 시스템을 제안하였다[9]. 다량의 얼굴과 배경 훈련영상들로부터 고차원 공간상의 분포를 각각 모델링 한 후 검출단계에서는 클래스 판단의 근거로서 테스트 데이터와 모델링 된 분포간의 거리를 사용하였다. 이 두 가지 이미지 기반 방식들은 훈련뿐만 아니라 검출단계에서도 적용되는 연산이 많아 실시간 시스템 구축에 적절치 못하다는 공통적인 문제점을 안고 있다.

최근 Viola와 Jones는 검출에 사용되는 복잡한 연산은 줄이고 동시에 높은 검출 성능을 보장하는 실시간 시스템에 관한 연구로 많은 관심을 끌었다[2]. 이들은 AdaBoost 알고리즘[1]을 얼굴 검출에 적용함으로써 320x240의 이미지에서 초당 14 프레임으로 동작하는 성능을 보이고 있다. 그러나 학습을 위해 필요한 이상적인 데이터 집합을 한 번에 모두 확보하기 어렵다는 문제는 모든 이미지 기반 방식에 적용되며 추가 데이터의 학습이 가능한 점진적 학습방법이 필요하다. 최근 이미 학습된 정보의 손실 없이 새로운 데이터로부터 추가적인 정보를 학습하는 방식인 Learn++ 알고리즘이 제안 되었으며, 학습을 통한 분류기 조합이라는 AdaBoost의 특징을 따른다. Learn++ 학습 알고리즘은 점진적으로 추가 되는 소규모 데이터 집합을 학습하여 중간단계의 분류기들을 생성하고, 이들의 조합으로 강한 분류기를 조직한다. 그러나 이러한 방식은 분류기 조직 과정에서 데이터 집합 각각의 특성을 반영할 수는 있으나, 전체 데이터 집합의 특성은 고려할 수 없는 단점이 있다. 최근 이와 같은 분류기의 부분적 최적화 문제 해결을 위한 방법[10]이 제안되었다. 그러나 전역적으로 최적화된 가중치 결정을 위해 각 클래스의 데이터가 정규 분포임을

가정하고 온라인 분류단계의 연산 량을 증가시킴으로써 실시간 얼굴 검출에 적용하기에는 적합하지 않다.

### 3. 얼굴 검출을 위한 점진적 학습 방법

본 연구에서 제안한 점진적 학습 기법은 다량의 훈련 예제 학습방식을 사용하는 얼굴 검출기의 설계에 적합하다. 얼굴 검출을 위한 학습과정의 목표는 얼굴과 얼굴이 아닌 영상을 구분하기 위한 기준을 찾는 것이다. 학습에 사용되는 데이터는 다양한 변화를 포함하는 얼굴 영상과 얼굴이 아닌 영상들의 집합이다. AdaBoost 학습 방식의 가장 큰 특징은 임의 추측방식(random guessing)의 분류성능(이진 분류 문제에서는 0.5)과 유사한 성능을 갖는 약한 분류기들의 조합을 통해 우수한 성능을 보장하는 강한 분류기를 조직할 수 있다는 것이다. 이 방식의 큰 문제점으로는 앞서 언급한 바와 같이 다양한 데이터 확보의 어려움과 다량의 훈련 집합 사용으로 인한 학습 소요시간의 장기화를 들 수 있다. 따라서 본 연구에서는 얼굴 검출을 위한 학습단계의 이러한 문제점을 보완하면서 AdaBoost 학습 방식을 통해 구성된 강한 분류기의 성능을 보장하는 학습 알고리즘을 적용하여 분류기를 구성하였다.

점진적으로 증가하는 새로운 훈련 데이터 집합의 학습을 통해 강한 분류기를 구성하는데 적합한 학습 알고리즘으로 Learn++ 알고리즘이 있다. 이는 AdaBoost를 기반으로 한 점진적 학습 방법으로서 두 알고리즘은 모두 약한 분류기의 조합으로 강한 분류기를 생성하는 공통적인 특성을 지닌다. Learn++는 이러한 특성을 바탕으로, 추가되는 소규모의 데이터 집합을 학습하기 위한 알고리즘이다. 본 연구에서는 얼굴 검출의 기준이 되는 분류기 생성을 위한 훈련영상 학습과정에 Learn++ 구조를 개선한 학습 알고리즘을 제안한다. 최종 분류기의 성능은 각 데이터 집합의 학습결과를 조합하는 방식에 영향을 받게 된다. 제안한 알고리즘은 각 훈련 데이터 집합 고유의 특성이 무시되거나 혹은 지나치게 영향을 받지 않게 한다는 점에서 기존의 방법과 다르다.

그림 1은 제안한 학습 알고리즘의 구조를 나타낸다. 그림 1의 (a) 블록 다이어그램은 전체적인 흐름을 나타낸다. 점진적으로 추가되는 훈련 데이터 집합에서 Modified Learn++ 알고리즘을 통해 해당 집합에 최적화된 분류기  $F_m$ 을 생성한다. 학습이 진행되는 동안 전체 데이터 집단을 대표하기 위한 검증 집합이 구성된다. 각 집단의 최종 분류기  $F_m$ 의 조합 가중치는 이러한 검증 집합을 기반으로 결정된다. 이러한 과정을 통해 결정된 분류기와 가중치의 조합으로 얼굴 검출을 위한 강한 분류기를 생성한다. 그림 1의 (b)는 각 데이터 집단의

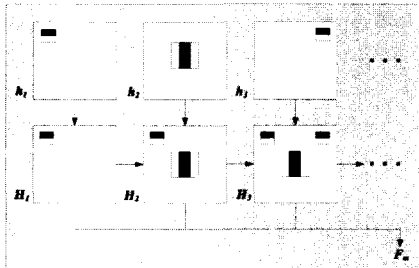
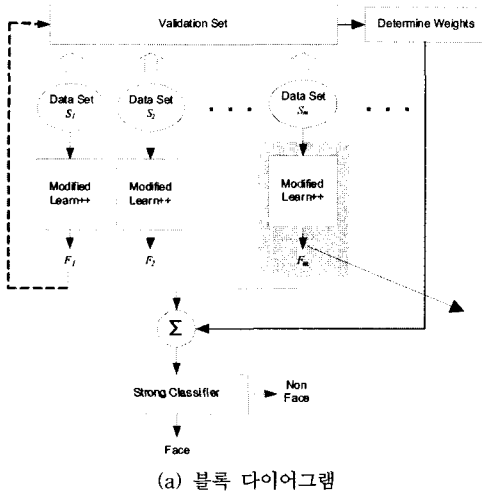


그림 1 학습 알고리즘의 흐름도

최종 분류기 생성 과정(그림 (a)에서 음영 처리된 부분)을 개념적으로 도식화 한 것이다. 하나의 특징으로 구성된 약한 분류기  $h_i$ 의 조합을 통해 합성 분류기  $H_i$ 를 생성하는 과정을 보인다. 그림에서 나타낸 사각형 모양의 특징들은 Viola와 Jones의 연구[2]에서 제안되었던 특징들로서 훈련영상의 integral image로부터 구한 Haar 웨이블릿(wavelet) 특징이다.

3.1 점진적 학습을 위한 약한 분류기 조합

그림 2는 그림 1의 (b)에서 행해지는 연산을 자세히 보여준다. 즉, 새롭게 추가되는 각 데이터 집합마다 수행되는 학습과정의 흐름을 나타낸다. 학습과정의 입력은 얼굴과 얼굴이 아닌 훈련 영상의 집합이며 학습을 통해 현재의 입력 데이터 집합을 가장 잘 분류하는 최종 분류기를 구하는 것이 목표이다. 알고리즘의 구성은 크게 세 부분으로 나누어 볼 수 있다. 첫 번째 부분은 훈련 집합의 학습에 필요한 데이터 및 파라미터 값을 설정하고 각 훈련영상의 가중치를 초기화 하는 부분이다. 두 번째는 현재 데이터의 분류에 가장 적절한 약한 분류기를 선택하고 이들을 조합해 가는 과정이다. 그리고 세 번째로는 각 데이터 집합의 학습 결과인 최종 분류기를

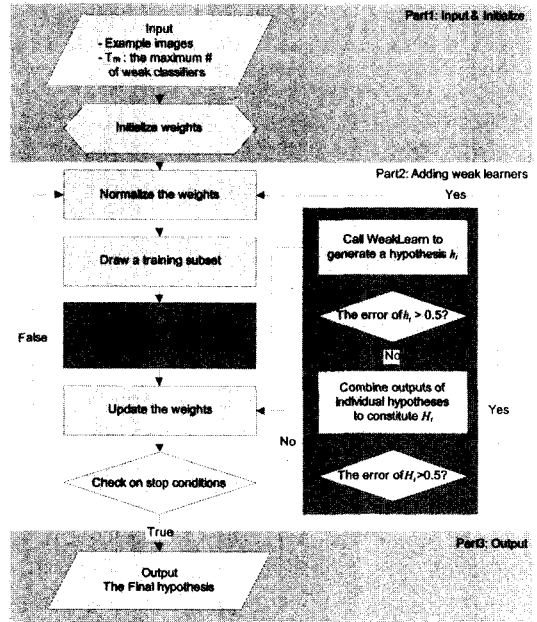


그림 2 얼굴 검출을 위한 Modified Learn++ 알고리즘

생성하는 부분이 있다.

• Part1: 입력 및 초기화

점진적 학습과정의 입력으로는 훈련 영상 집합  $S_m$ 과 반복 횟수 결정 파라미터인  $T_m$ 이 있다.  $m$ 번째로 추가된 데이터 집합  $S_m = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ 의 각 데이터  $x_i$ 는  $y_i = 1$ 인 경우 얼굴을,  $y_i = 0$ 인 경우 얼굴이 아닌 영상을 나타낸다.  $T_m$ 은 합성분류기를 구성할 약한 분류기의 최대 개수이며 이는 곧 Part2의 반복 횟수가 된다. 본 연구에서는  $T_m$ 의 값으로 30을 사용하였다. 데이터 집합을 구성하는 각 영상  $x_i$ 에는 가중치  $w(i)$ 가 할당된다. 초기화 단계에서 모든 영상은 동일한 가중치를 가지며 그 값은 얼굴과 얼굴이 아닌 클래스를 구성하는 데이터의 개수  $k$ 와  $l$ 에 의해 다음과 같이 결정된다.

$$w_1(i) = 1/2k, w_1(i) = 1/2l \text{ for } y_i = 1, 0, \text{ respectively (1)}$$

• Part2: 약한 분류기의 조합

두 번째 부분은 사전에 정의된  $T_m$ 회의 반복을 통해 약한 분류기를 선택하고 이들의 조합으로 합성분류기를 생성하는 과정이다. 데이터 집합  $S_m$ 으로부터 식 (2)를 통해 정규화된 가중치 분포  $D_t$ 를 기준으로  $(k/2 + l/2)$  개의 데이터로 구성된 훈련집합  $TR_t$ 를 선택한다.

$$D_t = w_t / \sum_{i=1}^N w_t(i), \text{ 단 } t=1, 2, \dots, T_m \text{ (2)}$$

훈련집합은 현재 영상이 얼굴인지 아닌지를 결정하는

최소 단위가 되는 하나의 가설, 즉 약한 분류기 결정을 위한 WeakLearn 알고리즘의 입력으로 사용된다. WeakLearn의 목적은 많은 특징들로부터 현재의 훈련데이터 분류에 가장 적합한 특징, 즉 하나의 분류기  $h_t$ 를 선별해 내는 것이다. 이 과정은 그림 3에서 보이는 바와 같다. WeakLearn의 결과인 약한 분류기는 훈련영상의 integral image로부터 구한 Haar 웨이블릿 특징들 중 분류에러가 최소가 되는 특징 하나를 선택한 것이다. 이는 Viola와 Jones의 연구[2]에서 제안되었던 방식이다.

WeakLearn을 통해 선택한 약한 분류기  $h_t$ 의 에러는 데이터 집합  $S_m$ 에서 식 (3)과 같이 계산한다.

$$e_t = \sum_{i=1}^N D_i(i) |h_t(x_i) - y_i| \quad (3)$$

이 과정에서는 약한 분류기의 학습에 사용되지 않은 데이터들이 성능 평가에 참여함으로써 선택된 약한 분류기의 타당성을 검증한다. 약한 분류기의 에러가 0.5를 초과한다면 훈련영상의 선택과정으로 돌아가 새로운 훈련영상 집합을 구성한다. 이진 분류 문제에서 약한 분류기의 신뢰도는 임의의 데이터를 근거 없이 추정하여 분류하는 경우의 분류확률(50%) 이상이어야 한다[14]. 그렇다고 해서 약한 분류기 각각의 성능을 개선하여 강한 분류기를 만들어야 하는 것은 아니다. Boosting 기반의 학습방식에서 최종 분류기를 구성하는 약한 분류기 각각의 성능을 개선하여 강한 분류기를 만드는 것은 과대적합(over fitting)의 문제를 야기할 수 있기 때문이다[1].

WeakLearn

1. 매  $k$ 번째 마다 주어진 하나의 특징  $f_k$ 와 임계값 (threshold)  $\theta$ 로부터 약한 분류기  $h_k(x)$ 를 훈련시킨다. ( $x$ 는 고정된 크기를 갖는 윈도우이며 실험에서는 20x20을 사용하였다)

$$h_k(x) = \begin{cases} 1 & |f_k(x)| > \theta \text{ 일 경우} \\ 0 & \text{그 외} \end{cases}$$

2.  $D_{TR}$ 에 대해  $h_k$ 의 에러를 계산한다.

$$\epsilon_k = \sum w_i |h_k(x_i) - y_i|$$

3. 가장 작은 에러값  $\epsilon_k$ 를 가진 분류기  $h_k$ 를 반환한다.

그림 3 WeakLearn

$t$ 회의 반복을 통해 생성된 약한 분류기들은 가중치기 반결합(weighted majority voting)[15] 방법을 사용하여 합성 분류기  $H_t$ 로 구성된다.

$$H_t(x) = \begin{cases} 1 & \sum_{j=1}^t \log(1/b_j) \cdot h_j(x) \geq \frac{1}{2} \\ \sum_{j=1}^t \log(1/b_j) \text{인 경우, 단 } b_j = e_j / (1 - e_j) & (4) \\ 0 & \text{그 외} \end{cases}$$

이는 각각의 약한 분류기들이 그들의 가중치  $\log(1/b_j)$ 를 기반으로 합성된 형태이다. 여기에서  $b_j$ 는 약한 분류기의 에러  $e_j$ 를 정규화 한 값이므로, 약한 분류기의 가중치는 분류기의 에러가 낮을수록 큰 값을 가진다.

합성 분류기  $H_t$ 의 에러  $E_t$ 는 약한 분류기의 에러를 구하는 것과 마찬가지로 다음과 같은 식을 사용하여 계산할 수 있다.

$$E_t = \sum_{i=1}^N D_i(i) |H_t(x_i) - y_i| \quad (5)$$

이때, 합성 분류기의 에러가 0.5를 넘는다면 다시 새로운 약한 분류기를 생성하기 위해 현재의 약한 분류기  $h_t$ 를 폐기하고 훈련영상을 선택하는 과정으로 돌아간다. 합성 분류기는 알고리즘의 최종단계에서 생성되는 강한 분류기를 구성하는 요소가 된다. 따라서 앞서 언급한 약한 분류기의 경우와 같은 이유로 데이터 집합  $S_m$ 에서 50% 이상의 신뢰도를 보여야 한다. 데이터 집합  $S_m$ 을 구성하는 모든 예제 각각의 가중치는 합성분류기의 정규화된 에러  $B_t$ 를 사용하여 식 (6)과 같이 갱신된다.

$$w_{t+1}(i) = w_t(i) B_t^{1 - |H_t(x_i) - y_i|}, \text{ 단 } B_t = E_t / (1 - E_t) \quad (6)$$

식 (6)에 따르면  $i$ 번째 데이터의 새로운 가중치는 합성 분류기가 옳은 판단을 내린 경우 그 값이 줄어들고 잘못된 판단을 내린 경우 동일하게 유지된다. 갱신된 가중치는 식 (2)에 의해 정규화 되므로 이러한 갱신과정의 효과는 반복  $t$ 에서 결정된 합성 분류기가 잘 분류하지 못한 데이터들의 가중치를 높이는 것이다. 갱신된 분포  $D_{t+1}$ 에 따라 새로운 훈련 집합이 선택된다. 즉,  $t+1$ 번째 반복의 목표는 이전의 반복을 통해 분류가 어려웠던 데이터들을 가장 잘 분류하는 새로운 분류기를 선택하여 합성 분류기에 추가하는 것이다. AdaBoost 알고리즘의 데이터 가중치 갱신과정은 이와 유사한 방식이나, 갱신의 기준이 바로 이전 단계의 약한 분류기 하나의 성능에만 의존한다는 점이 다르다.

약한 분류기를 선택하여 하나의 합성 분류기를 추가하는 part2의 과정은 사전에 정의한 약한 분류기 최대 개수인  $T_m$ 회 동안 반복된다. 그러나 다음의 종료 조건에 해당되는 경우에는  $T_m$ 을  $t-1$ 로 갱신하고 반복을 종료한다.

종료 조건:

1. 어떤 훈련 집합을 선택해도 약한 분류기의 에러가 항상 0.5를 초과
2. 어떤 훈련 집합을 선택해도 합성 분류기의 에러가 0.5를 초과
3. 합성분류기의 분류에러가 0인 경우, 즉 모든 데이터를 완벽하게 분류하는 경우

실제 학습 데이터를 적용해 본 결과 조건 2와 3에 해당하는 경우가 빈번히 발생하여 각 훈련 영상 집합들은 서로 다른  $T_m$  값을 가지게 된다. 이러한 학습과정을 통해 구한 최종 분류기  $F_m$ 의 에러는 식 (7) 과 같은 상한 값을 가진다(증명 과정은 [3]을 참조).

$$E_m \leq 2^T \cdot \prod_{i=1}^T \sqrt{E_i(1-E_i)} \quad (7)$$

**3.2 강한 분류기 구성을 위한 가중치 결정법**

얼굴 검출을 위한 강한 분류기  $C$ 는 훈련 데이터 집합  $S_m$ 의 최종 분류기  $F_m$ 을 기반으로 결정된다.  $C$ 는  $F_m$ 들을 가중치기반결합 방법에 의거하여 결합한 형태이다. 이때, 가중치의 결정 방식에 따라 분류기  $C$ 의 성능이 달라진다. 가중치 결정법은 그림 4에서와 같이 세 가지의 경우로 나누어 생각해 볼 수 있다.

=====  
Case1: uniform weight.

$$C = \begin{cases} 1 & \sum_{m=1}^M F_m(x) \geq \frac{1}{2} M \text{인 경우 } M \text{은 훈련집합의 전체개수} \\ 0 & \text{그외} \end{cases}$$

Case2: local optimum weight.  $\{x_1, \dots, x_m\}$

$$x_m = \log\left(\frac{1}{B_m}\right), B_m \text{은 최종 가설인 } F_m \text{의 정규 에러 값}$$

최종 가설  $F_m$ 의 에러 및 정규 에러 값은 각각,

$$E_m = \sum_{i=1}^M D(i) |F_m(x_i) - y_i|, B_m = E_m / (1 - E_m)$$

$$C = \begin{cases} 1 & \sum_{m=1}^M x_m F_m(x) \geq \frac{1}{2} \sum_{m=1}^M x_m \text{인 경우} \\ 0 & \text{그외} \end{cases}$$

Case3: global optimum weight.  $\{w_1, \dots, w_m\}$

$$w_m = \log\left(\frac{1}{B_m^v}\right), B_m^v \text{는 검증 집합 } V \text{에서의 최종 가설}$$

$F_m$ 의 정규화 된 에러

검증 집합  $V = \{(x_1^v, y_1^v), \dots, (x_K^v, y_K^v)\}$ 는 전체 데이터 집합  $S$ 에서 추출한 레이블이 표시된  $K$ 개의 데이터로 이루어져 있다.

검증 집합에서의 최종 가설  $F_m$ 의 에러 및 정규 에러 값은 각각,

$$E_m^v = \sum_{i=1}^K |F_m(x_i^v) - y_i^v|, B_m^v = E_m^v / (1 - E_m^v)$$

$$C = \begin{cases} 1 & \sum_{m=1}^M w_m F_m(x) \geq \frac{1}{2} \sum_{m=1}^M w_m \text{인 경우} \\ 0 & \text{그외} \end{cases}$$

=====  
그림 4 최종 분류기의 가중치 결정 방법

Case1은 각 데이터 집합  $S_m$ 의 모든 최종 분류기들에 같은 가중치를 할당하는 경우이다. 훈련 데이터 집합

의 특성이 강한 분류기의 구성에 영향을 주지 않는 특징이 있다. 따라서 각 데이터 집합간의 성질이 유사한 경우에만 성능을 보장할 수 있다.

Case2는 해당 훈련 집합의 분류성능에 따라 가중치를 결정하는 방법이다. 데이터 집합  $S_m$ 에서 최종 분류기  $F_m$ 의 분류 에러는 식 (8)과 같다.

$$E_m = \sum_{i=1}^M D(i) |F_m(x_i) - y_i| \quad (8)$$

가중치는 정규화된 에러  $B_m$ 을 기준으로 식 (9)와 같이 결정된다.

$$x_m = \log\left(\frac{1}{B_m}\right) \quad (9)$$

이 방법은 가중치가 해당 데이터 집합의 분류성능에만 의존 하므로 분류가 용이한 데이터 집합으로부터 선택된  $F_m$ 이 높은 가중치를 갖는 특징이 있다. Case1과 마찬가지로 각 데이터 집합간의 성질이 유사한 경우에 유용하며 이 경우 성능이 우수한 최종 분류기에 더 높은 가중치를 할당하므로 보다 우수한 성능의 강한 분류기를 기대할 수 있다. 그러나 Case1과 2는 훈련 데이터 집합간의 분류 난이도가 다르거나 새롭게 추가된 데이터 집합이 이미 학습된 데이터 집합과 이질적인 특성을 가질 때 좋은 성능을 유지할 수 없다. 특히 Case2에서는 데이터 집합간 분류 난이도의 차이가 클 경우, 상대적으로 분류가 어려운 집단으로부터 생성된 분류기들은 모두 무시되어 강한 분류기의 성능을 저하시킨다. 그러나 새롭게 추가될 데이터의 성질을 미리 예측할 수 없는 점진적 학습의 특성상 분류기가 개별 데이터 집합 고유의 성질에 영향을 받지 않도록 하는 것은 매우 중요한 문제이다. Case3은 이러한 문제 해결을 위해 제안된 방법이다.

식 (10)은 검증 데이터 집합  $V$ 에서 최종 분류기  $F_m$ 의 분류에러를 나타낸다.

$$E_m^v = \sum_{i=1}^K |F_m(x_i^v) - y_i^v| \quad (10)$$

가중치는  $F_m$ 의 분류에러인  $E_m^v$ 의 정규화된 에러  $B_m^v$ 를 기준으로 식 (11)과 같이 결정된다.

$$w_m = \log\left(\frac{1}{B_m^v}\right) \quad (11)$$

즉, 각 분류기의 가중치는 검증 데이터 집합  $V$ 에서의 분류성능에 의해 결정되는 것이다. 검증 데이터 집합은 학습을 진행하면서 사용한 원본 데이터의 일부를 보관한 것이다. 검증 집합을 구성할 데이터의 개수  $K$ 는 사전에 정의한다. 현재 훈련중인  $m$ 번째 데이터 집합에서 검증 집합에 추가할 데이터의 개수는  $K/m$ 개 이다. 각 데이터들은 동일한 발생 확률을 가지며 데이터 집합 각

각은 모집단을 반영한다. 이러한 경우 가장 적절한 표본 추출 방식은 각 데이터 집단의 요소들을 랜덤하게 추출하는 것이다. 각 집단으로부터 같은 개수의 데이터를 임의로 추출하는 방식은 미래의 데이터를 사전에 정확히 알 수 없는 점진적 학습 방식의 특성상 취할 수 있는 최선의 방법이다. 더욱이 이러한 샘플링 모집단에 대한 그릇된 사전 지식을 바탕으로 분류 했을 때 초래 되는 분류 에러에 대한 위험을 최소화 할 수 있다[16]. 이러한 검증 집합 기반의 가중치 결정 방식은 다른 방법들과 달리 전체 훈련 집합을 대표하는 새로운 부분 집합에서 각 데이터 집합의 최종 분류기  $F_m$ 의 신뢰도를 결정한다. 따라서 부분적인 데이터 집합의 영향은 줄이면서 전체적인 훈련 데이터를 기준으로 최적화된 가중치를 결정하는 것이 가능하다. 또한, 검증 집합은 FIFO 구조로 구성되어 학습이 진행됨에 따라 각 분류기의 가중치가 항상 새롭게 추가되는 데이터의 영향을 반영하여 갱신될 수 있도록 하였다.

4. 실험결과

훈련 집합을 구성하는 데이터는 얼굴과 얼굴이 아닌 클래스의 영상이다. 영상의 크기는 20x20이며 영상 내부에 표현 가능한 특징의 개수는 총 33,308개 이다. 얼굴 영상은 FERET 데이터베이스로부터 획득하였다. 영상에 얼굴 외의 영역이 포함되지 않도록 사전정보(ground truth)에 있는 눈과 입 중앙의 좌표를 기준으로 영상을 절단하여 20x20의 크기로 정규화 하였다. 사전정보에 좌표가 없는 영상의 경우, 눈 중앙과 입 중앙의 좌표를 영상에 직접 표시하여 절단의 기준으로 사용

하였다. 사용된 얼굴 데이터의 개수는 정면 얼굴 6,000장, 이질적 데이터 추가 실험을 위한 측면 얼굴 400장이 다. 얼굴이 아닌 훈련 데이터 집합은 인터넷을 통해 획득한 얼굴을 포함하지 않는 임의의 영상으로부터 6,400개의 부분영상을 취하여 구성하였다.

첫 번째 실험은 본 연구에서 사용된 점진적 학습 방법의 성능과 모든 데이터를 한 번에 학습하는 AdaBoost의 성능을 비교하기 위한 실험이다. 실험에 사용된 영상은 다양한 변화를 포함하는 정면 얼굴 영상 6,000장과, 임의의 얼굴이 아닌 영상 6,000장이다. 이중 훈련 데이터로는 각각 5,000장의 얼굴과 얼굴이 아닌 영상이 사용되었으며 훈련에 사용되지 않은 얼굴 영상 1,000장과 얼굴이 아닌 영상 1,000장을 테스트에 사용하였다. 훈련 집합과 테스트 집합을 각각 서로 다른 조합으로 선택하여 총 6회의 실험을 반복 하였다. 그림 5는 훈련에 사용된 정면 얼굴 데이터의 예이다.

표 2는 첫 번째 실험의 결과이다. AdaBoost와 점진적 학습법을 통해 생성된 강한 분류기의 분류성능을 비교하였다. 표의 수치는 분류 에러율(%)을 의미한다. 각 테스트의 강한 분류기를 구성하는 약한 분류기의 개수(특정 수)는 점진적 학습법에 의해 생성된 개수를 기준으로 하였다. 약한 분류기의 반복적 선택과정에서 3.1절에 언급된바 있는 종료조건을 만족하는 시점이 서로 다르므로 테스트 마다 선택된 약한 분류기의 개수가 달라진다. 점진적 학습법은 최종 분류기의 가중치 결정법에 따라 세 가지 방식으로 분류하여 실험하였다(3.2절 참고). 6회의 반복 실험결과 점진적 학습 알고리즘을 적용하여 구성한 분류기가 AdaBoost 학습법을 통해 생성된

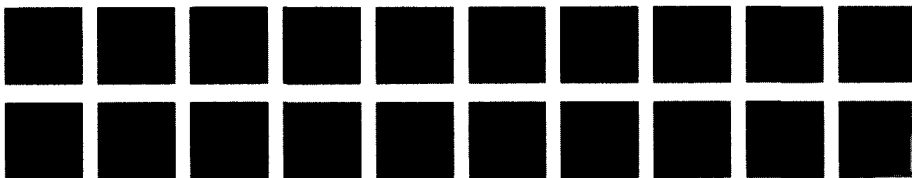


그림 5 훈련에 쓰인 정면 얼굴 이미지들의 예

표 2 AdaBoost와 점진적 학습법에 기반한 분류기의 검출 에러율 비교

| 테스트 / 특정수   | 방법 | AdaBoost | Case 1 (Uniform weight) | Case 2 (Local optimum weight) | Case 3 (Global optimum weight) |       |       |        |
|-------------|----|----------|-------------------------|-------------------------------|--------------------------------|-------|-------|--------|
|             |    |          |                         |                               | (200)                          | (400) | (600) | (1000) |
| Test1 / 535 |    | 3.85     | 2.85                    | 2.45                          | 2.4                            | 2.4   | 2.4   | 2.3    |
| Test2 / 494 |    | 5.45     | 5.3                     | 3.75                          | 3.9                            | 3.9   | 4     | 3.9    |
| Test3 / 456 |    | 3.75     | 3.65                    | 2.6                           | 2.6                            | 2.5   | 2.7   | 2.6    |
| Test4 / 514 |    | 5.3      | 5.05                    | 3.95                          | 4.4                            | 4.55  | 4.2   | 4.55   |
| Test5 / 471 |    | 6.4      | 5.75                    | 5.75                          | 5.6                            | 5.4   | 5.5   | 5.65   |
| Test6 / 525 |    | 4.15     | 3.05                    | 3.05                          | 3                              | 3     | 2.95  | 3      |

분류기보다 우수한 성능을 보임을 알 수 있다. 앞서 언급했듯이 첫 번째 실험에 사용된 훈련 데이터는 모두 정면 얼굴 영상이다. 이처럼 각 데이터의 집합간의 특성이 크게 다르지 않은 경우, Case 1, 2, 3에 해당하는 세 가지 가중치 결정법의 성능은 근소한 차이를 보인다. Case 1의 에러율이 Case2와 3에 비해 상대적으로 높은 것을 알 수 있는데, 이는 추가되는 각 데이터 집합의 특성이 전혀 반영되지 않았기 때문이다. Case3에서 괄호 안의 숫자는 검증 셋을 구성하는 데이터의 개수를 나타내며, 이 결과는 각 집단을 대표하는 데이터를 임의로 선택하는 경우 선택된 소집단의 대표성이 데이터의 개수에 크게 영향을 받지 않음을 보인다.

그림 6은 점진적 학습의 가중치 결정법 별로 데이터 집합의 추가에 따른 에러율 변화를 나타낸 그래프이다. 하나의 데이터 집합은 각각 랜덤하게 선택된 얼굴 영상 200장과 얼굴이 아닌 영상 200장으로 구성되며, 총 10,000장의 영상을 훈련하기 위해 25회의 추가가 이루어졌다. 그래프에서 보이는 에러율은 6회의 실험결과를 평균한 값이며, Case3은 검증 집합의 데이터가 400개인 경우를 기준으로 한 것이다. 세 가지 방법 모두 훈련 데이터의 개수가 증가함에 따라 점진적으로 에러가 감소

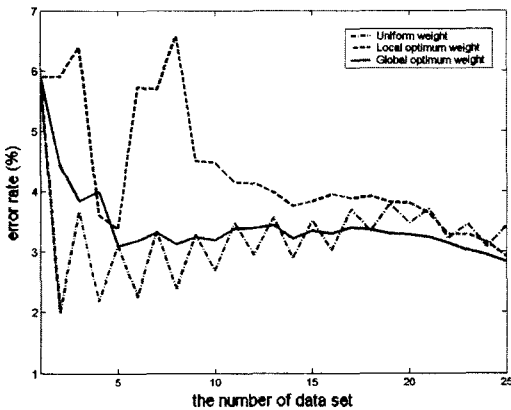


그림 6 데이터 집합 추가에 따른 에러율의 변화



(a) 26번째 훈련 영상을 구성하는 -60 회전된 얼굴 영상



(b) 27번째 훈련 영상을 구성하는 +60 회전된 얼굴 영상

그림 7 이질적인 데이터 추가학습에 사용된 영상의 예

함을 볼 수 있다. Local optimum weight 방법의 경우 데이터 집합의 개수가 적을 때 전체 데이터의 특성이 반영이 되지 않는 문제가 높은 에러율을 나타내게 하였다고 분석된다. 최종적으로 수렴한 에러 값은 Local optimum weight와 Global optimum weight의 경우가 유사하지만 가장 안정적인 감소 효과를 보이는 방법은 Global optimum weight를 사용한 경우이다.

두 번째 실험은 이질적인 데이터 집합을 추가 학습하는 경우 가중치 결정법에 따른 성능변화를 측정하기 위한 실험이다. 이질적인 데이터로는 회전된 얼굴영상을 사용하였다. 정면얼굴을 학습한 첫 번째 실험을 통해 생성된 6개의 분류기(테스트1~6의 결과)에 각각 그림 7과 같이  $\pm 60^\circ$ 로 회전된 얼굴 영상 집합을 추가 학습하였다. 테스트 데이터로는 첫 번째 실험에서와 마찬가지로 정면 얼굴 1,000장 얼굴이 아닌 영상 1,000장이 사용되었다.

그림 8은 이질적인 데이터 집합의 추가에 따른 에러의 변화율을 나타낸다. 26, 27번째 데이터 집합은 그림 7에서 보인 데이터와 같다. 각 에러율은 6회 실험 에러의 평균값이다. Local optimum weight 결정법의 에러가 가장 급격한 증가율을 보이는데, 이는 결합 분류기와 다른 학습 데이터와의 관계가 전혀 고려되지 않으므로 이질적인 데이터의 영향이 크게 증가하는 것이다. Uniform weight의 경우에도 데이터 집합의 특성이 반영되지 않으므로 에러가 증가하지만 모든 데이터 집합의 가중치가 동일하게 반영되므로 Local optimum weight에 비해 영향이 서서히 증가한다. 이와는 달리 Global optimum weight 결정법은 새롭게 학습된 최종 분류기의 가중치가 전체 훈련 집합을 대표하는 집단을 기준으로 판단되므로 이질적인 데이터 셋에 의한 갑작스런 에러 증가를 방지할 수 있다. 뿐만 아니라, 본 실험과 같은 경우, 측면 얼굴 데이터의 학습량을 늘려 가면 측면 얼굴 검출을 위한 분류기의 중요도를 높일 수 있다.



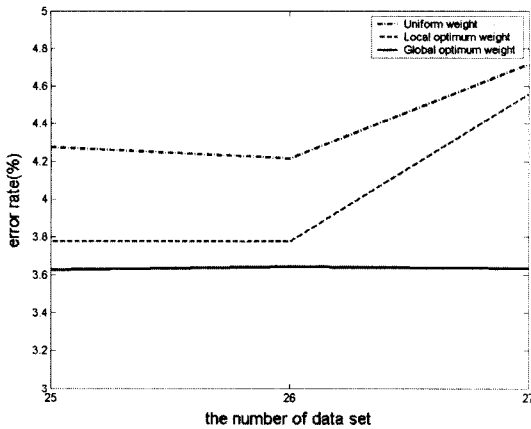


그림 8 이질적인 데이터 집합 추가에 따른 에러 변화율

### 5. 결론 및 추후과제

본 논문에서는 얼굴 검출에 적합한 점진적인 학습 알고리즘을 제안하였다. 검출을 위한 분류기 학습에 Learn++ 알고리즘의 구조를 개선하여 훈련 데이터의 추가 학습이 불가능한 AdaBoost의 문제점을 해결하였다. 또한 새롭게 제안한 Modified Learn++ 방법은 Learn++ 알고리즘이 안고 있는 분류기 가중치의 지역적 최적화 문제를 극복한 것이다. 전체 훈련 집합을 대표하는 검증 집합을 생성함으로써 부분적인 데이터 집합의 영향은 줄이고 전체적인 훈련 데이터를 기준으로 최적화된 가중치의 결정이 가능함을 보였다. 또한, 학습 데이터의 점진적 증가에 따라 분류기의 가중치를 갱신하여 새롭게 추가된 데이터 집합의 영향을 반영하도록 한다. 이러한 global optimum weight 결정 방식은 데이터 집합의 추가에 따른 분류 에러 감소를 가장 안정적 보장하며 이질적인 데이터 추가로 인한 갑작스런 에러 증가 방지효과가 있다. 제안한 학습 알고리즘은 현재 축면 얼굴 데이터의 학습과 분류기 조합 방식의 개선을 통해 다양한 포즈의 얼굴 검출에 사용될 분류기 학습법으로 확장이 진행 중이다.

### 참고 문헌

[1] Y. Freund and R. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences* Vol. 55, pp. 119-139, 1997.  
 [2] P. Viola and M. Jones, "Robust real time object detection," *In: IEEE ICCV Workshop on Statistical and Computational Theories of Vision*, 2001.

[3] R. Polikar, L. Udpa, S. Udpa and V. Honavar, "Learn++: An Incremental Learning Algorithm for Supervised Neural Networks," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 31, No. 4., pp. 497-508, 2001.  
 [4] T. Agui, Y. Kokubo, H. Nagashashi and T. Nagao, "Extraction of Face Recognition from Monochromatic Photographs Using Neural Networks," *Proceedings of Second International Conference on Automation, Robotics, and Computer Vision*, Vol. 1, pp. 18.8.1-18.8.5, 1992.  
 [5] R. Fe'raud, O.J. Bernier, J.-E. Villet and M. Collobert, "A Fast and Accurate Face Detector Based on Neural Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 1, pp. 42-53, 2001.  
 [6] S.-H. Lin, S.-Y., Kung and L.-J. Lin, "Face Recognition/Detection by Probabilistic Decision-Based Neural Network," *IEEE Transactions on Neural Networks*, Vol. 8, No. 1, pp. 114-132, 1997.  
 [7] H. Rowley, S. Baluja and T. Kanade, "Neural Network-Based Face Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 1, pp. 23-38, 1998.  
 [8] H. Rowley, S. Baluja and T. Kanade, "Human Face Detection in Visual Scenes," *Advances in Neural Information Processing Systems*, pp. 875-881, 1996.  
 [9] K.-K. Sung and T. Poggio, "Example-Based Learning for View-Based Human Face Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 1, pp. 39-51, 1998.  
 [10] R. Polikar, S. Krause and L. Burd, "Ensemble of classifiers based incremental learning with dynamic voting weight update," *Proceedings of the International Joint Conference on Neural Networks*, Vol. 4, pp. 2770 -2775, 2003.  
 [11] H.P. Graf, T. Chen, E. Petajan and E. Cosatto, "Locating Faces and Facial Parts," *Proceedings of First International Workshop Automatic Face and Gesture Recognition*, pp. 41-46, 1995.  
 [12] J. Cai and A. Goshtasby, "Detecting human faces in color images," *Image and Vision Computing*, Vol. 18, pp. 63-75, 1999.  
 [13] I. Craw, H. Ellis and J. Lishman, "Automatic Extraction of Face Features," *Pattern Recognition Letters*, Vol. 5, pp. 183-187, 1987.  
 [14] R. Duda, P. Hart and D. Stork, "Pattern Classification," *2nd Ed., New York: Wiley*, 2001.  
 [15] N. Littlestone and M. Warmuth, "Weighted majority algorithm," *Information and Computation*, Vol. 108, pp. 212-261, 1994.  
 [16] R. Ziemer, "Elements of engineering probability & statistics," *International Ed., Prentice Hall*, 1997.



박 지 영

2000년 성균관대학교 전기전자 및 컴퓨터공학부(학사). 2002년 성균관대학교 전기전자 및 컴퓨터공학부(석사). 2002년~현재 성균관대학교 정보통신공학부 박사과정. 관심분야는 컴퓨터비전, 시각기반 휴먼-로봇 인터페이스



이 준 호

1985년 서울대학교 전기공학과(학사). 1987년 The Pennsylvania State University, Department of Electrical Engineering(석사). 1994년 Purdue University, School of Electrical and Computer Engineering(공학박사). 1994년~1995년 University of California, Riverside 연구원. 1995년~1997년 한국과학기술연구원 선임연구원. 1997년~2001년 성균관대학교 전기전자컴퓨터공학부 조교수. 2001년~현재 성균관대학교 정보통신공학부 부교수. 관심분야는 컴퓨터비전 기반 HCI