

■ 2009년도 학생논문 경진대회 수상작

개방형 웹 서비스를 위한 증가적 얼굴 어노테이션 (Incremental Face Annotation for Open Web Service)

최 권 택 [†] 변 혜 란 ^{**}
(Kwontaeg Choi) (Hyeran Byun)

요 약 최근 Flickr, Facebook, Cyworld 처럼 사진 공유를 기반으로 하는 소셜 웹 서비스의 성공과 발달로 얼굴 검출/인식 기술을 이룬 서비스에 접목하려는 다양한 시도가 진행되고 있다. 그러나 인식을 향상에만 초점을 맞춘 기존의 일관처리 기반의 연구들은 수백만의 이용자가 수시로 접근하는 서비스에 적용하기 어렵다. 본 논문에서는 시간에 따라 증가하는 거대한 얼굴 영상 데이터베이스를 효과적으로 분류하기 위해 랜덤 사상(Random Projectio, RP) 비선형 회귀(Non-linear Regression) 그리고 REST(REpresentational State Transfer) 규약을 사용해 새로운 증가적 얼굴 어노테이션 방법을 제안하고자 한다. 다양한 비교실험 결과에서 제안된 방법은 향상된 인식률과 낮은 계산 복잡도 기록했다. 따라서 제안된 방법은 대규모 웹서비스에 적합한 얼굴 어노테이션 알고리즘이다.

키워드 : 얼굴 인식, 소셜웹서비스, 랜덤 사상

Abstract Recently, photo sharing and publishing based Social Network Sites(SNSs) are increasingly attracting the attention of academic and industry researches. Unlike the face recognition environment addressed by existing works, face annotation problem under SNSs is differentiated in terms of daily updated images database, a limited number of training set and millions of users. Thus, conventional approach may not deal with these problems. In this paper, we proposed a face annotation method for sharing and publishing photographs that contain faces under a social network service using random projection, non-linear regression and representational state transfer. Our experiments on several databases show that the proposed method records an almost constant execution time with comparable accuracy of the PCA-SVM classifier.

Key words : Face Recognition, SNS, Random Projeciton

· 이 논문 또는 저서는 2008년 정부(교육과학기술부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임(KRF-2008-313-D00998) 또한 이 연구의 주요 계산은 한국과학기술정보연구원(KISTI)이 무상으로 제공한 슈퍼 컴퓨팅 자원을 이용하여 수행되었음

[†] 학생회원 : 연세대학교 컴퓨터과학과
choikt@yonsei.ac.kr
^{**} 종신회원 : 연세대학교 컴퓨터과학과 교수
hrbyun@yonsei.ac.kr
(Corresponding author임)

논문접수 : 2009년 6월 1일
심사완료 : 2009년 6월 10일

Copyright©2009 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 소프트웨어 및 응용 제36권 제8호(2009.8)

1. 서론

얼굴 검출/인식 기술은 최근까지 매우 다양한 분야에 적용되었다[1]. 그러나 주로 보안 분야에서 활용되고 있을뿐 일상생활에 활용되지 못하고 있다. 특히 인식기의 작은 오류가 전체 보안 시스템의 치명적인 오류로 이어지기 때문에 관련 기술 및 응용분야가 매우 제한적이다. 최근 웹 2.0의 성공과 발달로 기존의 많은 응용프로그램들이 웹 서비스로 전환되고 있고, 얼굴 검출/인식 기술을 이룬 서비스와 접목하려는 다양한 시도가 진행되고 있다.

구글, 야후, 다음, 네이버 같은 대형 검색엔진은 이미 얼굴 검출기능을 이용해 인물사진을 검색할 수 있는 서비스를 지원하고 있다. 얼굴 인식 기술을 접목한 서비스로는 Riya.com과 Olaworks.com가 있다. 이 두 서비스

는 얼굴 사진 어노테이션을 통해 사진을 공유하고 검색할 수 있는 서비스를 제공한다. Myheritage.com은 주어진 사진과 유사한 저명인사를 찾아주는 서비스와 가계도를 공유하기 위해 얼굴 인식기술을 활용하였다. [2]에서는 소셜 네트워크 정보를 활용해 얼굴 인식 성능을 높이는 알고리즘이 제안되었고, [3]에서는 Facebook에서 추출한 실제 얼굴 영상을 이용해 기존 알고리즘의 성능을 평가하였다.

Flickr, Facebook, Cyworld 처럼 사진 공유를 기반으로 하는 소셜 네트워크 서비스(Social Network Service, SNS)의 폭발적인 성공은 공유되는 많은 사진들이 인물 사진이라는 점을 고려할 때 얼굴 검출/인식기술은 매우 중요한 요소가 된다. 얼굴 사진에 대한 어노테이션을 통해 전세계적으로 검색 과 공유를 지원하는 이러한 서비스는 일상 생활에서 얼굴 인식 기술이 적용될 수 있음을 의미한다. 또한 인식기의 오류는 수동적으로 사진에 어노테이션을 다시 부여하게 하는 약간의 불편함을 초래할 뿐 보호 정보 및 재산의 손실로 이어지지 않는다. 즉 소셜 웹 서비스에 결합된 얼굴 검출/인식 기술은 대중성과 애러러용이라는 특성을 가지기 때문에 기존의 보안 관련 얼굴 인식 기술이 겪는 한계를 극복할 수 있을 것으로 기대된다.

얼굴 인식 문제를 풀기 위해 다양한 방법이 제안되었고, 최근 인식 성능에서 큰 진전이 있는 것으로 보고되었다[4]. [1,4]에서 제시된 대부분의 얼굴 인식 알고리즘은 학습 데이터셋이 사전에 주어져서 일괄처리 방식의 최적화 알고리즘을 사용해 학습을 수행한다. 그러나 추가적인 학습 샘플이 발생할 경우 처음부터 다시 학습해야 하는 문제가 있다. 이러한 학습 알고리즘은 계산 비용이 매우 높기 때문에 소셜 네트워크 서비스처럼 매우 많은 사람들이 수시로 시스템에 접속해 사진을 업데이트하고 검색하는 경우 적합하지 않다.

좀더 합리적인 방법론은 새로운 학습데이터만을 이용해 기존의 학습 모델을 효율적으로 업데이트하는 방법이다. 이러한 방법은 증가적/온라인/적응적(Incremental/Online/Adaptive) 학습이라는 이름으로 많이 연구되었다. 특히 얼굴 인식에서는 IPCA(Incremental Principal Component), ILDA(Incremental Linear Discriminant Analysis), INMF(Incremental Non-negative Matrix Factorization)같은 특징 자체의 변화를 시간적으로 모델링하는 증가적 부공간(Incremental Subspace) 알고리즘이 연구되었고, ISVM(Incremental Support Vector Machine)과 INN(Incremental Neural Network)같은 특징분포의 변화를 시간적으로 모델링하는 분류기들도 연구되었다. 이들 방법은 추가적으로 발생하는 학습 샘플의 발생 빈도가 빈번하지 않고, 분류할 클래스수가 적을

때 일괄처리 방식에 근접하는 인식률을 얻으면서도 계산 비용을 매우 낮출수 있음을 보여주었다.

그러나 수백만명의 이용자가 수시로 데이터를 업데이트하는 상황을 고려하지 않았기 때문에 소셜 네트워크 서비스 상에서 공유 및 검색을 위한 얼굴 어노테이션 방법으로 적합하지 않다. 인식을 향상에만 연구의 초점을 맞추고 있는 기존의 얼굴 인식방법들은 대중성을 기반으로 하는 거대한 웹 서비스에 적용하기 쉽지 않다. 소셜 네트워크 서비스에 적용된 얼굴 인식은 PC기반의 얼굴 인식과는 다양한 관점에서 차이가 있고 표 1에 정리되어 있다.

표 1 PC와 SNS기반의 얼굴 인식 비교

	PC기반 얼굴 인식	SNS기반 얼굴 인식
주요용용	보안	공유
목적	매우 높은 신뢰도	높은 신뢰도 및 효율성
데이터특성	적용	매우 크고, 꾸준히 증가
형태	PC기반의 제품	웹기반의 서비스
사용자	제한된 소수	전세계 이용자

본 논문에서는 시간에 따라 증가하는 거대한 얼굴 영상 데이터베이스를 효과적으로 어노테이션하기 위해 랜덤 사상(Random Projection, RP), 비선형 회귀(Non-linear Regression) 그리고 REST(REpresentational State Transfer) 규약을 사용한 새로운 증가적 얼굴 분류기를 제안하고자 한다.

2. 문제 정의

동적이고 거대한 소셜 웹에서 얼굴 어노테이션 문제를 해결하기 위해 그림 1에서처럼 적어도 3개의 문제를 풀어야 한다. 보안 목적의 얼굴 인식과는 달리 개인의 사진 영상은 매일 꾸준히 사용자에 의해 업데이트 된다. 따라서 첫 번째 문제는 새로운 학습 데이터만을 사용해 기존의 학습된 모델을 효율적으로 재조정하는 것이다. 이를 증가적 샘플(Incremental Samples) 문제라 부르겠다. 두 번째 문제는 분류해야 할 사람의 수가 고정적이지 않고 끊임없이 증가하는 상황에서 새롭게 추가된 사람을 분류하는 문제이다. 이를 증가적 다중 분류(Incremental Multi-Class) 문제라 부르겠다. 마지막 문제는 한 대의 컴퓨터에서 수백만병을 처리할 수 있는 단일 분류기를 구성할 수 없기 때문에 분류기들을 네트워크 상에 분산 배치하여 대규모 서비스가 가능하도록 하는 것이다. 이를 증가적 서버 학습이라 부르겠다.

3. 제안하는 얼굴 어노테이션 방법

3.1 증가적 샘플 학습

[5,6]에서 처음으로 SNS 환경에서 시간에 따라 변화

는 얼굴 영상을 효과적으로 어노테이션 하는 방법이 제안되었다. 반복적인 알고리즘에 의한 최적화 방식을 사용하는 SVM(Support Vector Machine), 신경망, 베이저안 네트워크 같은 잘 알려진 분류기와는 달리, 최소제곱해(Least Square Solution)를 사용한 회귀(Regression) 알고리즘은 반복 연산없이 의사역행렬(Pseudo-inverse)에 의해 학습 파라미터를 구한다. 최소제곱해 α 는 아래와 같은 수식에 의해 구해진다[6].

$$\alpha = (\psi(X)^T \psi(X) + bI)^{-1} \psi(X)^T y \quad (1)$$

여기서 X 는 얼굴 영상에 대한 행벡터를 모아놓은 행렬, b 는 안정적인 역행렬을 구하기 위한 정규화값(regularization factor), y 는 0혹은 1로 이루어진 학습 레이블 벡터이다. $\psi(\cdot)$ 함수는 비선형 분포인 얼굴 영상을 표현하기 위한 다항식 기반의 비선형 함수이다[7].

특정 시간 t 에 추가적인 얼굴 샘플 X_t 와 이에 대한 레이블 벡터 y_t 가 주어지면 가중치 벡터 α_t 는 재귀적 최소해(Recursive Least Squares, RLS) 알고리즘에 의해 아래 수식으로 구해진다[8].

$$\alpha_t = \alpha_{t-1} + M_t^{-1} \psi(X_t)^T (y_t - \psi(X_t) \alpha_{t-1}) \quad (2)$$

여기서

$$M_t^{-1} = M_{t-1}^{-1} + M_{t-1}^{-1} \psi(X_t)^T (I + \psi(X_t) M_{t-1}^{-1} \psi(X_t)^T)^{-1} \psi(X_t) M_{t-1}^{-1} \quad (3)$$

새로운 학습 샘플이 주어졌을 때 모든 샘플을 가지고 α 를 추정하는 식 (1)에 비해 이전 α_{t-1} 와 새로운 학습 샘플 X_t 만을 사용해서 α_t 를 추정하기 때문에 SNS 환경에서 수시로 늘어나는 얼굴 영상을 학습하기에 매우 적합한 방법이다.

테스트 샘플 x 에 대해 $\psi(x) \alpha_t \geq 0.5$ 이면 양의 샘플이

되고 그렇지 않으면 음의 샘플이 된다. 따라서 식 (2)의 회귀모델은 이진 분류 문제만을 풀 수 있다.

3.2 증가적 다중 분류 학습

얼굴 인식과 같은 다중 클래스 문제를 풀기 위해 보편적으로 일대다(one-against-all) 분할 방법을 적용한다. k 개 클래스를 갖는 최소제곱해 벡터 α_t 를 모아 놓은 $\Theta_t = [\alpha_t^1, \alpha_t^2, \dots, \alpha_t^{k-1}, \alpha_t^k]$ 와 Y_t 지시자 행렬을 정의하고 이를 식 (5)에 대입하면 다중 클래스 문제를 풀기 위한 최소제곱해는 아래 수식에 의해 계산된다.

$$\Theta_{t-1} = \Theta_{t-1} + M_t^{-1} \psi(X_t)^T (Y_t - \psi(X_t) \Theta_{t-1}) \quad (4)$$

[5,6]에서는 식 (4)를 기반으로 개인별로 Θ_t 를 유지함으로써 다중 분류 문제를 해결 했지만, 요구되는 메모리 관점에서 여전히 증가하는 대규모 분류에는 적합하지 않다. 이러한 일대다 방식도 끊임없이 분류할 클래스가 증가하고, 이수가 수백만에 이르면 시간 및 메모리 복잡도가 급격히 증가하고 인식 성능이 매우 떨어진다.

예를 들어 백만명의 이용자를 가지고 있는 서비스의 경우 Θ_t 를 위한 메모리 크기는 $800M \approx 100 \times 1000000 \times 8$ (특정수 100, double 형 변수)가 된다. 새롭게 추가되는 학습데이터가 10개일 때 지시자 행렬 Y_t 의 크기도 $80M \approx 10 \times 1000000 \times 8$ 가 된다. 따라서 식 (4)는 실질적으로 계산되기 매우 힘들다. 사용자 수가 적더라도 매우 큰 매트릭스인 Θ_t 와 Y_t 는 업데이트 주기가 매우 짧은 상황에서 시스템에 큰 부하를 발생시킨다.

본 논문에서 이러한 문제를 해결하기 위해 부분 업데이트 방법을 제안한다. SNS 환경에서 이용자는 적은 수의 다른 이용자와 온라인 관계를 유지하기 때문에 거대한 Θ_t 와 Y_t 를 구성할 필요가 없다. 식 (4)에서의 매우 큰 메모리 요구 문제를 해결하기 위해 우선 거대한 Θ_t 는 파

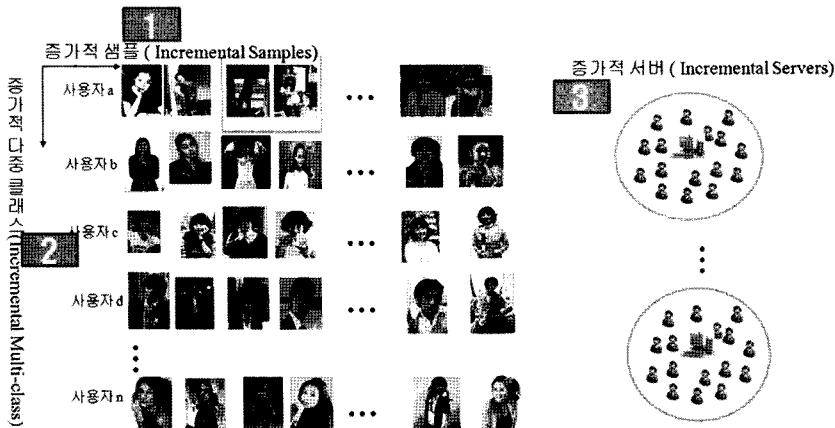


그림 1 문제 정의

일에 기록하고 사용자 i 와 관련된 사용자만의 학습 계수로 구성된 일시적인 Θ_i^t 를 아래 식 (5)처럼 계산한다.

$$\Theta_i^t = \Theta_{i-1}^t + M_t^{-1} \psi(X_i)^T (Y_i - \psi(X_i) \Theta_{i-1}^t) \quad (5)$$

예를 들어, 사용자 i 가 50명의 온라인 관계를 형성하고 있다면 Θ_i^t 의 크기는 $100 \times 50 \times 8$, Y_i^t 의 크기는 $10 \times 50 \times 8$ 그리고 P_t^{-1} 의 크기는 $100 \times 100 \times 8$ 이 된다. 따라서 학습 파라미터는 적은 메모리만을 사용해 효율적으로 계산될 수 있다.

식 (5)에서 주목할 점은 $(P_t^{-1})^i$ 가 아닌 P_t^{-1} 가 사용되었다. 식 (3)에서 알 수 있듯 이 항은 클래스 레이블과 상관없이 계산되기 때문에 사용자마다 계산할 필요 없이 시스템내의 모든 사용자에게 공통적으로 적용된다. 따라서 식 (5)는 매우 많은 사용자를 보유한 서비스 운영을 위해 메모리 소비를 급격하게 줄일 수 있다. 또한 Θ_i 대신 Θ_i^t 를 사용할 경우 인식률 향상을 이끌어 낼 수 있다. 특정 사용자와 관련된 적은수의 사용자만을 고려해 인식할 대상의 수를 줄일 수 있기 때문이다. 예를 들어 일반적으로 사진을 올릴 때 이벤트 단위로 모아서 올리는 경우, 나와 온라인 관계를 맺고 있는 전체 인원을 대상으로 인식하지 말고, 이벤트에 등장하는 사람만으로 Θ_i^t 를 구성할 수 있기 때문에 인식할 대상의 숫자가 적어지고 이는 인식률 향상으로 이어진다.

그러나 식 (5)는 새로운 클래스가 추가 될 때 이를 반영하지 못한다. 추가된 클래스를 구분하기 위한 학습 파라미터가 없기 때문이다. 일반성을 잃지 않고, 새로운 학습 샘플 X_i 가 새로운 클래스에 대한 학습 샘플로만 이루어졌다고 가정하면, 새로 추가된 클래스를 고려한 Θ_i^t 는 아래 수식처럼 다시 쓸 수 있다.

$$\Theta_i^t = [\Theta_{i-1}^t + M_t^{-1} \psi(X_i)^T (0 - \psi(X_i) \Theta_{i-1}^t) \alpha_0] \quad (6)$$

추가적인 샘플은 기존 클래스 모두에게 네거티브 샘플이 되기 때문에 식 (5)의 Y_i^t 는 0행렬이 된다. 그리고 α_0 벡터는 새로운 클래스를 위한 학습 계수로 아래 수식처럼 계산된다.

$$\begin{aligned} \alpha_0 &= (\psi(X_{1:t})^T \psi(X_{1:t}) + bI)^{-1} \psi(X_{1:t})^T y_{1:t} \\ &= M_t^{-1} (\psi(X_{1:t-1})^T \cdot 0 + \psi(X_t)^T \cdot 1) \\ &= M_t^{-1} \psi(X_t)^T \cdot 1 \end{aligned} \quad (7)$$

여기서 $X_{1:t}$ 표현은 $X_1 \sim X_t$ 까지의 모든 샘플을 의미하고, $y_{1:t}$ 표현은 $y_1 \sim y_t$ 까지의 모든 샘플에 대한 레이블 벡터를 의미한다. 추가적인 샘플은 기존 클래스 모두에게 네거티브 샘플이 되기 때문에 $y_{1:t-1}$ 의 원소는 모두 0이 되고, y_t 의 모든 원소는 1이 된다. 기존의 Θ_{i-1}^t 에 새로운 클래스를 위한 α_0 가 추가되기 때문에 Θ_i^t 의 열의

크기는 1증가한다. Θ_{i-1}^t 가 k 개의 분류 문제를 풀 수 있다면 Θ_i^t 는 $k+1$ 개의 분류 문제를 풀 수 있게 된다.

3.3 희소 랜덤 사상 기반의 특징 추출

지금까지의 수식에서는 고차원의 얼굴영상 데이터가 입력 특징으로 사용되었다. 32×32 크기의 영상 이미지가 사용되고, 다항식 차수 2차가 사용될 경우 $\psi(x)$ 의 차원은 $3075 = 1 + r + D(2r - 1) = 1 + 2 + 1024(2 \times 2 - 1)$ [7]에 이른다. 이는 상당히 고차원 데이터이기 때문에 계산 복잡도를 증가시키고 무엇보다도 학습 샘플이 작을 경우 역행렬과정에서 랭크(Rank)문제로 부정확한 역행렬이 구해지게 된다.

이러한 고차원의 얼굴 영상을 분석하기 쉬운 저차원 특징으로 차원 축소하는 PCA, LDA 그리고 ICA 같은 부공간 학습 방법은 오랫동안 연구되었다. 그러나 이러한 부공간 학습에 의한 차원 축소 알고리즘은 SNS환경에 적합하지 않다. 첫 번째 이유는 최적의 부공간을 추정하기 위해 일괄처리 기반의 알고리즘을 사용하는데 이는 계산 복잡도가 매우 높다. PCA와 LDA의 경우 증가적으로 부공간을 추정하는 IPCA, ILDA가 제안되어 있지만 SVD를 기반으로 하기 때문에 여전히 온라인 실시간 서비스를 위해서는 계산 복잡도가 매우 크다. 두 번째 이유는 차원축소 알고리즘이 많은 학습 샘플을 요구한다는 점이다. SNS 환경에서 사용자들이 적극적으로 얼굴 사진에 어노테이션을 한다고 가정할 수 없기 때문에 초기 학습샘플이 충분하지 않다고 보는게 타당하다. 기존의 특징 추출 알고리즘은 이러한 상황에 적합하지 않다.

이러한 문제를 해결하기 위해 본 논문에서는 학습과정이 필요없고, 계산 비용이 매우 적은 희소 랜덤 사상(Sparse Random Projection, SRP)[8]을 사용하였다. 이 알고리즘은 주로 데이터 마이닝 분야에서 대용량 데이터를 다루기 위해서 사용되었다. D 차원의 데이터를 d 차원으로 축소하기 위해 다양한 최적화 방법을 사용해 부공간을 추정하는 기존 알고리즘과는 달리 SRP에서 사상 행렬 $R \in R^{d \times D}$ 은 아래 수식에 의해 생성된다.

$$R_{i,j} = \sqrt{s} \begin{cases} 1, \text{확률 } \frac{1}{2s} \text{ 비율로} \\ 0, \text{확률 } 1 - \frac{1}{s} \text{ 비율로} \\ -1, \text{확률 } \frac{1}{2s} \text{ 비율로} \end{cases} \quad (8)$$

식 (8)에서 s 는 희소성을 의미한다. $s=3$ 일 때 R 행렬의 2/3는 0이 된다. [9]에서 s 값으로 1과 3을 이용했고, [10]에서는 $1 \sim \sqrt{D}$ 의 값을 사용하였다.

i.i.d.(independent and identically-distributed) 조건으로부터 랜덤하게 생성되는 기본적인 RP에 비해 직교

화(Orthogonalization)과정이 없고, \sqrt{s} 를 곱하는 실수 곱셈은 마지막 과정에 계산될 수 있기 때문에 사상 행렬 R 의 생성과 사상시의 계산 복잡도가 낮다. 그림 2는 랜덤하게 생성된 RP와 SRP의 몇 가지 베이스스를 보여준다. PCA 베이스스는 얼굴과 비슷하지만 RP기반의 베이스스는 랜덤분포를 갖는다. s 값으로 \sqrt{D} 를 사용할 때 그림 2(c)에서처럼 약 97%가 0이다. 희소성 s 는 인식률에 영향을 끼친다. 그러나 가장 적합한 값을 추정할 수 없을 뿐 아니라 데이터셋에 의존적이기 때문에 본 논문에서는 계산량을 많이 줄일 수 있는 \sqrt{D} 를 사용하였다.

앞서 지적했듯이 SNS 환경에서는 학습 샘플수가 적고, 다양한 카메라에 의해 제한없는 환경에서 얼굴 영상이 취득되기 때문에 얼굴 분포의 변화가 매우 다양하다. 이 문제는 얼굴 인식분야의 가장 어려운 문제 중 하나이다.



(a) PCA 베이스스



(b) RP 베이스스



(c) SRP 베이스스($s=3$)



(d) SRP 베이스스($s=32$)

그림 2 부공간 방법에 의해 추정된 다양한 베이스스

단일 분류기의 성능만으로 이 문제를 해결할 수 없기 때문에 본 논문에서는 다수의 SRP와 다수의 분류기를 SUM 룰[11]에 의해 선형적으로 결합하는 방식을 사용해 이러한 상황을 개선하고자 한다. SRP의 경우 같은 학습 데이터를 사용해도 무한대의 부공간으로 확장될 수 있기 때문에 어렵지 않게 C 개의 사상 행렬 R_c 를 생성할 수 있다. C 개의 분류기가 결합될 때 새로운 데이터 x 의 클래스는 아래와 같이 결정된다.

$$class(x) = \operatorname{argmax}_i \sum_{c=1}^C \Theta_{t,c}^i \psi(R_c x), \quad i = 1, 2, \dots, k \quad (9)$$

3.4 증가적 서버 학습

제한된 증가적 분류기는 사진 공유기반의 소셜 네트워크 서비스에서 발생하는 매우 커다란 데이터를 효과적으로 다룰 수 있다. 그러나 실 환경에서 매우 많은 사용자가 자주 사진을 업데이트하는 경우 식 (5)는 시간 및 메모리 복잡도 관점에서 하나의 컴퓨터에서 동작하기 매우 버거울 수 있다. 이 문제를 해결하기 위해 본 논문에서는 HTTP와 XML을 기반으로 하는 REST 규약을 사용해 컴퓨터 네트워크상에서 제안된 알고리즘들이 서로 통신할 수 있도록 시스템을 확장하고자 한다.

[12]는 원격 함수 호출(RPC)을 통해 대규모의 데이터를 다수의 컴퓨터를 사용해 학습할 수 있는 시스템을 제안하였다. 본 논문에서는 RPC대신 개방형 웹 서비스를 위해 REST 규약을 기반으로 OpenAPI를 제안한다. RPC에 비해 성능이 느리기는 하지만 그림 3에서처럼 다양한 서비스와 손쉽게 접목이 가능한 매시업(Mash-up)을 고려했기 때문이다.

제안하는 분산형 분류기는 그림 3처럼 구성되어 있다. 정면 얼굴 검출과[13] 인식 알고리즘은 웹 스크립트 언어인 ASP(Active Server Page)에서 호출하기 위해 ATL COM+로 작성되었고, ASP 모듈은 COM+ 객체를 호출하고 데이터베이스 시스템에 접근하는 등의 OpenAPI를 구현하는데 사용되었다. 각각의 인식 모델은 HTTP를 이용해 요청을 다른 서버에 전송하고, 해당 서버는 파라미터를 해석해 기능을 수행한 후 그 결과를 XML 형태로 되돌려 준다. 그리고 [6]에서 언급된, 서버 사이에서 발생할 수 있는 사용자의 중복 어노테이션을 막기 위해서 제안된 증가적 알고리즘을 사용해 REST 기반으로 협업 어노테이션이 될 수 있도록 하였다.

3.5 복잡도 분석

본 논문에서 제안한 알고리즘은 식 (3)과 (5)에 의해서 결정된다. 표 2는 두 수식에 대한 복잡도를 제시한다. r 은 다항식의 오더, n 는 샘플의 수, d 는 특징벡터의 차원, l 은 전개된 다항식의 개수, k 는 클래스 수를 의미한다. 전체 알고리즘이 선형대수만 사용하기 때문에 복잡도는 정밀하게 분석될 수 있다.

표 2에서 알 수 있듯, 가장 큰 영향을 끼치는 요소는 다항식 항의 수 $l(l=1+r+d(2r-1))$ 이다. $o(n^3)$ 항목이 있지만 새로 추가되는 샘플의 수(n)가 l 에 비해 훨씬 작기 때문에 l 이 계산 비용에 가장 큰 영향을 끼친다. 클래스 수 k 는 선형으로만 계산 비용을 발생시키고 수식 (5)에서만 필요하기 때문에 제안된 알고리즘은 클래스 수에는 많은 영향을 받지 않는다. 그리고 결합되는 분류기의 수 C 는 선형적으로 전체 계산 복잡도를 증가시킨다. 따라서 전체 성능에 가장 큰 영향을 끼치는 요인은 d 와 C 이다. 이 두 변수에 의한 계산 복잡도는 실험 부분에서 제시될 것이다.

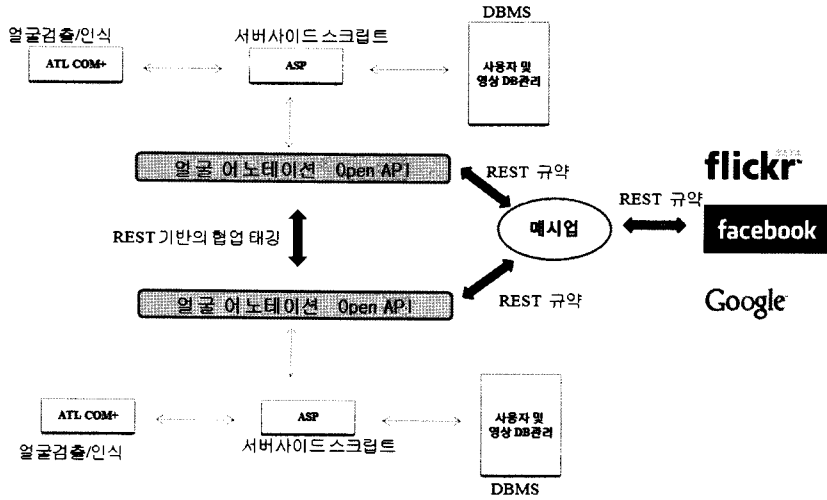


그림 3 REST 기반의 분산 분류기

표 2 복잡도 분석

수식	복잡도
수식(3)	$o(l^2n) + o(n^2l) + o(l^2n) + o(l^3) + o(n^3) + o(l^2)$
수식(5)	$o(lk) + o(l^2n) + o(lnk) + o(nk) + o(lnk)$

4. 실험

본 실험에서는 제안한 증가적 얼굴 어노테이션 알고리즘의 성능을 평가하기 위해 얼굴에 대한 다양한 변화가 있고, 샘플수가 많은 AR[14], EYALEB[15], PIE[16], YAHOO[17] 얼굴 데이터셋을 사용하였다(그림 4참고). 어노테이션 횟수에 영향을 미치는 인식률과 대규모 이용자의 신속한 처리에 영향을 미치는 계산효율성이 평가되었다.

통계적 성능 분석을 위해 가장 널리 사용되는 PCA 기반의 SVM 분류기와 비교하였다. 이 알고리즘은 최신 알고리즘은 아니지만 그동안 많은 논문에서 그 성능이 검증되었고, 온라인 학습 성능 비교를 위한 신뢰할 만한 분류기가 적기 때문에 베이스라인 분류기로 PCA-SVM을 선택하였다. 또한 본 논문의 목적이 얼굴 인식을 향상시키기 보다는 SNS환경에서 기존의 인식 성능을 유지하면서도 증가적 방식으로 얼굴 인식을 수행하는 효율적인 알고리즘을 제안하는 것이기 때문에 PCA-SVM 방법과 성능을 비교하는 것에 큰 무리가 없을 것으로 판단했다. 제안된 분류기는 SRP-RLS라 지칭하였다.

모든 실험에서 식 (1)에 있는 정규화 파라미터 b는 0.0001로 고정되었다. $\psi(\cdot)$ 에서 다항식 차수는 3차 이상에서 역행렬 문제가 발생해 2차만을 사용하였다. SVM의 경우 일괄방식을 적용하고 2차 다항식 커널과



(a) AR(AR, 126명×26장)



(b) Extended Yale B(EYALEB, 28명×64장)



(c) Yahoo News(YAHOO, 72명×20장)

그림 4 실험에 사용된 얼굴 데이터셋

RBF 커널이 사용되었다. SVM을 학습하기 위한 C 파라미터는 디폴트값을 사용하였고, RBF 커널의 파라미터로 1~20의 값을 사용하고 최대 인식률을 표시하였다. 입력 영상은 32×32크기로 매뉴얼하게 정규화시키고, R행렬에 사상시킨 후 0~1로 정규화시켜서 제안된 증가적 분류기와 SVM 분류기의 입력 값으로 사용하였다. 모든 실험은 10번 반복되어 평균 인식률과 CPU 사용시간이 측정되었다.

4.1 인식률 비교

일반적으로 얼굴 인식의 성능을 평가하기 위해 학습 샘플로 90%까지 사용해 인식률을 측정하지만 SNS 환경에서 학습 샘플수가 적을 수 있다고 가정했기 때문에,

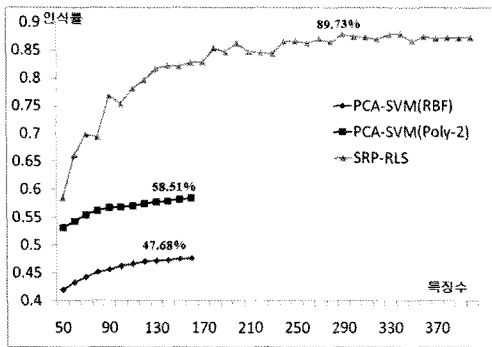
본 논문에서는 10%~30%까지의 학습 샘플을 사용해서 나머지 70%를 추정하는 방식을 사용하였다. 증가적 샘플과 증가적 다중 클래스 학습 알고리즘의 인식률을 평가하기 위해 인식할 대상의 50%, 한사람당 10%의 샘플을 초기 학습 데이터로 사용하였고, 모든 사용자에 대해서 30%의 샘플이 사용될 때까지 학습시켰다.

그림 5(a)는 AR 데이터셋에 대한 인식률 결과이다. 제안된 방법이 PCA-SVM보다 31.22% 높게 나타났다. 이 데이터셋은 선글라스와 스카프에 의한 가려짐이 있는 등 얼굴 영상의 변화량이 크기 때문에 30% 학습 데이터셋으로 나머지를 인식해야 하는 경우 데이터셋에 매우 의존적인 PCA-SVM의 인식률은 매우 저조함을 알 수 있다. 대다수의 유저가 활동적으로 어노테이션을 하지 않기 때문에 적은수의 샘플을 사용할 경우 얼굴 인식의 성능은 매우 중요하다. 또한 결과에서 볼 수 있듯, PCA-SVM의 경우 얼굴의 특징이 많이 사용되어도 인식 성능이 크게 향상되지 않고 포화된다. 반면 제안된 방법의 경우 특징수가 작을 때 성능이 저조하지만 특징수를 증가시키면 인식률 또한 증가하는 것을 볼 수 있다. 그림 5(b)는 EYALEB 데이터셋에서의 실험결과이다. AR 데이터셋과 마찬가지로 제안된 방법의 인식률이

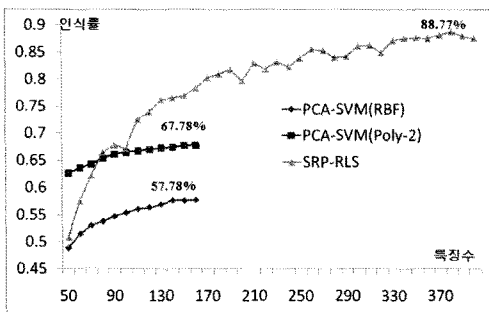
20.99% 높게 나타났다.

두 번째 실험으로 다수의 분류기를 결합한 분류기의 인식률을 평가하였다. 인식률에 영향을 주는 특징의 수와 결합되는 분류기의 수를 축으로 해서 인식률 결과를 그림 6에 나타내었다. 결과에서 볼 수 있듯 결합된 분류기의 성능이 단일 분류기보다 인식률이 높아짐을 알 수 있다. 특징 수와 결합되는 분류기 수의 증가는 인식률의 증가를 이끌지만 이는 시스템의 계산 복잡도를 높이기 때문에 적정수준에서 선택해야 한다. 그림 6(b)는 EYLABE 데이터셋에 대한 분류기 결합에 대한 인식률 결과이다.

세 번째 실험으로 θ_i 대신 θ_i^z 를 사용한 알고리즘의 인식률을 평가하였다. 특정 사용자와 관련된 적은수의 사용자만을 사용하기 때문에 최종 분류할 대상의 수를 줄일 수 있어 인식 성능이 향상된다. 표 3은 인식률을 비교한 결과이다. AR과 EYALEB 데이터셋의 경우 클래스 수에 상관없이 인식률이 매우 높게 나왔다. 하지만 YAHOO의 경우 클래스 수가 조금 증가해도 인식결과가 급격히 떨어진다. 이는 YAHOO 데이터셋의 특성 때문이다. 다른 데이터셋하고는 달리 YAHOO는 2년간 웹에 올려진 다양한 뉴스 이미지부터 추출된 영상이기 때문에 컨트롤된 환경에서 제작된 AR과 EYALEB 데이

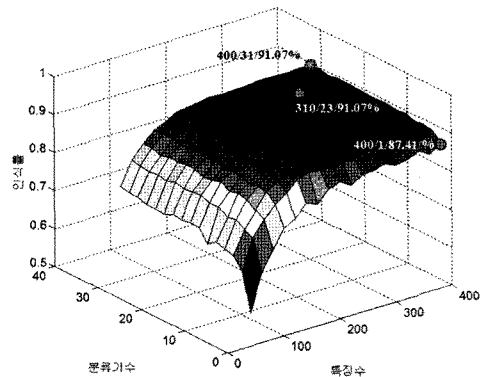


(a) AR

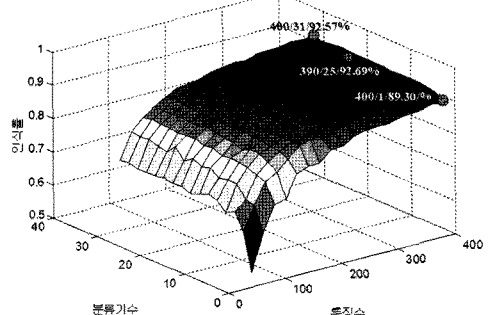


(b) Extended Yale B

그림 5 싱글 분류기에 대한 인식률 결과 (학습 데이터 30%, 테스트 데이터 70%가 사용됨)



(a) AR



(b) Extended Yale B

그림 6 결합 분류기에 대한 인식률 결과

표 3 Θ_i 를 사용한 인식률 비교

	AR(30%)	EYALEB (30%)	YAHOO		
			30%	50%	70%
사람수	인식률	인식률	인식률	인식률	인식률
3	98.25%	97.04%	73.05%	83.33%	96.67%
5	97.89%	91.56%	71.32%	80.00%	94.44%
7	95.49	91.75%	62.24%	67.14%	73.81%
11	92.34%	92.32%	40.26%	46.35%	56.06%

터셋 보다 실제 얼굴 분포를 더 잘 나타내기 때문이다. 따라서 이러한 환경에서는 인식할 클래스 수가 작아야 한다. YAHOO 데이터셋의 경우 학습 비율을 30%,50% 그리고 70%를 사용하였다. 이 실험에서 알 수 있듯 다양한 카메라를 사용해 제약없이 취득된 얼굴 영상은 매우 분류하기 어려움을 알 수 있다. 이 문제를 해결하기 위해 3가지 방법을 생각해 볼 수 있다. 첫 번째는 표 3에서처럼 학습샘플을 많이 주는 방법과 분류할 클래스 수의 수를 제한하는 방법이다. 마지막 방법은 여러장의 사진 정보들에 있는 다양한 문맥정보를 활용해 얼굴 인식률을 높이는 방법이다[17,18]. 본 논문에서는 앞의 두 가지 방법만을 사용하였다.

4.2 계산 복잡도 비교

보안을 위한 얼굴 인식 시스템과는 달리 SNS 환경에서 얼굴 인식이 치명적인 문제가 아니다. 인식률에 대한 약간의 성능차이는 이용자입장에서 크게 느껴지지 않는다. 보다 중요한 문제는 대규모 사용자를 위한 계산 효율성이다. 학습과 테스트에 대한 계산효율성을 평가하기 위해 많은 샘플을 제공하는 PIE 데이터셋을 사용하였다.

PCA-SVM의 경우 100개의 부공간이 사용되었고, 제안된 SRP-RLS는 400개의 특징을 사용하였다. 그림 5에서 나타나듯, PCA-SVM의 경우 많은 축을 사용하더라도 인식률이 높아지지 않지만, 제안된 SRP-RLS는 많은 축을 사용할 때 인식률이 높기 때문에 서로 다른 수의 축을 사용해 학습 시간을 비교하는 것이 타당하다.

그림 7에서 볼 수 있듯 제안된 학습 방식은 추가적인 샘플이 있을 때마다 상수시간을 기록하는 반면, SVM은 이전 데이터와 새로운 데이터를 이용해 다시 학습해야 하기 때문에 학습시간이 매우 오래 걸리고 크게 증가함을 알 수 있다.

다수의 분류기를 사용해 선형으로 결합하는 경우 그림 7에서 볼 수 있듯, 계산 비용도 선형적으로 증가한다. 즉 5개의 분류기를 결합한 SRP-RLS(x5)는 SRP-RLS(x1)보다 5배 더 많은 계산 비용을 요구한다. 모든 방법이 샘플하나의 레이블을 결정하는데 10ms이하의 매우 낮은 계산 비용이 발생하는 것으로 나와 결과는 포함시키지 않았다.

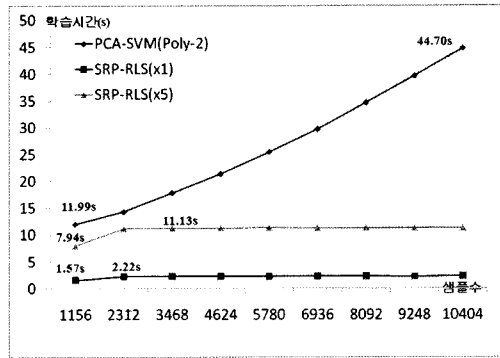


그림 7 학습 시간 비교

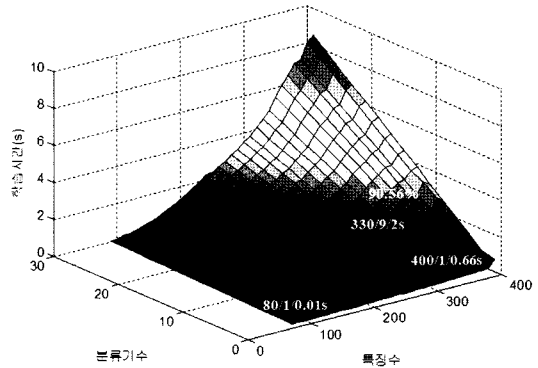


그림 8 특징 및 분류기 수에 따른 학습 시간 결과

다섯 번째 실험으로 전체 계산 비용에 가장 큰 영향을 미치는 특징수 d 와 결합되는 분류기 개수 C 에 따른 계산 복잡도를 비교하였다. 업데이트되는 샘플의 수는 50으로 고정시키고 나머지 변수는 이전 실험과 동일하다. 그림 8은 이에 대한 실험 결과이다. 그림 6의 실험에서 400개의 특징을 사용하고, 31개의 분류기가 결합될 때 가장 성능이 높게 나오지만 학습 시간이 8초 이상이 걸리기 때문에 이는 실시간 서비스를 위해 적절한 선택이 아니다. 특정 서비스가 2초 이내에 학습이 완료되려면 그림 6에서의 인식률을 고려해 가장 최적의 조합을 선택해야 한다. 계산 비용이 2초 이내여야 할 때 330개의 특징을 사용하고, 9개의 분류기를 결합하는 지점에서 인식률이 가장 높았다(90.5%). 일반적으로 특징의 수를 가변적으로 변경시킬 수 없지만 결합되는 분류기의 수는 동적으로 조절이 가능하다. 따라서 사용자가 적을 때 결합되는 분류기의 수를 많이 사용하고, 사용자가 많을 때 분류기의 수를 적게 업데이트하는 경험에 의한 전략을 사용하면, 인식성능과 계산 비용을 적절히 조절할 수 있게 된다.

여섯 번째 실험으로 Θ_i 를 사용했을 때의 계산 비용

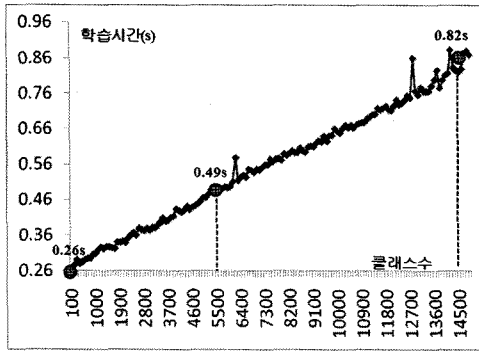


그림 9 Θ_1 를 이용한 다중 클래스 분류 성능 결과

을 평가하기 위해 분류할 클래스의 수를 100개~50000개 까지 늘려가며 성능을 측정하였다. 복잡도 분석에서 지적했듯이 그림 9에서 보면, 클래스의 수가 커져도 계산 비용은 그리 크지 않다. 그러나 클래스의 수가 15000일 때 메모리 할당 오류가 발생해 더 이상 학습이 불가능했다. 사용된 컴퓨터의 메모리는 4G였다. 이 수치는 일반 보안 목적에는 충분히 큰 수치지만, 소셜 네트워크 서비스를 운영하기에는 매우 적은 수치이다. Θ_1 에서는 업데이트해야하는 최소한의 공간만을 사용하기 때문에 이러한 문제가 발생하지 않는다.

마지막 실험으로 Microsoft WAS 웹 스트레스 툴을 사용해 분산 분류기의 성능을 평가하였다. 표 4는 새로 추가되는 학습 샘플의 크기에 따라서 초당 HTTP 요청 수를 보여주고 있다. 400개의 특징을 사용하고, 1개의 분류기만을 사용하여 학습 시간을 측정하였다. 결과표에서 알 수 있듯, 10개의 영상이 추가될 때 1초에 대략 7개의 요청을 처리할 수 있는 것으로 평가되었다.

표 4 웹 스트레스 테스트 시뮬레이션

재학습 샘플 수	10	50	80	110	150
HTTP호출	6.71	5.43	4.61	4.01	3.35

5. 결론

본 논문에서는 소셜 네트워크 서비스 환경에서 효율적으로 얼굴 영상에 어노테이션할 수 있는 알고리즘을 제안하였다. 보안을 목적으로 하는 PC기반의 환경과는 달리 소셜 네트워크 서비스와 결합된 얼굴 인식 기술에서는 증가적 샘플, 증가적 다중 클래스, 그리고 증가적 서버 문제가 발생한다. 이러한 환경에서 시간에 따라 증가하는 거대한 얼굴 영상 데이터베이스를 효과적으로 분류하기 위해 랜덤 사상, 비선형 회귀 그리고 REST 규약 사용해 새로운 증가적 얼굴 분류기를 제안하였다. 다양

한 데이터셋을 이용한 실험 결과로부터 제안한 방법은 작은 수의 학습 샘플에도 불구하고 높은 인식률을 가지고 있고, 매우 많은 사용자를 빠르게 처리할 수 있었다. 소셜 네트워크 서비스와 얼굴 인식 기술의 결합은 대중성과 여러 허용성이라는 특성 때문에 기존의 보안 분야에 적용된 얼굴 인식 기술의 단점을 벗어난 새로운 응용 분야를 창출할 수 있을 것으로 기대된다. 향후 사진 영상에 있는 다양한 문맥정보를 활용해 얼굴 인식률을 끌어올리는 연구를 진행하고자 한다.

참고 문헌

- [1] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, "Face Recognition: A Literature Survey," *ACM Computing Surveys*, pp.399-458, 2003.
- [2] Z. Stone, T. Zickler, and T. Darrell, "Autotagging Facebook: Social network context improves photo annotation," *CVPR Workshops*, pp.1-8, 2008.
- [3] B. C. Becker, and E. G. Ortiz, "Evaluation of Face Recognition Techniques for Application to Facebook," *Automatic Face and Gesture Recognition*, 2008.
- [4] P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe, "FRVT 2006 and ICE 2006 Large-Scale Results," *NISTIR 7408*, 2007.
- [5] 최권택, 변혜란, "소셜 네트워크 환경에서의 얼굴 어노테이션", *한국정보과학회 제35회 추계 학술 발표 대회*, 2008.
- [6] K. Choi, K. Toh, and H. Byun, "A Collaborative Face Recognition Framework on a Social Network Platform," *Automatic Face and Gesture Recognition*, 2008.
- [7] K.-A. Toh, Q.-L. Tran, and D. Srinivasan, "Benchmarking a reduced multivariate polynomial pattern classifier," *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol.26, no.6, pp.740-755, 2004.
- [8] Q. L. Tran, K.-A. Toh, and D. Srinivasan, "Adaptation to changes in multimodal biometric authentication," *IEEE Conference on Cybernetics and Intelligent Systems*, 2004.
- [9] D. Achlioptas, "Database-friendly random projections," *ACM Symposium on the Principles of Database Systems*, pp.274-281, 2001.
- [10] P. Li, T. Hastie, and K. W. Church, "Very sparse random projections," *KDD*, pp.287-296, 2006.
- [11] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas, "On Combining Classifiers," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.20, no.3, pp.226-239, 1998.
- [12] T.N. Do, and F. Poulet, "Classifying one billion data with a new distributed SVM. algorithm," *IEEE International Conference on Computer Science, Research, Innovation and Vision for the*

- Future, pp.59-66, 2006.
- [13] P. Viola, and M. Jones, "Rapid object detection using a boosted cascade of simple features," *IEEE Conf. on Computer Vision and Pattern Recognition*, 2001.
- [14] A. Martinez and R. Benavente. "The AR Face Database," 1998.
- [15] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol.23, no.6, pp.643 - 660, 2001.
- [16] T. Sim, S. Baker, and M. Bsat. "The CMU Pose, Illumination, and Expression Database," *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol.25, no.12, pp. 1615-1618, 2003.
- [17] T. L. Berg, A. C. Berg, J. Edwards, and D. A. Forsyth, "Who's in the picture?," *Advances in Neural Information Processing Systems 17*, pp. 137-144, 2005.
- [18] M. Naaman, R. B. Yeh, and H. Garcia-Molina, "Leveraging context to resolve identity in photo albums," *JCDL*, 2005.



최 권 택

2001년 한림대학교 컴퓨터공학과 졸업
 2006년 연세대학교 컴퓨터과학과 석사 졸업. 2006년~현재 연세대학교 컴퓨터과학과 박사과정. 관심분야는 패턴인식, 얼굴 인식, 머신비전



변 혜 란

1980년 연세대학교 수학과 졸업(이학사)
 1983년 연세대학교 대학원 수학과 졸업(이학석사). 1987년 University of Illinois, Computer Science(M.S.). 1993년 Purdue University, Computer Science(Ph.D.). 1994년~1995년 한림대학교 정보공학과 조교수. 1995년~1998년 연세대학교 컴퓨터과학과 조교수. 1998년~2003년 연세대학교 컴퓨터과학과 부교수. 2003년~현재 연세대학교 컴퓨터과학과 교수. 관심분야는 인공지능, 영상인식, 영상처리