

논문 2007-44CI-3-14

사전 클러스터링을 이용한 LDA-확장법들의 최적화

(On Optimizing LDA-extensions Using a Pre-Clustering)

김 상 운*, 구 범 용**, 최 우 영**

(Sang-Woon Kim, Byum-Yong Koo, and Woo Young Choi)

요 약

얼굴인식 등과 같은 고차원 패턴인식에서 학습패턴의 수가 패턴 차원에 비해 매우 적을 경우 희소성 문제(the Small Sample Size problem)가 발생한다. 최근 이 문제를 해결하기 위하여 LDA, PCA+LDA, Direct-LDA 등을 비롯한 다양한 LDA-확장 법이 제안되었다. 본 논문에서는 LDA-확장 법으로 차원을 축소하기 전에 학습 패턴을 사전 클러스터링 하여 서브 클래스 수를 증가시키는 방법으로 LDA-확장에 기반을 둔 식별기의 성능을 향상시키는 방법을 제안한다. LDA (또는 Direct-LDA)에서 축소된 특징공간의 차원은 학습패턴의 클래스 수로 제한되기 때문에 LDA의 식별 성능을 향상시킬 수 있도록 학습패턴을 사전에 클러스터링하여 서브 클래스의 수를 증가시키는 방법이다. 즉, 학습패턴의 특성 공간(the eigen space)은 레인지 공간(the range space)과 널 공간(the null space)으로 구성되며, 레인지 공간의 차원은 클래스 수의 증가에 따라 증가한다. 따라서 변환 행렬을 구성할 때 클래스의 수를 늘려 널 공간을 최소화하게 되면 이 공간에 기인한 정보의 손실을 최소화 할 수 있다. 제안 방법을 X-OR 형태의 인공데이터와 AT&T와 Yale 벤치마크 얼굴영상 데이터베이스를 대상으로 실험한 결과 본 방법의 효용성을 확인하였다.

Abstract

For high-dimensional pattern recognition, such as face classification, the small number of training samples leads to the Small Sample Size problem when the number of pattern samples is smaller than the number of dimensionality. Recently, various LDA-extensions have been developed, including LDA, PCA+LDA, and Direct-LDA, to address the problem. This paper proposes a method of improving the classification efficiency by increasing the number of (sub)-classes through pre-clustering a training set prior to the execution of Direct-LDA. In LDA (or Direct-LDA), since the number of classes of the training set puts a limit to the dimensionality to be reduced, it is increased to the number of sub-classes that is obtained through clustering so that the classification performance of LDA-extensions can be improved. In other words, the eigen space of the training set consists of the range space and the null space, and the dimensionality of the range space increases as the number of classes increases. Therefore, when constructing the transformation matrix, through minimizing the null space, the loss of discriminative information resulted from this space can be minimized. Experimental results for the artificial data of X-OR samples as well as the bench mark face databases of AT&T and Yale demonstrate that the classification efficiency of the proposed method could be improved.

Keywords : 패턴인식, 얼굴인식, PCA, LDA, Direct-LDA

I. 서 론

통계적 패턴인식^[1,2]에서의 식별은 패턴을 특징공간에 사상한 다음, 이 공간에서 테스트 패턴의 클래스를 결정하는 방법이다. 그러나 얼굴인식에서 영상화소를 특징으로 사용하면 특징들 사이에 정보가 중복되고 처리 시간이 증가하는 차원의 유허로움(the curse of dimensionality)^[1-4]이 발생한다. 따라서 패턴인식에서

* 정회원, 명지대학교 컴퓨터공학과
(Dept of Computer Engineering, Myongji University),

** 학생회원, *** 정회원, 명지대학교 전자공학과
(Dept of Electronic Engineering, Myongji University),

※ 이 논문은 2006년 정부(과학기술부)의 재원으로
한국과학재단의 지원을 받아 수행된 연구임
(F01-2006-000-10008-0).

접수일자: 2006년11월14일, 수정완료일: 2007년5월3일

패턴의 수에 비해 특징이 고차원일 경우에 식별 성능을 향상시키기 위해서는 패턴 특징을 축소시키는 처리가 필요하다. 일반적으로 샘플 수와 특징 수의 관계를 정의하는 것은 매우 어려운 일이지만, 원하는 식별 성능을 구현하기 위해서는 클래스내의 샘플 수(n)와 특징(d) 사이에는 $n/d > 10$ 의 관계를 유지하도록 권고하고 있다.^[11,5]

식별 성능을 향상시키기 위한 특징 공간의 축소 변환 방법으로는 PCA(Principal Component Analysis)^[2-4,16]와 LDA(Linear Discriminant Analysis)^[2-4,17]가 널리 알려져 있다. PCA는 전체 패턴의 분포를 가장 잘 근사시킬 수 있는 변환으로 패턴표현이나 압축에 적합한 반면, LDA는 각 클래스별 패턴의 분포에 대한 분리도를 최대로 하는 변환으로 식별에 적합한 차원 축소법이다. 그런데 얼굴 인식이나 멀티미디어 검색 등의 응용에서 패턴차원은 고차원인 반면, 학습을 위한 샘플 수는 충분치 못한 경우가 많다. 이 때, 패턴 샘플의 수가 차원에 비하여 적을 경우 변동행렬(scatter matrix)이 특이행렬이 되기 때문에 LDA의 변별력이 떨어지는 희소성 문제(undersampled problem)^[8,9]가 발생한다¹⁾. 최근 이러한 희소성 문제를 해결하기 위해 PCA+LDA^[6], Direct-LDA^[10,11]등과 같은 LDA-확장 법을 비롯하여 LDA/GSVD(Generalized Singular Value Decomposition)법^[8], DCV(Discriminative Common Vector)법^[12], RDA(Relational Discriminant Analysis)법^[13,14]등의 선형 방법과 커널함수(kernel function)를 이용하는 비선형 방법^[15,16] 등 다양한 방법이 제안되었다.

PCA+LDA는 2-단계 LDA라고도 하며, 식별목적에 적합한 차원 축소를 위해 LDA를 수행하기 전에 먼저 PCA를 수행하는 방법이다^[5]. 이때 PCA에서 축소되는 차원은 LDA 단계에 사용하는 클래스-내 변동행렬을 비-특이행렬이 되도록 제한한다. 그러나 PCA+LDA에서는 PCA단계에서 클래스 구조를 무시하는 전체 패턴의 변동행렬만을 이용하기 때문에 식별에 유용한 정보가 유실될 수 있다.

한편, Direct-LDA는 기존의 LDA에서와는 달리 클래스-내 변동행렬(within-class scatter matrix, 이하 간단히 S_w 로 표기함)과 클래스-간 변동행렬(between-class scatter matrix, 이하 S_b 로 표기함)을 동시에 대각화(simultaneous diagonalization)^[2,8]시키는 변환 행렬을

찾는 방법이다. Direct-LDA의 주요 내용은 중요한 식별정보를 포함하는 S_w 의 널 공간(the null space)를 최대한 유지하면서 비교적 중요하지 않은 식별정보를 포함하는 S_b 의 널 공간을 제거하는 것이다²⁾. 이는 먼저 S_b 를 대각화한 후 S_w 를 대각화함으로써 이루어진다.

그러나 S_b 의 널 공간 역시 식별에 중요한 정보를 포함할 수 있고, S_w 를 S_b 의 레인지 공간으로 사상할 때 S_w 의 널 공간에 해당하는 식별정보를 잃어버릴 가능성이 있다. 또한 Direct-LDA에서 기본적으로 사용하고 있는 Fisher의 판별기준은 데이터 분포가 단일 모형을 대상으로 한 것으로, 다수의 가우시언 분포가 중첩되어 있는 모형인 경우에는 적합하지 않다^[17].

이와 같은 문제를 해결하기 위하여, 우선 널 공간을 최소화하여 정보의 유실을 막는 방법을 생각할 수 있다. 그런데 레인지공간의 차원 m 과 샘플 패턴의 클래스 수 c 와는 $m \leq c-1$ 의 관계가 성립하기 때문에 m 을 증가시키기 위해서는(즉, 널 공간을 줄이기 위해서는) 클래스 수 c 를 크게 하여야 한다. 그런데 가우시언 분포가 비선형으로 중첩되어 있는 다중 모형(multimodal)인 경우, 각 클래스에 존재하는 다중 클러스터를 하나의 서브 클래스(sub-class)로 간주하여³⁾(클래스 수를 증가시켜) 레인지 공간의 차원 수를 늘리면 널 공간에 기인한 정보 유실을 방지할 수 있다.

이러한 점에 착안하여, 문헌 [2]에서는 두 클래스의 중심이 겹치면서 각 클래스가 대각선 방향으로 분포하는 다중 모형의 경우 두 클래스를 4개의 서브 클래스로 취급하고, 두 분포가 서로 물려 있는 “바나나 분포”의 경우에는 각 클래스를 3개의 클러스터로 나누어 6개의 서브 클래스로 취급하는 방안을 제안하였다(문헌 [2]의 그림 10.5 참조). 최근 문헌 [17]에서는 비선형 다중 모형 분포를 갖는 클래스를 서브 클러스터로 분할하기 위하여 Leave-one-out-test와 Stability와 같은 판별기준을 이용하는 방안을 제안하였다. 이 방법은 효과적으로 다중 모형 분포를 갖는 클래스의 비선형 구조를 찾아낼 수는 있으나, 처리 시간이 길다는 문제가 있다.

본 논문에서는 기존의 LDA(또는 LDA-확장 법)에

1) 희소성 문제를 소규모 샘플 문제(small sample size problem)라고도 혼용한다.

2) 고유치 문제에서 0(또는 0에 가까운) 고유치에 해당하는 고유벡터들의 공간을 널 공간(the null space)이라 하고, 그 나머지 고유벡터의 공간을 레인지 공간(the range space)이라 한다.

3) 실제의 클래스와 구별하기 위하여 클러스터를 기준으로 설정한 클래스를 서브 클래스(sub-class)라 한다.

클러스터링 기법을 적용하여 클래스 수를 늘림으로써, 추출할 수 있는 특징의 수를 늘릴 수 있는 방법을 제안한다. 이 때, 클러스터링 할 클래스를 선정하기 위하여 학습 데이터를 대상으로 클래스 별로 군집 내부거리 (intra-distance)^[4]를 계산하여 클러스터링 할 클래스의 우선순위를 결정한다. 그리고 높은 순위의 클래스부터 클러스터링 하여 해당 클래스를 구성하는 서브 클래스를 구하여 식별 성능을 향상시키는 방법을 제안한다.

제안한 방법은 우선순위가 높은 몇몇 클래스만을 대상으로 클러스터링을 수행하는 방법으로 간단하게 클래스의 비선형 구조를 찾아 인식률을 향상시킬 수 있는 장점이 있다. 본 논문의 구성은 다음과 같다. 제 II장에서는 기존의 LDA-확장 법을 간단히 소개한다. 제 III장에서는 제안 방법을 설명하고, 기존 방법과의 차이점을 비교한다. 제 IV장에서는 인공데이터 및 실제의 벤취마크 데이터를 대상으로 한 기존 방법과 제안 방법의 실험 결과를 고찰하고, 제 V장에서 결론을 맺는다.

II. 기존의 방법들

1. LDA

LDA는 패턴을 부분 공간으로 사상하였을 때 사상된 패턴들의 클래스별 분리도를 최대로 하는 축들로 구성된 공간을 탐색하는 것이다. 이를 위해 다음과 같이 클래스-내 분산과 클래스-간 분산을 정의한다.

$$S_w = \sum_{i=1}^c \sum_{x \in C_i} (x - m_i)(x - m_i)^T \quad (1)$$

$$S_b = \sum_{i=1}^c n_i (m_i - m)(m_i - m)^T \quad (2)$$

여기서 c : 클래스의 수, C_i : i 번째 클래스에 속하는 데이터 셋, n_i : 각각 i 번째 클래스의 데이터 개수, m_i : i 번째 클래스의 평균, m : 전체 평균이다.

식 (1), (2)로부터 Fisher의 판별함수 J 는

$$J(W) = \frac{W^T S_b W}{W^T S_w W} \quad (3)$$

로 정의하며, J 를 최대로 하는 W 는 고유치 문제

$$S_b w_i = \lambda_i S_w w_i \quad (4)$$

의 해로, p 번째까지의 고유벡터들을 나열하여

$$W = [w_1, w_2, \dots, w_p] \quad (5)$$

로 구할 수 있다. 여기서 p 는 부분 공간의 차원이다. 따라서 변환행렬 W 를 이용하여 학습패턴 $\{x_i\}_{i=1}^n$ 의 차원을 축소시키는 변환 식은 다음과 같다.

$$y_i = W_{lda}^T x_i \quad (6)$$

2. PCA+LDA

PCA는 부분공간으로 사상된 패턴을 원 공간으로 복원시켰을 때, 재구성 오차가 최소가 되는 공간을 구하는 방법으로, 전체 패턴의 분산 행렬은 다음과 같다.

$$S_t = \sum_{i=1}^n (x - m)(x - m)^T \quad (7)$$

PCA+LDA에서 PCA는 학습패턴을 S_w 가 특이 행렬이 되지 않는 부분공간으로 사상시킨 다음, LDA는 이 공간에서 클래스별 분리도를 최대화 시키는 부분공간을 다시 탐색한다. PCA+LDA에 대한 변환 행렬은

$$W_{pca+lda} = W_{lda} W_{pca} \quad (8)$$

이고, 변환행렬을 이용한 변환 식은 다음과 같다.

$$y_i = W_{pca+lda}^T x_i \quad (9)$$

3. Direct-LDA

Direct-LDA는 식 (3)의 Fisher 판별함수에서 사용되는 두 개의 행렬 S_w 와 S_b 를 동시에 대각화시키는 변환 행렬을 찾는 것이다. 이때 S_w 와 S_b 가 특이 행렬일 때 S_w 의 널 공간이 가장 중요한 식별 정보를 가진다고 판단한다^[10]. 따라서 S_w 의 널 공간을 보존하기 위해 먼저 S_b 를 대각화하여 S_b 의 널 공간을 제거한 후 S_w 를 대각화하는 순서를 따른다. 유도 과정은 다음과 같다.

먼저 S_b 를 대각화하기 위해서

$$V^T S_b V = \Lambda \quad (10)$$

과 같이 S_b 로부터 고유치 Λ 와 고유벡터 V 를 구한다. 여기서 $V^T V = I$ 이고, Λ 는 대각요소를 내림차순으로 정렬한 대각행렬이다. 이 때, S_b 가 특이 행렬일 경우

$$Y^T S_b Y = D_b > 0 \quad (11)$$

과 같이 S_b 의 고유벡터 V 들 중에서 0이나 0에 가까운 고유치에 해당하는 고유벡터들을 무시하여 새로운 부분공간 Y 를 구성한다. 여기서 D_b 는 V 에서 0이 아닌 고유치만을 갖는 $p \times p$ 의 부분 행렬(sub-matrix)이고, Y 는 D_b 에 대응하는 고유벡터로서 $d \times p$ 의 행렬이다. 단, $p = c - 1$, d : 특징 공간의 차원, c : 클래스의 수이다.

다음, S_b 를 백색화(whitening)^[2]하기 위해서 우측 항이 단위행렬 I 가 되도록

$$(YD^{-1/2})^T S_b (YD^{-1/2}) = I \quad (12)$$

로 하면 $Z^T S_b Z = I$ 가 되는 백색화 변환행렬 $Z = YD_b^{-1/2}$ 를 구할 수 있다.

다시 Z 를 이용하여 S_w 를 S_b 의 부분공간 Y 로 사상한 후(사상된 S_w 를 S_w' 라고 함), $S_w' (= Z^T S_w Z)$ 을

$$U^T S_w' U = D_w \quad (13)$$

과 같이 대각화한다. 여기서 D_w 는 S_w' 의 고유치 대각행렬이고, U 는 S_w' 의 고유벡터로 $U^T U = I$ 가 성립한다. 그런데 $S_w' = Z^T S_w Z$ 이므로

$$(ZU)^T S_w ZU = D_w \quad (14)$$

가 되고, 다시

$$A^T S_w A = D_w, (A = ZU) \quad (15)$$

가 되어 S_w 와 S_b 를 동시에 대각화시키는 결합된 형태의 행렬 $A (= ZU)$ 를 구할 수 있다.

끝으로, LDA단계에서 Fisher 판별함수의 분모 항을 먼저 대각화시키기 위해서 식 (15)를

$$(AD_w^{-1/2})^T S_w AD_w^{-1/2} = I \quad (16)$$

로 하면 $W_{dda}^T S_b W_{dda} = I$ ($W_{dda} = AD_w^{-1/2}$)가 된다. 따라서 최종 변환행렬 W_{dda} 는

$$W_{dda} = ZUD_w^{-1/2} \quad (17)$$

가 되고, 차원이 축소된 벡터는 다음과 같다.

$$y_i = W_{dda}^T x_i \quad (18)$$

III. 제안 방법

1. 사전 클러스터링

패턴 벡터들의 분포형태가 가우시언 분포가 비선형적으로 결합된 다중 모형 분포(multimodal)일 때 Fisher 판별함수는 효과적으로 작동하지 못한다. 이러한 문제를 극복하기 위하여 본 연구에서는 각 클래스를 구성하는 클러스터들을 찾아서 서브 클래스를 구성하여 S_b 를 계산함으로써 식별 성능의 저하를 방지한다.

서브 클래스의 수는 학습패턴의 모든 클래스에 대한 클러스터링 결과를 참조하여 결정하는 방법, 클래스의 군집 내부거리 등을 참조하여 일부 클래스만을 클러스터링한 결과를 참조하여 결정하는 방법, 학습패턴 전체를 대상으로 Leave-One-Out-Test나 Stability 등의 판별함수^[4]을 이용하는 방법 등이 있다. 본 논문에서는 군집 내부거리 등을 참조하여 일부 클래스만을 클러스터링하는 방법을 이용한다.

학습데이터의 클래스 간에 비선형적 분포가 될 수 있는 가능성은 한 클래스 내에서 패턴 벡터들의 분산정보를 포함하는 군집 내부거리로 측정할 수 있다. 한 클래스를 구성하는 패턴 벡터들의 군집 내부거리는

$$D_{intra}^2 = \frac{1}{m(m-1)} \sum_{j=1}^m \sum_{i=1, i \neq j}^m \sum_{k=1}^d (x_{jk} - x_{ik})^2 \quad (19)$$

로 정의할 수 있다^[4]. 여기서 m : 클래스를 구성하는 샘플의 수, d : 특징 공간의 차원이다.

클래스 별로 패턴벡터의 분산을 고려한 군집 내부거리를 구한 후, 군집 내부거리로 클래스들을 내림차순으로 정렬한다. 그리고 정렬된 클래스 중에서 상위 N 개의 클래스만을 대상으로 클러스터링하여 찾은 클러스터를 서브 클래스로 생성한다. 본 논문의 실험에서는 클러스터링을 위해 k -means 알고리즘을 이용한다.

2. 제안 알고리즘

제안 방법은 클러스터링단계, 특징추출단계, 식별단계로 나눌 수 있다. 각 단계별 내용은 다음과 같다.

단계 1(사전 클러스터링단계): 내부 군집거리가 큰 순서대로 클래스를 나열한 다음, 가장 큰 것부터 N 개의 클래스를 클러스터링 하여 클래스별로 2개의 클러스

터를 구한다. 생성된 클러스터들을 하나의 개별적인 서브 클래스로 보고, 클러스터링하지 않은 클래스와 함께 학습패턴집합의 클래스로 이용한다.

단계 2(특징추출단계): 새롭게 할당된 클래스(서브 클래스) 정보를 가진 학습패턴집합을 대상으로 기존의 차원 축소 알고리즘을 적용한다. Direct-LDA의 경우 먼저 $\{x_i\}_{i=1}^n$ 로부터 S_w, S_b 를 계산한 다음, S_b 을 이용하여 변환행렬 Z 를 구성한다. 그리고 $Z^T S_w Z$ 로부터 변환행렬 W_{dda} 를 구성하여 차원 축소된 $\{y_i\}_{i=1}^n$ 을 생성한다.

단계 3(식별단계): 새롭게 생성된 특징공간상에 학습 패턴을 투영하여 표준패턴을 추출한 다음, 평가패턴의 클래스를 결정하는 식별처리를 수행한다. 이 때 NN 규칙을 포함하여 다양한 식별 기를 학습할 수 있다.

3. 제안 방법과 기존 방법과의 차이점 고찰

(1) 패턴벡터들의 분포형태가 다중 모형일 때 Fisher 판별함수는 식별에 적합하지 않다. 사전 클러스터링을 통하여 각 클래스를 2개(본 실험에서는 문제의 단순화를 위해 2개의 클러스터만 고려함)의 클러스터로 나누어 차원을 축소하여도 클래스의 구조는 유지된다.

(2) 새로운 서브 클래스는 Direct-LDA의 학습에서 S_w 와 S_b 를 계산할 때 영향을 미친다. 기존의 Direct-LDA에서 최대의 $rank(S_b)$ 는 $c-1$ 인 반면 제안 방법에서는 사전 클러스터링을 통하여 서브 클래스 c' 를 생성할 때 최대의 $rank(S_b)$ 는 $c'-1$ ($c' > c$)로 증가하여 S_b 의 레인지 공간의 차원이 증가한다. 이는 S_b 의 레인지 공간을 $d-c'+1$ 로 높여 무시되는 S_b 의 널 공간을 줄일 수 있고 S_w 가 S_b 의 증가된 레인지 공간으로 사상될 때 벡터들이 중복될 가능성을 줄여 준다.

IV. 실험

1. 실험 데이터

제안한 방법을 평가하기 위해 다양한 실험을 하였다. 먼저 3차원 2-클래스의 인공데이터 "XOR3"를 생성하여 실험하였다. 이때 두 클래스는 각각 중심이 $\{(-2, -2)^T, (2, 2)^T\}$ 와 $\{(2, -2)^T, (-2, 2)^T\}$ 인 두개의 부분 집합으로 나누어 표준편차 1인 정규분포의 데이터를 100개씩 모두 400개를 생성하였다(그림 1 참조). 여기에 다시 4개의 부분집합별로 각각 중심이 0이면서

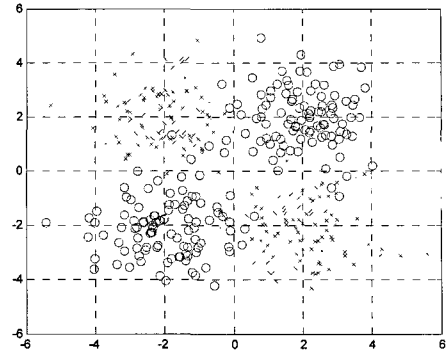


그림 1. 2차원 인공데이터
Fig. 1. 2-dimenaional artificial data.

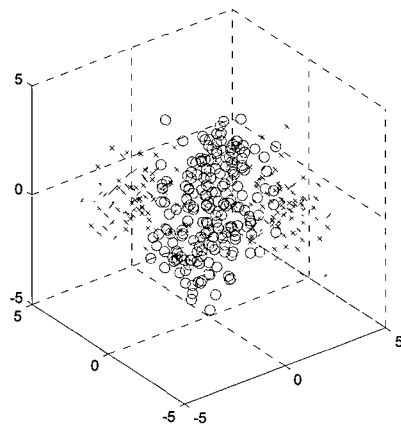


그림 2. 실험용 3차원 인공데이터 XOR3
Fig. 2. Experimental 3-dimenaional artificial data XOR3.

표 1. 실험 데이터의 특성
Table 1. Characteristics of experimental data.

실험 데이터 종류	벡터 차원	클래스 수	클래스별 패턴 수	패턴벡터 총 수
XOR3	3	2	200	400
AT&T	2576	40	10	400
Yale	4880(644)	15	11	165

정규분포를 따르는 랜덤 수를 생성하여 세 번째 특징으로 추가한 3차원 데이터를 생성하였다(그림 2 참조). 여기서 두 클래스는 각각 x와 o 기호로 표시하였다.

또한 실제의 벤취마크 데이터로 얼굴 데이터베이스 AT&T와 Yale를 이용하였다. AT&T⁴⁾는 92x112 크기의 영상을 1/4 크기로 축소하여 이용하였고 Yale⁵⁾은

4) <http://www.cl.cam.ac.uk/Research/DTG/attarchive/facedatabase.html>
5) <http://www1.cs.columbia.edu/~belhumeur/pub/images/yalefaces>

80x61 크기와 28x23 크기의 얼굴영역 만을 추출하여 이용하였다^[11]. 본 논문의 실험에서 이용한 데이터 셋의 특성은 표 1과 같다.

2. 실험 방법

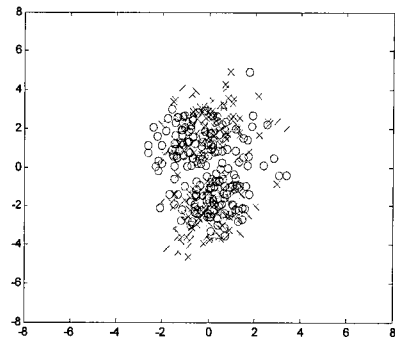
제안 방법의 식별성능을 평가하기 위한 실험 방법은 학습패턴의 생성, 사전 클러스터링, 특징추출, 최-근방 이웃(nearest neighbor: NN) 식별의 단계로 이루어 졌다. 먼저 학습패턴집합을 생성하는 단계는 n 개의 입력 패턴집합으로부터 한 패턴을 제외한 후, 제외한 패턴을 입력패턴으로 사용하고 나머지 $n-1$ 개의 패턴은 학습 패턴집합으로 사용하는 Leave-one-out 성능평가 방법(이하, L법이라 함)을 사용하였다. 이는 입력 패턴집합으로부터 1에서 n 번째 패턴을 차례로 하나씩 평가패턴(측정 패턴)으로 선정하는 방식으로, 서로 다른 학습패턴집합을 생성하고 각 학습패턴에 대해 사전-클러스터링단계에서부터 NN 식별단계까지의 과정을 n 번 반복한 다음, 결과를 평균하여 인식 율을 계산하였다.

3. 실험 결과

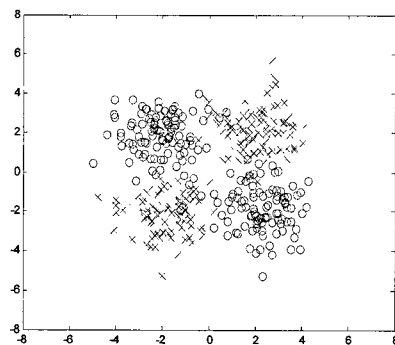
XOR3, AT&T, Yale에 대한 실험결과는 다음과 같다. 먼저, XOR3를 2차원 특징공간으로 사상시킨 결과는 그림 3과 같다. 여기서는 2개의 클래스로부터 총 4개의 클러스터(네 개의 서브 클래스)를 만들어 내는 경우만을 실험하였다. 그림 3(a), (b)는 기존의 LDA와 제안한 방법을 이용하여 3차원 XOR3를 2차원 데이터로 축소한 결과이다. 또한 그림 3(c)는 제안한 방법으로 Direct-LDA를 수행한 결과이다⁶⁾.

기존의 LDA로 패턴 벡터의 차원을 축소할 경우 그림 3(a)에서와 두 클래스의 특징공간이 서로 겹치게 된다. 이와는 달리 사전 클러스터링을 적용하였을 경우 그림 3(b), (c)에서와 같이 클래스간의 겹치는 영역이 상대적으로 줄어들게 된다. 따라서 쉽게 클래스를 식별할 수 있음을 알 수 있다.

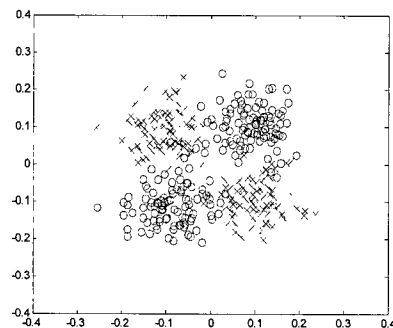
AT&T와 Yale에 대한 실험결과는 다음과 같다. 먼저 클러스터링 할 클래스를 결정하기 위하여 군집거리를 측정하였다. 예를 들어, AT&T를 대상으로 우선순위 10위내의 클래스번호는 각각 1, 35, 16, 31, 28, 11, 10, 36, 20, 2이었고, 클래스별 군집내부거리는 1.00 0.98 0.95 0.94 0.91 0.84 0.81 0.80 0.73 0.73이었다. 그리고 Yale의



(a)



(b)



(c)

그림 3. 2차원으로 사상된 XOR3 인공데이터. (a) LDA, (b) 제안한 LDA, (c) 제안한 Direct-LDA
Fig. 3. XOR3 artificial data mapped into 2-dimensional data. (a) LDA, (b) proposed LDA, and (c) proposed Direct-LDA.

경우 클래스번호 3, 15, 6, 1, 8, 9, 7, 5, 4, 12의 내부 군집거리는 각각 1.00, 0.99, 0.98, 0.92, 0.88, 0.87, 0.72, 0.68, 0.67, 0.65이었다. 여기서 군집내부거리는 정규화한 값이다. 즉, AT&T의 1번과 35번 클래스에 포즈나 배경, 또는 조명에 변화가 큰 얼굴영상들이 모여 있음을 알 수 있다. 따라서 이들 영상을 클러스터링하여 서브 클래스로 나누어 처리하였다.⁷⁾

⁶⁾ XOR3에 PCA+LDA를 적용할 경우 PCA와 LDA단계에서 각각 차원 축소가 이루어지게 되어 1차원 데이터가 된다.

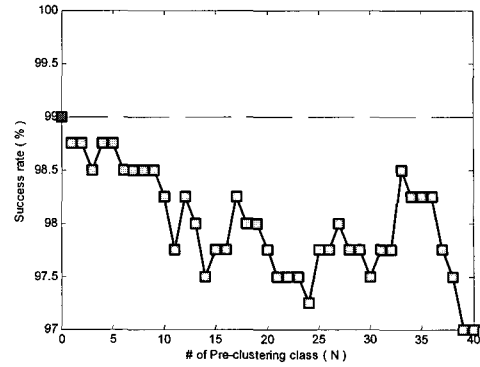
⁷⁾ 제안 방법은 간단한 반면 특별한 분포 구조를 처리하는 데에는 제한이 있다. 예를 들어, 한 클래스의

또한, 사전 클러스터링 할 클래스의 수 N 을 결정하기 위하여 (즉, 서브 클래스의 수를 결정하기 위한 클러스터를 구하기 위하여) LDA, PCA+LDA, Direct-LDA에 대한 식별 율의 변화를 관찰하였다. 클래스별로 군집 내부거리를 구하여 정렬한 다음, 가장 거리가 큰 클래스부터 클러스터링한 결과를 대상으로 차원을 축소하였다. 이와 같은 과정을 1부터 c 까지 모든 클래스에 대하여 반복하면서 NN법으로 식별성능을 측정하였다.

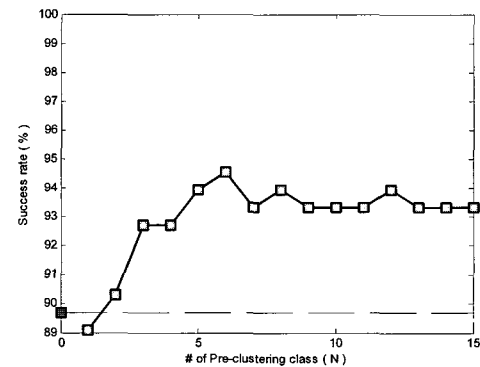
AT&T와 Yale를 대상으로 기존의 Direct-LDA와 제안한 방법을 실험한 결과는 그림 4와 같다. 여기서 수평축(x-축)은 클러스터링을 수행한 클래스의 개수이고, 수직축(y-축)은 L법으로 측정된 식별 율(%)에 해당한다. 또한 그림에서 수평 점선은 기존 방법에 대한 식별 율이고⁸⁾, 실선은 제안한 방법의 식별 율이다.

먼저 AT&T의 경우 그림 4(a)에서와 같이, 기존의 Direct-LDA법으로 차원을 축소한 다음 NN법으로 식별하였을 경우 식별 율은 99.00(%)인데 반하여 하나의 클래스를 클러스터링한 다음 Direct-LDA법으로 차원을 축소하는 제안 방법의 식별 율은 98.75(%)이었다. 다시 클러스터링 클래스 수를 늘려감에 따라(서브 클래스 수를 증가시켜 감에 따라) 식별 율은 증감을 반복하면서 전체적으로 감소하였다. 그러나 Yale 데이터의 경우 그림 4(b)에서와 같이, (내부 군집거리가 가장 큰) 첫 번째 클래스를 클러스터링하였을 경우에는 89.00(%)이었던 식별 율이 클러스터링 클래스의 수를 늘려감에 따라 95%정도까지 증가하였다.

이상과 같은 분석 결과로부터, AT&T의 경우 클러스터링하는 클래스의 수를 (식별 율이 최대가 되는) $N=1$ 로 설정하였고, Yale 데이터의 경우에는 $N=6$ 으로 설정하였다. 그 밖에 LDA와 PCA+LDA의 경우에도 그림 4의 Direct-LDA법과 같은 분석 방법으로 클러스터링 할 클래스의 수를 결정하였다. 실험 결과로부터, AT&T의 경우 $N=3$, $N=2$ 로 설정하였고, Yale 데이터의 경우에는 $N=11$, $N=3$ 으로 각각 설정하였다(이를 위한 그림은 생략하였다). 여기서 AT&T의 N 이 Yale 데이터의 N 보다 적은 이유는 AT&T 얼굴영



(a)



(b)

그림 4. 두 얼굴영상 데이터베이스에 대한 식별 율(%). (a) AT&T에 대한 Direct-LDA의 식별 결과, (b) Yale에 대한 Direct-LDA의 식별 결과

Fig. 4. Classification rates(%) for two facial image data bases. (a) results of Direct-LDA for AT&T, (b) results of Direct-LDA for Yale.

상은 Yale 영상에 비하여 조명상태나 얼굴 표정, 배경 영상 등이 양호하기 때문인 것으로 보여 진다. 즉, AT&T의 경우 클러스터링에 의한 클러스터가 잘 형성되지 않는 반면에 Yale 데이터의 경우에는 비교적 분명한 클러스터가 형성되었기 때문이다.

이상과 같이 클러스터링할 클래스의 개수 N 을 결정하여 사전 클러스터링을 수행한 다음, LDA, PCA+LDA 및 Direct-LDA로 차원을 축소하였다. XOR3와 AT&T, Yale를 대상으로 기존 방법과 제안 방법의 식별 율을 실험한 결과는 표 2와 같다. 여기서 사전 클러스터링을 적용하지 않은($N=0$) 기존의 차원 축소방법의 식별 율과 사전 클러스터링을 적용하는 제안 방법의 식별 율을 그에 해당하는 N 값과 함께 표기하였다. 또한, 첫 번째 칸(NN 걸림)은 차원 축소를 하지 않은 원형 데이터에 대한 최-근방 식별 율(%)이다.

표 2에서 XOR3를 대상으로 LDA를 수행할 경우 평균 인식률이 60.25에서 91.25(%)로 크게 상승하였다.

데이터들이 sparse하게 분포되어 있으면서 uni-modal 분포를 가질 수 있는데, 이 경우에도 군집거리가 크므로 클러스터링 할 클래스로 선택될 가능성이 있다. 이점을 지적한 심사위원께 감사드린다.
⁸⁾ 제안 방법에서는 사전 클러스터링할 클래스의 수에 따라 식별 율이 변하였다. 그러나 기존의 방법에서는 한번만 차원 축소가 이루어지기 때문에 식별 율을 수평으로 표시하였다.

표 2. 실험 데이터 셋에 대한 인식률(%)⁹⁾
Table 2. Classification rates(%) for experimental data sets.

실험 데이터 종류	NN	기존의 방법			제안한 방법		
		LDA	PCA+LDA	Direct-LDA	LDA	PCA+LDA	Direct-LDA
XOR3	95.00	60.25	91.75	-	91.25 (N=2)	91.75 (N=2)	92.25 (N=2)
AT&T	97.75	97.75	95.25	99.00	98.50 (N=3)	95.50 (N=2)	98.75 (N=1)
Yale	79.39	95.15	93.94	89.70	95.15 (N=11)	95.76 (N=3)	94.55 (N=6)

또한, AT&T와 Yale의 경우에도 전체적으로 식별율이 향상하였다. 예를 들어 Yale에 대한 Direct-LDA의 경우, 기존 방법의 89.70(%) 인식율이 제안 방법에서는 94.55(%)로 증가하였다. 그러나 AT&T의 경우에는 99.00(%)의 인식율이 제안 방법에서는 98.75(%)에 머무르고 있으며, 이는 AT&T의 클래스 별 얼굴영상이 서로 비슷하여 명확한 클러스터를 형성하지 못하기 때문인 것으로 보인다.

끝으로, 제안 방법은 AT&T가 Yale에 비하여 식별 성능이 저조하였다. 그 이유는 두 데이터 모두 포즈와 표정이 다른 반면, 배경과 조명 상태에서는 Yale이 AT&T와 비교하여 큰 차이를 보이고 있기 때문이다. 즉 특징 벡터는 그레이 레벨 값으로 구성되기 때문에 배경 및 조명 상태가 다를 경우 식 (19)에서와 같이 그레이 레벨은 클래스 내부거리의 계산에 큰 영향을 미치게 된다. 실험 결과는 이러한 상황을 반영한 것으로 제안 방법은 조명상태와 배경 등에 기인한 비선형 구조를 해결하는데 유용할 것으로 보여 진다.

3. 수행 시간

LDA, PCA+LDA 및 Direct-LDA에 대하여 사전 클러스터링을 적용하지 않은 상태와 적용한 상태에서 각 알고리즘의 단계별로 처리시간을 측정한 결과는 표 3과 같다. 여기서는 AT&T 만을 대상으로 수행시간을 측정하였으며, 전체 데이터에 대한 leave-one-out 실험에

표 3. 처리 시간의 비교 (sec)
Table 3. Comparison of processing CPU-time (sec).

단계	기존의 방법			제안한 방법		
	LDA	PCA+LDA	Direct-LDA	LDA	PCA+LDA	Direct-LDA
0	0	0	0	9.19	9.66	9.25
1	7.05	7.21	7.67	7.11	6.28	6.98
2	2082.28	141.97	0.14	2085.35	129.75	1.02
3		2.53			0.09	
4		2.42	0.76		0.03	
5		3.25			0.03	
6	0.28	0.14		0.16	0.08	0.16
합	2083.62	157.32	8.57	2101.81	145.92	17.31

소요되는 시간을 측정한 값이다. 또한, 각 알고리즘의 수행을 처리내용에 따라 단계별로 나누어 측정하였다. LDA의 경우 전체 과정을 공분산 계산 단계, 변환행렬 구성단계, 변환 행렬을 이용한 차원 축소단계로 나누어 측정하였다. 그리고 PCA+LDA는 여섯 단계로 나누어 측정하였고, Direct-LDA는 네 단계로 나누어 측정하였다. 그리고 모든 알고리즘에서 단계 0은 사전 클러스터링을 수행하는 단계이다. 측정에 이용한 컴퓨터는 Pentium 4 CPU 3.00GHz, 1.50GB RAM이고 시뮬레이션 응용프로그램은 MATLAB 7.0이다.

표 3의 측정결과는 제안 방법의 각 단계별 처리 시간은 기존 방법과 큰 차이가 없음을 보이고 있다. 예를 들어, LDA, PCA+LDA, Direct-LDA의 기존 방법의 처리 시간은 각각 2083.62, 157.32, 8.57(sec)이며, 제안 방법에서는 2101.81, 145.92, 17.31(sec)로 증가하여 서로 비슷한 계산 특성을 보이고 있다. 또한, Direct-LDA를 수행할 경우 기존 방법에서는 8.57(sec)이었던 처리시간이 제안 방법에서는 사전 클러스터링 시간을 포함하여 17.31(sec)임을 보이고 있다.

V. 결 론

본 논문에서는 클래스별로 사전 클러스터링을 통하여 생성된 클러스터를 가상의 클래스로 취급하여 LDA, PCA+LDA, Direct-LDA 등의 차원 축소 법을 적용하는 방법을 제안하였다. 3차원 인공 데이터 XOR3와 실생활 벤치마크 데이터 AT&T, Yale를 대상으로 제안 방법을 평가하는 실험을 하였다. 실험 결과, 제안 방법의 식별율이 기존의 방법보다 전체적으로 우수함을 확인하였다. 먼저 XOR3와 같은 2-클래스 문제의 경우 기존의

⁹⁾ XOR3에 대한 PCA+LDA의 경우 PCA에서 3차원을 2차원으로 축소한 다음 LDA에서는 차원 축소 없이 변환시킨 결과에 대한 식별율이다. 반대로 PCA에서 차원 축소 없이 변환시킨 다음 LDA에서 3차원을 2차원으로 축소시킬 경우 식별율은 91.25(%)이었다. 또한 Direct-LDA의 식별율을 공란으로 표기한 이유는 식 (2)의 S_b 계산에서 클래스의 평균과 전체 평균이 겹치게 되는 경우 변환행렬을 구할 수 없었기 때문이다.

LDA에서는 특징을 1차원으로 밖에 뽑을 수 없는데 반해 제안한 방법은 보다 높은 차원으로 특징을 추출하여 인식률을 향상시킬 수 있다. 또한 Direct-LDA의 경우 비교적 얼굴 위치, 조명 상태, 표정 변화의 정도가 낮은 AT&T에 대해서는 제안 방법을 적용하였을 때 성능향상이 없거나 크지 않았다. 그러나 조명이나 배경의 변화가 큰 Yale 데이터에 대해서는 식별성능이 큰 폭으로 향상되었다. 따라서 고차원 얼굴 영상 데이터의 분포가 비선형적 다중 모형일 경우에 제안방법의 사전 클러스터링이 차원 축소에 효과적임을 확인하였다.

참 고 문 헌

- [1] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: a review", *IEEE Trans. Pattern Anal. Intell.* vol. 22, no. 1 pp. 4-37, March 2000.
- [2] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Second Edition, Academic Press, San Diego, pp. 452-459, 1990.
- [3] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, Wiley, New York, 2001.
- [4] 김상운, *MATLAB으로 배우는 패턴인식 및 학습*, 홍릉과학출판사, 서울, 2005. 1.
- [5] A. K. Jain and B. Chandrasekaran, "Dimensionality and sample size consideration in pattern recognition", *Handbook of Statistics*, P. R. Krishnaiah and L. N. Kanal, Eds., vol. 2, pp. 835-855, 1982.
- [6] M. Turk and A. P. Pentland, "Face recognition using eigenfaces," *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*. pp. 586-591, 1991.
- [7] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 7, pp. 711-720, May 1997.
- [8] J. Ye, R. Janardan, C. H. Park, and H. Park, "An optimization criterion for generalized discriminant analysis on undersampled problems", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 26, no. 8, pp. 982-994, Aug. 2004.
- [9] L. Chen, H. Liao, M. Ko, J. Lin, and G. Yu, "A new LDA-based face recognition system which can solve the small sample size problem", *Pattern Recognition*, vol. 33, pp. 1713-1726, 2000.
- [10] P. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data with application to face recognition", *Pattern Recognition*, vol. 34, pp. 2067-2070, 2001.
- [11] H. Gao and J. W. Davis, "Why direct LDA is not equivalent to LDA", *Pattern Recognition* vol. 39, pp. 1002-1006, 2006.
- [12] H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana, "Discriminative common vectors for face recognition", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 27, no. 1, pp. 4-13, Jan. 2005.
- [13] 김상운, "Prototype Reduction Schemes와 Mahala-nobis거리를 이용한 Relational Discriminant Analysis," *대한전자공학회논문지-CI*, vol. 43, no. 1, pp. 9-16, 2006. 1.
- [14] S. -W. Kim, "On solving the small sample size problem using a dissimilarity representation for face recognition", *Lecture Note in Computer Science*, vol. LNCS-4179, pp. 1174-1185, 2006.
- [15] G. Baudat and F. Anouar, "Generalized Discriminant Analysis Using a Kernel Approach" *Neural Comput.*, vol. 12, pp. 2385-2404, 2000.
- [16] S. -W. Kim and B. J. Oommen, "On optimizing kernel-based Fisher discriminant analysis using prototype reduction schemes", *Lecture Note in Computer Science*, vol. LNCS-4109, pp. 826-834, 2006.
- [17] M. Zhu and A. M. Martinez, "Subclass discriminant analysis", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 28, no. 8, pp. 1274-1286, Aug. 2006.



김 상 운(정회원)
 1978년 한국항공대학 학사.
 1980년 연세대학교 대학원 석사.
 1988년 동 대학원 공학박사.
 1989년~현재 명지대학교 컴퓨터
 공학과 교수.

<주관심분야 : 패턴인식, 미디어처리>



구 범 용(학생회원)
 2005년 명지대학교 전자공학과
 학사
 2007년 명지대학교 전자공학과
 석사
 <주관심분야 : 영상처리, 패턴인
 식>



최 우 영(정회원)
 1985년 서강대학교 전자공학과
 학사
 1987년 서강대학교 전자공학과
 석사
 1992년 서강대학교 전자공학과
 박사

1990년~1991년 생산기술연구원 HDTV 연구단
 연구원

1992년~현재 명지대학교 전자공학과 교수

1996년~1997년 University of Iowa, Dept. of
 Biomedical Eng. 방문교수

<주관심분야 : 영상처리, 컴퓨터 비전, 패턴인식,
 컴퓨터그래픽스>