
주 요소와 독립 요소 분석의 통합에 의한 얼굴 인식

류재홍* · 김강철* · 임창균*

Face Recognition By Combining PCA and ICA

Jaehung Yoo* · Kang Chul Kim* · Chang-Gyoon Lim*

요 약

기존의 독립 요소 방법에 의한 얼굴인식에서는 주 요소 해석법으로 고유치 크기에 의해 특징을 추출하고 감소된 차원에서 특징 개선을 위한 독립 요소 해석법의 학습을 수행한다. 제거된 특징 공간 내에 필요한 요소가 있는 경우를 고려하지 못한 것이다. 새로운 방법은 독립 요소 해석에 의한 학습을 먼저 시행하고 분리된 데이터를 4차 중심 모멘트에 의한 축적계수(cumulant)인 커토시스(kurtosis)의 절대값 크기에 의하여 특징을 추출한다. 하지만 독립 요소 방법은 효과적으로 노이즈를 제거하지 못한다. 두 방법의 결합효과는 주 요소 해석법을 노이즈 필터로 사용할 때 극대화 될 수 있다. 즉 주 요소 해석법을 백색화와 노이즈 필터로 하고 독립 요소 해석법을 특징 추출 방법으로 사용하는 것이다. 실험 결과는 새로운 방법론이 기존의 방법론보다 우수함을 보여준다.

ABSTRACT

In a conventional ICA(Independent Component Analysis) based face recognition method, PCA(Principal Component Analysis) first is used for feature extraction, ICA learning method then is applied for feature enhancement in the reduced dimension. It is not considered that a necessary component can be located in the discarded feature space. In the new ICA(NICA), learning extracts features using the magnitude of kurtosis (4-th order central moment or cumulant). But, the pure ICA method can not discard noise effectively. The synergy effect of PCA and ICA can be achieved if PCA is used for noise reduction filter. Namely, PCA does whitening and noise filtering. ICA performs feature extraction. Experiment results show the effectiveness of the new ICA method compared to the conventional ICA approach

키워드

Face Recognition, PCA, ICA, Feature Extraction

I. 서 론

얼굴 영상의 특징은 경계선, 등고선, 영상 영역 등의 국부적 시각 특징, 히스토그램, 모멘트 등의 통계적 특징, 푸리에의 변환 등의 계수적 특징, 주성분 해석법(PCA-Principal Component Analysis), 독립요소 해석법(ICA-Independent Component Analysis) 등의 대수적 특징으로

나누어진다. 이중에서 대수적 특징의 하나인 주성분 해석법을 이용한 얼굴인식방법은 1990년대 초부터 지속적으로 연구되어 왔다[1]. 하지만 얼굴 인식을 단지 2차 중심 모멘트인 분산에 한정하는 것은 고차원 모멘트에 포함된 특징을 포괄할 수는 없는 한계를 가지고 있다. 최근 들어 신호처리 분야에서는 독립 요소 해석법을 사용하여 혼합된 신호를 분리하는 연구가 활발히 진행되고 있다. 얼굴

인식 등 영상 처리 분야에서도 독립 요소 해석법을 이용한 연구가 진행되어 오고 있다[2,3].

얼굴 인식에 이용한 기존의 독립 요소 해석법은 주로 주 요소 해석법으로 고유치 크기에 의해 특징을 추출하고 감소된 차원에서 독립 요소 해석법을 사용한 신경회로망 방식의 학습법을 적용한다[4,5]. 특징 추출에서 제외된 차원에 필요한 독립 요소가 온전히 속한 경우는 전혀 고려하지 않았다. 새로운 방법론은 독립 요소 해석에 의한 학습을 먼저 시행하고 분리된 데이터를 4차 중심 모멘트에 의한 축적계수(cumulant)인 커토시스(kurtosis)의 절대값 크기에 의하여 특징을 추출한다. 두 방법의 시너지 효과는 주 요소 해석법을 노이즈 필터로 사용하고 독립 요소 해석법을 특징 추출 방법으로 사용할 때 극대화 될 수 있다[6,7].

II. 관련 연구

영상의 크기를 픽셀의 수로 N 이라 하고 영상의 개수가 n 인 데이터에서 동일 좌표로 구성된 픽셀 벡터 x_i , ($i=1, \dots, N$)와 영상 행렬 $X = [x_1, x_2, \dots, x_N]$ 을 정의한다. 한 영상은 영상행렬의 한 행에 나타난다. 이 행렬의 크기는 $n \times N$ 이 된다. 한편 영상 행렬을 X^T 로 정의하는 경우 각 열은 영상 벡터가 되는데 본 논문에서는 다루지 않는다[1].

영상에 대한 센터링 즉 평균 벡터가 영이 되도록 한다.

$$m_x = \frac{1}{N} \sum_{i=1}^N x_i \quad (1)$$

$$x_i = x_i - m_x \quad (2)$$

센터링 후 공분산(Covariant) 행렬은 상관성(Correlation) 행렬과 동일하게 다음과 같다.

$$C = \frac{1}{N-1} X \cdot X^T \quad (3)$$

주 요소 해석은 공분산을 고유 벡터 행렬 R 과 고유치 대각선 행렬 L 로 분해한다.

$$C = R \cdot L \cdot R^T \quad (4)$$

여기서 $R = (r_1, r_2, \dots, r_n)$

$$L = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n),$$

$(\lambda_1 > \lambda_2 > \dots > \lambda_n)$ 이다.

주 요소 해석법은 또한 주어진 데이터 행렬을 고유 벡터 행렬과 주 요소 행렬로 분해한다.

$$X = R \cdot P \quad (5)$$

주 요소 해석법에서 주어진 데이터의 분산인 고유치를 내림차순으로 정렬하여 처음의 m 개의 고유치에 해당하는 고유 벡터 공간으로 주어진 데이터를 추산하면 다음과 같다.

$$X_{est} = R_m \cdot P_m \quad (6)$$

여기서 $R_m = (r_1, r_2, \dots, r_m)$, $m \leq n$ 이다. 행렬의 크기는 $(n \times N) = (n \times m)(m \times N)$ 이다. 감소된 주 요소 행렬은 주 기축 영상(principal basis image)이고 다음과 같다[5].

$$P_m = R_m^T \cdot X \quad (7)$$

주 요소 해석법에 의한 얼굴 인식은 식 (6)에서 각 영상의 표현이 R_m 의 각 행 벡터에 주어지므로 훈련 데이터와 평가 데이터에 대하여 R_m 식을 구하여 특징 벡터 행렬로 삼는다.

$$R_m = X \cdot P_m^T \cdot L_m^{-1} \cdot \frac{1}{N-1} \quad (8)$$

$$R_{m \text{ test}} = X_{test} \cdot P_m^T \cdot L_m^{-1} \cdot \frac{1}{N-1} \quad (9)$$

영상에 대한 독립 요소를 구하기 위하여 독립 요소 해석법을 적용한다. 독립 요소 해석의 일반적인 식은

$$X = A \cdot S \quad (10)$$

$$Y = W \cdot X \quad (11)$$

와 같다. X 는 관측 데이터이고 A 는 알 수 없는 혼합 행렬(unknown mixed matrix)이며, S 는 알 수 없는 원래의 데이터(unknown source matrix)이다. Y 는 S 를 순렬 행렬과 스케일 행을 곱한 결과로 간주한다.

주 요소 해석법에 구한 감소된 주 요소 행렬 P_m 을 독립 요소 해석법에 적용한다. P_m 의 분산을 같게 하기 위하여 백색화 필터는 크기가 $m \times m$ 이고 다음과 같다.

$$W_z = 2 \cdot E(P_m \cdot P_m^T)^{-1/2} \quad (12)$$

백색화 필터 과정을 거친 후 최대 엔트로피 방법을 통하여 학습을 한다[2].

$$P_{\Pi m} = W_z \cdot P_m$$

$$z = \frac{1}{1 + e^{-y}}$$

$$\frac{\partial}{\partial z} \frac{\partial z}{\partial y} = (1 - 2 \cdot z) = z'$$

$$\Delta V = (I + z' \cdot y^T) \cdot V \quad (13)$$

여기서 $W_I = V \cdot W_z$ 라 두면

$$Y = W_I P_m \quad (14)$$

가 된다. 식 (6)과 식 (14)에서 독립 요소 공간으로 주어진 데이터를 추산하면 다음과 같다.

$$X_{est} = R_m \cdot W_I^{-1} \cdot Y \quad (15)$$

독립 요소 해석법에 의한 얼굴 인식은 식 (15)에서 각 영상의 표현이 $R_m \cdot W_I^{-1}$ 의 각 행 벡터에 주어지므로 훈련 데이터와 평가 데이터에 대하여 특징 벡터 행렬로 삼는다.

$$B_{train} = R_m \cdot W_I^{-1} \quad (16)$$

$$B_{test} = R_{test} \cdot W_I^{-1} \quad (17)$$

III. 주 요소와 독립 요소 분석 통합과 얼굴인식 알고리즘

특이치 분해법(SVD-Singular Value Decomposition)에 의하여 혼합 행렬 A 는 크기가 $n \times m$ 이고 분리 행렬 W 는 크기가 $m \times n$ 일 때 다음과 같이 분해된다.

$$A = R \cdot L^{1/2} \cdot V^T \quad (18)$$

$$W = V \cdot L^{-1/2} \cdot R^T \quad (19)$$

여기서 R 는 크기가 $n \times n$ 인 직교 행렬이고 $L^{1/2}$ 는 $n \times m$ 대각 행렬이며 V 는 $m \times m$ 직교 행렬이다. 혼합 행렬 A 의 랭크가 r 일 때 실제 크기는 각각 $n \times r$, $r \times r$, $r \times m$ 가 된다. 식 (10)과 (11)을 다시 쓰면 사각 혼합 행렬 A 에 대하여

$$X = A \cdot S \quad (20)$$

$$Y = W \cdot X \quad (21)$$

관찰된 데이터 행렬 X 는 크기가 $n \times N$ 이고 원천 데이터의 행렬 S 는 크기가 $m \times N$ 이다. 특이치 분해하면 다음과 같다.

$$X = \sqrt{N-1} \cdot R \cdot L^{1/2} \cdot \Psi^T \quad (22)$$

$$S = \sqrt{N-1} \cdot V \Psi^T \quad (23)$$

여기서 Ψ^T 는 $X^T X$ 의 고유 벡터 행렬이다. 원천 데이터의 행렬 S 는 정규화된 데이터이다.

식 (19)에서 주 요소 해석법에 의한 전처리 과정 중 백색화 필터 $L^{-1/2} \cdot R^T$ 에 의한 정규화가 자연스럽게 내재되어 있음을 발견 할 수 있다.

$$X_1 = L^{-1/2} \cdot R^T \cdot X \quad (24)$$

$$X_1 = \sqrt{N-1} \cdot \Psi^T \quad (25)$$

$$X_1 = V^T \cdot S \quad (26)$$

$$Y = S = V \cdot X_1 \quad (27)$$

주어진 행렬 X_1 에 관하여 분리 행렬 V 를 최대 엔트로피 방법을 통하여 학습을 한다.

학습을 한 후 출력 행렬 Y 의 행 벡터의 커토시스의 절대값을 내림차순 정렬을 한다.

$$kurt(Y) = |E(Y^4) - 3 \cdot [E(Y^2)]^2| \quad (28)$$

정렬한 순서를 가지고 출력 행렬 Y 와 분리 행렬 V 를 정렬한다.

새로운 독립 요소 해석 방법론에 의한 얼굴인식 알고리즘을 소개하면 다음과 같다.

먼저 주 요소 해석법을 사용하여 노이즈 제거를 행한다. 주 요소 해석법에 의하여 구해진 고유치를 내림차순 정렬을 하고 난 후 총 분산(total variance)의 95%를 커버하는 부분 공간(subspace)의 차원 m 을 정한다.

$$\sum_{i=1}^m \lambda_i \geq 0.95 \sum_{i=1}^n \lambda_i \quad (29)$$

다음은 백색화 필터 처리를 하게 된다.

$$X_1 = L_m^{-1/2} \cdot R_m^T \cdot X \quad (30)$$

$$Y = V \cdot X_1 \quad (31)$$

여기서 $k \leq m$, Y 의 크기가 $k \times N$, W 의 크기가 $k \times n$, W 의 초기치를 $L_k^{-1/2} \cdot R_k^T$ 이면 화이트 전 비정규화 데이터에 대한 유일한 학습법인 DICA(Direct ICA) 방법이 된다. 기존의 방법은 모두 정규화 데이터에 대하여 학습하고 있다[8,9].

여기서 $k \leq m$, Y 의 크기가 $k \times N$, V 의 크기가 $k \times n$, V 의 초기치가 작고 랜덤 행렬이고 식 (30)에서 행렬의 크기가 다음과 같으면 NDICA(New Direct ICA) 방법이 된다.

$$(n \times N) = (n \times m)(m \times n)(n \times N) \quad (32)$$

여기서 Y 의 크기가 $m \times N$, V 의 크기가 $m \times m$, V 의 초기치가 작고 랜덤 행렬이고 식 (30)에서 행렬의 크기가 다음과 같으면 NICA(New ICA) 방법이 된다.

$$(m \times N) = (m \times m)(m \times n)(n \times N) \quad (33)$$

여기서 $k \leq m$ Y 의 크기가 $k \times N$, V 의 크기가 $k \times k$, V 의 초기치가 항등 행렬이고 식 (30)에서 행렬의 크기가 다음과 같으면 기존의 ICA방법이 된다[5].

$$(k \times N) = (k \times k)(k \times n)(n \times N) \quad (34)$$

정규화된 관찰 행렬 X_1 에 대하여 독립 요소 분리 학습을 실행 한 후 학습되어진 출력벡터 Y 를 커토시스의 절대값으로 내림차순 정렬을 한다. 각 영상의 표현이 식 (19)의 A 의 각 행 벡터에 주어지므로 훈련 데이터와 평가 데이터에 대하여 특징 벡터 행렬로 삼는다.

$$X \cdot Y^T = X \cdot X^T \cdot W^T = (N-1)A \quad (35)$$

$$B_{train} = X \cdot Y_k^T \quad (36)$$

$$B_{test} = X_{test} \cdot Y_k^T \quad (37)$$

여기서 Y_k 은 커토시스 절대값이 큰 순서대로 선택한 k 개의 기축 영상 행렬로 크기가 $k \times N$ 이다. 훈련 데이터와 평가 데이터로부터 정규화된 거리 함수(normalized metric)인 여현각(cosine angle)으로 1-NN(Nearest Neighbor)방법을 사용하여 얼굴 인식률을 평가한다.

IV. 실험 및 결과 분석

실험데이터는 벤치마크 DB중 하나인 Cambridge 대학의 Olivetti 연구실(Olivetti Research Laboratory : ORL)의 얼굴 영상 데이터 베이스를 사용하였다. 총 400개의 얼굴 영상으로 1명당 10개씩 40명의 정보가 있다. 본 실험은 각 개인 당 5개의 영상을 학습데이터로 5개를 평가데이터로 사용하였다.

그림 1은 10개 중 5개를 선택할 조합의 수 252개를 이용하여 학습 데이터를 선정하고 총 분산의 95%를 커버하는 차원수 $m=94$ 에 대한 PCA 인식률을 나타낸다. 최대 인식률을 가진 학습 데이터 번호는 1, 2, 5, 6, 9이다. 이 학습 데이터 번호를 가지고 주 요소 해석, 독립 요소 해석에 대한 추가 실험을 수행한다.

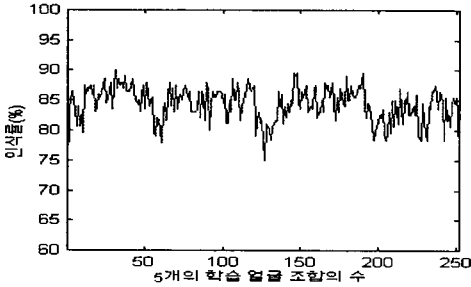


그림 1. PCA 인식률 (m=94)
Fig. 1 PCA Performance (m=94)

그림 2는 특징 요소의 개수(m)의 변화에 대한 주 요소 해석법(PCA), 기존의 독립 요소 해석법(ICA), 직접 독립 요소 해석법(DICA)의 인식률변화 그래프이다. 기존의 독립 요소 해석법(ICA)에서 보면 요소의 개수가 증가해도 급격한 인식률 감소가 나타나는 등 건전성(robustness)을 볼 수 없다. 인식률 감소된 것은 학습하기 전 주 요소해석에서 요소를 감소하면서 독립 요소의 학습을 행하므로 주요한 독립 요소 데이터가 제외되어 있는 상태에서 독립 요소 학습이 되어진 것으로 볼 수 있다. 한편 본 논문에서 제안한 DICA는 일관성 있게 PCA보다 인식률이 높다.

그림 3은 노이즈 필터 없이 순수하게 ICA를 행한 경우 커토시스 절대값에 따른 특징 추출 방법이 전혀 효과를 발휘하지 못함을 보여준다.

그림 4는 본 논문에서 제안하는 방법들(NICA, NDICA)에 대한 실험 결과이다. 노이즈 필터링 후 ICA를 행한 경우 커토시스 절대값에 따른 특징 추출 성능이 그림 2의 기존 ICA의 인식률 불건전성과 그림 3의 순수 NICA의 특징 추출 방법의 비효율성을 동시에 극복함을 알 수 있다. 여기서 독립 요소의 개수 $k = 1, \dots, m; m = 94$ 이다.

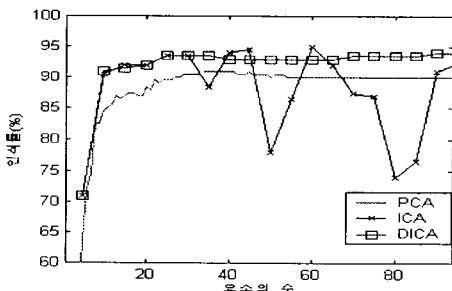


그림 2. PCA, ICA, DICA 인식률
Fig. 2 PCA, ICA, DICA Performance

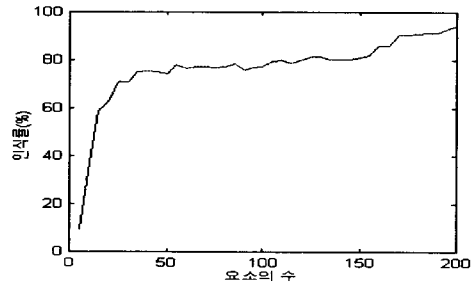


그림 3. 순수 NICA 인식률
Fig. 3 Pure NICA Performance

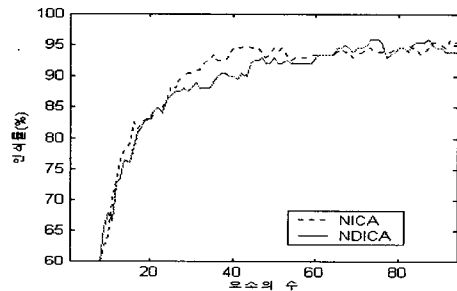


그림 4. NICA, NDICA 인식률
Fig. 4 NICA, NDICA Performance

표 1. 최고 인식률 통계자료
Table 1. Peak Performance Statistics

| | 인식률(%) | 요소의 수 |
|----------|--------|-------|
| PCA | 90 | 50 |
| ICA | 95 | 60 |
| DICA | 94 | 90 |
| PureNICA | 94.5 | 200 |
| NICA | 96 | 92 |
| NDICA | 96 | 73 |

표 1은 각 알고리즘의 최대 인식률을 요소 개수와 함께 종합 비교한 것이다.

V. 결론

본 논문에서는 기존의 독립 요소 해석 방법을 이용한 얼굴 인식 방법을 분석하고 새로운 독립 요소 해석 방법을 제시하였다. 기존의 독립요소 해석법은 건전성이 없고 또한 새롭게 제안한 독립 요소 해석 방법인 커토시스

의 절대값을 이용한 방법은 주 요소를 모두 사용할 경우 노이즈 영향으로 필요한 독립 요소를 효과적으로 선택하지 못한다. 두 방법의 시너지 효과는 주 요소 해석법을 노이즈 필터로 사용하고 독립 요소 해석법을 특징 추출 방법으로 사용할 때 극대화될 수 있다. 향후 연구과제는 비선형 독립 요소 해석의 응용분야에 새 방법론을 적용하는 것이다.

참고문헌

- [1] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71-86, 1991.
- [2] A. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, Vol. 7, No. 6, pp. 1129-1159, 1995
- [3] A. Bell, T. J. Sejnowski, "Independent Components of Natural Scenes are edge filters," *Vision Research*, Vol. 37, No. 23, pp. 3327-3338, 1997
- [4] J. Hurri, A. Hyvärinen, J. Karhuunen, and E. Oja, "Image Feature Extraction Using Independent Component Analysis," *Proc. of IEEE NORSIG'96*, 1996
- [5] M. S. Bartlett, H. M. Lades, and T. J. Sejnowski, "Independent Component Representations for Face Recognition," *Proc. of SPIE on Human Vision and Electronic Imaging III*, Vol. 3299, pp. 528-539, 1998.
- [6] 류재홍, 고재홍, "새로운 독립 요소 해석법," 한국 퍼지 및 지능 시스템학회, 2000년도 추계학술대회 학술 발표 논문집, Vol. 10. No. 2, pp. 305-309, 2000.
- [7] J. Sarela and R. Vigario, "Overlearning in marginal distribution-based ICA: analysis and solutions," *Journal of machine learning research*, Vol. 4,, pp. 1447-1469, 2003.
- [8] S. Amari, "Natural Gradient Learning for Over-and Under-Complete Bases in ICA," *Neural Computation*, Vol. 11, No. 8, pp. 1875-1883, 1999.
- [9] M. Joho, H. Mathis, and R. H. Lambert, "Overdetermined Blind Source Separation: Using More Sensors Than Source Signals In A Noisy Mixture," *Proc. Int. Conf. on Independent Component Analysis and Blind Signal Separation ICA 2000*, pp. 81-86, 2000.

저자소개

류 재 홍(Jae Hung Yoo)



Wayne State University,
전산학과, 박사
현재 전남대학교 공학대학
컴퓨터공학과, 부교수

※관심분야: 컴퓨터그래픽스, 인공신경망, 패턴인식

김 강 철(Kang Chul Kim)



경상대학교 전자공학과, 공학박사
현재 전남대학교 공학대학 컴퓨터
공학과 부교수

※관심분야: VLSI 및 임베디드시스템 설계

임 창 균(Chang-Gyoon Lim)



Wayne State University,
컴퓨터공학과 졸업, 박사
현재 전남대학교 공학대학
컴퓨터공학과, 부교수

※관심분야: 인공지능, 임베디드 소프트웨어, 유비쿼터스 응용