

얼굴 인식을 위한 분류기법에 관한 비교연구

이동훈^o 박영태

경희대학교 전자공학과

dhlee@kvision.khu.ac.kr, ytpark@khu.ac.kr

A comparative study of classification methods for Face Recognition

Donghun Lee^o Youngtae Park

Dept. of Electronic Engineering, Kyunghee University

요 약

본 논문에서는 얼굴 영상의 변화를 보상할 수 있는 전처리 과정으로서 기하학적 특징에 기반한 순수 얼굴 영역 검출 기법을 도입하고 분류기로서 간단한 정합 기법을 사용한 얼굴 인식 기법을 제안한다. 실험결과를 보면 제안한 기법은 NNC, BPN, SVM 분류 기법에 비해 월등한 성능을 가진다.

1. 서 론

얼굴 인식을 위한 주요기법인 PCA+LDA에 의한 mapping 기법에 적용 가능한 세 가지 주요 분류기로서 최근접 이웃 분류기(nearest neighborhood classifiers; NNC)와 오류 역전파 알고리즘이라고도 불리는 역전파(BP) 알고리즘, 그리고 구조적 리스크 최소화법을 통해 특징 벡터 공간에서의 최적의 결정경계영역을 찾아내는 것으로서 이진 분류 문제를 푸는 방법으로 이용되는 SVM(support vector machines)분류기[1]등을 들 수 있다. NNC분류기의 가장 큰 장점은 training과정이 필요없다는 점에 있다. 실제 test 할 데이터가 입력되는 순간 기존에 저장되어 있던 여러 patterns과의 euclidean distance를 비교하는 방법이기 때문이다. 반면에 pattern을 저장하는데 많은 메모리가 요구된다는 단점이 있다. 역전파 알고리즘은 계층적 구조를 가지는 신경망에서 학습 알고리즘으로 효과적으로 사용되고 있다[2]. 그러나 이 역전파 알고리즘은 학습시에 발생하는 문제점이 있는데 Local maxima에 빠져서 학습이 이루어지지 않는 경우가 생기는 것이다. SVM은 1995년에 Vladimir Vapnik에 의해 이원 패턴 인식 문제를 해결하기 위해 제안된 학습방법으로 부정예제로부터 긍정예제를 분류해 낼 수 있는 결정면(decision surface)을 찾아내는 분류모형이다. SVM은 이원패턴 분리를 위한 알고리즘으로 개발되었기 때문에 k 개의 클래스를 학습시키기 위해서는 여러 개의 SVM분류기를 조합하여 다원 분류기로 확장을 하게 된다. 이 목적을 위한 2가지 구조가 있는데, 첫 번째 구조인 승자독식방법(one-against-all strategy)은 가장 간단하고 효율적인 조합방법으로 두 범주의 결정함수를 이용하여 각각의 k 범주에 대해 이원결정함수를 구축함으로써 k 개의 범주로 확장하는 방법이다. 비록 승자독식 방법은 간단하고 효율적인 방법이지만, 경계면을 동시에 만족시킬 수 있는 최적의 다원 범주 결정면(optimal multicategory decision boundaries)를 구축하지 못하기 때문에 애매한 classification을 수행하게 된다[3]. 이러한 단점을 보완하기 위해 두 범주간의 경계선을 직접적으로 처리하는 두 번째 구조인 쌍 단위 분류방

법(one-against-one strategy)이 제안되었다. 쌍 단위 분류방법은 각 쌍마다 분류기를 만드는 방법으로 $\frac{k(k-1)}{2}$ 개의 모든 조합에 대하여 SVM분류기를 학습

시키는 것이다. 본 논문에서는 SVM을 다중클래스 분류기로 확장하기 위하여 두 번째 구조인 쌍 단위 분류 방법(one-against-one strategy)을 채택하였다.

PCA+LDA에 의한 mapping기법과 템플릿 정합 기법[4] 모두 얼굴 영역의 회전, 이동, 표정, 그리고 조명 조건의 변화에 민감한 특성을 가진다. 따라서 신뢰도가 높은 얼굴인식 기법의 구현을 위해선 순수한 얼굴영역의 위치를 정확히 추출하는 전처리 과정이 필수적으로 요구된다.

본 논문에서는 PCA+LDA에 의한 mapping기법에서의 NNC, BPN, SVM 분류기 및 템플릿 정합 기법의 인식 특성을 비교하고 강건한 전처리 과정으로서 기하학적 특징에 기반한 순수 얼굴 검출 기법[5]을 도입한 후 분류기법으로서 간단한 템플릿 정합 기법을 사용하여 월등한 성능의 얼굴인식 알고리즘을 구현할 수 있음을 보인다.

2. NNC, BPN, SVM, 템플릿 정합기법

2.1 NNC

클래스 개수를 N_C 라 하면 0이 아닌 eigenvalue는 최대 N_C-1 개가 존재한다. 따라서 N_C-1 개의 특징벡터에 대해 간단한 nearest neighbor rule을 사용하여 특징 벡터를 분류한다. i 번째 특징벡터를 $F_i = (f_{i1}, f_{i2}, \dots, f_{i, N_C-1})^T$ 로 표기하고 각각의 특징량을 정규화한 특징벡터는

$$\bar{f}_{ij} = (f_{ij} - \mu_j) / \sigma_j \quad (1)$$

로 구성된다. μ_j 와 σ_j 는 각각 j 번째 특징량의 평균과 표준편차를 나타낸다. 클래스 별로 계산된 클러스터 센터를 $\{\eta_1, \eta_2, \dots, \eta_{N_C}\}$ 로 표기하면 입력 얼굴영상은 다음과 같이 분류된다.

$$classification = \arg \min_k \|\bar{F} - \eta_k\| \quad (2)$$

2.2 BPN

역전파 알고리즘은 입력층과 은닉층의 연결강도를 변경시키는 신호로 델타 신호를 사용하는데 델타신호는 연결강도와 출력층에서의 목표치와 출력치와의 오차인 E와의 관계로부터 다음의 식과 같다.

$$\Delta w_{ij} = -\alpha \nabla E = -\alpha \frac{\partial E}{\partial w_{ij}} \quad (3)$$

$$\delta_{yi} \equiv -\frac{\partial E}{\partial (NET_i)} \quad (4)$$

여기서 Δw_{ij} 는 i번째 입력과 j번째 출력사이의 연결강도 변화량이고 α 는 양수이다.

전체 오차 E는 다음과 같이 구할 수 있다.

$$E = E + \frac{1}{2} \|d - y\|^2 \quad (5)$$

여기서 d는 목표치이고 y는 출력치이다.

다층 신경망에서 입력층과 은닉층에 전파되는 오차신호는 다음과 같다.

$$\delta_y = \frac{1}{2} (d - y)(1 - y^2) \quad (6)$$

$$\delta_z = \frac{1}{2} (1 - Z^2) \sum_{i=1}^m \delta_{y_i} w \quad (7)$$

2.3 SVM

2개의 클래스를 $y_i \in \{-1, 1\}$ 로 표기하고, 각 클래스에 속한 training vector인 x_1, x_2, \dots, x_m 가 각각 y_1, y_2, \dots, y_m 에 대응된다면, hyperplane은 다음과 같이 주어진다.

$$(w \cdot x) + b = 0 \quad (w \in R^n, b \in R) \quad (8)$$

training vectors이 margin이 최대이고 error없이 분리되어지는 optimal separating hyperplane(OSH)[6]는 다음의 식을 만족하는 w와 b를 갖는다.

$$\min x_i(w \cdot x_i + b)y_i = 1, i = 1, \dots, l \quad (9)$$

$$(w \cdot x_i + b)y_i \geq 1 \quad (10)$$

training vector x으로부터 OSH와의 거리는 다음과 같다.

$$d(w, b; x) = \frac{|w \cdot x + b|}{|w|} \quad (11)$$

식(11)으로 margin은 $\frac{2}{|w|}$ 임을 알 수 있다. 따라서

OSH는 아래의 식을 최소화시키는 w를 갖는다.

$$\Phi(w) = \frac{1}{2} |w|^2 \quad (12)$$

식(12)은 Lagrange functional에 의해 다음의 식과 같다.

$$L(w, b, \alpha) = \frac{1}{2} w \cdot w - \sum_{i=1}^N \alpha_i [y_i(w \cdot x_i + b) - 1] \quad (13)$$

여기서, α_i 는 Lagrange multipliers이다.

위식은 w, b 에 대해 최소화되어야 하고 음이 아닌 α_i 에 대해 최대화되어야 한다. 위식의 최적조건은 아래의 식과 같다.

$$\frac{\partial}{\partial w} L(w, b, \alpha) = 0, \quad \frac{\partial}{\partial b} L(w, b, \alpha) = 0 \quad (14)$$

식(13)은 아래와 같이 dual problem으로 변환할 수 있다.

$$\max_{\alpha} W(\alpha) = \max_{\alpha} \left\{ \min_{w, b} L(w, b, \alpha) \right\} \quad (15)$$

식(15)에서 $\bar{\alpha}$ 는 다음의 식으로 구할 수 있다.

$$\bar{\alpha} = \arg \min_{\alpha} \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j x_i \cdot x_j \quad (16)$$

여기에서, α_i 는 아래의 식을 만족한다.

$$\alpha_i \geq 0, \quad i = 1, \dots, l, \quad \sum_{i=1}^l \alpha_i y_i = 0 \quad (17)$$

식(16),(17)에서 OSH는 다음과 같이 계산된다.

$$\bar{w} = \sum_{i=1}^l \bar{\alpha}_i y_i x_i \quad (18)$$

$$\bar{b} = -\frac{1}{2} \bar{w} \cdot [x_r + x_s] \quad (19)$$

x_r 과 x_s 는 support vector를 나타내고 아래의 식을 만족한다.

$$\bar{\alpha}_r, \bar{\alpha}_s > 0, \quad y_r = 1, \quad y_s = -1 \quad (20)$$

새로운 벡터 x가 입력될 경우 결정함수는 다음과 같다.

$$f(x) = \text{sign}(\bar{w} \cdot x + \bar{b}) \quad (21)$$

2.4 템플릿 정합기법

얼굴 표정의 다양함을 고려하여 템플릿 영상의 자동 생성기법을 다음과 같이 제안한다. k번째 클래스에 속한 학습영상의 집합을 $I^k = \{I_1^k, I_2^k, \dots, I_n^k\}$ 로 표기한다.

템플릿 생성기법:

Step 1: $m = 1$ (m 은 템플릿 개수를 의미한다.)

Step 2: 모든 화소의 좌표 (x, y) 에 대해 화소값들 $\{I_1^k(x, y), I_2^k(x, y), \dots, I_n^k(x, y)\}$ 의 median 값을 산출하여 m 번째 템플릿 영상 T_m 을 생성한다.

Step 3: T_m 과 각 영상 I^k 의 거리(distance)를 다음과 같이 계산한다.

$$\delta_i = \frac{1}{N_{\Delta}} \min_{\Delta} \sum_{(x,y)} |T_m(x, y) - I_i(x, y + \Delta)| \quad (22)$$

여기서 N_{Δ} 는 T_m 과 I_i 영상에서의 중첩된 영역에 속한 화소수를 의미한다.

Step 4: 템플릿 영상의 생성에 참여한 모든 영상의 δ 값이 Q보다 작다면 템플릿 생성과정을 마친다.

Step 5: 일부 영상의 δ 값이 Q보다 작다면, 이들 영상의 다음 단계의 템플릿 생성과정에서 m 을 증가시키고 Step 2의 과정을 되풀이한다.

Step 6: δ 값이 Q보다 작은 영상이 존재하지 않을 경우, δ 값이 가장 큰 영상으로 T_m 을 대체한다. m 을 증가시키고 Step 2의 과정을 되풀이한다.

식(22)에서는 수직 좌표축의 변화의 가능성을 고려하여 δ_i 의 계산과정에서 수직축으로 일정 구간만큼 이동시키면서 최소의 δ_i 값을 계산한다. 또한 Q 값에 따라 생성된 템플리트 개수가 결정된다. 템플리트 영상의 생성 후, 입력 영상은 거리 δ 가 가장 작은 클래스로 분류된다.

3. 실험결과 및 고찰

3.1 영상 데이터베이스와 전처리 과정

실험에서는 그림 1에 보인바와 같이 AR-face 데이터베이스를 사용하였다. 기하학적 특징에 기반한 순수 얼굴 영역 검출기법을 적용하여 각 영상으로부터 회전과 이동 변화가 보상된 순수 얼굴 영역을 추출하였다. 그림 2에 그 예를 보인다. 조명 변화를 보상하기 위해 DC-notch filter를 사용하였다. 얼굴 영역 검출 기법에 관한 자세한 내용은 [5]에서 다루고 있다.



그림 1. AR-face 데이터베이스의 예



그림 2. 검출된 순수 얼굴영역의 예

3.2 고해상도 영상에서의 실험결과(30x30)

각 클래스마다 10개의 순수 얼굴 영상을 추출하여 30x30 크기로 정규화 하였다. PCA+LDA 기법에서의 NNC, BPN, SVM 분류기 및 템플리트 정합 기법의 성능을 비교하기 위해 leave-one-out 방식을 사용해 예러울을 측정하였고 그 결과를 표 1에 보인다. 템플리트 정합의 성능은 N_c 에 무관하게 99%의 높은 인식률을 보이며 성능이 월등함을 알 수 있다. 이는 전처리 과정에서 얼굴 영상의 변화가 이미 정확히 보상되었기 때문이다.

3.3 영상의 변화에 따른 실험결과

회전, 이동변화에 따른 인식 알고리즘의 강건성을 유추하기 위해, 전처리 과정을 하지 않은 AR-face 데이터베이스로 실험하였다. 그림 1에 대한 실험 결과를 표 2에 제시한다. 실험결과는 템플리트 정합 기법이 PCA+LDA 기법에서의 NNC, BPN, SVM 분류기법에 비해 인식성능이 떨어짐을 보인다. 따라서 표 2에서 제시한 바와 같이 템플리트 정합 기법에서 고 정확도의 얼굴인식 기법을 구현하기 위해서는 영상의 미세한 변화를 보상할 수 있는 전처리 과정의 선택이 필수적이다.

표 1. NNC,SVM,BPN, 템플리트 정합의 Leave-One-Out 예러울 (30x30 영상)

(N_c = Number of classes, N_s = Total number of samples, N_f = Number of features)

N_c	N_s	N_f	NNC	SVM	BPN		Template Matching
					200회 반복	500회 반복	
6	60	5	5(8.3%)	5(8.3%)	1(1.7%)	1(1.7%)	0(0%)
10	100	9	4(4%)	6(6%)	3(3%)	3(3%)	0(0%)
14	140	13	2(1.4%)	1(0.7%)	7(5%)	3(2.1%)	0(0%)
20	200	19	2(1%)	4(2%)	7(3.5%)	5(2.5%)	2(1%)
25	250	24	5(2%)	5(2%)	1(0.4%)	1(0.4%)	2(0.8%)
30	300	29	5(1.7%)	9(3%)	14(4.7%)	12(4%)	2(0.67%)
35	350	34	6(1.7%)	11(3.1%)	13(3.7%)	12(3.4%)	2(0.57%)
40	400	39	7(1.8%)	8(2%)	42(10.5%)	27(6.8%)	2(0.5%)

표 2. NNC,BPN,SVM, 템플리트 정합의 예러울 비교

N_c	N_s	N_f	NNC	SVM	BPN		Template Matching
					200회 반복	500회 반복	
6	60	5	2(3.3%)	4(6.7%)	0(0%)	0(0%)	18(30%)
10	100	9	2(2%)	4(4%)	7(7%)	2(2%)	37(37%)
14	140	13	2(1.4%)	2(1.4%)	10(7.1%)	9(6.4%)	64(45.7%)
20	200	19	5(2.5%)	5(2.5%)	2(1%)	1(0.5%)	104(52%)
25	250	24	8(3.2%)	9(3.6%)	4(1.6%)	3(1.2%)	122(48.8%)
30	300	29	9(3%)	9(3%)	11(3.7%)	8(2.7%)	142(47.3%)
35	350	34	10(2.9%)	8(2.3%)	29(8.3%)	27(7.7%)	169(48.3%)
40	400	39	14(3.5%)	10(2.5%)	26(6.5%)	20(5%)	179(44.8%)

4. 결론

얼굴 인식기법에서 전처리 과정을 도입하지 않은 템플리트 정합 기법은 PCA+LDA mapping 기법에서의 NNC, BPN, SVM 분류 기법보다 얼굴 영역의 미세한 회전과 이동 변화에 민감하게 반응하여 높은 인식율을 유지하기 어려움을 실험을 통해 확인하였다. 따라서 기하학적 특징에 기반한 순수 얼굴영역 검출기법과 같이 회전 및 이동 변화를 보상할 수 있는 강건한 전처리 과정을 도입할 경우, 본 논문에서 제안한 간단한 템플리트 정합 기법이 다른 분류 기법에 비해 월등한 성능을 가짐을 확인하였다.

5. 참고문헌

[1] Colin Campbell, An Introduction to Kernel Methods, Radial Basis Funtion Networks: Design and Applications, Springer Verlag, Berlin, 2000.
 [2] Rumelhart. D. E., Hinton. G. E., & Williams. R. J: Learning representations by error propagation. Parallel distributed processing vol. I, pp. 318-362, 1986.
 [3] M. Pontil and A. Verri. Support vector machines for 3-d object recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence, 20:637-646, 1998.
 [4] R. Brunelli, T. Poggio, Face recognition: Features versus templates, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 15, pp.1042-1052, 1983.
 [5] 이대호, 박영태, 기하학적 특징에 기반한 순수 얼굴영역 검출기법, 한국정보과학회 논문집, 2002.
 [6] V. N. Vapnik. Statistical learning theory. John Wiley & Sons, New York, 1998.