

LITFE를 이용한 얼굴영상 인식

서석배, 이경화, 김영호, 김대진*, 강대성

동아대학교 전기전자컴퓨터공학부

포항공과대학교 컴퓨터공학과*

Face Image Recognition using the LITFE

Seok-Bae Seo, Kyung-Hwa Lee, Young-Ho Kim, Daijin Kim*, Dae-Seong Kang

School of Electrical, Electronic, and Computer Eng., Dong-A Univ.

Dept. of Computer Eng., POSTECH*

sbseo@electro.donga.ac.kr

Abstract

본 논문에서는 얼굴영상의 특징추출에 적합한 LITFE (Linear Interpolated Triangle Feature Extraction)를 이용하여 얼굴영상을 인식하는 알고리즘을 제안한다. LITFE는 얼굴의 위치정보를 보존하면서 영상 분할이 가능한 특징추출 알고리즘으로, PCA (Principal Component Analysis)의 신경회로망적 접근방법인 GHA (Generalized Hebbian Algorithm)와 병행하면 얼굴의 특징을 효과적으로 추출하여 인식할 수 있는 장점이 있다.

1. Introduction

먼저 LITFE로 추출한 특징을 학습시키는 신경망 기반 PCA 알고리즘인 GHA에 대해 알아본 다음, 선형 보간법을 영상에 적용하는 방법을 설명한다^{[1]-[4]}

1.1. GHA (Generalized Hebbian Learning)

GHA는 PCA를 신경회로망을 이용하여 구현한 알고리즘의 한 종류로 순방향 단층구조이며 자율학습의 특성이 있다. 학습된 GHA 웨이트는 가중치 특성이 있으며, 이를 이용하면 입력데이터를 보다 작은 차원으로 축소시킬 수 있다^[5] 응용분야는 패턴 인식을 위한 특징추출, 신경회로망 학습을 위한 입력 데이터 차원 감소 등이며 실제 영상 데이터를 압축에 적용되어 15:1 이상의 높은 압축 알고리즘을 구현할 수 있었다. 본 논문에서는 웨이트들이 가중치에 의해 내림차순 정렬되는 GHA 특성을 이용하여 데이터에 대한 특징을 중요도에 따라 추출한다. GHA를 다음의 3단계로 요약하여 설명한다.

Step 1. 신경회로망의 연결강도 w_{ji} 를 무작위의 작은 실수 값으로 초기화하고 학습률 η 와 반복회수 n 을 설정한다. 여기서 $i=0, 1, \dots, p-1$, $j=0, 1, \dots, m-1$ 로 p 와 m 은 각각 입력노드와 출력노드의 수를 나타낸다.

Step 2. 연결강도 w_{ji} 를 반복회수 n 만큼 식 (1), 식 (2)를 이용하여 학습시킨다.

$$y_j(n) = \sum_{i=0}^{p-1} w_{ji}(n) x_i(n) \quad (1)$$

$$\Delta w_{ji}(n) = \eta [y_j(n) x_i(n) - y_j(n) v(n)] \quad (2)$$

여기서, $v(n) = \sum_{k=0}^{m-1} w_{kj}(n) y_k(n)$ 이다.

Step 3. Step 2에서 구한 연결강도 w_{ji} 를 이용하여 압축을 위한 계수 C 를 생성한다. C 는 식 (3)으로 나타내며 각각의 인자는 식 (4)로 구할 수 있다. 가중치는 j 가 적을수록 높으며, c_0 는 다른 계수에 비하여 매우 큰 값을 지닌다.

$$C = \{c_0, c_1, \dots, c_{m-1}\} \quad (3)$$

$$c_j = \sum_{i=0}^{p-1} w_{ji}(n) x_i(n) \quad (4)$$

최종 수정된 w_{ji} 는 학습입력 x_i 에 대한 고유벡터의 특징이 있으며, 이를 이용하면 입력보다 낮은 차원으로 데이터를 표현할 수 있다. Sanger는 w_{ji} 가 학습입력의 처음 m 개의 주성분 방향으로 수렴함을 증명하였다^[6]. 여기서 m 은 p 보다 작은 정수로 입력벡터의 축소된 차수를 의미한다. 학습이 성공적으로 수행되면 p 차원으로 표현된 학습입력의 직교성 (orthogonal)이 증가되어 낮은 차원으로 학습입력을 효과적으로 표현할 수 있다.

1.2. Linear Interpolation

가장 간단한 예로, 10×8 영상에서 8개의 삼각형을 16개로 늘이는 방법을 설명한다. 그림 1에서처럼 세로의 화소를 2개로 나누되 나머지가 생길 때는 한 화소를 추가한다. 즉 ABCD가 있는 열에서는 AB와 CD로 나누고 합칠 때도 동일한 방법을 적용한다. 그리고 열의 수가 홀수인 EFG 분할의 경우는 $F=F'=F''$, $J=J'=J''$ 로 하고, 원 영상을 복원하기 위해서 다시 합칠 경우 $F=(F'+F'')/2$, $J=(J'+J'')/2$ 로 한다. ABCD와 EFG의 경우처럼, 열의 화소수가 홀수인가 짝수인가에 따라 중복되는 화소가 교번해서 나타나므로, 화소의 수가 증가하더라도 동일한 알고리즘을 적용할 수 있다. 그림 9에서 좌우 양쪽은 TFE에 대한 데이터의 일부분이고, 중앙의 분리된 삼각형은 16개의 삼각특징을 갖는 LITFE의 일부이다. 그래서 스케일의 영상의 경우 A부터 J의 값은 0과 255사이이다^[7].

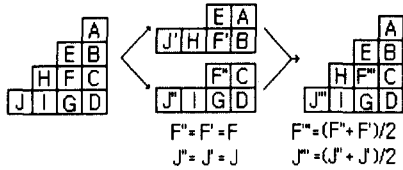


그림 1. 보간법에 의한 삼각형 분할

II. Linear Interpolated Triangular Feature Extraction

TFE는 크기를 재조정된 영상에 대하여 정보를 보존하며 영상을 분할하는데 성공하였지만, 삼각형태가 8개로 제한되어 영상의 세밀한 분할을 할 수 없는 단점이 있다^[1]. 본 논문에서는 그 단점을 보완하기 위해서 TFE에 영상처리의 선형보간법(Linear Interpolation)을 적용한 LITFE (Linear Interpolated Triangle Feature Extraction)를 소개하고 이를 얼굴 영상 압축 및 인식에 이용한다.

그림 2에는 32×30 영상에 LITFE를 적용할 경우의 예로 보간이 일어나는 위치를 어두운 부분으로 나타내었다. 영상의 가로길이를 N_r , 특징추출 시 시작화소를 1이라 할 때, 보간이 일어나는 개수 N_{intp} 와 n 까지 합 S_n 은 식 (5), 식 (6)으로 나타낼 수 있으며, i 번째 보간이 일어나는 화소의 위치 $P_{intp}(i)$ 는 식 (7)에 나타내었다.

$$N_{intp} = \text{round}\left(\frac{N_r}{4}\right) \quad (5)$$

여기서, round 는 반올림연산으로, 만약 $\text{round}(1.5) = 2$ 이다.

$$S_n = \sum_{i=0}^{n-1} (n - i) \quad (6)$$

$$P_{intp}(i) = S_{(N_r/2)} - \sum_{j=1}^i 4(j-1) \quad (7)$$

여기서 $i = 1, 2, \dots, N_{intp}$

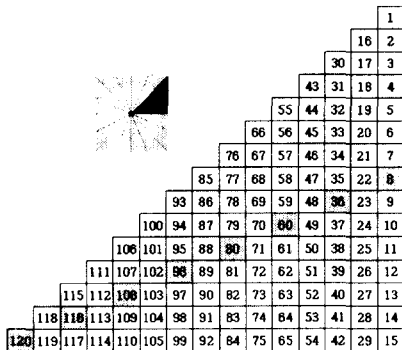


그림 2. 32×30 영상에서 보간의 위치

이상의 식에 의해서 32×30 영상의 삼각특징에서 보간이 일어나는 화소개수 $N_{intp}=8$, 해당 삼각형의 화소수 $S_{15}=120$ 이고 위치는 $P_{intp}(i)$ 는 각각 120, 116, 108, 96, 80, 60, 38, 8 이다. 10×8 영상의 경우는 $N_{intp}=2$, $S_5=10$ 이며, $P_{intp}(1)=10$, $P_{intp}(2)=6$ 임을 알 수 있다. TFE에서 8개의 영역 모두 동일한 데이

터 취득 순서를 가지므로 위의 숫자는 TFE의 8개의 삼각형, 즉 LITFE의 16개의 삼각형에 동일하게 적용된다.

제안하는 LITFE의 성능을 객관적으로 평가하기 위해서 기존의 알고리즘과 비교하는 과정이 필요하다. 본 논문에서는 영상압축 및 인식에 널리 사용되는 블록기반 특징추출 알고리즘과 LITFE를 비교하여 제안하는 알고리즘이 인식뿐만 아니라 압축에도 유용하게 쓰일 수 있다는 것을 증명할 것이다.

블록기반 알고리즘의 경우 가로 세로의 길이가 짝수이고 정방형의 형태가 일반적으로 쓰이고 있다. 따라서 LITFE을 위한 입력영상에서 좌우한 열(column)을 복사해서 추가함으로써 정방형의 영상으로 만든 다음 이를 블록 기반의 입력영상으로 사용한다.

먼저 블록기반특징추출의 경우 전체의 영상을 16개의 동일한 정사각형 영역으로 분할하고, 제안하는 알고리즘은 삼각형의 형태로 16개로 분할한다. 그림 3은 두 특징추출 방법의 형태를 비교한 것으로 임의의 영상에 적용하여 보았다. 두 얼굴영상에서 개수는 16개로 동일하지만 삼각형의 꼭지점으로 갈수록 얼굴 중심정보가 위치하는 LITFE가 비교적 유사한 형태가 있는 균일한 특징을 추출할 수 있음을 볼 수 있다.



그림 3. 실제영상에 적용한 예

III. Experimental Results

3.1. 입력 영상

본 실험에 사용한 입력데이터는 112×92 크기의 그레이 스케일 영상으로, 40명에 대하여 10개의 영상씩 총 400개로 구성되어있다. 실험을 용이하게 하기 위해서 정면 얼굴을 위주로 1인당 2개의 영상을 선택하였고, 크기도 LITFE에 적합하도록 32×30으로 재조정하였다. 영상의 각 화소는 0-255 값을 갖는 그레이 스케일이며, 대부분의 영상에서 얼굴이 중앙에 위치하고 90% 이상의 얼굴이 포함되어 있다.

3.2. Feature extraction

그림 4는 블록기반 특징추출과 제안하는 LITFE를 비교한 결과이다. 블록기반 특징추출의 경우에는 정방형의 사각형 형태로 특징을 추출하지만 제안하는 방법은 영상의 중심을 기준으로 부채꼴 형태로 배열된 삼각형이 특징추출의 형태임을 볼 수 있다.



블록 기반 LITFE

그림 4. 특징추출방법 비교

그림 5는 32×30 크기의 영상에 대하여 식 (7)에서 설명한 보간이 일어나는 화소 $P_{inp}(i)$ 의 위치를 나타낸 결과이다. 그림 5에서 8개의 각 TFE에서 120, 116, 108, 96, 80, 60, 38, 8이 보간이 일어나는 위치임을 알 수 있다.

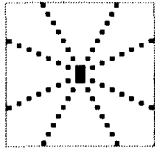


그림 5. 32×30 영상에서 보간의 위치

3.3. Comparison of feature extraction

본 논문에서 제안하는 압축방법의 성능을 평가하기 위해서, 두 특징추출 알고리즘에 K-means 알고리즘을 적용하여 비교하였다. 영상의 크기는 블록기반 특징추출은 32×32로 8×8 크기의 작은 블록으로 분할하였고 TFE는 32×30에 대한 영상을 16개의 삼각형으로 분할하였다. 그림 6은 K=100에서 블록기반 특징추출방법과 TFE에 의한 결과를 타나낸 것이다.

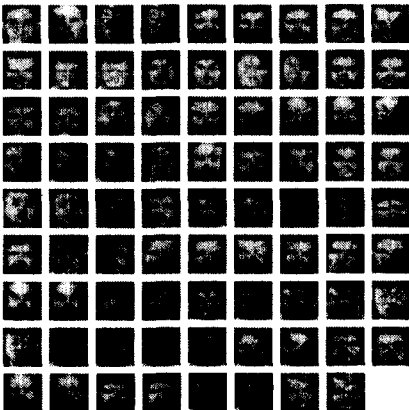
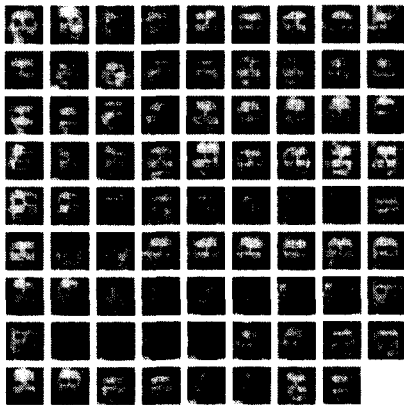


그림 6. 블록기반 특징추출과 LITFE 비교1 (K=100)

표 1은 K=10부터 500까지 증가시키며 LITFE와 블록기반 특징추출의 PSNR을 비교한 결과이다. 표 I에는 80개의 얼굴영상에 대해서 해당하는 K개수로 압축한 다음 그에 대한 평균 PSNR과, 80개 중 최저 PSNR과 최고 PSNR을 각각 표시하였고, 해당하는 그림의 번호를 Note에 나타내었다. 그림에 대한 번호

는 그림 6에서 좌측상단이 00이고, 우측과 아래로 내려가며 번호가 증가하는 순서로 번호를 설정하여 우측하단이 79가 된다. 전반적으로 제안한 알고리즘인 LITFE의 성능이 우수함을 알 수 있다.

표 I. K에 따른 PSNR 비교

Number of K	LITFE			Block based		
	PSNR [dB]	Note	PSNR [dB]	Note		
10	Avg. 18.63		Avg. 18.23			
	Min. 15.87	31	Min. 14.98	31		
	Max. 20.89	08	Max. 20.13	08		
50	Avg. 20.04		Avg. 19.91			
	Min. 18.07	20	Min. 17.41	31		
	Max. 21.93	00	Max. 22.69	08		
100	Avg. 20.71		Avg. 20.66			
	Min. 18.29	20	Min. 17.67	31		
	Max. 24.10	02	Max. 22.78	00		
200	Avg. 21.73		Avg. 21.79			
	Min. 18.80	20	Min. 19.04	20		
	Max. 26.64	11	Max. 25.65	04		
500	Avg. 24.69		Avg. 24.80			
	Min. 21.06	38	Min. 20.97	70		
	Max. 32.95	18	Max. 32.30	20		

그림 7은 21.09[dB]의 동일 PSNR에서 압축결과에 대한 비교이다. 블록기반 특징추출 방법을 이용한 결과(좌)는 원영상(중앙)에 비하여 얼굴의 형태가 많이 왜곡된 반면, 제안한 삼각특징추출인 LITFE에 기반한 결과(우)에서는 그 현상이 대폭 줄었음을 볼 수 있었다. 그리고 블록화 현상 또한 제안한 알고리즘이 훨씬 적음을 알 수 있다. 따라서 얼굴의 형태를 보존하는 관점에서도 기존의 블록기반 특징추출보다 LITFE가 뛰어난을 알 수 있었다.



블록기반 원영상 LITFE
그림 7. 동일 PSNR에서의 성능비교

3.4 GHA의 계수를 이용한 영상정합

본 실험에서 GHA 계수 C 의 p 차 베이스까지 사용하는 정합기준 $Dist_C$ 를 식 (8)로 정의하여, 이 값이 크면 클수록 질의 영상과 다른 영상으로 분류한다. 그리고 입력 영상에 대한 GHA 계수를 C_{query} 라하고, 중요도에 따라 다른 비트로 근사화된 GHA 계수 C 를 C_{approx} 라 정의한다. 식 (9)는 GHA의 계수를 근사시킨 C_{approx} 를 계산하는 식이다. 여기서 $T_{[b_1, b_2]}(n)$ 은 n 값을 근사화 하는 식으로 b_1 은 정수 자리의 비트를 b_2 는 소수자리의 비트를 나타낸다. 만약 n 값이 (-)이면 부호를 나타내어야 하므로, 1 비트를 부호에 할당함으로써 절대값을 가지는 수에 비교해 하나의 비트를 늘여서 C 를 표현해야 한다. b 는 각각의 c 에 할당한 비트 수로 본 실험에서는 부호 비트를 고려하지 않았을 때 최대 8 비트 최소 2 비트이다. 예를 들어 입력 n 이 소수인 경우 $T_{[0.2]}(0.9) = 0.750$, $T_{[0.2]}(-0.9) = -0.750$, $T_{[0.3]}(0.9) = 0.875$ 이며, 정수인 경우는 $T_{[2.0]}(6) =$

3, $T_{\{2,0\}}(-6) = -3$, $T_{\{3,0\}}(6) = 6$ 이 된다. 부호 비트는 전체 비트에 한 비트를 추가하여 부호를 표시하면 된다.

$$Dist_C = f(C_{query}, C_{aprox}) \tag{8}$$

$$= \sum_{i=0}^{k-1} (c_{query_i} - c_{aprox_i})^2$$

$$C_{aprox} = T_{\{b1, b2\}}(C) \tag{9}$$

표 II.는 영상정합을 위하여 각 GHA의 계수에 할당된 비트의 수이다. 할당된 수치의 특별한 의미는 없고 단지 베이스의 중요도에 따라 가변적인 비트를 할당하는 것에 목적이 있다. 즉 작은 번호를 가지는 중요한 베이스에 높은 비트를 상대적으로 중요하지 않는 비트에는 낮은 비트를 할당하였고, 마지막 3개의 베이스는 부호를 고려하지 않고 절대값을 이용하였다. 중요도 순서로 10개, 즉 상위 10개의 베이스만 사용하여도 학습에 참여한 80개 영상에 대해서 100%의 인식률을 보였다.

표 II. GHA 계수의 할당 비트 수

GHA 계수(C)	할당된 비트		
	b1	b2	부호
C1	0	8	1
C12	0	7	1
C13	0	6	1
C14	0	5	1
C15	0	4	1
C6,C17	0	3	1
C8,C9,C10	0	2	0

3.5 제안하는 시스템 : LITFE와 GHA를 이용한 얼굴영상 인식시스템

이상의 결과를 종합하여 제안하는 얼굴영상의 인식을 위한 시스템을 그림 8에 나타내었다. 먼저 모든 영상에 대해서 GHA 계수를 계산한 결과와 영상을 저장해둔 Image 데이터베이스를 생성한다. 다음으로 질의 영상이 들어오면 LITFE에 의해서 16개 또는 8의 배수의 특징을 추출하고 각각에 대한 계수 C를 구한다. 마지막으로 데이터베이스에서 가장 유사한 계수를 찾아 해당하는 영상을 결과영상으로 출력한다.

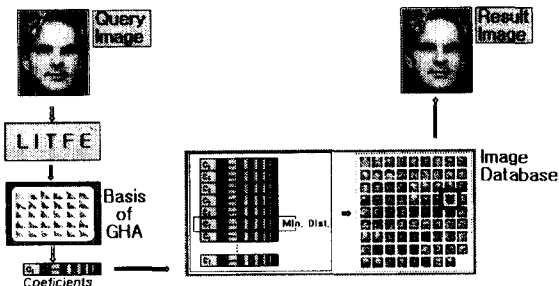


그림 8. 제안하는 영상인식시스템

제안하는 시스템의 특성을 요약하면 검색과 저장의 용이하도록 영상을 분할하여 특징을 추출하면서도 위치정보를 포함하는 LITFE의 장점과 적은 수의 계수로도 영상을 효과적으로 표현할 수 있는 신경회로망에 기반한 PCA인 GHA의 특성을 조합한 것이다.

V. Conclusions

본 논문에서는 얼굴영상 인식을 위한 시스템을 제안하였다. 특징추출 부분에는 영상을 8의 배수로 분할하면서 얼굴에 대한 위치정보와 얼굴의 윤곽을 저장하고 있는 LITFE를 제안하였고, 얼굴의 형태와 이목구비의 정보를 반영할 수 있는 GHA 알고리즘을 도입하여 효과적인 인식시스템을 설계하였다.

본 논문에서 강조하고 싶은 내용은 얼굴영상의 인식을 위한 특별한 시스템에 관한 설계방향이다. 기존의 시스템의 경우 좌우대칭이며 전체적인 형태와 이목구비의 위치정보가 중요한 얼굴영상을 일반 영상과 같이 취급함으로써 효과적인 시스템을 설계하지 못하였다. 한편으로 영상 전체를 데이터로 하여 얼굴의 이목구비와 형태를 압축에 반영하려는 시도는 있었으나, 이 방법은 영상의 크기가 커짐에 따라 늘어나는 베이스의 수에 의하여 특성이 저하되는 치명적인 단점을 극복하지 못하였다. 본 논문에서는 일반적인 영상처리와 얼굴영상 전체를 특징으로 이용하는 방법의 장점만을 취한 TFE를 소개하였고 이를 보완한 LITFE 알고리즘을 제안하였다. 그리고 얼굴을 인식을 위한 효과적인 알고리즘인 GHA 성능을 실험을 통하여 평가하였다.

References

- [1] 서석배, 김대진, 강대성, "삼각특징추출과 GHA를 이용한 얼굴영상 압축알고리즘," 전자공학회논문지 제38권 S편 제1호, 2001.1.
- [2] Kyung-Hwa Lee, Seok-Bae Seo, Daijin Kim, and Dae-Seong Kang, "Face Image Compression using Generalized Hebbian Algorithm of Non-Parsed Image," International Technical Conference on Circuits/Systems, Computers and Communications, Vol. 2, pp 847-850, 2000.
- [3] Oyoung Lee, Hyeyoung Park, and Seungjin Choi, "PCA vs. ICA for Face Recognition," International Technical Conference on Circuits/Systems, Computers and Communications, Vol. 2, pp 873-876, 2000.
- [4] Home page of The AT&T Laboratories at Cambridge <<http://www.cam-orl.co.uk/facedatabase.html>>
- [5] Simon Haykin, "Neural Networks," Prentice Hall, 1999, pp.419-422.
- [6] Sanger, T. D., "Optimal unsupervised learning in a single-layer linear feed-forward neural network," Neural networks 12, pp.459-473, 1989.
- [7] Rafael C. Gonzalez and Richard E. Woods, "Digital Image Processing," Addison Wesley, 1993.