

1차 모멘트와 주요성분분석을 이용한 얼굴표정 인식

Recognizing Facial Expression Using 1-order Moment and Principal Component Analysis

조용현, 홍성준⁰

대구가톨릭대학교 컴퓨터정보통신공학부

Yong-Hyun Cho, Seung-Jun Hong⁰

School of Computer and Information Comm. Eng., Catholic Univ. of Daegu

E-mail : yhcho@cu.ac.kr

요 약

본 논문에서는 영상의 1차 모멘트와 주요성분분석을 이용한 효율적인 얼굴표정 인식방법을 제안하였다. 여기서 1차 모멘트는 영상의 중심이동을 위한 전처리 과정으로 인식에 불필요한 배경의 배제와 계산시간의 감소로 인식성능을 개선하기 위함이다. 또한 주요성분분석은 얼굴 표정의 특징인 고유영상을 추출하는 것으로, 이는 2차의 통계성을 고려한 중복신호의 제거로 인식성능을 개선하기 위함이다. 제안된 방법을 각각 320*243 픽셀의 48개(4명*6장*2그룹) 얼굴표정을 대상으로 Euclidean 분류척도를 이용하여 실험한 결과 전처리를 수행하지 않는 기존 방법보다 우수한 인식성능이 있음을 확인하였다.

1. 서론

사람과 사람의 대면상황 하에서 정보가 전달되는 다양한 채널 중에는 얼굴표정이 큰 부분을 차지한다. 주된 이유는 인간의 내적 정서상태가 주로 얼굴 표정을 통하여 표현되기 때문이다. 이러한 표정인식 시스템 개발을 위해서는 안경, 머리 모양, 표정 등에 의해 수시로 바뀌는 얼굴 모양과 주변 환경의 다양한 변화를 고려해야 한다[1-4].

표정은 주어진 얼굴에 포함된 특징요소들의 변형을 통하여 나타난다. 얼굴의 특징요소들은 표정을 결정하는데 서로 다른 비중을 지니며, 하나의 특징요소도 표정에 따라 상대적인 중요도가 달라질 수 있다. 특히 초기의 얼굴표정 연구는 연구자의 주관적 판단이나 일반 사람들의 평정자료에 근거를 두고 정서상태를 표정으로 표현하는 주안점을 두고 이루어 졌다[4]. 그 중에서도 체계적인 얼굴근육의 움직임을 부호화하여 각 정서상태의 표현에 연결시키는 작업들로서 Ekman과 Friesen이 개발한 Facial Action Coding System(FACS)을 들 수 있다[4]. FACS는 얼굴근육의 움직임을 상세하게 기록한 후 이를 토대로 행복, 놀람, 공포, 분노, 혐오, 슬픔의 6가지 기본정서를 판별할 수 있도록 구성되어 있다.

최근 표정인식을 위한 연구가 여러 가지 방법으로 수행되어져 왔다[4]. 표정인식에 대한 기존 연구

들을 살펴보면 크게 동영상에서의 표정인식과 정지 영상에서의 표정인식 두 가지 범주로 나눌 수 있다. 동영상에서의 표정인식은 주로 움직임 영역을 추출하며, 정지영상에서는 구체적인 특징요소를 추출한다. 이들 얼굴들에 대한 표정인식 방법들에서 model-based 기법들은 얼굴의 명백한 실질적인 모델들을 이용하는 방법이며, feature-based 기법들은 코, 눈썹, 미간 등과 같은 얼굴의 기하학적 특징들을 이용한다. 또한 holistic 기법은 영상 픽셀로부터 직접 표정을 분류하는 방법이다.

표정인식에는 holistic analysis에 바탕을 둔 연구들이 주로 이용되며, 주요성분분석(principal component analysis : PCA)[2]과 이들에 바탕을 둔 기법들이 연구되고 있다[3,4]. 이들 방법들은 얼굴 속에 포함된 진폭위상만을 나타내는 2차원 통계성을 이용한다. 특히 PCA는 얼굴영상을 표현하는 고유벡터를 추출하여 벡터 공간으로 투영시킨 가중치를 이용하는 방식으로, 전체의 얼굴영상들을 적은 수의 벡터를 이용하여 표현할 수 있으며 빠르고, 계산이 간단하며, 점진적 변화에 덜 민감하다.

본 논문에서는 얼굴영상으로부터 구조적 표정특징을 추출하기 위하여 PCA를 이용하고, 전처리로 1차 모멘트[6]에 의한 중심이동을 조합한 기법을 제안한다. 여기서 전처리는 인식에 불필요한 얼굴영역

을 배제시켜 계산량을 줄이고, 아울러 특징추출 시간의 감소를 위함이다. 제안된 기법을 320*243 픽셀의 48개(4명*6표정*2그룹) 얼굴표정[7]을 대상으로 Euclidean 표정 분류척도를 이용하여 분류성능을 비교·검토하였다.

2. 중심이동과 고유영상 추출

일반적으로 영상의 배경은 형상에 비해 비교적 변화가 적다. 그러므로 배경을 제거함으로써 주어진 영상의 크기보다 작은 크기의 유효한 영상을 추출하는 전처리를 수행하면 보다 빠르고 정확한 인식이 가능하게 된다. 표정인식에서도 성능과 효율을 높이기 위해 이러한 전처리 과정을 적용할 수 있다.

여기에서는 영상의 1차 모멘트 평형[6]을 이용하여 유효영상의 중심을 찾기 위하여 1차 모멘트 평형을 이용한 전처리 기법을 제안한다. 또한 전처리된 얼굴의 고유영상을 추출하기 위해 주요성분분석을 이용한다.

2.1 영상의 중심이동

입력되는 일반적인 표정영상을 X축 또는 Y축으로 평행 이동시켜 이루어지는 중심이동은 차 영상의 1차 모멘트 평형[6]에 의해 유효영상의 중심을 계산함으로써 가능하다. 이를 위해 원 영상의 차 영상에 대한 화소의 변화평균값을 구하고, 그 값을 이용하여 유효한 변화의 판별을 위한 논리필터링을 수행한다. 이때 논리필터링은 차 영상의 값이 판별값보다 크거나 같으면 유효한 화소로 1의 값을 갖도록 하며, 그렇지 않으면 0의 값을 갖도록 함으로써 균일한 형상의 영상을 생성한다. 그림 1은 논리필터링 후 1차 모멘트를 구하기 위한 임의의 영상 단면을 나타낸 것이다.

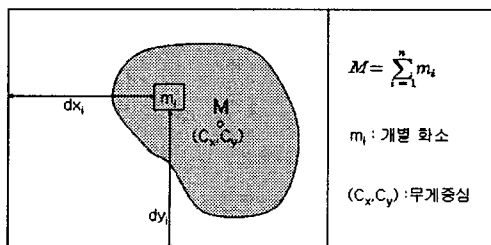


그림 1. 1차 모멘트를 위한 영상단면

여기서 1차 모멘트 평형은 각 개별 화소에 대한 모멘트 값의 합이 영상 전체의 중심거리와 영상 전체의 화소 값의 곱과 동일함을 의미한다[11]. 이때 중심계산을 위한 1차 모멘트 평형을 수식으로 표현하면 다음의 식 (1)과 같다.

$$M \times C_x = \sum_{i=1}^n (m_i \times dx_i), \quad M \times C_y = \sum_{i=1}^n (m_i \times dy_i) \quad (1)$$

여기서 m_i 는 i 번째 픽셀의 논리값으로 유효한 화소이면 1, 그렇지 않으면 0이며, M 은 유효영상의 논리값에 대한 합이다. 이때 유효영상의 중심좌표 (C_x, C_y) 는 식 (2)와 같이 계산된다.

$$C_x = \frac{\sum_{i=1}^n dx_i}{M}, \quad C_y = \frac{\sum_{i=1}^n dy_i}{M} \quad (2)$$

따라서 1차 모멘트 평형으로부터 계산된 유효영상의 좌표 (C_x, C_y) 를 중심으로 보다 작은 영상을 선택하여 입력영상으로 이용한다. 즉, 임의로 입력되는 원 영상에서 계산된 중심좌표로 이동시켜 유효한 영상을 선택하면 인식에 불필요한 배경을 배제시킬 수 있어 차원이 감소된 영상을 얻을 수 있다. 이렇게 하면 영상의 특징인 고유영상의 추출을 위해 이용되는 PCA의 입력 데이터를 줄일 수 있어 인식시간과 인식률을 개선시킬 수 있다.

2.2 PCA에 의한 고유영상 추출

PCA는 입력 데이터의 공분산에 의해 결정되는 정규직교기저로 표현되며, 가우스 변수의 모든 정보는 공분산행렬 내에 포함되기 때문에 가우스 분포의 입력 데이터에 대해서는 적당한 것이다. 일반적으로 영상 내 각 픽셀사이의 2차원 기저영상은 주요한 특징이며, 이의 추출에 PCA를 이용하였다.

자기상관행렬 $R_{xx} = \langle xx^T \rangle$ 를 가진 평균이 영인 입력벡터 $x = [x_1, x_2, \dots, x_n]^T$ 에 대해서 생각해 보자. 여기서 T 는 전치를 나타내며, $\langle \cdot \rangle$ 는 기대치를 나타낸다. 또한 $\hat{w}_1, \hat{w}_2, \dots, \hat{w}_m$ 이 R_{xx} 의 고유벡터와 직교되는 연결가중치 벡터라 할 때, $\hat{w}_1 = [\hat{w}_{11}, \hat{w}_{12}, \dots, \hat{w}_{1n}]^T$ 는 가장 큰 고유치 λ_1 과 일치하며, $\hat{w}_2 = [\hat{w}_{21}, \hat{w}_{22}, \dots, \hat{w}_{2n}]^T$ 는 두 번째로 큰 고유치 λ_2 , 그리고 $\hat{w}_n = [\hat{w}_{n1}, \hat{w}_{n2}, \dots, \hat{w}_{nn}]^T$ 는 가장 작은 고유치 λ_n 과 각각 일치한다. 이상의 관계를 행렬방정식으로 나타내면 식 (3)과 같다.

$$R_{xx}\hat{w}_j = \lambda_j\hat{w}_j, \quad (j=1,2,\dots, n) \quad (3)$$

여기서 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ 이다. 주어진 입력벡터 x 를 위한 첫 번째 m 개의 주요 특징을 나타내는 고유벡터 y 는 다음의 선형변환식 (4)로 나타낼 수 있다.

$$y = \hat{W}x \quad (4)$$

여기서 $\hat{W} = [\hat{w}_1, \hat{w}_2, \dots, \hat{w}_m]^T \in \mathbb{R}^{m \times n}$ 이며, 이 식에서 연결가중치행렬 \hat{W} 의 행은 가장 큰 고유치와 일치하는 상관행렬 R_{xx} 의 고유벡터임을 의미한다.

다시 말하면, 입력데이터 공간에서 k차원의 주요 특징을 나타내는 부공간은 R_{xx} 의 k개 주요 고유벡터에 의해 구성된 부공간으로 정의된다. 결국 PCA는 $\langle \|\hat{w}_j^T x\|_2^2 \rangle$ 가 최대인 고유벡터 $\hat{w}_1, \hat{w}_2, \dots, \hat{w}_k$ 의 방향을 찾는 것이다. 일반적으로 얻어지는 고유값은 크기에 따라 정렬하고 고유벡터도 해당 고유값의 위치대로 정렬한다. PCA에서 순서대로 정렬된 고유값의 뒤쪽은 0에 가까운 값을 가지게 되어 이를 삭제할 수 있다. 이는 고유벡터의 작은 값들을 고려하지 않음으로써 입력 데이터의 차원을 줄이기 위함이다.

한편 PCA를 좀 더 효과적으로 수행하기 위해 신호의 영 평균(zero-mean)을 수행한다[5]. 이는 신호의 1차적 통계성을 고려한 정규화로 PCA를 간략화하기 위함이다. 신호의 영 평균은 신호벡터 x 에서 평균값 x^* 를 뺀 차로 $x = x - x^*$ 이다.

따라서 중심이동의 전처리된 얼굴영상을 대상으로 PCA를 수행하면 2차의 고유영상을 추출할 수 있다. 중심이동과 PCA를 조합한 제안된 얼굴인식 과정을 요약하면 그림 2와 같다. 학습과정에서 추출된 고유벡터는 얼굴의 특징벡터로 2차원의 속성을 가지는 기저영상이며 시험영상의 판별을 위해 저장된다. 한편 식별과정에서 유사성 판별을 위한 거리 척도로는 city-block, Euclidean, negative angle들이 이용된다. 본 논문에서는 Euclidean의 척도를 이용하였다.

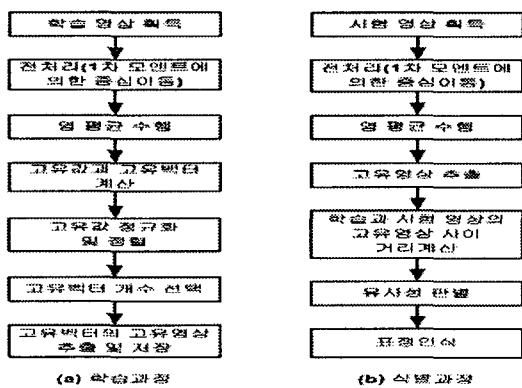


그림 2. 제안된 얼굴영상 인식과정

3. 실험 및 결과분석

중심이동과 PCA에 의한 고유영상을 이용한 제안된 표정인식 방법의 성능을 평가하기 위해 320*243 픽셀의 48개(4명*6표정*2그룹) 영상을 대상으로 실험하였다. 그림 3은 4명을 대상으로 좌에서 우로 각

각 행복, 보통, 슬픔, 놀람, 링크, 졸음의 6가지 표정을 가지는 학습과 시험영상들이다. 여기서 시험영상은 학습영상에 비해 얼굴의 크기가 변하든지 또는 위치가 변화된 영상들이다. 실험은 펜티엄IV-3.0G 컴퓨터에서 Matlab 6.5로 구현하였다.

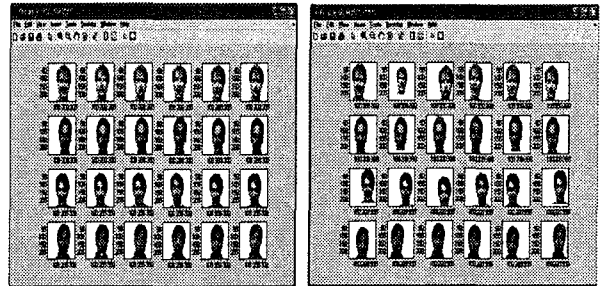


그림 3. 4명의 6가지 학습 및 시험 표정영상

한편 그림 4는 학습과 시험영상을 대상으로 1차 모멘트를 구하여 중심 이동시킨 후 배경을 삭제한 전처리된 영상들이다. 여기서 전처리된 영상은 원래 77,760(320*243) 픽셀에서 19,520(160*122) 픽셀로 약 3.98배정도 차원을 감소시킨 영상들이다. 이렇게 하면 불필요한 배경의 제거에 따른 계산시간의 감소와 인식성능의 개선을 얻을 수 있다.

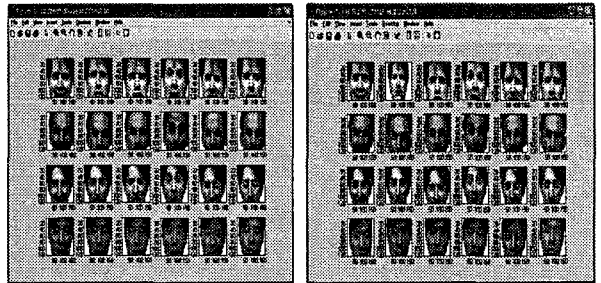


그림 4. 전처리된 학습 및 시험 표정영상

그림 5는 전처리된 24개의 학습영상을 대상으로 주요성분분석을 이용하여 추출된 24개의 고유영상과 이들로부터 복원된 학습영상을 나타낸 것이다. 이때 복원된 영상과 전처리된 학습영상과의 전체오차는 약 0.56정도로 아주 유사함을 알 수 있다.

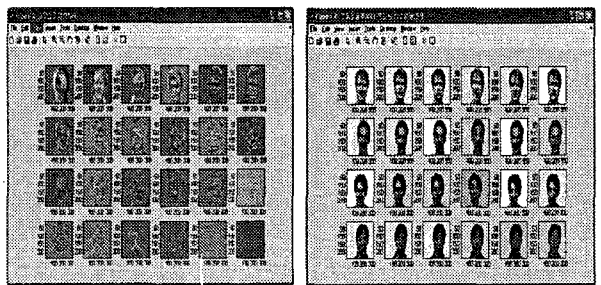


그림 5. 전처리된 학습영상의 고유 및 복원 영상

<표 1>은 고유영상을 24개로 하였을 때, 학습표정 24개와 시험표정 24개를 대상으로 거리척도에 따른 인식결과를 각각 나타낸 것이다. 기존의 방법은 중심이동의 전처리되지 않은 얼굴영상을 대상으로 PCA를 행한 것이고, 제안된 방법은 중심이동의 전처리된 얼굴표정을 대상으로 PCA를 행한 것이다. 여기서 학습표정과 시험표정 사이의 유사성을 측정하기 위한 척도로 Euclidean의 척도를 이용하였다. 또한 오차평균은 올바르게 인식된 영상을 대상으로 학습표정의 기저영상과 시험표정의 기저영상 사이의 오차의 절대치 평균이다. 표에서 보면 2가지 방법에서 학습된 표정영상은 모두 100%의 인식률을 보이나, 시험 표정영상은 각각 약 8.33%와 83.3%의 인식률을 보인다. 따라서 중심이동의 제안된 방법은 기존방법에 비해 평균 약 10배 정도 개선된 인식률을 가진다. 또한 24개의 학습표정과 24개의 시험표정에 대하여 제안된 방법은 기존방법에 비해 각각 오차평균에서 평균적으로 약 3.46배와 약 29배 정도의 작은 값을 가진다. 이는 중심이동의 제안된 방법은 그렇지 않은 기존방법에 비해 인식을 위한 식별과정에서 좀 더 정확하게 얼굴을 분류함을 보여 준 것이다. 오차평균의 개선정도에서 보면 학습표정보다 시험표정에서 더욱 더 크게 나타나 전처리의 우수성을 재확인할 수 있다. 한편 24개의 시험표정에서 기존방법은 22개의 표정이 오인식되었으며, 오인식의 결과는 3-놀람(16번)과 4-행복(19번)의 학습표정이 각각 10회와 5회 나타났다. 하지만 제안된 방법에서는 각각 2-졸음(12번)과 4-웁크(23) 학습표정이 각각 3회와 1회 나타났다. 따라서 16번과 19번 및 12번의 학습표정은 다른 표정에 비해 상대적으로 많은 공통특징을 가진 것으로 추측된다. 한편 기존방법의 경우 중앙에서 좌우로 위치가 이동되거나 영상의 크기가 변한 것에 대해 매우 심각한 영향을 받으나 제안된 방법은 상대적으로 이들에 대해 강건한 인식성능이 있음을 알 수 있다. 결국 중심이동의 전처리는 크기나 위치의 변화에도 매우 우수한 인식성능이 있음을 확인할 수 있다.

<표 1> 24개의 학습 및 시험표정에 대한 인식결과

거리 척도	입력 영상	기존 방법 (PCA)			제안된 방법 (전처리+PCA)		
		인식률 (%)	오차 평균	오인식 영상	인식률 (%)	오차 평균	오인식 영상
Euclidean	학습 표정	100	4.95e-012	-	100	1.43e-012	-
	시험 표정	8.33	1.16e+04	(3),(15), (16),(19), (21),(24)	83.3	0.04e+01	8,9,11(12), 23(22)

* i(j) : i = 인식되어야 할 학습얼굴번호,
j = 오인식된 학습 얼굴번호

3. 결론

본 논문에서는 영상의 중심이동과 고유영상을 이용한 효율적인 얼굴표정 인식방법을 제안하였다. 여기서 중심이동은 얼굴영상의 1차 모멘트로부터 얻어지는 전처리 과정으로 인식에 불필요한 배경의 배제와 계산시간의 감소로 인식성능을 개선하기 위함이다. 또한 고유영상은 얼굴표정의 특징으로 주요 성분분석을 이용하여 추출하였으며, 이는 2차의 통계성을 고려한 중복신호의 제거로 인식성능을 개선하기 위함이다.

제안된 방법을 각각 320*243 픽셀의 48개(4명*6 표정*2그룹) 얼굴표정을 대상으로 Euclidean 분류척도를 이용하여 실험한 결과, 우수한 인식성능이 있음을 확인하였다. 제안된 방법은 전처리를 수행하지 않은 기존방법에 비해 인식률과 인식 정확성에서 개선된 성능이 있음을 확인하였다.

향후 제안된 방법을 좀 더 큰 규모의 인식문제와 다양한 분야에 적용하는 연구가 지속적으로 이루어져야 할 것이다.

5. 참고문헌

[1] J. T. Tou and R. C. Gonzalez, 'Pattern Recognition Principles,' Addison-Wesley Pub., London, 1974

[2] B. A. Draper, K. Baek, M. S. Barlett, and J. R. Beveridge, "Recognizing Faces with PCA and ICA," Computer Vision and Imaging Understanding 91, pp.115-137, 2003

[3] S. H. Jeng, H. Y. M. Liao, C. C. Han, M. Y. Chern, and Y. T. Liu, "Facial Feature Detection Using Geometrical Face Model: An Efficient Approach", Pattern Recognition, Vol. 31, No. 3, pp.273-282, 1998

[4] Y. L. Tian, T. Kanada, and J. F. Cohn, "Recognizing Action Units for Facial Expression Analysis," IEEE. Trans. on Pattern Analysis and Machine Intelligence, ,Vol.23, No.2. pp. 1-19, Feb. 2001

[5] K. I. Diamantaras and S. Y. Kung, Principal Component Neural Networks : Theory and Applications, Adaptive and learning Systems for Signal Processing, Communications, and Control, John Wiley & Sons, Inc., 1996

[6] 김상철역, '재료역학' 청문출판사, 1992년 4월

[7] "Yale Face Databases," <http://cvc.vale.edu/projects/valefaces/valefaces.html>