

# 동적얼굴영상으로부터 감정인식에 관한 연구

이명원\*, 궤근창\*\*

\*조선대학교 제어계측공학과

\*\*조선대학교 제어계측로봇공학과

e-mail: [mailsana@naver.com](mailto:mailsana@naver.com), [kwak@chosun.ac.kr](mailto:kwak@chosun.ac.kr)

## A Study on Emotion Recognition from a Active Face Images

Myung-Won Lee\*, Keun-Chang Kwak\*\*

\*Dept of Control and Instrumentation Engineering, Chosun University

\*\*Dept of Control, Instrumentation, and Robot Eng., Chosun University

### 요 약

본 논문에서는 동적얼굴영상으로부터 감정인식을 위해 벡터 표현 보다는 직접적인 텐서 표현으로 특징들을 추출하는 텐서 기반 다선형 주성분분석(MPCA: Multilinear Principal Component Analysis) 기법을 사용한다. 사람 6가지의 얼굴 표정을 사용하는데 한 사람의 각 표정마다 5프레임으로 묶어서 텐서 형태로 취하여 특징을 추출하고 인식한다. 시스템의 성능 평가는 CNU 얼굴 감정인식 데이터베이스를 이용하여 특징점 개수와 성능척도에 따른 실험을 수행하여 제시된 방법의 유용성에 관해 살펴본다.

### 1. 서론

감정인식에 대한 연구들은 많은 연구자들에 의해서 이루어지고 있다. 감정에 대한 인식방법은 음성을 이용하거나 생체신호를 이용, 표정을 이용하는 방법이 있다.

음성에서 감정 특징(Tone, Pitch, Formant Frequency, 말의 빠르기)을 추출하여 추출된 특징을 이용하여 화난 상태, 기쁜 상태 혹은 다른 상태 등을 알 수 있게 연구되고 있다[1].

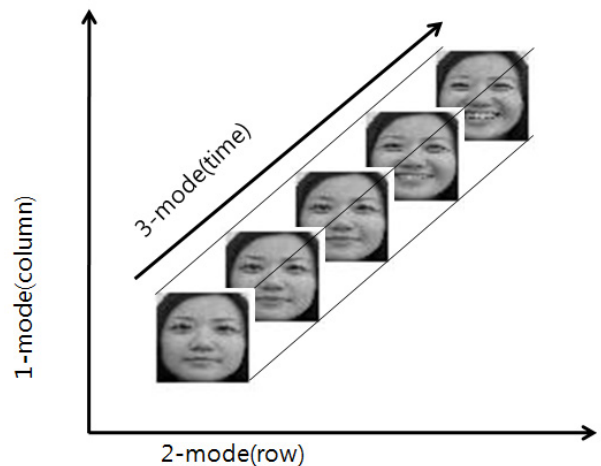
생체 신호는 근전도(EMG: Electromyography)센서나 피부온도, 심전도(ECG: Electrocardiogram), 호흡센서 등을 이용할 수 있다. 하지만 얼굴이나 몸에 부착을 해야만 한다는 단점이 있다[2].

얼굴 표정은 감정을 전달함에 있어서 중요한 역할을 하는 생체인식 분야 중 하나로, 연구 방법으로는 크게 모델 기반과 영상 기반 방법으로 분류할 수 있다[3]. 모델 기반 방법은 얼굴의 특징점의 위치나 움직임 등을 기반으로 하는 방법이고 영상 기반 방법은 본래 영상의 화소 값을 이용하는 방법이다. 기존의 많은 방법들은 2차원 영상을 고차원 벡터로 표현하여 특징을 추출한다[4-10]. 그러나 본 논문에서는 직접적인 텐서 표현, 즉 그레이스케일 이미지의 행, 열, 시간을 포함하여 특징들로 추출하는 3차 텐서 기반 다선형 주성분 분석기법(Multilinear Principal Component Analysis) 기법[11]을 이용한 동영상기반 얼굴인식을 수행한다.

본 논문의 구성은 다음과 같다. 2장에서는 MPCA에 대해서 설명하고 3장에서는 MPCA기법을 이용한 얼굴 표정 인식에 대한 성능평가를 보여주고 4장에서는 결론을 맺는다.

### 2. 텐서 기반 다선형 주성분 분석

본 논문에서 제안된 표정 인식에 대한 입력은 그레이스케일 이미지의 행과 열 그리고 시간 축을 포함한 3차원 텐서이다. 그림 1은 제안된 3차 텐서를 시각적으로 나타내고 있다. 그림에서 보는 바와 같이 일반적인 표정인식 기법과는 달리, 시간 축을 통하여 하나의 표정을 하나의 스냅 샷에서 결론짓지 않고 여러 프레임을 하나의 표정으로 인식하게 한다. 이를 위해 특징들을 추출하기 위해 텐서 기반 다선형 주성분 분석 기법을 이용한다. 다선형 주성분



(그림 1) 3차원 텐서 표현

분석 기법은 주성분 분석 기법의 다선형 확장이며 Lu[11]에 의해 걸음세 인식에 성공적으로 적용되었다. 본 논문

서는 이 알고리즘을 동영상 표정인식에 사용하였다.

MPCA 알고리즘을 간단히 설명하면 다음과 같다.

텐서들은 전형적으로  $A$ 로 표시하고, 텐서의 요소는  $N$ 에 의해 표시된다.  $N$ 은 텐서 대상의 차수를 나타내며 각 지수는 하나의 모드를 정의한다. 텐서의 요소는 괄호에서 지수를 가지고 표시되며,  $N$ 차 텐서는  $A \in R^{I_1 \times I_2 \times \dots \times I_N}$ 이다. 그것은  $N$ 개의 지수들  $i_n$ 에 의해 표시되고, 각  $i_n$ 은  $A$ 의  $n$ -모드를 할당한다.  $A \times_n U$ 에서 행렬  $U \in R^{I_n \times I_n}$ 에 의한 텐서  $A$ 의  $n$ -모드 곱은 다음과 같은 텐서로 표현되어진다.

$$(A \times_n U)(i_1, \dots, i_{n-1}, j_n, j_{n+1}, i_{n+1}, \dots, i_N) = \sum_{i_n} A(i_1, \dots, i_N) \cdot U(j_n, i_n) \quad (1)$$

위 식에 근거해서, 텐서는 다음 식과 같이  $N$ 개의 투영행렬  $U^{(1)}, U^{(2)}, \dots, U^{(N)}$ 에 의해 다른 텐서로 투영되어질 수 있다.

$$Y = X \times_1 U^{(1)T} \times_2 \dots \times_N U^{(N)T} \quad (2)$$

$n$ -모드 투영행렬  $U^{(n)T}$ 에 의한  $X$ 의  $n$ -모드 벡터의 투영은  $n$ -모드 벡터와  $U^{(n)T}$ 의 행 사이의 내적으로써 계산되어진다. MPCA는 다음 식과 같이 텐서기반 스캐터 척도를 최대화하도록 한다.

$$\Psi_Y = \sum_{m=1}^M \|Y_m - \bar{Y}\|_F^2 \quad (3)$$

$$\bar{Y} = \frac{1}{M} \sum_{m=1}^M Y_m \quad (4)$$

이 문제는  $N$ 개의 선형 최적화 문제로 분해되어져서 해결되어질 수 있다. 모든 다른 투영행렬이 주어졌을 때, 행렬  $\tilde{U}^{(n)}$ 은 행렬의 가장 큰  $P_n$ 개의 고유치에 대응하는 고유벡터들로 구성하는  $\Psi_Y$ 를 최대화한다. 얻어진 투영행렬  $\tilde{U}^{(n)}$ 은  $\prod_{n=1}^N P_n$ 로써 보여질 수 있다. 그러나, 그것들의 모두가 인식을 위해 유용하지는 않기 때문에 클래스 분별력을 통해 선택되어질 수 있다. 클래스 분별력에 의해 순서화된 특징벡터는 다음과 같이 계산되어진다[11].

$$Y_m = X_m \times_1 \tilde{U}^{(1)T} \times_2 \tilde{U}^{(2)T} \dots \times_N \tilde{U}^{(N)T} \quad (5)$$

위 식으로부터 학습영상과 검증영상의 특징벡터들이 구해지고 아래의 식으로부터 맨하탄 거리, 유클리디안 거리, 코사인 유사도에 의해 얼굴 감정 성능을 비교한다.

$$d(X, Y) = \sum_{i=1}^n |x_i - y_i| \quad (6)$$

$$d(X, Y) = \|x - y\|^2 = \sum_{i=1}^n (x_i - y_i)^2 \quad (7)$$

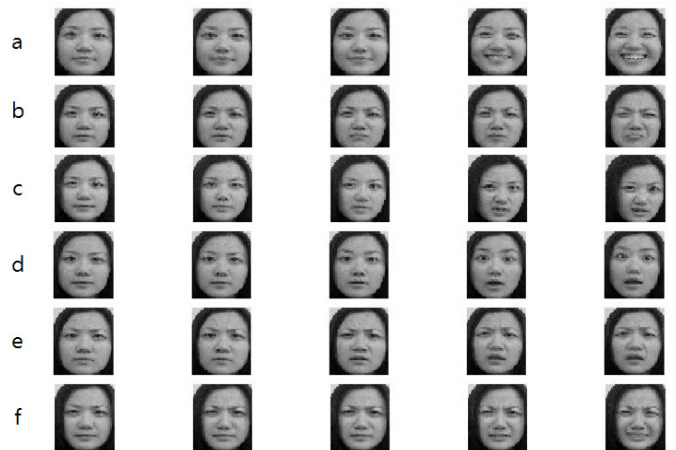
$$d(X, Y) = \cos\theta = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (8)$$

### 3. 실험 및 결과

본 논문에서 제시한 텐서기반 다선형 주성분 기법의 성능을 평가하기 위해, CNU 얼굴감정인식 데이터베이스를 사용하였다[6]. 이 데이터베이스는 카메라로부터 얼굴 동영상을 촬영하고 하나의 표정에 대해 15프레임씩 구성되어 있고 기쁨, 슬픔, 화남, 놀람, 혐오, 공포와 같은 6가지 감정 표정을 10명씩 구축하였다. 이 중에서 학습은 5프레임씩 180장(900장)을 사용하였고 검증도 5프레임씩 180장(900장)을 사용하였다. 그림 2는 10사람의 6개의 표정을 나타낸다. 그림 3은 한 사람에게 대한 6가지의 표정이다. a, b, c, d, e, f 별로 하나의 텐서, 즉 하나의 특징이 5장의 프레임이 모여서 이루어진다.



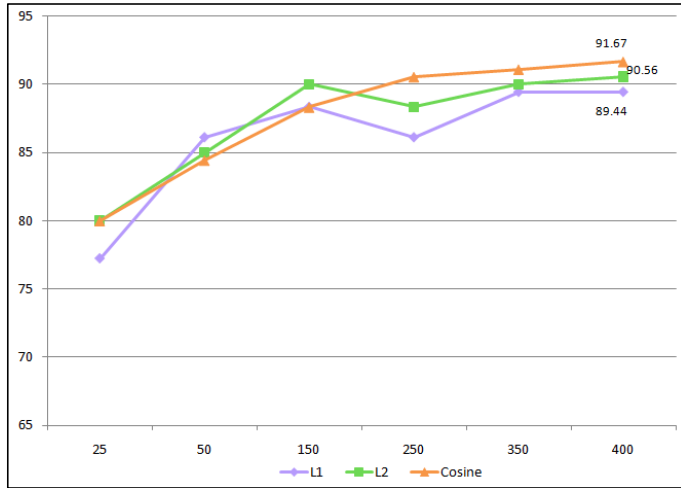
(그림 2) 얼굴 표정 데이터베이스



(그림 3) 한사람에 대한 6가지 표정

<표 1> 검증 결과

P의 갯수	L1	L2	Cosine
25	77.22	80	80
50	86.11	85	84.44
150	88.33	90	88.33
250	86.11	88.33	90.56
350	89.44	90	91.11
400	89.44	90.56	91.67



(그림 4) 특징점의 개수에 따른 인식률

표 1은 맨하탄 거리방법(L1), 유클리디안 거리방법(L2)과 코사인 유사도를 통하여 검증에 대한 인식률을 특징점 (P)의 개수에 따라 분류한 것이다. 그 결과 특징점의 개수가 400개일 때 가장 좋은 인식률을 보여주며, 그림 4는 그 결과를 한눈에 보이기 위해 차트로 보여준다.

특징점의 개수를 50단위로 계속 올려서 실험을 해보았지만 특징점의 개수가 400일 때보다 미치지 못한 실험 결과를 보여주어서 그 결과는 표와 그림에서 생략한다.

#### 4. 결론

논문은 텐서 기반의 다선형 주성분 분석기법을 이용한 동영상 기반 표정인식에 대해 다루었다. 유클리디안 거리와 코사인 유사도 방법을 통하여 특징점의 개수가 400일 때 가장 좋은 인식률을 가져왔다. 차후 다른 방법과 비교하여 벡터 특징들을 가지고 하는 것과 텐서 형태의 특징을 가지고 있을 때의 효율성 비교를 하겠다.

#### 감사의 글

본 연구는 교육과학기술부와 한국연구재단의 지역혁신인력양성사업으로 수행된 연구결과임.

#### 참고문헌

- [1] 심귀보, 박창현, “음성인식으로 감정인식 요소 분석,” 퍼지 및 지능 시스템학회 논문지, 제 11권, 제 11호, pp. 510-515, 2001.
- [2] 황세희, 박창현, 심귀보, “감성 인식을 위한 생체 정보 수집,” 한국지능시스템학회 학술발표 논문집, pp. 353-356, 2005.
- [3] B. Fasel, Juergen Luetttin, “Automatic facial expression analysis: a survey,” Pattern Recognition, vol. 36, pp. 259-275, 2003.
- [4] 주종태, 박경진, 고희은, 양현창, 심귀보, “얼굴표정을 이용한 감정인식 및 표현 기법,” 한국지능시스템학회 학술발표 논문집, pp. 295-298, 2007.
- [5] 한수정, 곽근창, 고현주, 김승석, 전명근, “ICA-factorial 표현법을 이용한 얼굴감정인식,” 한국퍼지 및 지능 시스템학회 논문지, 제13권, 제3호, pp. 371-376, 2003.
- [6] 이대중, 이경아, 고현주, 전명근, “정지영상과 동영상의 융합모델에 의한 얼굴 감정인식,” 한국지능시스템학회 논문지, 제15권, 제5호, pp. 573-580, 2005.
- [7] 강효석, 백재호, 김은태, 박민용, “감정확률을 이용한 동적 얼굴표정의 퍼지 모델링,” 한국지능시스템학회 논문지, 제19권, 제1호, pp. 1-5, 2009.
- [8] M. Turk and A. Pentland, “Face recognition using eigenfaces,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 586-591, 1991.
- [9] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, “Eigenfaces vs fisherface: Recognition using class specific linear projection,” IEEE Trans. Pattern Anal. Mach. Intell. vol. 19, no. 7, pp. 711-720, 1997.
- [10] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, “Face recognition by independent component analysis,” IEEE Trans. on Neural Networks, vol. 13, no. 6, pp. 1450-1464, 2002.
- [11] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, “MPCA: Multilinear principal component analysis of tensor objects,” IEEE Trans. on Neural Networks, vol. 19, no. 1, pp. 18-39, 2008.