

Feature Extraction Method of 2D-DCT for Facial Expression Recognition

Dong-Ju Kim[†] · Sang-Heon Lee[†] · Myoung-Kyu Sohn[†]

ABSTRACT

This paper devises a facial expression recognition method robust to overfitting using 2D-DCT and EHMM algorithm. In particular, this paper achieves enhanced recognition performance by setting up a large window size for 2D-DCT feature extraction and extracting the observation vectors of EHMM. The experimental results on the CK facial expression database and the JAFFE facial expression database showed that the facial expression recognition accuracy was improved according as window size is large. Also, the proposed method revealed the recognition accuracy of 87.79% and showed enhanced recognition performance ranging from 46.01% to 50.05% in comparison to previous approaches based on histogram feature, when CK database is employed for training and JAFFE database is used to test the recognition accuracy.

Keywords : Facial Expression Recognition, Overfitting, Hidden Markov Model

얼굴 표정인식을 위한 2D-DCT 특징추출 방법

김 동 주[†] · 이 상 현[†] · 손 명 규[†]

요 약

본 논문에서는 2D-DCT와 EHMM 알고리즘을 이용하여 과적합에 강인한 얼굴 표정인식 방법을 고안하였다. 특히, 본 논문에서는 2D-DCT 특징추출을 위한 윈도우 크기를 크게 설정하여 EHMM의 관측벡터를 추출함으로써, 표정인식 성능 향상을 도모하였다. 제안 방법의 성능평가는 공인 CK 데이터베이스와 JAFFE 데이터베이스를 이용하여 수행되었고, 실험 결과로부터 특징추출 윈도우의 크기가 커질수록 표정 인식률이 향상됨을 확인하였다. 또한, CK 데이터베이스를 이용하여 표정 모델을 생성하고 JAFFE 데이터베이스 전체 샘플을 테스트한 결과, 제안 방법은 87.79%의 높은 인식률을 보였으며, 기존의 히스토그램 특징 기반의 표정인식 접근법보다 46.01~50.05%의 향상된 인식률을 보였다.

키워드 : 얼굴표정인식, 과적합, 은닉 마르코프 모델

1. 서 론

얼굴 표정인식 접근 방법은 크게 모델 기반과 영상 기반의 방식으로 나눌 수 있다[1]. 대표적인 모델 기반 방식의 예로는 AAM(Active Appearance Models)를 이용하여 이로부터 얻어진 특징점들의 움직임을 분석하여 인식하는 방법이 있다[2]. 영상 기반 표정인식 방법은 얼굴 표정의 특징 변화를 추출하기 위하여 얼굴 영상에 필터 또는 필터뱅크를 적용하여 수행하는 방식으로 대표적인 예로는 PCA (Principal Component Analysis)[3], ICA(Independent Component Analysis)[4],

가버 웨이블릿(Gabor Wavelet)[5]을 이용한 분석 방법이 있다. 더불어, 최근의 영상기반 표정인식 방법은 조명과 회전 에 강인한 LBP(Local Binary Pattern)[6] 및 LDP(Local Directional Pattern)[7] 변환의 블록 히스토그램 특징 벡터와 SVM(Support Vector Machine) 분류기를 이용한 접근 방법이 연구되고 있다.

기존의 다양한 표정인식 접근 방법과는 다르게, 본 논문에서는 얼굴인식 분야에 성공적으로 적용되었던 2D-DCT (Two-Dimensional Discrete Cosine Transform)를 이용한 특징추출과 표정인식을 위해 통계적 모델링 기법인 EHMM (Embedded Hidden Markov Model)을 사용하였다. 일반적으로, 기존의 2D-DCT와 EHMM을 이용한 얼굴인식 분야에서는 개개인을 구분할 수 있는 유효한 얼굴 특징을 추출하기 위하여 2D-DCT 윈도우 크기를 8×8 또는 16×16을 사용한다. 또한, 얼굴인식과 유사하게 윈도우 크기를 16×16으로 설정하여 LBP 특징을 추출하고, EHMM 알고리즘을 이용하여

※ 본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2013년도 문화콘텐츠 산업기술 지원사업의 연구결과와 교육과학기술부에서 지원하는 대구경북 과학기술원 기관고유사업에 의해 수행되었음(13-IT-03).

† 정 회 원: 대구경북과학기술원 IT융합연구부 선임연구원

논문접수: 2013년 9월 23일

수정일: 1차 2013년 11월 5일

심사완료: 2013년 11월 6일

* Corresponding Author: Sang-Heon Lee(pobylee@dgist.ac.kr)

표정 인식을 수행한 기존 연구가 있다[8]. 하지만, 표정인식에서는 개개인을 구분하기 위한 국부 영역에서의 얼굴 특징이 아닌, 각 표정을 잘 표현 할 수 있는 전역 범위에서의 특징 추출이 요구된다. 이러한 아이디어를 바탕으로, 본 논문에서는 각 표정별 특징추출 과정에서 얼굴인식에서 사용한 윈도우 크기보다 더 큰 윈도우를 적용하여 특징 벡터를 추출하고, 통계적 모델링 기법을 이용하여 표정을 분류하는 방법을 제안한다.

2. 표정인식

2.1 특징 추출

표정인식에 사용하는 특징 파라미터는 EHMM의 입력으로 사용되는 관측 벡터로, 본 논문에서는 각 표정에 대한 전역 특성을 표현하기 위하여 2D-DCT의 저차 계수를 특징 파라미터로 사용하였다. $P \times L$ 크기의 영상에 대한 2D-DCT는 Equation (1)과 같이 표현된다.

$$F(u,v) = \frac{2}{\sqrt{PL}} C(u)C(v) \left[\sum_{i=0}^{P-1} \sum_{j=0}^{L-1} f(i,j) \times \cos \frac{(2i+1)u\pi}{2P} \times \cos \frac{(2j+1)v\pi}{2L} \right] \quad (1)$$

$$\text{단, } C(u), C(v) = \begin{cases} \frac{1}{\sqrt{2}} & \text{for } u,v=0 \\ 1 & \text{for } u,v \neq 0 \end{cases}$$

$f(i,j)$: $P \times L$ 영상

$F(u,v)$: 2D-DCT 계수

너비 W 와 높이 H 인 얼굴 영상은 특징벡터 추출을 위하여 $P \times L$ 크기의 블록으로 나누어진다. P 와 L 은 영상의 폭과 높이에 대한 블록 크기로서, 2D-DCT를 계산하는 윈도우 크기이다. 영상의 폭과 높이에 대한 중첩의 크기를 각각 Q 와 M 이라고 하면, 얼굴 영상으로부터 추출되는 블록의 개수 T 는 Equation (2)와 같이 계산된다.

$$T = \left(\frac{W-Q}{P-Q} \right) \times \left(\frac{H-M}{L-M} \right) \quad (2)$$

표정인식에 사용하는 관측벡터 열의 개수는 Equation (2)로 계산되는 블록의 개수와 같으며, 다음과 같은 방법으로 계산된다. $P \times L$ 윈도우는 얼굴 영상의 왼쪽에서 오른쪽으로, 위에서 아래방향으로 스캔한다. 여기에서 이웃한 윈도우 사이의 중첩은 수평방향으로 Q , 수직방향으로 M 만큼 이동하고, $P \times L$ 윈도우를 2D-DCT 변환하여 관측벡터 열을 생성한다.

2.2 표정 모델 생성 및 인식

본 논문에서는 표정인식을 위한 알고리즘으로 EHMM을 사용하였다. EHMM은 일차원의 HMM을 2차원 구조로 나타내기 위하여 일반화한 방법으로, super-states와 embedded-states의 집합으로 구성된다. 여기에서 각각의 super-state는 하나의 일차원 HMM을 포함하며, 관측확률이 없고 상태전이 확률만을 갖는다. 다음은 EHMM의 구성요소를 나타낸다[9].

1) Super-state 모델 구성

- N_0 : super-state의 개수
- Π_0 : super-state 초기상태 분포
- A_0 : super-state 상태 전이 확률 매트릭스

2) Embedded-state 모델 구성

- N_1^k : k 번째 super-state의 embedded-state 개수
- Π_1^k : k 번째 super-state의 embedded-state에 대한 초기 상태 분포
- A_1^k : k 번째 super-state의 embedded-state에 대한 상태 전이 확률 매트릭스
- B_1^k : k 번째 super-state의 embedded-state에 대한 관측 확률 매트릭스

연속 EHMM에서 embedded-state의 관측확률은 Equation (3)과 같이 표현된다.

$$b_i^{(k)}(O_{t_0 t_1}) = \sum_{m=1}^M c_{im}^{(k)} N(O_{t_0 t_1}, \mu_{im}^{(k)}, U_{im}^{(k)}), \quad 1 \leq i \leq N_1^k \quad (3)$$

Equation (3)에서 $O_{t_0 t_1}$ 는 얼굴영상의 가로축 관측지점 t_0 와 세로축 관측지점 t_1 에서의 관측 벡터이다. M 은 mixture 개수이며, $c_{im}^{(k)}$ 는 k 번째 super-state의 i 번째 embedded-state에서 m 번째 mixture의 계수이다. $N(O_{t_0 t_1}, \mu_{im}^{(k)}, U_{im}^{(k)})$ 는 평균 벡터 $\mu_{im}^{(k)}$ 와 공분산 매트릭스 $U_{im}^{(k)}$ 을 갖는 가우시안 확률밀도 함수를 나타낸다.

표정 모델 학습은 처음으로 2D-DCT 관측벡터를 추출하고, 추출한 관측벡터를 세로축과 가로축 방향으로 균일하게 분할하여 EHMM 모델 파라미터를 초기화한다. 초기화된 모델 파라미터들은 doubly embedded Viterbi segmentation 알고리즘과 segmental K-means 알고리즘을 이용하여 재 추정되고, 모델 파라미터들이 수렴할 때까지 재 추정 과정을 반복한다. 인식 과정에서는 입력 영상에 대한 관측벡터와 doubly embedded Viterbi 알고리즘을 이용하여 각 표정별 모델에 대한 확률값을 계산하고, 가장 높은 확률값을 갖는 모델을 선택하여 표정을 인식하게 된다.

3. 실험 및 결과

제안하는 표정인식 방법의 성능평가는 CK(Cohn-Kanade AU-Coded Facial Expression) 데이터베이스[10]와 JAFFE (Japanese Female Facial Expression) 데이터베이스[11]를 이용하여 수행되었다. CK 데이터베이스는 100명의 학생에 대한 7가지 표정(무표정, 기쁨, 화남, 놀람, 공포, 슬픔, 혐오)을 갖는 이미지 시퀀스로 구성되어 있으며, 이중에서 320개의 시퀀스를 실험에 사용하였다. JAFFE 데이터베이스는 213장의 일본인 여성 모델 10명으로 구성되어 있으며, 한 사람에 대하여 7가지 표정(무표정, 기쁨, 화남, 놀람, 공포, 슬픔, 혐오) 이미지로 구성되어 있다. Fig. 1은 110×150 크기로 정규화 된, CK 데이터베이스와 JAFFE 데이터베이스의 샘플 영상 예시를 나타내고 있다.

제안 방법의 성능 비교는 최신의 영상기반 표정인식 방법인 LBP와 SVM을 이용한 방법[6]과 LDP와 SVM을 이용한

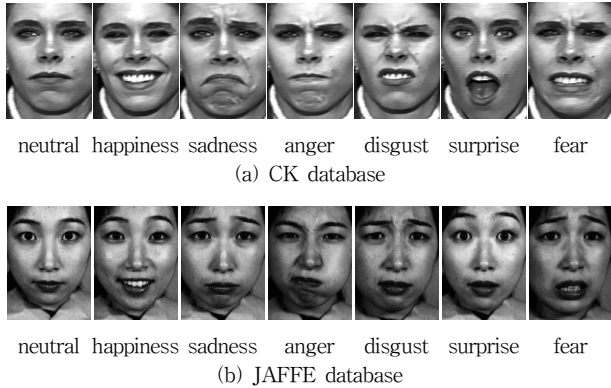


Fig. 1. Sample facial expression images

방법[7]을 이용하여 실험을 수행하였다. 이와 같이 이진패턴을 이용한 일반적인 표정인식 접근 방법은 Fig. 2와 같은 과정으로 수행된다. 처음으로 얼굴 영상은 LBP 또는 LDP 연산을 이용하여 이진패턴으로 변환되고 구분된 몇 개의 영상 블록에 대한 히스토그램을 추출하여 이들을 연결한 벡터를 표정을 구분하기 위한 특징으로 사용한다. 다음으로, 추출한 히스토그램 특징 벡터는 SVM 분류기를 통하여 표정을 인식한다.

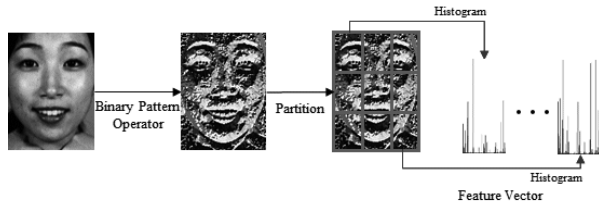


Fig. 2. Feature extraction procedures of histogram-based facial expression recognition approaches

CK 데이터베이스와 JAFFE 데이터베이스 상에서, 이와 같은 최신의 표정인식 방법을 사용한 실험 결과를 Table 1과 Table 2에 각각 나타내었다. 단, 학습과 테스트 데이터셋을 랜덤 샘플링을 통하여 구분하는 기존 연구[6][7]와는 다르게, 본 논문에서는 개인별 표정 영상을 순차적으로 선택하여 데이터 셋을 구성하였다. 이러한 방법을 통하여 본 실험에서는 데이터 셋을 5개의 그룹으로 나누고, 한 개 그룹을 테스트에, 나머지 4개 그룹을 학습에 사용하는 5-fold 교차 검증을 수행하여 인식 결과를 도출하였다. 더불어, 표정인식 결과는 LBP와 LDP 연산에 따라 구분하였으며, 히스토그램 특징 벡터 추출에 사용한 서브 블록의 개수에 따라 인식 결과를 구분하여 나타내었다. 실험결과, CK 데이터베이스의 경우, LBP는 76.35%, LDP는 74.80%의 최대 인식률을 보였으며, JAFFE 데이터베이스의 경우에는 LBP가 61.39%, LDP가 60.56%의 최대 인식률을 보였다. 이로부터, 7×6 블록으로 나누어 특징을 추출한 경우가 3×3 및 5×5 블록에서 특징을 추출한 경우보다 더 좋은 인식 결과를 보임을 확인하였다.

전역 범위에서의 특징추출 효과를 검증하기 위하여, 본 논문에서는 2D-DCT 특징추출 윈도우 크기를 가변적으로 변화하면서 표정인식 결과를 조사하였다. 특징 추출 윈도우 크기는 가로와 세로의 범위를 각각 8부터 32픽셀까지 변화하

면서 인식 결과를 조사하였으며, CK 데이터베이스와 JAFFE 데이터베이스를 이용한 실험 결과를 Table 3과 Table 4에 각각 나타내었다. 실험 결과, 2D-DCT 윈도우 크기가 커질수록 인식률도 높아지는 것을 알 수 있으며, 특히, 너비가 32이고 높이가 24일 경우, CK는 73.51%, JAFFE는 66.69%의 가장 높은 인식률을 보였다. 비록, CK 데이터베이스의 경우에 제안 방법이 기존 히스토그램 특징을 이용한 방법보다 2.84% 낮은 인식률을 보였지만, JAFFE 데이터베이스 경우에는 5.30%의 더 좋은 인식 결과를 보였다.

Table 1. Recognition results of previous approaches for CK database

Feature descriptor	Number of regions		
	3×3	5×5	7×6
LBP	69.94%	74.25%	76.35%
LDP	66.12%	74.02%	74.80%

Table 2. Recognition results of previous approaches for JAFFE database

Feature descriptor	Number of regions		
	3×3	5×5	7×6
LBP	51.96%	60.47%	61.39%
LDP	51.96%	56.72%	60.56%

Table 3. Recognition results by different window size for CK database

Height \ Width	8	16	24	32
	8	58.66%	66.21%	67.18%
16	63.98%	68.69%	69.37%	71.02%
24	65.39%	69.45%	71.71%	73.51%
32	67.82%	69.43%	72.39%	73.36%

Table 4. IV Recognition results by different window size for JAFFE database

Height \ Width	8	16	24	32
	8	47.55%	56.79%	57.52%
16	52.73%	56.82%	59.04%	60.50%
24	52.60%	62.50%	63.33%	66.69%
32	55.46%	61.49%	59.11%	62.95%

다음으로, 제안 방법의 과적합 문제에 대한 강인함을 검증하기 위하여, 본 논문에서는 두 데이터베이스를 동시에 이용한 성능 비교 실험을 수행하였다. 일반적으로, 기존의 히스토그램 특징과 SVM 분류기를 이용한 표정인식 방법은 학습 과정에 사용한 데이터베이스에 정합된 모델을 생성하게 되고, 이로 인해 상이한 데이터베이스를 테스트에 이용하였을 때, 과적합 문제가 발생하게 된다. 제안 방법이 과적합에 강한 방법인지의 여부를 검증하기 위하여 본 논문에서는 CK 데이터베이스의 전체 샘플을 학습에 사용하고, JAFFE 데이터베이스의 전체 샘플을 테스트에 사용하는 실험을 수행하였다. 실험 결과, 기존의 LBP 및 LDP 연산과 SVM을 사용한 표정인식 방법에 대한 인식 결과를 Table 5에 나타내었으며, 제안하는 표정인식 방법의 인식 결과를 Fig. 3에 나타내었다. Table 5로부터 LBP와 SVM을 이용한 방법은 41.78%, LDP와 SVM을 이용한 방법은 37.74%의 최대 인식률을 보임을 알 수 있었으며, 이러한 결과는 CK 데이터베이스

Table 5. Recognition results of previous approaches

Feature descriptor	Number of regions		
	3×3	5×5	7×6
LBP	24.88%	38.96%	41.78%
LDP	36.62%	38.49%	37.74%

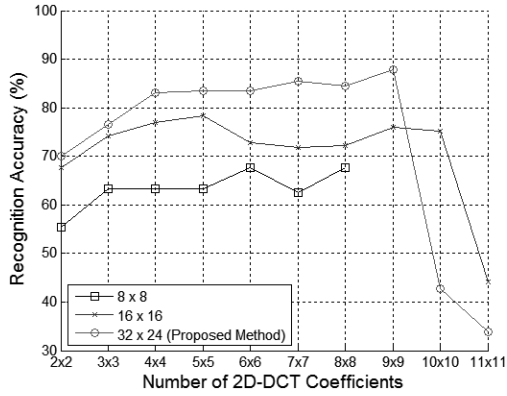


Fig. 3. Recognition results of proposed approach

스를 학습에 사용하고 JAFFE 데이터베이스를 테스트에 이용하여 실험한 인용논문 [6]의 결과인 41.3% 인식률과 비슷한 성능임을 알 수 있다. 반면, 본 논문에서 제안하는 방법은 2D-DCT 윈도우 크기가 32×24이고 저차원 계수의 크기가 9×9일 경우, 87.79%의 인식률을 보였다. 결론적으로, 제안 방법은 기존의 히스토그램 기반의 방법보다 46.01~50.05%의 성능향상을 보였으며, 이로부터 제안방법이 과적합에 강한 표정인식 접근법임을 확인하였다. 더불어, (Fig. 3)의 실험에서 윈도우 크기가 8×8, 16×16일 경우, 8×8은 67.61%, 16×16은 78.28%의 최대 인식률을 보였으며, 이로부터 32×24의 윈도우를 사용하는 제안방법이 기존 윈도우 크기를 사용한 경우보다 9.51~20.18%의 성능 향상이 나타나 는 것을 확인 할 수 있었다.

4. 결 론

본 논문에서는 과적합에 따른 인식 성능 저하의 문제점을 해결하고자, 2D-DCT를 관측벡터로 사용하는 EHMM 기반의 표정인식 방법을 제안하였다. CK 데이터베이스 전체 샘플을 학습 과정에 이용하고 JAFFE 데이터베이스 전체 샘플을 테스트에 이용한 성능평가 실험에서 제안방법은 87.79%의 인식률을 보였으며, 또한, 기존의 표정인식 접근 방법보다 46.01%~50.05%의 인식 성능 향상 결과로부터, 제안 방법이 과적합에 강한 표정인식 방법임을 확인하였다.

Reference

[1] B. Fasel and J. Luetin, "Automatic facial expression analysis: a survey," Pattern Recognition, Vol.36, pp.259-275, 2003.
 [2] B. Abboud, F. Davoine, and M. Dang, "Facial expression recognition and synthesis based on an appearance model," Signal Processing: Image Communication, Vol.19, No.8, pp.723-740, 2004.

[3] M. Turk and A. Pentland, "Eigenfaces for recognition," Journal of Cognitive Neurosci, Vol.3, No.1, pp.71-86, 1991.
 [4] M. Bartlett, H. Lades, and T. Sejnowski, "Independent component representations for face recognition," In Human Vision and Electronic Imaging III, Vol.3299, pp.528-539, 1998.
 [5] T. S. Lee, "Image representation using 2D Gabor wavelets," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.18, No.10, pp.959-971, 1996.
 [6] C. Shan, S. Gong, and P. W. McOwan, "Facial Expression Recognition based on Local Binary Patterns: A Comprehensive Study," Image and Vision Computing, Vol.27, No.6, pp.803-816, 2009.
 [7] T. Jabid, M. H. Kabir, and O. Chae, "Robust Facial Expression Recognition based on Local Directional Pattern," ETRI Journal, Vol.32, No.5, pp.784-794, 2010.
 [8] J. Cao and C. Tong, "Facial expression recognition based on LBP-EHMM," 2008 Congress on Image and Signal Processing, pp.371-375, 2008.
 [9] A. Nefian and M. Hayes, "An Embedded HMM-based Approach for Face Detection and Recognition," Proc. IEEE Int. Conf. On Acoustics, Speech and Signal Processing, Vol.6, 1999.
 [10] T. Kanade, J. Cohn, and Y. Tian, "Comprehensive Database for Facial Expression Analysis," IEEE Int. Conf. Autom. Face Gesture Recog., pp.46-53, 2000.
 [11] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic Classification of Single Facial images," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.21, No.12, pp.357-1362, 1999.



김 동 주

e-mail : radioguy@dgist.ac.kr
 1998년 충북대학교 전파공학과(학사)
 2000년 충북대학교 전파공학과(석사)
 2010년 성균관대학교 정보통신공학부(박사)
 2011년~현 재 대구경북과학기술원
 IT융합연구부 선임연구원

관심분야: 얼굴인식, HCI, 영상신호처리, 패턴인식



이 상 헌

e-mail : pobylee@dgist.ac.kr
 1993년 경북대학교 전자공학과(학사)
 1996년 경북대학교 전자공학과(석사)
 2013년 경북대학교 전기전자공학과(박사)
 2005년~현 재 대구경북과학기술원
 IT융합연구부 선임연구원

관심분야: 컴퓨터비전, 영상신호처리

손 명 규



e-mail : smk@dgist.ac.kr
 1997년 경북대학교 전자공학과(학사)
 1999년 서울대학교 전기공학과(석사)
 2005년~현 재 대구경북과학기술원
 IT융합연구부 선임연구원

관심분야: HCI, 패턴인식, 컴퓨터비전