

Facial Expression Recognition Using SIFT Descriptor

Dong-Ju Kim[†] · Sang-Heon Lee^{**} · Myoung-Kyu Sohn^{**}

ABSTRACT

This paper proposed a facial expression recognition approach using SIFT feature and SVM classifier. The SIFT was generally employed as feature descriptor at key-points in object recognition fields. However, this paper applied the SIFT descriptor as feature vector for facial expression recognition. In this paper, the facial feature was extracted by applying SIFT descriptor at each sub-block image without key-point detection procedure, and the facial expression recognition was performed using SVM classifier. The performance evaluation was carried out through comparison with binary pattern feature-based approaches such as LBP and LDP, and the CK facial expression database and the JAFFE facial expression database were used in the experiments. From the experimental results, the proposed method using SIFT descriptor showed performance improvements of 6.06% and 3.87% compared to previous approaches for CK database and JAFFE database, respectively.

Keywords : Facial Expression Recognition, Scale Invariant Feature Transform, Support Vector Machine

SIFT 기술자를 이용한 얼굴 표정인식

김 동 주[†] · 이 상 현^{**} · 손 명 규^{**}

요 약

본 논문에서는 SIFT 기술자를 이용한 얼굴 특징과 SVM 분류기로 표정인식을 수행하는 방법에 대하여 제안한다. 기존 SIFT 기술자는 물체 인식 분야에 있어 키포인트 검출 후, 검출된 키포인트에 대한 특징 기술자로써 주로 사용되나, 본 논문에서는 SIFT 기술자를 얼굴 표정인식의 특징벡터로써 적용하였다. 표정인식을 위한 특징은 키포인트 검출 과정 없이 얼굴영상을 서브 블록 영상으로 나누고 각 서브 블록 영상에 SIFT 기술자를 적용하여 계산되며, 표정분류는 SVM 알고리즘으로 수행된다. 성능평가는 기존의 LBP 및 LDP와 같은 이진패턴 특징기반의 표정인식 방법과 비교 수행되었으며, 실험에는 공인 CK 데이터베이스와 JAFFE 데이터베이스를 사용하였다. 실험결과, SIFT 기술자를 이용한 제안방법은 기존방법보다 CK 데이터베이스에서 6.06%의 향상된 인식결과를 보였으며, JAFFE 데이터베이스에서는 3.87%의 성능향상을 보였다.

키워드 : 얼굴표정인식, SIFT, SVM

1. 서 론

표정은 인간의 감정을 전달할 수 있는 중요한 수단으로 얼굴을 이용한 표정인식은 인간의 감정 상태를 알아낼 수 있는 효과적인 방법 중 하나이다. 표정은 인간의 감정을 표현하는 수단이므로 몇 개의 표정만으로 다양한 감정을 표현하기는 어렵다. 이에, Ekman[1]은 인간의 감정을 놀람, 화남, 공포, 혐오, 슬픔, 기쁨의 6가지 기본 정서로 구분하였으며, 이후 얼굴을 이용한 표정인식은 이러한 6가지 표정을 구분하는 문제를 다

루고 있다. 일반적으로, 얼굴 표정인식은 얼굴검출, 특징추출, 그리고 표정분류의 단계로 수행된다. 표정인식의 첫 단계인 얼굴영역 검출 단계는 얼굴의 크기가 개인별, 표정별로 달라지므로, 정밀한 얼굴영역 검출이 요구되며, 본 논문에서는 ASM (Active Shape Model)[2, 3]을 이용하여 얼굴 영역을 획득하였다. 특징추출 단계는 표정인식에 중요한 역할을 차지하며, 본 논문에서는 기존의 객체 인식, 중복성 검출 및 객체 추적에 사용되는 SIFT(Scale Invariant Feature Transform)[4, 5] 기술자를 이용하여 특징벡터를 생성하고, 이를 SVM(Support Vector Machine)[3, 6] 분류기로 표정인식을 수행하는 방법에 대하여 제안한다. 제안 방법의 성능평가는 CK(Cohn-Kanade AU-Coded Facial Expression) 표정 데이터베이스[7]와 JAFFE (Japanese Female Facial Expression) 표정 데이터베이스[8] 상에서 기존의 LBP(Local Binary Pattern)[9, 10] 및 LDP (Local Directional Pattern) [11, 12] 등의 이진패턴 특징을 이용한 표정인식 방법과 성능비교를 수행하였다.

※ 본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2015년도 문화콘텐츠산업 기술 지원사업의 연구결과와 교육과학기술부에서 지원하는 대구경북과학기술원 기관교육사업에 의해 수행되었습니다(16-IT-03).

† 정 회 원 : 동서대학교 컴퓨터공학부 조교수

** 정 회 원 : 대구경북과학기술원 IT융합연구부 선임연구원

Manuscript Received: June 16, 2015

First Revision: September 8, 2015

Accepted: September 9, 2015

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

2. 제안 방법

2.1 얼굴 정규화

본 논문에서는 표정인식을 위한 정밀한 얼굴영역 획득을 위하여 ASM 특징점(Landmark)을 이용하였다. 일반적으로, ASM은 통계적인 모델을 이용하여 특징점을 추출하는 방법으로 특징점 검출 속도가 빠르고, 형태 모델의 여러 가지 변형을 잘 표현할 수 있으며, 조명변화에 덜 민감한 방법으로 알려져 있다[2, 3]. ASM에서 형태 모델을 만드는 학습 과정은 학습 데이터로부터 라벨링된 특징점들을 이용하여 평균 형태를 구한다. 이러한 형태 모델은 Equation (1)로 표현된다.

$$x \approx \bar{x} + Pb \tag{1}$$

여기서 \bar{x} 는 평균 형태 벡터이고, P 는 형태의 공분산에 대한 고유벡터 열로 구성된 행렬이며, b 는 형태 파라미터를 원소로 하여 구성된 벡터이다. 학습된 형태 모델을 이용하여 특징점을 추출하는 과정은 초기의 모양을 형태 모델과 입력 영상을 이용하여 조정한다. 다음으로 b 를 구하여 제한 조건 범위 밖에 있으면 재조정 한 후, b 의 모든 값들이 제한조건 범위를 만족할 때까지 반복하여 입력영상에 대한 형태를 구한다. Fig. 1은 ASM을 이용하여 검출한 얼굴 특징점과 특징점의 위치 정보를 이용하여 검출된 정밀 얼굴영역을 보여준다.

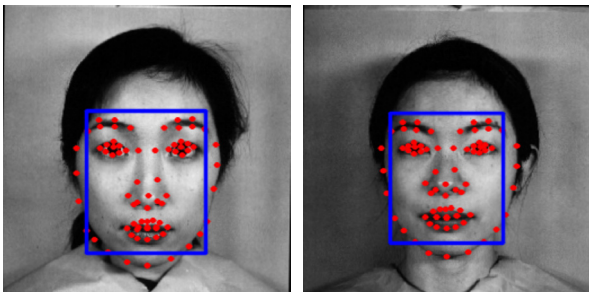


Fig. 1. ASM-based face normalization

2.2 표정 인식

본 논문에서는 표정인식을 위한 특징으로 SIFT 기술자를 사용하였다. SIFT 알고리즘은 David Lowe[4, 5]에 의해 처음 제안되었으며, 영상 회전, 스케일 변화, 조명 변화에 매우 강인성을 가진 특징 추출 알고리즘이다. 일반적으로 객체 인식, 중복성 검출 및 객체 추적 등의 분야에 주요 특징으로 사용되는 SIFT 특징 추출은 (1) 스케일 공간 극값 검출 (Scale Space Extrema Detection), (2) 키포인트 지역화(Key Point Localization), (3) 방향성 배치(Orientation Assignment) 와 (4) 기술자 생성(Description Generation) 등의 단계로 수행된다. 물체인식에 있어 코너와 같은 구분되는 특징점 탐지는 스케일 공간에서의 극점 탐색과 잡음 극점 제거를 통하여 의미있는 키포인트를 검출하고, 검출된 키포인트에 대

하여 SIFT 기술자를 이용하여 특징벡터를 추출한다.

그러나, 본 논문에서는 객체 인식 및 정합 등을 위한 키포인트 검출 과정을 수행하지 않고, SIFT 기술자만을 얼굴 영상에 적용하여, 이를 표정인식 특징벡터로 사용하였다. 즉, Fig. 2와 같이 너비 W 와 높이 H 인 영상은 특징 추출을 위하여 $N \times N$ 크기의 블록으로 나누어진다. 영상의 폭과 높이에 대한 중첩의 크기를 O 라 하면, 얼굴영상으로부터 추출되는 블록의 개수 T 는 Equation (2)와 같이 계산된다.

$$T = \left(\frac{W-O}{N-O}\right) \times \left(\frac{H-O}{N-O}\right) \tag{2}$$

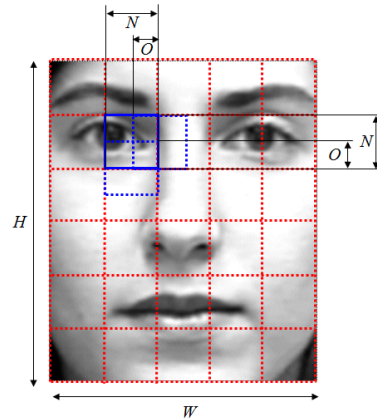


Fig. 2. Block extraction in a face image

다음으로, 얼굴영상에 대한 특징벡터는 $N \times N$ 크기의 각 서브 블록마다 SIFT 기술자를 적용하고, 모든 블록의 특징을 결합함으로써 구성된다. 한 서브 블록에 대한 SIFT 특징 추출의 첫 과정은 Equation (3)으로 표현되는 블러링 영상 $L(x, y)$ 을 생성하는 것이다. 여기서, $I(x, y)$ 와 $G(x, y, \sigma)$ 는 각각 얼굴영상과 가우시안 함수를 의미한다. 다음으로 SIFT 특징 기술자는 그레디언트(Gradient) 벡터의 크기(Magnitude)와 방향(Orientation)을 Equation (5)와 (6)으로 각각 계산하고, 한 서브 블록을 다수의 그리드 영역으로 나누어, 모든 영역에서 8 방향의 빈(Bin)을 갖는 방향 히스토그램을 계산한다. 이러한 방향 히스토그램을 계산을 모든 서브 블록에 대하여 수행하고, 이를 연결한 벡터를 전체 영상에 대한 SIFT 특징으로 사용한다.

$$L(x, y) = G(x, y, \sigma) * I(x, y) \tag{3}$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{4}$$

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \tag{5}$$

$$\theta(x, y) = \tan^{-1} \left(\frac{L(x+1, y) - L(x-1, y)}{L(x, y+1) - L(x, y-1)} \right) \tag{6}$$

Fig. 3은 서브 블록의 크기 N 이 8일 경우, SIFT 기술자에 의한 방향 히스토그램 계산 방법의 일례를 나타낸다. 한 서브 블록은 4×4 픽셀 크기의 2×2 그리드로 구분되고, 각 그리드에 대하여 방향 히스토그램을 계산한다. 다음으로 2×2 그리드에 대한 방향 히스토그램을 연결하여, 한 서브 블록에 대하여 $32(2 \times 2 \times 8\text{bin})$ 차원을 갖는 벡터를 생성한다. 최종적으로 전체 얼굴영상에 대한 SIFT 특징벡터는 $T \times 32$ 차원을 갖는다. 일반적으로 객체인식 및 정합 분야에서 SIFT 기술자는 N 을 16을 사용하고, 한 서브 블록을 4×4 그리드로 나누어 히스토그램을 계산함으로써 $128(4 \times 4 \times 8\text{bin})$ 차원을 갖는 특징을 사용한다. 그러나 본 논문에서는 서브 블록의 크기 N 을 다양하게 변화하여 표정 특성을 잘 표현할 수 있는 최적의 윈도우 패치를 찾음으로써 표정인식 시스템을 구성하였다.

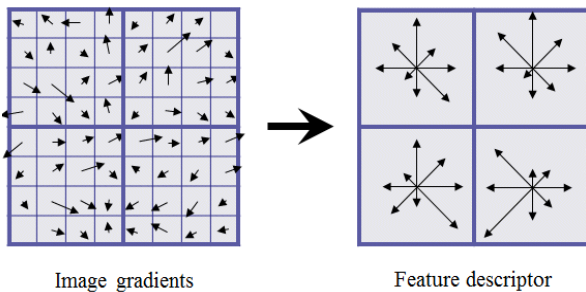


Fig. 3. SIFT descriptor

다음으로 SIFT 기술자를 이용하여 특징 벡터를 추출하고, 얼굴 표정인식은 SVM 알고리즘으로 수행된다. SVM[3, 6]은 인식 성능과 처리 속도 성능이 매우 뛰어나 최근에 패턴인식, 얼굴인식, 문자인식, 데이터마이닝 등의 다양한 분류 문제에 적용되고 있으며, 이에 본 논문에서도 표정 분류를 위하여 SVM을 이용하였다. SVM은 최초 이진패턴 분류 문제를 해결하기 위해 제안된 지도학습 방법 중 하나로, 일반적으로 분류하려는 부류의 결정 초평면 사이에 존재하는 여백을 최대화하여 분류기의 일반화 능력을 극대화하는 방법이다.

3. 실험 및 결과

본 논문에서 제안하는 SIFT 기술자 기반 표정인식 방법은 CK 표정 데이터베이스[7]와 JAFFE 표정 데이터베이스[8] 상에서 기존의 LBP[9, 10] 및 LDP[11, 12] 등의 이진패턴 특징을 SVM으로 인식하는 방법과 성능비교를 수행하였다. CK 표정 데이터베이스는 100명의 학생에 대한 7가지 표정을 갖는 이미지 시퀀스로 구성되어 있으며, 이 중에서 320개의 시퀀스를 실험에 사용하였다. JAFFE 표정 데이터베이스는 213장의 일본인 여성 모델 10명으로 구성되어 있으며, 한 사람에 대하여 7가지 표정 이미지로 구성되어 있다. Fig. 4는 ASM 특징점 정보를 이용하여 정규화를 수행한 CK 표정 데이터베이스 및 JAFFE 표정 데이터베이스의 샘플 영

상을 나타낸다. 더불어, 본 논문에서는 각 데이터베이스에 대한 데이터 셋을 5개의 그룹으로 나누고, 한 개 그룹을 테스트에, 나머지 4개 그룹을 학습에 사용하는 5-fold 교차 검증을 수행하여 실험 결과를 도출하였다.

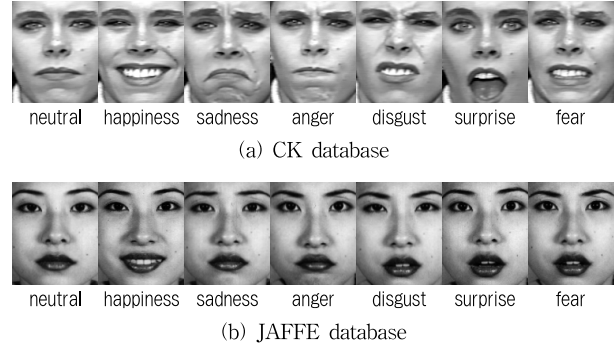


Fig. 4. Sample facial expression images

실험결과, CK 데이터베이스와 JAFFE 데이터베이스 상에서 LBP와 LDP 등의 이진패턴 특징을 SVM으로 인식하는 기존 방법의 표정인식 결과를 각각 Table 1과 Table 2에 나타내었다. 또한, 제안하는 SIFT 기술자를 이용한 표정인식 방법에 대한 실험 결과를 각각 Table 3과 Table 4에 나타내었다. 제안 방법의 실험 결과는 표정 특성을 잘 표현할 수 있는 최적의 윈도우 크기를 찾기 위하여 서브 블록의 크기 N 을 16~48까지 변화하고 O 는 $N/2$ 으로 고정하였으며, 한 서브 블록에서의 그리드는 4×4 로 고정하여 인식률을 도출하였다. 여기서, 한 서브 블록에서의 그리드 개수가 고정되므로 N 의 크기가 커질수록 전체 영상에서의 SIFT 특징 차원이 작아지는 것을 직관적으로 알 수 있다. 실험 결과로부터 LBP 특징을 사용하는 기존 방법에서 CK 데이터베이스는 77.63%의 최대 인식률을 보였으며, JAFFE 데이터베이스의 경우에는 69.74%의 최대 인식률을 보였다. 반면, SIFT 특징을 이용한 제안 방법은 CK 데이터베이스에서는 83.69%, JAFFE 데이터베이스의 경우에는 73.61%의 최대 인식률을 보였다. 이로부터 제안하는 표정인식 방법은 기존 방법보다 CK 데이터베이스의 경우 6.06%, JAFFE 데이터베이스의 경우 3.87%의 향상된 인식 성능을 확인할 수 있었다. 더불어, CK 데이터베이스와 JAFFE 데이터베이스를 이용한 실험 결과를 함께 고려할 때, 제안 방법에서의 SIFT 특징추출을 위한 최적의 서브 블록 윈도우 크기는 32×32 와 40×40 임을 인식 결과로부터 알 수 있었다.

SIFT 특징 추출에 있어 한 서브 블록에 대하여 4×4 그리드 사용하면, 전체 영상에 대한 특징 벡터 차원은 $T \times 4 \times 4 \times 8$ 을 갖는다. 여기서, T 는 전체 서브 블록 개수, 8은 8방향의 빈을 갖는 방향을 의미한다. 고차원의 특징 벡터는 표정인식 시스템의 수행 속도와 밀접한 연관성을 갖으며, 이에 다음 실험에서는 SIFT 특징 차원 축소에 따른 효과를 살펴보고자 한다. 특징 차원 축소는 한 서브 블록에서의 그리드를 4×4 에서 2×2 로 바꾸어 수행되었으며, 이는

전체 특징 벡터의 차원을 4배 줄이는 효과를 갖는다. 이에 위와 같은 실험 환경에서 CK 데이터베이스와 JAFFE 데이터베이스를 이용한 표정인식 실험 결과를 Table 5와 Table 6에 각각 나타내었다.

실험 결과, CK 데이터베이스 경우에 최대 인식률은 83.34%, JAFFE 데이터베이스의 경우는 73.16%를 보였으며, 이러한 결과는 4×4 그리드를 사용할 경우보다 CK 데이터베이스는

0.35%, JAFFE 데이터베이스는 0.45%의 다소 저하된 인식 성능을 확인할 수 있다. 그러나 위와 같은 소폭의 인식 성능 손실 대비, SIFT 특징 차원 축소에 따른 계산 복잡도를 크게 개선한 점을 고려할 때, 2×2 그리드를 사용하여 SIFT 특징을 추출하는 방법은 4×4 그리드를 사용하는 경우보다 실시간 표정인식 시스템에서 보다 효율성이 높을 것으로 사료된다.

Table 1. Recognition results of previous method for CK database

	LBP (5×5)	LBP (7×6)	LDP (5×5)	LDP (7×6)
Set1	69.84%	76.58%	69.04%	73.41%
Set2	75.59%	73.22%	72.04%	71.65%
Set3	80.93%	78.98%	78.98%	75.48%
Set4	76.06%	78.76%	78.37%	81.46%
Set5	84.10%	80.62%	81.39%	82.17%
Average	77.30%	77.63%	75.96%	76.83%

Table 2. Recognition results of previous method for JAFFE database

	LBP (5×5)	LBP (7×6)	LDP (5×5)	LDP (7×6)
Set1	57.14%	66.66%	54.76%	64.28%
Set2	69.04%	73.80%	71.42%	73.80%
Set3	73.80%	61.90%	64.28%	50.00%
Set4	64.28%	61.90%	66.66%	61.90%
Set5	75.55%	84.44%	62.22%	68.88%
Average	67.96%	69.74%	63.86%	63.77%

Table 3. Recognition results of proposed method for CK database when using 4×4 grid per each sub-block

SIFT	16×16	24×24	32×32	40×40	48×48
Set1	89.68%	84.12%	86.90%	86.11%	82.14%
Set2	81.10%	81.49%	81.88%	79.92%	78.34%
Set3	85.21%	85.99%	84.43%	81.32%	83.26%
Set4	81.85%	81.85%	84.55%	83.01%	79.92%
Set5	80.62%	81.00%	79.06%	83.72%	82.55%
Average	83.69%	82.89%	83.36%	82.81%	81.24%

Table 4. Recognition results of proposed method for JAFFE database when using 4×4 grid per each sub-block

SIFT	16×16	24×24	32×32	40×40	48×48
Set1	52.38%	54.76%	61.90%	71.42%	66.66%
Set2	76.19%	78.57%	83.33%	71.42%	76.19%
Set3	61.90%	61.90%	69.04%	76.19%	71.42%
Set4	76.19%	73.80%	80.95%	69.04%	73.80%
Set5	66.66%	73.33%	71.11%	80.00%	64.44%
Average	66.66%	68.47%	73.26%	73.61%	70.50%

Table 5. Recognition results of proposed method for CK database when using 2×2 grid per each sub-block

SIFT	16×16	24×24	32×32	40×40	48×48
Set1	85.71%	84.52%	76.98%	82.53%	72.61%
Set2	79.92%	79.52%	80.31%	81.88%	77.55%
Set3	80.93%	84.04%	79.76%	78.98%	76.26%
Set4	81.08%	83.39%	83.01%	78.76%	75.28%
Set5	82.94%	85.27%	81.78%	84.10%	77.13%
Average	82.11%	83.34%	80.36%	81.25%	75.76%

Table 6. Recognition results of proposed method for JAFFE database when using 2×2 grid per each sub-block

SIFT	16×16	24×24	32×32	40×40	48×48
Set1	52.38%	57.14%	57.14%	64.28%	64.28%
Set2	78.57%	80.95%	78.57%	83.33%	78.57%
Set3	64.28%	71.42%	69.04%	66.66%	69.04%
Set4	71.42%	73.80%	78.57%	73.80%	71.42%
Set5	77.77%	80.00%	82.22%	77.77%	71.11%
Average	68.88%	72.66%	73.10%	73.16%	70.88%

4. 결 론

본 논문에서는 ASM 특징점을 이용하여 얼굴 정규화를 수행하고, 정규화 영상을 SIFT 특징과 SVM 분류기를 이용하여 표정인식을 수행하는 방법에 대하여 제안하였다. CK 데이터베이스와 JAFFE 데이터베이스에서 제안 방법의 성능 평가는 다양한 서브 블록 윈도우의 크기에 따라 기존의 이진패턴 특징 기반의 접근법과 비교하여 수행되었다. 실험 결과, 제안방법은 CK 데이터베이스에서 기존의 방법보다 6.06%의 성능향상을 보였고, JAFFE 데이터베이스에서는 기존 방법보다 3.87%의 성능향상을 보였으며, 이로부터 제안 방법의 효용성을 확인하였다.

References

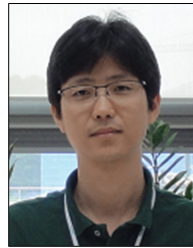
- [1] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *Journal of Personality and Social Psychology*, Vol.17, No.2, pp.124-129, 1971.
- [2] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models - their training and application," *Computer Vision and Image Understanding*, Vol.61, pp.38-59, 1995.
- [3] H. T. Le and N. T. Vo, "Face alignment using active shape model and support vector machine," *International Journal of Biometrics and Bioinformatics*, Vol.4, No.6, pp.224-234, 2012.
- [4] David G. Lowe, "Object recognition from local scale-invariant features," *Proceedings of the International Conference on Computer Vision*, Vol.2. pp.1150-1157, 1999.
- [5] M. T. Carlos, P. B. Marcos, and B. A. Jesus, "Fused intra-bimodal face verification approach based on scale-invariant feature transform and a vocabulary tree," *Pattern Recognition Letters*, Vol.36, pp.254-260, 2014.
- [6] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, Vol.20, No.3, pp.273-297, 1995.
- [7] T. Kanade, J. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," *IEEE International Conference Automatic Face Gesture Recognition*, pp.46-53, 2000.
- [8] 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.
- [9] 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.
- [10] W. L. Chao, J. J. Ding, and J. Z. Liu, "Facial expression recognition based on improved local binary pattern and class-regularized locality preserving projection," *Signal Processing*, Vol.117, pp.1-10, 2015.
- [11] 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.
- [12] F. Zhong and J. Zhang, "Face recognition with enhanced local directional patterns," *Neurocomputing*, Vol.119, No.7, pp.375-384, 2013.



김 동 주

e-mail : kkb0320@gdsu.dongseo.ac.kr
1998년 충북대학교 전파공학과(공학사)
2000년 충북대학교 전파공학과(공학석사)
2010년 성균관대학교 전기전자 및
컴퓨터공학과(공학박사)
2011년~2015년 대구경북과학기술원
IT융합연구부 선임연구원

2015년~현재 동서대학교 컴퓨터공학부 조교수
관심분야: 얼굴인식, HCI, 영상신호처리, 패턴인식



손 명 규

e-mail : smk@dgist.ac.kr
1997년 경북대학교 전자공학과(공학사)
1999년 서울대학교 전기공학과(공학석사)
2005년~현재 대구경북과학기술원
IT융합연구부 선임연구원
관심분야: HCI, 패턴인식, 컴퓨터비전



이 상 현

e-mail : pobbylee@dgist.ac.kr
1993년 경북대학교 전자공학과(공학사)
1996년 경북대학교 전자공학과(공학석사)
2013년 경북대학교 전기전자공학과(공학
박사)
2005년~현재 대구경북과학기술원
IT융합연구부 선임연구원

관심분야: 컴퓨터비전, 영상신호처리