Journal of Korea Multimedia Society Vol. 20, No. 5, May 2017(pp. 758-768)
https://doi.org/10.9717/kmms.2017.20.5.758

공간 계층적 구조 기반 지역 기술자 활용 얼굴인식 기술

김경태†, 최재영††

Using Spatial Pyramid Based Local Descriptor for Face Recognition

Kyeong Tae Kim†, Jae Young Choi††

ABSTRACT

In this paper, we present a novel method to extract face representation based on multi-resolution spatial pyramid. In our method, a face is subdivided into increasingly finer sub-regions (local regions) and represented at multiple levels of histogram representations. To cope with misaligned problem, patch-based local descriptor extraction has been also developed in a novel way. To preserve multiple levels of detail in local characteristics and also encode holistic spatial configuration, histograms from all levels of spatial pyramid are integrated by using dimensionality reduction and feature combination, leading to our spatial-pyramid face feature representation. We incorporate our proposed face features into general face recognition pipeline and achieve state-of-the-art results on challenging face recognition problems.

Key words: Face Recognition, Spatial Pyramid, Local Descriptor, Histogram Representation, Feature Combination

1. 서 론

얼굴인식은 생체인식, 인간-컴퓨터 상호작용, 비디오 감시 등 다양한 응용 프로그램 사용으로 패턴인식(patter recognition)과 컴퓨터비전(computer vision) 분야에서 다양하고 활발한 연구가 진행되고 있다. 대부분의 패턴인식 시스템과 같이 얼굴인식 시스템은 일반적으로 (1) 얼굴 특징 추출기(face feature extractor)와 (2) 분류기(classifier)로 구성된다. 다양한 분류기 모델 중에서, 최근접 분류기(nearest neighbors(NN))와 최근접 분류기를 변형시킨 분류기 기술들이 얼굴인식에 많이 사용되고 있다[1].

분류기 학습 이외에도, 얼굴 특징 표현은 얼굴인식 과정에서 중요한 역할을 한다. 효과적인 얼굴 표현은 얼굴 자세, 표정, 중첩(occlusion), 오정렬(misalignment), 해상도 변화[36] 등에 대한 문제에 강인해야 한다. 일반적으로 얼굴 표현은 전역 특징(holistic feature)과 지역 특징(local feature)이라는

※ Corresponding Author : Jae Young Choi, Address: (17035) 81 Oedae-ro, Mohyeon-myeon, Cheoin-gu, Yongin-si, Gyeonggi-do, Korea, TEL : +82-31-330-4906, FAX : +82-31-330-4906, E-mail : jychoi@hufs.ac.kr
Receipt date : Feb. 26, 2017, Revision date : Apr. 16, 2017
Approval date : Apr. 25, 2017
† Division of Computer and Electronic Systems Engineering, Hankuk University of Foreign Studies
(E-mail : kyeongtae.kim@hufs.ac.kr)
†† Division of Computer and Electronic Systems Engineering, Hankuk University of Foreign Studies
※ This work was supported by Hankuk University of Foreign Studies Research Fund.
※ This research was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No. 2016R1E1A2020509).
※ This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2015R1D1A1A01057420).

두 가지 범주로 나눌 수 있다[2]. 전역 특징 표현은 전체적인 얼굴의 일반적인 특성을 나타내는 것이고, 반대로 지역 특징 표현은 얼굴의 부분영역들을(예: 눈, 코, 입 등) 상세한 특성으로 반영한 것이다.

전역 또는 지역 특징을 기반으로 한 얼굴 표현에 대해 많은 연구가 이루어졌지만 얼굴인식에 가장 적합한 얼굴특징 표현이 무엇인지에 관해서는 미해결 문제로 남아있다. 최근 BoW(Bag-of-Words)모델[3, 4]을 활용한 영상 인식 및 분류 기술들이 활발하게 사용되고 있다. BoW 모델에서는 일반적으로 낮은 단계의 기술자(low-level descriptor)를 클러스터링(clustering) 기술에 적용하여 비주얼워드(visual words)의 집합으로 인코딩(encoding)한다. BoW 표현은 비주얼객체분류(visual object categorization)[3, 5]에 광범위하게 사용되었지만, 얼굴인식 성능 향상을 위한 BoW 기반 얼굴인식 알고리즘 개발연구에 관해서는 현재 미흡하다. 지금까지 개발된 일반적인 BoW 모델을 얼굴인식에 직접 적용하면 얼굴의 공간적 구성(spatial configuration)이 제거되어 얼굴 표현의 정보를 신뢰성 있게 표현하는데 문제점이 생긴다.

위에서 언급한 한계점을 극복하기 위한 BoW모델[6, 8]과 얼굴영상의 공간정보(spatial information)를 통합하는 연구 내용들은 아직 매우 미흡한 실정이다. 제안 방법과 가장 관련성이 높은 기존 연구로서 Li et at. [6, 7]은 얼굴영상을 여러 개의 블록으로 분할하고 각 블록에 대해 SIFT(Shift-Invariant Feature Transformation) 기술자[25]를 추출하는 Block-Based Bag-of-Words (BBoW) 방법을 사용하였다. SIFT 기술자[25]는 각각의 블록마다 비주얼단어들 집합으로 양자화(quantized)된다. 모든 지역 블록들에서 추출된 코드워드(codeword) 분포를 나타내는 히스토그램 벡터들을 결합하고 이것의 결과로 생성되어 확장된 히스토그램을 SVM(Support Vector Machine)에 적용하여 얼굴을 식별한다. 하지만, 기존 연구들은 다음과 같은 한계가 있다. 첫 번째, 블록 기반 사전(dictionary) 생성(즉, 특정 블록에 대한 하나의 사전(dictionary))이 필요하다. 대응되는 블록(예: 눈 또는 입)이 얼굴 외형 변화, 특히 얼굴 자세 변화 및 중첩(occlusion)에 영향을 받으면, 각각의 사전(dictionary)에 존재하는 판별 정보는 손실 될 것이다. 두 번째로 이와 같은 접근방법으로 모든 블록의 수와 크기가 고정되는데, 이는 얼굴 자르기(face cropping) 중에 발생하는 오정렬 오류로 인해 얼굴인식 성능이 저하될 수 있다. 위에서 언급된 한계성을 극복하기 위해 제안방법에서는 얼굴영상을 여러 개의 하위 얼굴 영역들(facial subregions)로 분리한 후, 각 하위 영역들로부터 패치영상들을 추출하고 추출된 패치들의 지역 기술자(local descriptor)를 계산한다. 기존연구 [8]은 얼굴 구성 요소(예: 눈, 코, 입)에서 밀집형 SIFT(dense-SIFT) 기술자를 추출하고 BoW 사전(dictionary)를 구성하는데 활용하였다. 이 방법에 BoW 사전의 크기는 개별적인 얼굴 구성(facial component) 특성들이 얼마나 차별적 정보(discriminant information)를 포함하고 있는지에 따라 결정된다. 하지만 이 방법은 얼굴영상의 기준점들(fiducial points)의 정확한 검출이 요구된다. 따라서 인식성능이 민감하게 변화하는 기준점 위치 검출 결과에 지나치게 의존한다는 단점이 있다.

본 논문은 얼굴인식을 위해 새로운 방법으로 BoW 기반 얼굴특징표현(face feature descriptor)을 제시한다. 제안한 방법에는 공간 계층적 프레임워크(spatial pyramid framework(SPF))[5]를 통해 전역(global)과 지역(local) 판별 특징을 모두 활용할 수 있다. 공간 계층적 프레임워크(SPF)를 이용하여 얼굴 영상을 하위 영역(혹은 채널)들로 나누고, 각각의 하위 영역들로부터 패치영상(patch image) 집합이 추출된다. 그 다음, 각각의 패치영상에서 지역 기술자(local descriptor)[9]를 추출하고, 각각의 하위 영역(sub-region)들에 포함된 패치영상들로부터 추출된 지역 기술자들의 특성을 고려하여 공간적 히스토그램(spatial histogram)을 계산한다. 공간적 계층 단계(level of spatial pyramid)에서는, 모든 얼굴 하위 영역(sub-region)들로부터 추출된 코드워드(codeword)분포 정보를 나타내는 다중 히스토그램들을 결합한다. 여러 단계의 얼굴해상도(face resolution)로 지역 정보를 찾아내는 동시에 공간구성 정보를 인코딩하기 위해, 모든 공간-계층 단계로부터 추출된 히스토그램들을 결합하고, 이를 제안하는 '공간 계층적 얼굴 특징'이라 하며 얼굴인식에 활용한다.

2. 공간 계층적 얼굴 특징 기반 얼굴인식

2.1 패치기반 지역 기술자 추출

실제 얼굴인식에서, 검출된 얼굴 영상은 완벽하게

정렬이 될 수 없기 때문에 자른 얼굴 영상에서 공간적인 오정렬(misalignment)이 존재할 수 있다. 이 문제를 해결하기 위해 얼굴영상에서 패치영상들(patch images)을 추출하고 패치영상들 집합으로 얼굴 영상(하위 영역)을 표현한 후, 각 패치에 대해 지역 기술자(local descriptor)를 계산한다. $N \times M$ 크기 얼굴영상 $\mathbf{I}$로부터 다음과 같이 패치영상들 집합을 추출한다.

$$\{\boldsymbol{P}_{i,j} | i \in (n/2, N-n/2), j \in (n/2, M-n/2)\} \quad (1)$$

여기서 패치영상 $\boldsymbol{P}_{i,j}$의 사이즈는 $n \times n$이고 각각의 픽셀은 (i, j)위치에서 추출된다. 식 (1)에서 서로 다른 위치의 패치들을 일부 겹치게 추출하여 지역 기술자(local descriptor)를 추출하기 때문에, 얼굴 정렬과정에서 발생할 수 있는 오정렬 오류에 강인할 수 있다. Fig. 1은 제안한 패치기반 지역 기술자 추출 방법으로 오정렬 오류에 강인함을 나타낸다. Fig. 1(c)에서 볼 수 있듯이, 제안하는 방법에서 얻은 거리 값들 대부분이 기존의 키포인트 기반 접근법(key-point-based approach)을 사용하여 계산된 거리 값들과 비교했을 때 더 적은 값들로 분포되어 있다. 이러한 실험적 관찰결과는 얼굴인식에서 빈번하게 발생하는 오정렬(misalignment) 혹은 얼굴표정(expression)과 중첩(occlusion)에 의해 얼굴 영상 내의 특성들이 국부적으로 손상되어도, 제안하는 패치 기반 지역 기술자 추출 방법이 얼굴영상들 사이의 올바른 유사성(face similarity)을 판정할 수 있음을 의미한다.

2.2 공간 계층적 히스토그램 벡터 계산

BoW 영상 인코딩(encoding)과정은 네 가지 핵심 요소[20]인 지역 기술자 추출, 사전 학습(dictionary learning), 기술자 인코딩, BoW표현을 사용한다. 본 연구에서 제안한 방법은 [5]의 공간 계층적 프레임워크(spatial pyramid framework)를 기반으로 수행 한 후 2.1절에서 설명한대로 영상 패치에서 지역 기술자

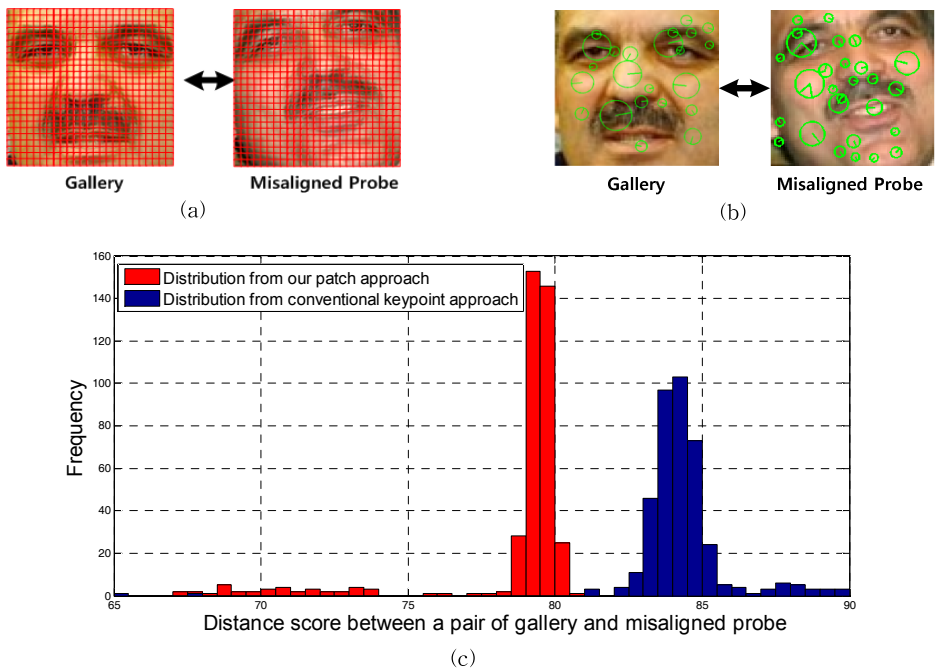

Fig. 1. Demonstration for the robustness (or tolerance) of our patch-based local descriptor extraction against misalignment between the pre-registered gallery and probe(test) images from the same subject. (a) Computing SIFT descriptors using our patch approach. (b) Computing SIFT descriptors using conventional keypoint approach [15, 16]. (c) Distributions of the distances [17] in the SIFT feature space to evaluate the degree of similarity between a pair of gallery and probe. These distributions are generated using a total of 800 pairs of face images.

를 추출한다. 사전 학습(dictionary learning)단계에서 패치 영상의 지역 기술자를 각각 K-means 군집화(clusteirng) 알고리즘을 이용하여 비주얼워드(코드워드(codeword))로 변환된다. 패치 샘플들의 지역 기술자들은 K개의 클러스터로 분할하여 K개의 비주얼워드로 구성된 하나의 사전(dictionary)를 구성한다. 이 비주얼 사전(visual dictionary)은 이후 소프트 할당(soft assignment)[16] 방법으로 각 지역 기술자를 코드화된 벡터로 인코딩하는데 사용한다. 그 다음으로 공간 계층적 히스토그램 표현은 공간 피라미드의 각 하위 영역(채널) 내의 코드화 된 히스토그램 벡터에서 생성된다.

Fig. 2에서 도시 된 바와 같이, 각 하위 영역(sub-region) 내의 패치 집합은 비주얼워드들의 개수를 세어 공간적 히스토그램(spatial histogram)으로 나타낼 수 있다. 레벨(level) l에서 $M = 2^l \times 2^l$ 개수의 하위 영역들이 존재한다. 레벨 l의 공간 히스토그램 벡터 $\boldsymbol{h}_l^{(i)}$는 [5]에서 사용 된 것과 동일한 방식으로 i-번째 하위 영역으로 계산한다. 여기서 $\boldsymbol{h}_l^{(i)} \in R^K$이고 K는 비주얼워드들의 개수이다. 최종적으로 레벨 l의 하위 영역들을 계산한 모든 히스토그램들이 다음과 같이 결합된다.

$$\boldsymbol{h}_l = \left[(\boldsymbol{h}_l^{(1)})^T (\boldsymbol{h}_l^{(2)})^T \cdots (\boldsymbol{h}_l^{(M)})^T \right]^T \quad (2)$$

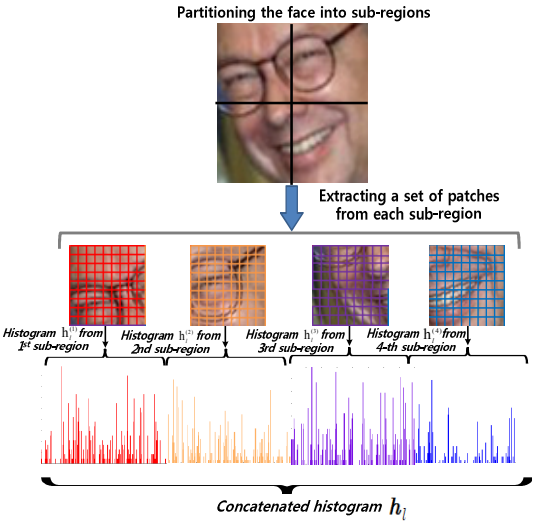

Fig. 2. Illustrating the creation of a concatenated histogram $\boldsymbol{h}_l$ at level l (corresponding to l=1). Note that a spatial histogram $\boldsymbol{h}_l^{(i)}$ for each sub-region is computed from a set of corresponding patches.

여기서 $\boldsymbol{h}_l \in R^{KM}$ 그리고 T는 행렬의 전치 연산자이다. 결합된 히스토그램(concatenated histogram) $\boldsymbol{h}_l$은 레벨 l에서 공간 계층적 얼굴 특징(spatial pyramid face representation)을 형성한다.

2.3 저차원 특징결합 기반 공간 계층적 얼굴 특징 추출

서로 다른 레벨의 공간 계층(spatial pyramid)에서 추출한 히스토그램(식 (2))들이 결합되어 서로 다른 판별정보를 최대한 활용하고, 우수한 얼굴인식 성능을 얻기 위해 특징결합 기술[19] 사용을 제안한다. 프로토타입(prototype)으로 등록한 얼굴 영상인 갤러리(Gallery) 집합을 $\boldsymbol{G}$로 표시하며, $\boldsymbol{G}$에 포함된 얼굴영상을 $\boldsymbol{I}^g$로 표시한다(즉, $\boldsymbol{I}^g \in \boldsymbol{G}$). 또한, $\boldsymbol{I}^p$는 얼굴인식의 대상이 되는 얼굴 영상이며, 프로브(probe)로 표시한다. $\boldsymbol{I}^g$와 $\boldsymbol{I}^p$에 대한 레벨 l에서 연결된 히스토그램은 각각 $\boldsymbol{h}_l^g$와 $\boldsymbol{h}_l^p$로 표시되며, 여기서 $l = 0, \ldots, L$이고 L은 공간피라미드의 레벨을 나타낸다.

여기서 중요한 것은 모든 레벨을 대상으로 $\boldsymbol{h}_l$를 결합을 하면 각각의 하위 영역으로부터 히스토그램을 병합하고 형성하여 고차원의 특징 벡터를 산출한다는 것이다. 예를 들어, $L = 2$와 $K = 300$(코드워드(codeword)의 수)의 설정은 25,500차원 히스토그램 벡터를 생성한다. 이러한 큰 특징 벡터를 직접 적용하는 것은 고차원으로 인한 얼굴인식 성능이 저하가 될 수 있으며, 계산을 비효율적으로 만든다. 이러한 문제를 극복하기 위해 우리는 간단하면서 효과적인 방법인 저차원 특징 추출 기술 [21, 22] 사용을 활용한다. 저차원 특징 추출기를 Φ_l(즉, PCA[23])로 표시하자. Φ_l는 공간 피라미드 레벨 l에서 계산한 결합 히스토그램은 훈련 집합(training set)을 활용하여 형성 되며, 모든 훈련 얼굴 영상에서 계산된다. $\boldsymbol{h}_l^g$와 $\boldsymbol{h}_l^p$의 저차원 특징은 다음과 같이 얻어진다.

$$f_l^g = \Phi_l(h_l^g) \quad (3)$$

여기서 $f_l^g, f_l^p \in R^D$이다. 다음으로 $L+1$ 개의 상보적인(complementary) 저차원 특징들을 열(column) 방향으로 결합하여 다음과 같이 공간 계층적 얼굴 특징을 추출한다.

$$f^g = \left[(f_l^g)^T \cdots (f_L^g)^T \right]^T \text{ and } f^p = \left[(f_l^p)^T \cdots (f_L^p)^T \right]^T \quad (4)$$

여기서 $f^g, f^p \in R^D$이고 $D = \sum_{l=0}^{L} D_l$이다.

$\boldsymbol{I}^p$에 대해 얼굴인식 작업(식별 또는 인증)을 수행하기 위해 f^p와 대응하는 가장 유사한 f^{g^*}를 아래와 같이 최근접 분류기[14]를 활용하여 찾는다.

$$g^* = \arg\min_{\boldsymbol{I}^g \in \boldsymbol{G}} d(f^p, f^g) \tag{5}$$

여기서 $d(\bullet)$는 거리를 계산하는 메트릭 함수(metric function)이며 g^*는 얼굴영상 $\boldsymbol{I}^p$의 인식된 신원(identity)을 나타낸다.

3. 실 험

3.1 실험 환경

본 연구에서 제안한 공간 계층적 얼굴 특징 기반 얼굴인식 프레임워크(framework)는 실험 평가로 많이 사용되는 CMU-PIE[10], FERET[11], FRGC 2.0 [35] 및 LFW[12]의 얼굴 영상 데이터베이스를 사용하였다. 두 눈의 좌표를 기준으로 120×120픽셀로 잘라 얼굴 영역을 형성하고 각각 잘린(cropped) 얼굴 영상은 두 눈의 위치를 기준으로 정렬하였다. 반면, 본 실험에서 LFW 영상[34]은 정렬이 되지 않는 얼굴 영상이 많이 발생한다. 따라서 LFW 영상을 Viola-Jones 얼굴 검출기 [13]에 먼저 감지시키고 감지된 얼굴 영역은 120×120픽셀로 다시 조정하였다. 실질적인 얼굴인식 시나리오를 가정하기 위해 LFW 영상들에 경우 정렬 없이 얼굴 검출 출력을 실험에 직접 사용하였다. Fig. 3은 자른 얼굴 영상들의 일부를 보여준다.

본 실험에서, 많이 사용하는 기술자인 SIFT[18]를 지역 기술자 추출에 사용되었다. 패치들 간의 중첩을 허용하기 위해, 지역 패치는 패치 영상의 크기가 8×8픽셀인 얼굴 영상에 2픽셀씩 겹치도록 하였다. 사전 학습(dictionary learning)을 위해 표준 Lloyd k-means를 구현하였고 사전(dictionary)의 크기는 [5]에서 권장하는 대로 400으로 설정하였다. BoW코딩을 구현하기 위해 지역 기술자를 비주얼워드(Visual words)로 매핑(mapping)하였고 비주얼워드를 공간적 히스토그램에 축척하기 위해 랜덤화 된 kd-tree 매칭[16, 24]이 사용하였다(또는 비주얼워드(visual word)는 kd-tree의 랜덤화된 포레스트(randomized forest)에 저장된다[24]). 1×1, 2×2, 3×3의 3개 레벨(level)을 가지는 공간 계층(즉, $L=2$)으로 하였다. 식 (3)에서 나타낸 Φ_l를 구성하기 위해 RLDA[21]와 KDDA[22]를 저차원 특징을 추출하는데 사용하였다. 최근접 분류기(nearest neighbors(NN))를 구현하기 위해 유클리드 거리(Euclidean distance)가 RLDA와 KDDA에 사용되었다. 본 연구에서 3가지의 얼굴 데이터베이스에 대해 인증(identification) 실험을 수행하고 성능 평가를 위해 누적 매칭 특성(Cumulative Matching Characteristic (CMC)) 곡선[11]을 사용하였다. 일반적으로 얼굴인식 성능은 [11, 14]에서 사용된 저차원 특징의 개수에 의존한다. 따라서 정확한 비교를 위해 [21, 22]에서 best found correct recognition rate(BstCRR))이 인식율(identification rate)로 선택되었다. 각 데이터 집합에서 수집 된 정면 영상 집합은 무작위로 훈련 집합과 프로브(probe) 집합으로 나누었으며, 두 개의 집합 간에 중복된 영상들은 없다. 또한 사전(dictionary)는 훈련 집합에서 무작위로 선택된 인식 대상자(class)당 2개의 영상으로만 학습되었다. 앞서 언급한 무작위 분할

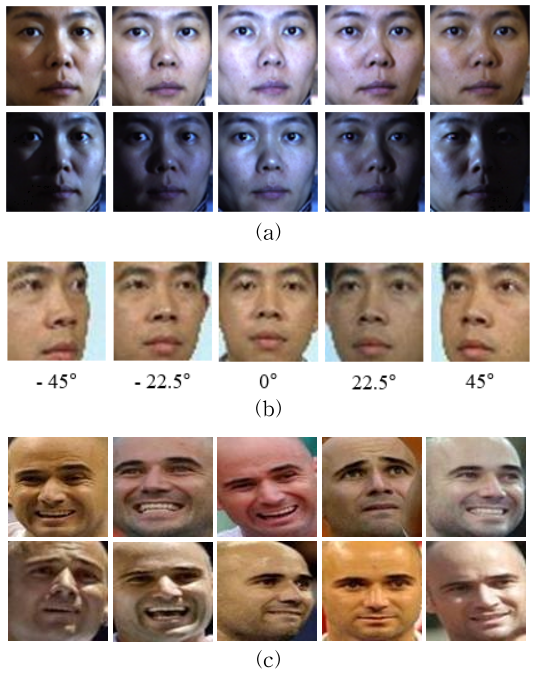

Fig. 3. Examples of cropped facial images (a) from CMU-PIE, (b) FETET, and (c) LFW DBs. Note that every faical image from CMU-PIE and FERET is manually cropped using eye coordinate information and aligned using a fixed template [11], whereas all facial images from LFW are directly obtained using face detector output without any alignment process [34].

및 선택을 10회 반복하였고, 결과는 모든 10회 실행하여 평균과 표준편차로 나타내었다. 갤러리(gallery) 집합은 인식 대상자당 한 개의 정면 또는 최소 정면에 가까운 영상으로 구성되었다.

제안한 공간 계층적 얼굴 특징을 BBoW얼굴 특징[7, 9]뿐만 아니라 기존 BoW("CBoW"로 약칭) 기반 얼굴 특징[4, 25]들과 비교하였다. CBoW방법을 구현하기 위해 기존의 SIFT 기술자의 권고에 따라 각 얼굴 영상의 키포인트(key-point)를 먼저 검출한다. 그 다음 각 키포인트(key-point)의 크기와 방향에 따라 각 얼굴 영상의 SIFT 기술자를 계산한다. 마지막으로, 얼굴 영상을 표현하기 위해 이러한 특징을 코드워드(codeword)로 벡터 양자화(vector quantize)하였다. 또한, BBoW방법을 구현할 때 우리는 논문[6, 7]에 제시된 매개 변수를 동일하게 사용하였다. 구체적으로, 5×5 블록 분할, 밀집형 SIFT(dense-SIFT) 기술자를 계산하기 위한 4×3 표본 격자 크기, 각 블록 K=75 의 사전(코드북(codebook))크기가 사용되었다. 신뢰성 있는 비교를 위해 CBoW와 BBoW에서 얻은 얼굴 특징들을 섹션 2.3에서 설명한 방법과 유사한 과정으로 저차원 특징 추출과 최근접 분류기(nearest neighbors(NN))에 적용하여 인식성능을 비교했다.

3.2 CMU-PIE 데이터베이스 활용 실험 비교

광범위한 조명 변화(illumination variation)에 대해 제안하는 공간 계층적 얼굴 특징의 우수성을 검증하기 위해 CMU-PIE 데이터베이스가 사용되었다. CMU-PIE에는 68명의 인식 대상자들에 대해 2,856장(인식 대상자 당 42장의 영상)의 정면 영상으로 구성되어 있다. 각 인식 대상자에 대한 얼굴 영상이 '42개의 조명 밝기 변화' 조건[10]을 가지고 있다. 무작위로 분할을 통해 훈련 집합은 6영상 × 68개의 인식 대상자로 구성되었고 나머지 2,448개의 영상들은 프로브(probe) 집합을 형성하는데 사용하였다. Fig. 3(a)는 이 실험에서 사용된 영상의 예를 보여준다. 얼굴 영상이 조명과 그림자로 인해 많은 밝기 변화가 매우 가변적임을 관찰할 수 있다.

Table 1은 rank-1 인식율(identification rate)을 보여준다. 우리의 공간 계층적 얼굴 특징은 RLDA와 KDDA에서 각각 91.80%, 95.91%의 성능을 나타내었다. Table 1에서 제안한 공간 계층적 얼굴 특징이 BoW 및 BBoW 얼굴 표현보다 우수한 것을 볼 수 있으며, 이것은 공간 계층적 얼굴 특징이 조명 변화에 대해 강인함을 보여준다.

3.3 FERET 데이터베이스 활용 실험 비교

우리는 자세 변화(pose variation)에 대해 공간 계층적 얼굴 특징의 유용성(usefulness)을 평가하였다. FERET 데이터베이스에서 300명의 인식 대상자를 대상으로 총 2,508장의 얼굴 영상을 수집하였고 두 눈의 정규화를 위해 확실하게 식별될 수 있는 회전된 얼굴 영상만 선별하였다. 사용된 얼굴 영상은 -45°에서 45°까지 다양한 얼굴 자세 각도가 포함되어있다. (Fig. 3(b) 참고). 무작위로 분할하여 훈련 집합은 1,200장의 영상(300명의 인식 대상자당 4장의 영상)으로 구성하였고, 프로브(probe) 집합에는 300명의 인식 대상자의 나머지 영상인 1,308장의 영상으로 구성하였다.

비교 결과는 Table 2와 같다. 제안하는 공간 계층적 얼굴 특징은 RLDA와 KDDA에 대해 BoW와 BBoW표현보다 우수하다는 것을 알 수 있다. 특히 우리가 제안한 공간 계층적 얼굴 특징은 자세 변화가 있는 얼굴 영상을 올바르게 인식한다는 점에서 CBoW접근 방식보다 훨씬 우수하다. 구체적으로, 제안하는 방법은 RLDA와 KDDA에서 각각 CBoW보다 9.39%와 9.14%정도 더 높은 인식율(identification rates)을 달성했다. 이러한 이유는 공간 계층적 얼굴 특징이 여러 단계에서 로컬 패치의 외형 변화를 포착하고 시각적 변화가 얼굴영상에 나타날 때 얼굴의

Table 1. Comparisons of average rank-1 identification rates (in percent) on CMU-PIE DB to show the robustness of our method against severe illumination variation

Low-dimensional feature extraction	Face Representation		
	CBoW [4, 25]	BBoW [6-7]	Proposed method
RLDA	73.61±1.15	85.77±2.19	91.80±2.79
KDDA	79.05±0.83	89.68±1.01	95.91±0.91

Table 2. Comparisons of average rank-1 identification rates (in percent) on FERET DB to show the robustness of our method against pose variation

Low-dimensional feature extraction	Face Representation		
	CBoW [4, 25]	BBoW [6-7]	Proposed method
RLDA	78.71±1.98	82.10±1.92	88.10±1.92
KDDA	83.38±0.70	86.15±1.23	92.52±0.30

공간 구조 불변성(spatial structure invariance)을 유지할 수 있기 때문이다. 이처럼 전체 얼굴 영역에서 추출된 코드워드(codeword) 히스토그램 패턴과 비교할 때 얼굴 자세 변화가 있는 얼굴 영상의 경우에도 해상도의 각 레벨에서 특정 지역영역에서 유도된 판별 히스토그램 패턴이 보존될 가능성이 높다.

3.4 LFW 데이터베이스 활용 실험 비교

비제약된(unconstrained) 얼굴인식 환경에서 취득된 얼굴 영상 데이터베이스인 Labeled Faces in the Wild (LFW)을 사용하여 추가적으로 실험을 하였다. 웹에서 수집된 13,000장 이상의 얼굴 영상이 포함되어 있으며, 자세와 표정의 변화가 크고, 중첩이 많이 발생되어 있다[12]. LFW 평가 프로토콜은 원래 실제 얼굴인식 환경에 대한 얼굴인식 검증을 위해 만들어진 것이다. 그러나 [34]를 참조하여, 우리는 얼굴 인식 실험을 수행하기 위해 LFW영상을 사용하였고 섹션 3.1에서 설명했듯이, 자른 얼굴 영상을 수집하기 위해 얼굴 검출기를 사용하였다[34]. Fig. 3(c)는 LFW 영상과 잘린 얼굴 영상의 예를 보여준다. Fig. 3(c)에서 볼 수 있듯이, 오정렬과 중첩된 영상은 실험에서 제외하지 않았다. 이를 통해 실제 응용에서 얼굴 인식을 수행하는 것과 비슷한 실험환경으로 공간 계층적 얼굴 특징의 유효성(effectiveness)을 평가할 수 있다. 본 실험에서 LFW 데이터베이스에 대한 실험을 위해 423명의 인식 대상자를 선택하여 5,604장의 얼굴 영상 데이터 집합을 형성하였다. 인식 대상자 당 얼굴 영상 수는 6장에서 19장 사이이다. 훈련 집합의 경우 1,269장 영상(인식 대상자 당 3장의 영상)과 프로브(probe) 집합은 나머지 영상인 4,335장의 영상으로 두 개의 영상 집합으로 나눠진다. Fig. 4는 비교를 위해 사용된 얼굴 특징들에 대한 CMC 곡선을 보여준다. CBoW와 BBoW의 성능이 얼굴 검출기로 출력된 얼굴영상들을 인식 대상으로 할 경우 성능이 크게 저하된다는 것을 확인 할 수 있다. 제안한 공간 계층적 얼굴 특징이 기존의 BoW 기반 얼굴 특징과 비교하였을 때 우수한 인식율을 보여준다. 특히 KDDA의 경우 공간 계층적 얼굴 특징은 rank-1 인식율 성능에서 CBoW보다 15.28 %, BBoW보다 10.66% 만큼 향상하였다.

3.5 최신 얼굴인식 방법들과의 성능 비교

제안하는 방법과 최근에 개발된 얼굴인식 방법들의 성능을 신뢰성 있게 비교하기 위해, 얼굴인식 성능 평가 표준 방법인 'FRGC 2.0 실험 4' 프로토콜을 활용하였다[35]. 실험 4는 비제약된 조명(uncontrol-

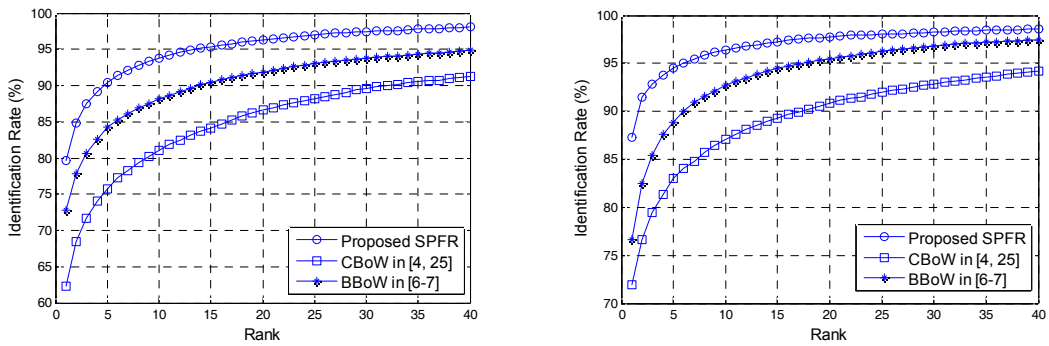

Fig. 4. CMC curves on the LFW database. (a) RLDA. (b) KDDA.

Table 3. Comparisons with other state-of-the-art face recognition methods on the "FRGC 2.0 Experiment 4" using the verification rate (%) at 0.1% FAR. Note that KDDA [22] is used for our method

Method	Verification rate (ROC III) when FAR = 0.1%
Gabor + DCT + LDA [27]	89.00
LGBP + LGXP + Block-based LDA [28]	85.20
LBPH + Gabor + KDR (X2) [29]	88.10
Hybrid Fourier feature + LDA [30]	81.14
Gabor + LBP + KDCV [31]	83.60
Multiscale LPQ (MLPQ) + Kernel Fusion [32]	90.36
MDML-DCPs + PLDA + Linear SVM [33]	93.39
Proposed method	92.54

led illumination) 변화를 가지는 정면 얼굴 영상들의 인식 성능을 평가하도록 설계되어 있다. FRGC 2.0 실험 4에서 타겟(target) 집합은 16,038장의 조명에 대한 제약된 영상으로 구성되어 있으며, 쿼리(query) 집합은 조명 변화가 있는 8,015장의 영상으로 구성되어 있다. 공정한 비교를 위해 최근에 개발된 다른 얼굴인식 방법들의 인식 결과들은 논문[27, 33]에서 직접 인용하였다. 또한 최근에 발표된 연구에서 0.1% 오수락율(False Acceptance Rate, FAR)에 대한 인증율(verification rate) 성능을 보고하였기 때문에 다른 얼굴 기술자와 직접적으로 비교하기 위해 우리 역시 0.1% 오수락율(FAR)에서 인증율(verification rate)을 도출하였다. 따라서 제안한 방법과 다른 최신 얼굴인식 방법들과 공평하고 신뢰성 있게 비교 할 수 있다. Table 3에서 볼 수 있듯이, 다른 최신 방법들과 비교했을 때 제안방법이 더 높은 성능을 보이거나 혹은 동등한 성능을 보임을 알 수 있다. 특히, 제안 방법은 인증율 92.54%을 달성하였고 이는 최근의 보고된 가장 높은 인증율 93.39%[33]와 거의 유사한 성능이다. 이런 실험결과들은 제안 방법이 향상된 얼굴인식 성능을 실현하는데 유용하며 최신 얼굴인식 방법들을 한 단계 진보시킬 수 있는 잠재성을 가지고 있음을 증명한다.

4. 결 론

본 논문에서는 패치 기반의 지역 기술자와 얼굴 영상이 계층화 된 하위 영역을 이용하여 얼굴 표현을 추출하는 방법을 제시하였다. 우리의 공간 계층적 얼굴 특징은 여러 단계의 공간 계층을 사용하여 지역 특징(지역 얼굴 영역 내의 상세한 특성)과 전체적인 얼굴의 전역 특징(예: 얼굴의 공간 구성)을 얻을 수 있었다. 실험 결과를 통해 제안한 공간 계층적 얼굴 특징이 차별적이고 안정된 얼굴 표현을 추출 할 수 있음을 입증하였고, 인식 성능을 향상시킬 수 있는 가능성을 보여주었다. 향후 연구에서는 많이 사용되어지는 LBP[14]혹은 Gabor[14]와 같이 다른 지역 텍스처 기술자(local texture descriptors)를 포함하여 얼굴 인식 프레임워크(framework)를 확장할 계획이다. 향후 연구 계획은 제안하는 얼굴인식 프레임워크가 모든 종류의 지역 기술자를 통합 할 수 있다는 점에서 유용할 것이다.

REFERENCE

[1] J.T. Chien and C.C. Wu, "Discriminant Waveletfaces and Nearest Feature Classifiers for Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 12, pp. 1644-1649, 2002.

[2] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical Ensemble of Global and Local Classifiers for Face Recognition," *IEEE Transactions Image Processing*, Vol. 18, No. 8, pp. 1885-1886, 2009.

[3] J. Sivic and A. Zisserman, "Efficient Visual Search Cast as Text Retrieval," *IEEE Transactions Pattern Analysis and Machine Intelligence*, Vol. 31, No. 4, pp. 591-606, 2009.

[4] L. Fei-Fei and P. Perona, "A Bayesian Hierarchical Model for Learning Natural Scene Categories," *Proceeding of 2005 IEEE*

Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 524-531, 2005.

[5] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories," *Proceeding of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 2169-2178, 2006.

[6] Z.S. Li, J. Imai, and M. Kaneko, "Robust Face Recognition Using Block-based Bag of Words," *Proceeding of International Conference on Pattern Recognition*, pp. 1285-1288, 2010.

[7] Z.S. Li, J. Imai, and M. Kaneko, "Block-Based Bag of Words for Robust Face Recognition under Variant Conditions of Facial Expression," *Illumination, and Partial Occlusion, IEICE Transactions on Fundamentals*, Vol. 94, No. 2, pp. 533-541, 2011.

[8] Le An, M. Kafai, and B. Bhanu, "Face Recognition in Multi-Camera Surveillance Videos using Dynamic Bayesian Network," *Proceeding of International Conferenceon Distributed Smart Cameras*, pp.1-6, 2012.

[9] K. Mikolajczyk and C. Schmid, "A Performance Evaluation of Local Descriptors," *IEEE Transactions Pattern Analysis and Machine Intelligence*, Vol. 27, No. 10, pp. 1615-1630, 2005.

[10] T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression Database," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 12, pp. 1615-1618, 2003.

[11] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, "The FERET Evaluation Methodology for Face Recognition Algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 10, pp. 1090-1104, 2000.

[12] G.B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*, Technical Report 07-49, University of Massachusetts, Amherst, Vol. 1, No. 2, pp. 3, 2007.

[13] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. I-511-I518, 2001.

[14] J.Y. Choi, Y.M. Ro, and K.N. Plataniotis, "Color Local Texture Features for Color Face Recognition," *IEEE Transactions on Image Processing*, Vol. 21, No. 3, pp. 1366-1380, 2012.

[15] K. Mikolajczyk and C. Schmid, "Scale and Affine Invariant Interest Point Detectors," *International Journal of Computer Vision*, Vol. 60, No. 1, pp. 63-86, 2004.

[16] A. Vedaldi and B. Fulkerson, "VLFeat-An Open and Portable Library of Computer Vision Algorithms," *Proceedings of the 18th ACM International Conference on Multimedia*, pp. 1469-1472, 2010.

[17] S. Hua1, G. Chen, H. Wei, and Q. Jiang, "Similarity Measure for Image Resizing Using SIFT feature," *EURASIP Journal on Image and Video Processing*, Vol. 6, pp. 1-11, 2012.

[18] D.G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91-110, 2004.

[19] P. Gehler and S. Nowozin, "On Feature Combination for Multiclass Object Classification," *Proceeding of 2009 IEEE 12th International Conference on Computer Vision*, pp. 221-228, 2009.

[20] M. Brown, G. Hua, and S. Winder, "Discriminative Learning of Local Image Descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 1, pp. 43-

57, 2011.

[21] S.J. Lee, C.M. Oh, and C.W. Lee, "Improved Face Recognition based on 2D-LDA using Weighted Covariance Scatter", *Journal of Korea Multimedia Society,* Vol. 17, No. 12, pp. 1446-1452, 2014.

[22] J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, "Face Recognition Using Kernel Direct Discriminant Analysis Algorithms," *IEEE Transactions on Neural Networks*, Vol. 14, No. 1, pp. 117-126, 2003.

[23] M.A. Turk and A.P. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71-86, 1991.

[24] Leung, Thomas, Malik, and Jitendra, "Representing and Recognizing the Visual Appearance of Materials Using Three-dimensional Textons," *International Journal of Computer Vision*, Vol. 43, No. 1, pp. 29-44, 2001.

[25] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual Categorization with Bags of Keypoints," *Proceeding of ECCV Workshop on Statistical Learning in Computer Vision*, Vol. 1, No. 1-22, pp. 1-2, 2004.

[26] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 12, pp. 2037-2041, 2006.

[27] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical Ensemble of Global and Local Classifiers for Face Recognition," *IEEE Transactions on Image Processing*, Vol. 18, No. 8, pp. 1885-1896, 2009.

[28] S. Xie, S. Shan, X. Chen, and J. Chen, "Fusing Local Patterns of Gabor Magnitude and Phase for Face Recognition," *IEEE Transactions on Image Processing*, Vol. 19, No. 5, pp. 1349-1361, 2010.

[29] X. Tan and B. Triggs, "Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions," *IEEE Transactions on Image Processing*, Vol. 19, No. 6, pp. 1635-1650, 2010.

[30] W. Hwang, H. Wang, H. Kim, S.C. Kee, and J. Kim, "Face Recognition System Using Multiple Face Model of Hybrid Fourier Feature under Uncontrolled Illumination Variation," *IEEE Transactions on Image Processing*, Vol. 20, No. 4, pp. 1152-1165, 2011.

[31] X. Tan and B. Triggs, "Fusing Gabor and LBP Feature Sets for Kernel-based Face Recognition," *Proceeding of International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 235-249, 2007.

[32] C. Chan, M. Tahir, J. Kittler, and M. Pietikainen, "Multiscale Local Phase Quantisation for Robust Component-based Face Recognition using Kernel Fusion of Multiple Descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 5, pp. 1164-1177, 2013.

[33] C. Ding, J.H Choi, D. Tao, and L.S. Davis, "Multi-directional Multi-Level Dual-Cross Patterns for Fobust Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 38, No. 3, pp. 518-531, 2016.

[34] S. Liao and A.K. Jain, "Partial Face Recognition: An Alignment Free Approach," *Proceeding of International Joint Conference on Biometrics*, pp. 1-8, 2011.

[35] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the Face Recognition Grand Challenges," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 947-954, 2005.

[36] S.I. Choi, "Feature Generation Method for Low-Resolution Face Recognition", *Journal of Korea Multimedia Society*, Vol. 18, No. 9, pp. 1039-1046, 2015.

김 경 태

2016년 중원대학교 학사
2017년~현재 한국외국어대학교 컴퓨터.전자공학부 석사
관심분야 : 머신러닝, 패턴인식, 영상처리

최 재 영

2011년 KAIST 전기및전자공학과 박사
2008년~2009년 토론토대학 연구원
2011년~2012년 토론토대학 연구원
2012년~2013년 펜실베니아대학 연구원
2013년~2014년 삼성전자 책임연구원
2014년~2016년 중원대학교 의료공학과 조교수
2016년~현재 한국외국어대학교 컴퓨터.전자공학부 조교수
관심분야 : 딥 러닝 기반 안면인식, 머신러닝, 패턴인식, 영상처리