Ensemble of Deep Convolutional Neural Networks With Gabor Face Representations for Face Recognition

Jae Young Choi, *Member, IEEE*, and Bumshik Lee, *Member, IEEE*

***Abstract*— Most DCNN-based FR approaches typically employ grayscale or RGB color images as input representations of DCNN architectures. However, other effective face representation methods have been developed and incorporated into current practical FR systems. In light of this fact, the focus of our study is to employ Gabor face representations in the design of DCNN-based FR frameworks to improve FR performance. To this end, we develop a novel "Gabor DCNN (GDCNN) ensemble" method that effectively applies different and multiple Gabor face representations as inputs during the training and testing phases of a DCNN for FR applications. The proposed GDCNN ensemble method primarily consists of two parts: 1) GDCNN ensemble construction and 2) GDCNN ensemble combination. The goal of the former part is to build an ensemble of GDCNN members (i.e., base models), each learned with a particular type of Gabor face representation. The objective of the latter part is to adaptively combine multiple FR outputs of individual GDCNN members. We perform extensive experiments to evaluate our proposed method on four public face databases (DBs) using the associated standard evaluation protocols. Experimental results demonstrate that our approach exhibits significantly better FR performance than typical DCNN-based approaches that rely only on grayscale or color face images as input representations. In addition, the feasibility of our proposed GDCNN ensemble has been successfully demonstrated by making comparisons with other state-of-the-art DCNN-based FR methods.**

***Index Terms*— Deep convolutional neural network (DCNN), Gabor face representations, Gabor DCNN (GDCNN) ensemble, face recognition (FR), confidence based majority voting.**

I. Introduction

FACE recognition (FR) has been an active area of pattern recognition and computer vision research, owing to its numerous and practical applications such as biometric identification, human-computer interaction, video surveillance, *etc*. FR generally comprises two tasks [1]: face identification and verification. The goal of face identification is to identify an unknown face in an image, while face verification involves verifying the claimed identity of a face. It must be noted that our study focuses on the development of a face identification algorithm.

Recent state-of-the-art FR methods are characterized by deep convolutional neural networks (DCNNs) [11]–[15]. DCNNs are known to characterize large data variations and learn compact and discriminative feature representations when the size of the training data is sufficiently large. Note that almost all DCNN-based FR methods perform training and testing by inputting facial images in the form of *grayscale* or *RGB* image representations [10]–[15]. Meanwhile, numerous effective face representation methods (such as Local Binary Pattern (LBP) and Gabor texture representations [2]–[5]) have been developed and applied to FR systems. In particular, it has been widely accepted that Gabor face representations [4], [5] exhibit good capabilities in dealing with facial appearance variations such as illumination, resolution, and pose changes. In [52], it was observed that some filters from shallow layers in specific DCNNs (e.g., ImageNet) are similar to Gabor filters; however, a typical DCNN approach (based on grayscale or RGB color input images) **may not explicitly capture the properties of several Gabor face representations** such as spatial localization, orientation selectivity, and spatial frequency selectivity in output feature maps. Moreover, such an approach **may not enhance the robustness of learned Gabor features** in terms of transition, scale changes, and rotations. To our knowledge, such properties have not been thoroughly explored in popular DCNN architectures devised for FR. This motivates us to examine the **explicit use of different and complementary Gabor face representations** to learn more robust features in the design of DCNN-based FR framework. To this end, we propose a novel "Gabor DCNN (GDCNN) ensemble" FR method that **exploits various Gabor face representations as inputs during training and testing phases of the DCNN ensemble**.

Several studies have suggested integrating Gabor filters with DCNN architectures. In particular, our study is partly related to the ideas presented in [53], [54] in which only a single "fixed" Gabor filter was incorporated into the first or second convolution layer to extract Gabor features, reducing the training complexity of DCNNs. However, **we assume a different approach by utilizing multiple and different**

Manuscript received July 10, 2019; accepted November 17, 2019. Date of publication December 18, 2019; date of current version January 28, 2020. This work was supported by the Hankuk University of Foreign Studies Research Fund, in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant 2018R1D1A1A09082615, in part by the Marine Industry Technology Development Project, named "the Development of Artificial Intelligence based Satellite Image Restoration and Prediction Technology" under Grant 20190228. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Raja Bala. *(Corresponding author: Bumshik Lee.)*

J. Y. Choi is with the Pattern Recognition and Machine Intelligence (PMI) Laboratory, Division of Computer & Electronic Systems Engineering, Hankuk University of Foreign Studies - Global Campus, Yongin 17305, South Korea.

B. Lee is with the Multimedia Information Processing Laboratory, Department of Information and Communications Engineering, Chosun University, Gwangju 61452, South Korea (e-mail: bslee@chosun.ac.kr).

Digital Object Identifier 10.1109/TIP.2019.2958404

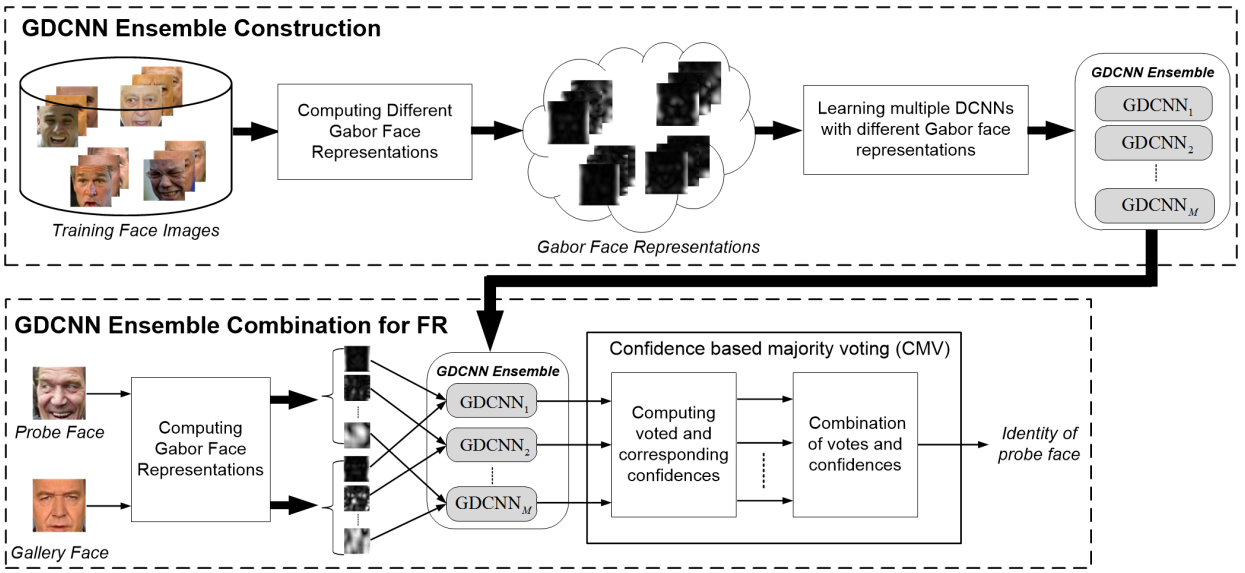


Fig. 1. Overview of the proposed Gabor DCNN (GDCNN) ensemble method for FR.

kinds of Gabor face representations (instead of using a single fixed Gabor filter) **as inputs to numerous DCNNs within the ensemble**, with the aim to (a) discover various discriminative patterns in the associated output feature maps for FR beyond grayscale or RGB image inputs (b) enhance the robustness of learned activation features to variations in illumination, expression, pose, etc. (see Fig. 8 for justifying this advantage).

Figure 1 presents an overview of the proposed Gabor DCNN (GDCNN) ensemble method for FR. As shown in Figure 1, our proposed method consists of two primary components: (a) GDCNN ensemble construction and (b) GDCNN ensemble combination. In GDCNN ensemble construction, different types of Gabor filters obtained by altering parameters such as scales and orientations are used to derive multiple and different Gabor face representations, each of which include a two-dimensional (2D) image formation. Each different Gabor face representation is then separately employed to learn a corresponding DCNN, generating a so-called *GDCNN ensemble*. The objective of our GDCNN ensemble combination is to combine multiple FR outputs obtained from individual GDCNNs in an effective and adaptive manner. For this purpose, we develop a novel confidence-based majority voting (CMV) scheme that is designed to consider the number of votes for an identity (class) label (received from the ensemble of GDCNNs) and the associated confidence values of these votes during the combination of FR outputs. By fusing the number of votes and corresponding confidence values, the identity of a given probe (test) face image can be determined. We performed extensive and comparative experiments to test our proposed GDCNN ensemble method on four public face databases (DBs): FERET [6], CAS-PEAL-R1 [7], LFW [8], and MegaFace [56]. The experimental results demonstrate that our proposed approach considerably improves FR performance compared to the commonly used approaches that learn a DCNN with grayscale or color face images as input representations. In addition, our method achieves competitive or better FR performance than that of the state-of-the-art FR approaches.

The remainder of this paper is organized as follows: Section II reviews previous studies on deep learning-based face recognition. Section III describes the proposed GDCNN ensemble framework for FR and Section IV explains its implementation details. In Section V, we present extensive and comparative experimental results that demonstrate the effectiveness of the proposed method. Finally, a discussion of the findings and the concluding remarks are presented in Section VI.

II. Related Work

There has been significant progress in the development of face recognition owing to the ground-breaking success of the applications of deep learning in computer vision and pattern recognition. Prior to deep learning, FR methods were primarily based on the well-designed handcrafted feature extraction process. Most FR methods that employ handcrafted features locally extract shallow features from facial images. In particular, LBP [3] and Gabor wavelets [4], [5], [9] are some of the most popular examples of the handcrafted features of FR. In particular, the features of *Gabor wavelets* have been proven to be highly discriminative for FR because of the different levels of locality [5], [9].

Recent, state-of-the-art FR methods have been dominated by deep convolutional neural networks (DCNN). We briefly review several recent studies on DCNNs for FR. A pioneering concept, called DeepFace, that uses DCNNs for face verification problems was proposed in [11]. The Siamese network architecture, which consists of two identical DCNNs whose inputs are a pair of two face images to be distinguished, was employed. Two high-level features extracted from the two DCNNs are used for metric learning based on the L_2-norm distance of the two extracted features. DeepID [12], [13] takes

advantage of more than 200 DCNNs to form the so-called multi-scale DCNNs for FR. Thanks to such a sophisticated structure, DeepID exhibits state-of-the-art performance for face verification and identification on the public datasets, including LFW [8]. FaceNet [14] was developed by Google researchers who proposed the so-called triplet loss on sampled triplet face images, including a pair of images from the same person (subject) and an image from different persons. The aim of triplet loss is to make the images from the same person appear closer than the ones from different persons in terms of the Euclidean distance for the purpose of face verification. The authors of VGG-Face [15] implemented triplet loss on very deep networks and trained the deep networks on the dataset collected by their proposed protocol with approximately 2.7 million images spanning 2622 celebrities. The aforementioned studies essentially demonstrate the usefulness of the DCNN models in attaining high-level FR performance.

It is important to note that most current DCNN-based FR approaches focus only on the use of grayscale or *RGB* color face images as input representations. As such, the manner in which DCNNs handle different, multiple input representations (for improved FR) still remains an open problem in the field of FR. The possibility of employing different input representations for learning DCNNs have been recently considered in several research works [16], [17]. The well-known LBP texture representations for training a set of DCNNs with different architectures for emotion recognition was proposed in [16]. Instead of using raw grayscale values, Gabor filtered features of a face image were combined with a convolutional neural network for face detection in [17].

Differing from other recent DCNN-based FR approaches [11]–[15], the primary contributions of our work are summarized as follows:

- Instead of using grayscale or color input representations for FR, we propose the *use of Gabor face representations* to learn an ensemble of DCNNs and to execute DCNN-based ensemble FR.
- We construct an ensemble of Gabor DCNNs (GDCNNs) using multiple and various types of Gabor face representations as inputs, which is different from previous approaches [53], [54] of equipping a "fixed" Gabor filter (kernel) within a DCNN architecture. Our approach can be useful for learning different and complementary DCNN models for a given FR task.
- It must be noted that the ensemble combination approaches in FR with deep learning [55] are limited to simple averaging or majority voting. In light of this fact, **making an optimal combination of the outputs of DCNNs in an ensemble is yet to be explored**. In this context, we exploit the effective "weighted fusion" mechanism, particularly, when multiple and complementary FR outputs are obtained from our proposed GDCNN ensemble, which results in significantly enhanced FR performance on challenging face datasets (see the results in Fig. 6).

In what follows, we explain in detail our proposed method.

III. Proposed Gabor DCNN (GDCNN) Ensemble Framework for FR

In this section, we first present our proposed approach of GDCNN ensemble construction based on different Gabor face representations, and then present a detailed discussion on the proposed combination approach of multiple FR outputs. Finally, we present the implementation details of the proposed method.

A. Computing Gabor Face Representations

To compute Gabor face representations of a given face image $\mathbf{I}$, Gabor filters [4], [5], [18] can be used to detect the amplitude of the spatial frequencies of pixel gray values. Without loss of generality, we assume that $\mathbf{I}$ is a 2D grayscale image. A set of 2-D Gabor filters is defined as follows [4]:

$$\Psi_{u,v}(z)=\frac{\|k_{u,v}\|^2}{\sigma^2}e^{(-\|k_{u,v}\|^2\|z\|^2/2\sigma^2)}\times\left[e^{ik_{u,v}z}-e^{-\sigma^2/2}\right] \quad (1)$$

where u and v define the orientation and the scale of the Gabor filters, $z=(x,y)$ denotes the pixel location, $||\cdot||$ denotes the norm operator, $k_{u,v}=k_v e^{i\Phi_u}$, $k_v=k_{\max}/f^v$, $\Phi_u=\pi u/8$, $k_{\max}$ denotes the maximum frequency, and f denotes the spacing factor between the filters in the frequency domain [4]. Note that the Gabor filters in Eq. (1) can assume a variety of different forms, along with different U orientations and V scales, exhibiting different types of Gabor face representations.

Using the Gabor filter $\Psi_{u,v}(z)$, Gabor face representations with a particular u and v are obtained using the following operation [4], [5]:

$$\text{Gabor}_{(u,v)}(z)=\mathrm{I}(z)*\Psi_{u,v}(z) \quad \text{for } 0\le u\le U-1,\quad 0\le v\le V-1 \quad (2)$$

where $*$ denotes the convolution operator and $\text{Gabor}_{(u,v)}(z)$ is a Gabor face representation (with orientation u and scale v) of a given face image $\mathbf{I}$. Figure 2 presents a visualized illustration of the process of obtaining Gabor face representations of a face image. Assuming that U orientations and V scales are given, we obtain a total of UV Gabor filters, yielding a set of UV Gabor face representations $\left\{\text{Gabor}_{(u,v)}(z)|0\le u\le U-1, 0\le v\le V-1\right\}$. For better readability, we will simply denote UV as M in the remaining text. To construct the GDCNN ensemble, a set of M Gabor face representations $\{\text{Gabor}_k(z)\}_{k=1}^{M}$ are individually applied to pretrained DCNN models as inputs.

B. Gabor DCNN (GDCNN) Ensemble Construction

Before presenting a description of the proposed GDCNN ensemble construction, we provide a brief review of the DCNN model for completeness. DCNN is quite useful in several applications, particularly, in image related tasks such as image classification, recognition, segmentation, etc. [19]. The primary components of a typical DCNN model include a convolution layer, a pooling layer, and a fully connected (FC) layer. The convolution layer extracts a feature map by

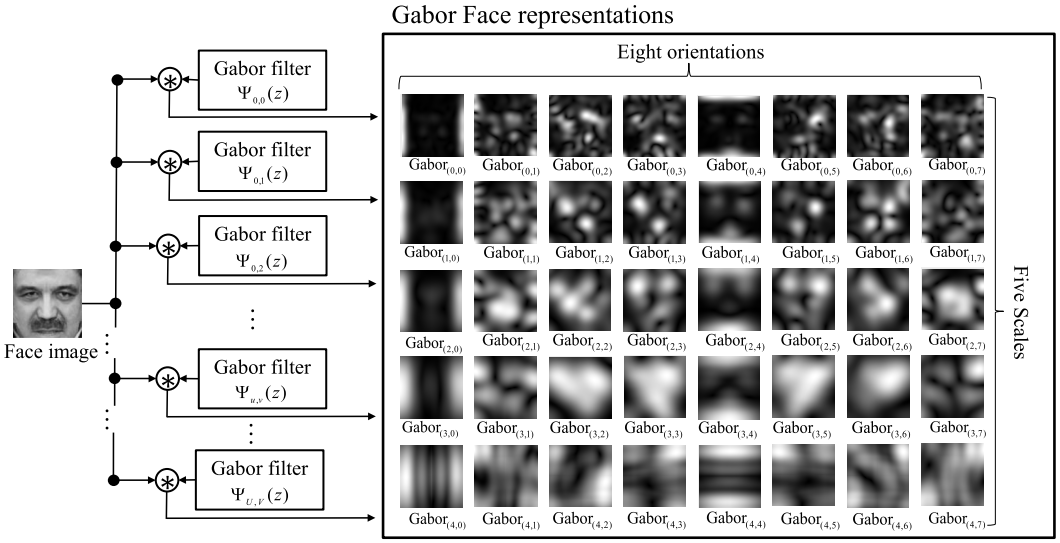


Fig. 2. Illustration of Gabor face representations, denoted by $\text{Gabor}_{(u,v)}$ [shown in Eq. (2)], with eight orientations ($U = 8$) and five scales ($V = 5$) of a face image, producing a total of $UV = 40$ Gabor face representations. Note '$*$' denotes the convolution operator.

performing a convolution operation on the input map as follows [19], [20]

$$\mathrm{O}_j = \max\left(0, b_j + \sum_i \Phi_{ij} * \mathrm{X}_i\right) \tag{3}$$

where X_i and O_j are the i-th input map and the j-th output map, respectively, Φ_{ij} is the convolution kernel between X_i and O_j, b_j is a bias term, and $*$ denotes convolution.

Based on the feature (activation) map extracted from the convolution layer, the pooling layer aims to extract strong features while ignoring weak features using dimension reduction and input space abstraction via sub-sampling. The FC layer refers to a layer in which the computation of any element in the output requires all elements in the input. The goal of the FC layer is to use all distributed representations (features) in the current layer to build features with stronger capabilities in the next layer. In DCNN, the last layer usually includes a softmax layer for obtaining the classification or recognition results. Once a DCNN model has been learned for one specific task, it can be generalized to other tasks by fine-tuning this pre-trained model on the target datasets [21].

To create GDCNN ensemble members, we propose the use of different and complementary Gabor face representations. To this end, M Gabor face representations $\{\text{Gabor}_k\}_{k=1}^{M}$ are obtained by transforming an original *RGB* (or grayscale) face image into a set of Gabor face representations with different combinations of scale and orientation parameters as shown in Eq (2), which generates a total of M individual training sets−each corresponding to one of the M available Gabor face representations. Each training set with a specific Gabor face representation is used as the input for learning an associated GDCNN, denoted by $\mathbf{M}_k\,(k = 1, \ldots, M)$, as an ensemble member. Let $\mathbf{T}_k (k = 1, \ldots, M)$ denote one of the M available training sets, organized for learning a particular GDCNN member, $\mathrm{X}_{\text{gabor}}^{(i)} \in \mathbf{T}_k$ denote the i-th training Gabor face representation sample, $\mathrm{y}_i = [y_i^{(1)}, y_i^{(2)}, \ldots, y_i^{(G)}]$ denote a true label vector of $\mathrm{X}_{\text{gabor}}^{(i)}$ with only one element being 1 at the true class (identity) label position and the others being 0, and G denote the total number of identity classes to be recognized. A corresponding GDCNN $\mathbf{M}_k$ is then learned by minimizing the following "softmax" loss function [19][1]:

$$-\frac{1}{|\mathbf{T}_k|}\sum_{\mathrm{X}_{\text{gabor}}^{(i)} \in \mathbf{T}_k} \mathrm{y}_i \cdot \log(\hat{y}_i) \tag{4}$$

where $|\cdot|$ denotes the cardinality of the set and $\cdot$ is the dot product of the two vectors. In Eq. (4), $\hat{y}_i$ denotes the probability estimation of GDCNN on $\mathrm{X}_{\text{gabor}}^{(i)}$ over a total of G identity classes:

$$\begin{aligned}\hat{y}_i &= \begin{bmatrix} p\,(y_i^{(1)} = 1|\mathrm{X}_{\text{gabor}}^{(i)};\ \mathbf{W}) \\ p\,(y_i^{(2)} = 2|\mathrm{X}_{\text{gabor}}^{(i)};\ \mathbf{W}) \\ \vdots \\ p\,(y_i^{(G)} = G|\mathrm{X}_{\text{gabor}}^{(i)};\ \mathbf{W}) \end{bmatrix} \\ &= \frac{1}{\sum_{m=1}^{G} e^{\mathrm{W}_m^{\mathrm{T}} O(\mathrm{X}_{\text{gabor}}^{(i)})}} \begin{bmatrix} e^{\mathrm{W}_1^{\mathrm{T}} O(\mathrm{X}_{\text{gabor}}^{(i)})} \\ e^{\mathrm{W}_2^{\mathrm{T}} O(\mathrm{X}_{\text{gabor}}^{(i)})} \\ \vdots \\ e^{\mathrm{W}_G^{\mathrm{T}} O(\mathrm{X}_{\text{gabor}}^{(i)})} \end{bmatrix}\end{aligned} \tag{5}$$

where $\mathbf{W}$ denotes the weight connecting the last FC layer to the softmax layer, $O(\mathrm{X}_{\text{gabor}}^{(i)})$ denotes the output of the last FC layer for input $\mathrm{X}_{\text{gabor}}^{(i)}$, W_m denotes the m-th column vector of $\mathbf{W}$, $\mathrm{W}_m^{\mathrm{T}} O(\mathrm{X}_{\text{gabor}}^{(i)}), m = 1, \ldots, G$, are inputs of the softmax layer, and T denotes the transpose operator.

Figure 3 presents a visualization of our proposed construction approach of GDCNN ensemble members using different Gabor face representations. Our approach may be beneficial

[1]Note that any other loss functions such as CosFace [60] can also be applied to the construction of our GDCNN ensemble. Please refer to the results in Table IV.

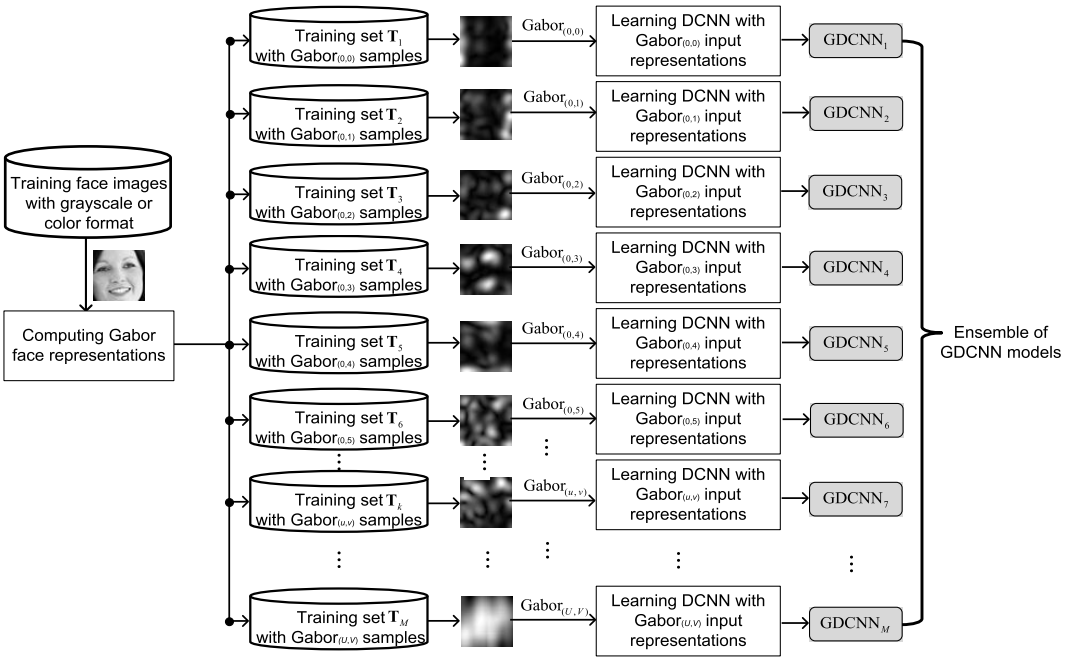


Fig. 3. Visualized illustration for generating an ensemble of GDCNN models using different Gabor face representations. For illustration purposes, a training face sample and its corresponding Gabor face representation Gabor$_{(u,v)}$ are depicted. The subscript (u,v) indicates the parameters used to obtain numerous Gabor$_{(u,v)}$ in the following format: (*orientation*, *scale*). Note that $UV = M$.

for learning different and complementary discriminant deep features from DCNN models because learning DCNNs on various Gabor face representations can increase their representative capability in terms of FR performance improvements. To the best of our knowledge, our proposed approach is the first in which FR performance is improved using an **ensemble of DCNNs learned with different Gabor face representations**.

As demonstrated in our experiment (see Table I and II) in Section V, using Gabor face representations can be advantageous for boosting FR performances, which is beyond the case of using only *RGB* (or grayscale) as the input space for learning DCNNs. An ensemble of M GDCNN models is applied to a combination module that adaptively fuses the recognition outputs of all GDCNN ensemble members together.

C. GDCNN Ensemble Combination for FR

The proposed GDCNN ensemble combination is developed based on majority voting [22], which is the most popular ensemble combination method [23], [24]. In conventional majority voting, every classifier in an ensemble casts a vote for one class label, and the final output class label is the one that receives the maximum number of the votes. Differing from this approach, our method accounts for the *votes* of identity (class) labels and the corresponding *confidences* to adaptively combine the outputs of all GDCNN members for improved recognition performance. A Key advantage of our method is to supply more power to the stronger GDCNN ensemble members with higher FR performance when they cast their votes.

We now present the proposed GDCNN ensemble combination procedure. Let $\{\mathbf{G}^{(n)}\}_{n=1}^{G}$ be a gallery set consisting of prototype enrolled face images of G distinct identities. Let P be an unknown face image that must be recognized, which is denoted as a probe. In addition, we denote Gabor face representations of $\{\mathbf{G}^{(n)}\}_{n=1}^{G}$ and P as $\{\mathrm{G}_{\mathrm{gabor}}^{(n)}\}_{n=1}^{G}$ and $\mathrm{P}_{\mathrm{gabor}}$, respectively. In the proposed method, to compute "the number of votes" and the corresponding "confidence values" for a particular identity label, the posterior probability must be computed for each GDCNN:

$$p(\ell(\mathrm{P}_{\mathrm{gabor}}) = \ell(\mathrm{G}_{\mathrm{gabor}}^{(n)})|\mathbf{M}_k; \mathbf{W}_k) = \frac{e^{\mathrm{W}_{k,n}^{\mathrm{T}}\, O(\mathrm{P}_{\mathrm{gabor}})}}{\sum_{m=1}^{G} e^{\mathrm{W}_{k,m}^{\mathrm{T}}\, O(\mathrm{P}_{\mathrm{gabor}})}} \quad \text{for } n = 1, \ldots, G \tag{6}$$

where $\ell(\cdot)$ denotes a function that returns an identity label of an input Gabor face representation, $\mathbf{W}_k$ denotes the weight of the last FC layer of k-th GDCNN $\mathbf{M}_k$, $e^{\mathrm{W}_{k,n}^{\mathrm{T}}\, O(\mathrm{P}_{\mathrm{gabor}})}$ denotes the total input into a softmax layer of each GDCNN, which is computed using the output $O(\mathrm{P}_{\mathrm{gabor}})$ of the FC layer and the associated weight $\mathrm{W}_{k,n}$, connecting the FC layer to the softmax layer; readers can refer to [19], [25] for more detailed calculations. In Eq. (6), $p(\ell(\mathrm{P}_{\mathrm{gabor}}) = \ell(\mathrm{G}_{\mathrm{gabor}}^{(n)})|\mathbf{M}_k; \mathbf{W}_k)$ denotes the probability that the identity label of $\mathbf{P}$ is determined as the same identity label of $\mathbf{G}^{(n)}$, assuming that $\mathrm{P}_{\mathrm{gabor}}$ is forwarded to the k-th GDCNN for FR tasks. In the remainder of this paper, for simplicity, we will refer to the $p(\ell(\mathrm{P}_{\mathrm{gabor}}) = \ell(\mathrm{G}_{\mathrm{gabor}}^{(n)})|\mathbf{M}_k; \mathbf{W}_k)$ as "confidence $c_k^{(n)}$" [26] which measures the degree of belief to which the vote for an identity label of $\mathbf{P}$ (received from $\mathbf{M}_k$) is equal to the correct vote (as its true identity label) of the n^{th} gallery face $\mathbf{G}^{(n)}$.

Based on the confidence values $c_k^{(n)}, k = 1, \ldots, M$, we calculate the number of votes (as FR outcome) for a particular identity label. Let $N_{\text{vote}}^{(n)}$ be the total number of votes given to the identity label of $\mathbf{G}^{(n)}$, received from all available GDCNNs $\{\mathbf{M}_k\}_{k=1}^{M}$:

$$N_{\text{vote}}^{(n)} = \sum_{k=1}^{M} \Delta_k^n(\mathrm{P}_{\text{gabor}}, \mathrm{G}_{\text{gabor}}^{(n)}) \quad \text{for } n = 1, \ldots, G \tag{7}$$

and

$$\Delta_k^n(\mathrm{P}_{\text{gabor}}, \mathrm{G}_{\text{gabor}}^{(n)}) = \begin{cases} 1 & \text{if } n = \arg\max\limits_{i=1}^{G} c_k^{(i)} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where $\Delta_k^n(\mathrm{P}_{\text{gabor}}, \mathrm{G}_{\text{gabor}}^{(n)})$ is an indicator function that returns one when the maximum of the posterior probability values between $\mathrm{P}_{\text{gabor}}$ and $\mathrm{G}_{\text{gabor}}^{(n)}$ $(n = 1, \ldots, G)$ is achieved at $i = n$, and returns zero otherwise. To account for the belief that $N_{\text{vote}}^{(n)}$ can correctly recognize a given probe face image, we compute the so-called "total confidence value" associated with $N_{\text{vote}}^{(n)}$:

$$C_{\text{conf}}^{(n)} = \sum_{k=1}^{M} \Delta_k^n(\mathrm{P}_{\text{gabor}}, \mathrm{G}_{\text{gabor}}^{(n)})\ c_k^{(n)}. \tag{9}$$

It must be noted that in Eq. (9), $C_{\text{conf}}^{(n)}$ is the sum of confidence values $c_k^{(n)}(k = 1, \ldots, M)$ for the vote of the identity label of $\mathbf{G}^{(n)}$, where the vote has been received from each $\mathbf{M}_k$. As such, $C_{\text{conf}}^{(n)}$ denotes the total degree of belief that the identity label of $\mathbf{P}$ is assigned to that of $\mathbf{G}^{(n)}$.

By using the value of $C_{\text{conf}}^{(n)}$, we can enhance the importance of $N_{\text{vote}}^{(n)}$, which is produced by a subset of GDCNN ensemble members with high reliability in terms of contributing their individual FR outcomes more to the final FR result. This motivated us to propose the combined use of "total number of votes" $N_{\text{vote}}^{(n)}$ and "total confidence value" $C_{\text{conf}}^{(n)}$ to finally determine the identity of a probe face. In our method, the gallery identity label that achieves the highest combined value of $N_{\text{vote}}^{(n)}$ and $C_{\text{conf}}^{(n)}$ is used to identify a given probe face:

$$\ell(\mathbf{P}) = \ell(\mathbf{G}^{(n*)}) \text{ and}$$
$$n^* = \arg\max_{n=1}^{G} (N_{\text{vote}}^{(n)} \times C_{\text{conf}}^{(n)}). \tag{10}$$

IV. Implementation Details

Two successful and publicly available DCNN models (architectures), namely VGG-Face [15] and Lightened CNN [27], were used to implement the proposed GDCNN ensemble method. These two DCNN models were chosen because they have been determined to be successful for FR in the wild while being publicly available. The former DCNN model includes a very deep architecture and the latter is computationally efficient.

To construct an ensemble of GDCNNs as described in Section III.B, we used 40 different Gabor face representations (8 orientations × 5 scales) as inputs to individual DCNNs (i.e., "VGG-Face" or "Lightened CNN"), generating 40 GDCNN ensemble members. Each GDCNN member was implemented using MatConvNet [28] and pre-trained on the CASIA-WebFace dataset [29] comprising 494,414 face images of 10,575 subjects downloaded from the website [30]. For this purpose, we created a large-scale Gabor face representation dataset consisting of approximately 19.8M (19,776,560 = 494,414 facial images × 40 Gabor representations) Gabor images obtained from 10,575 subjects. In terms of storage costs, approximately 37.6 gigabytes of memory for a 120 x 120 Gabor representation image is required. We used the standard batch size of 128 for the training phase. The hyper-parameters of each GDCNN are the same as used by [31]: momentum 0.9; weight decay 5×10^{-4}; initial learning rate 10^{-2}, which is decreased by a factor of 10 when the validation error stops decreasing (specifically, when the number of errors increases for more than three consecutive times). Overall, each GDCNN was trained using three decreasing learning rates as suggested by [31].

Fig. 4. Examples of cropped facial images from the FETET DB. It must be noted that every facial image was manually cropped using eye coordinate information and aligned using a fixed template [6]. Best viewed in color.

V. Experiments

We evaluated our proposed GDCNN ensemble method on four widely used face datasets, including the FERET [6], CAS-PEAL-R1 [7], LFW [8], and MegaFace [56] datasets. In the following three subsections, we present comparative experimental results to demonstrate the effectiveness of the proposed method for FR under both controlled and unconstrained environments.

A. Evaluation of the FERET Dataset

The FERET DB consists of 13,539 facial images corresponding to 1,565 subjects. In our experiments, we followed the standard FERET evaluation protocol [6]. The standard FERET data set contains the data partition for recognition tests as follows. A gallery set was composed of 1,196 subjects with one image per subject and four different probe sets (**fb**, **fc**, **dup1**, and **dup2**) were used in the recognition stage. The **fb** probe set includes 1,195 images of subjects taken at the same time as gallery images with a different facial expression. The **fc** probe set includes frontal-view images captured under different illumination conditions. The **dup1** images were captured within one year of the gallery images, and **dup2** images were captured at least one year after the gallery images were acquired. Figure 4 presents some cropped example images from the FERET dataset. In terms of performance, we used the rank-one identification accuracy, i.e., the identification rate of the top response as being correct. Additional details related to the FERET testing protocol can be found in [6].

TABLE I

COMPARISONS OF RANK-1 IDENTIFICATION RATE BETWEEN OUR GDCNN ENSEMBLE AND THE STATE-OF-THE-ART FR METHODS OVER THE COURSE OF THE FREET TESTING PROTOCOL [6]. NOTE THAT IN THE PROPOSED GDCNN ENSEMBLE, A TOTAL OF 40 GABOR FACE REPRESENTATIONS ARE USED AS INPUTS TO "VGG-FACE" OR "LIGHTENED CNN" ARCHITECTURES

Method	Probe set			
	fb	fc	dup1	dup2
LBP [3]	97.0	79.0	66.0	64.0
LGBP [39]	98.0	97.0	74.0	71.0
HGPP [40]	97.5	99.5	79.5	77.8
LLGP [41]	99.0	99.0	80.0	78.0
DT-LBP [42]	99.0	100.0	84.0	80.0
DLBP [43]	99.0	99.0	86.0	85.0
POEM [44]	97.6	95.0	77.6	76.2
GOM [45]	99.9	100	95.7	93.1
LGXP [18]	99.0	99.0	94.0	93.0
Baseline DCNN (VGG-Face [15]+RGB color input)	**99.9**	**100**	95.8	93.5
Baseline DCNN (Lightened CNN [27]+RGB color input)	**99.9**	**100**	94.9	92.2
Our GDCNN ensemble (VGG-Face [15]+Gabor input)	**99.9**	**100**	**99.7**	**99.1**
Our GDCNN ensemble (Lightened CNN [27]+Gabor input)	**99.9**	**100**	98.5	97.9

The bold values denote the best result of FR methods in each probe set.
The results of other methods are from the original paper.

Table I presents comparative FR results over the course of the FERET testing protocol. It must be noted that we first make direct comparisons with other state-of-the-art results recently reported by other existing studies on the FERET DB. As such, all comparison results have been directly cited from recently published papers [3], [18], [39]–[45]. In addition, for completeness, we reported FR performances obtained using a so-called "baseline DCNN model" learned with the same "VGG-Face or Lightened CNN" architectures, but using *RGB color* or *grayscale* face images as input representations is the most typical approach when employing DCNNs for FR [10]–[15].

From the results in Table I, we have the following observations. First, DCNN-based FR approaches outperform all previous hand-crafted FR approaches by a large margin. Second, the results of our proposed method using Gabor face representations are much better than those obtained for the baseline DCNN approach that uses *RGB* color input representations for "VGG-Face" and "Lightened CNN" architectures. Finally, our GDCNN ensemble method achieves the best performances compared with all other results listed in Table I; particularly, our GDCNN ensemble yields excellent FR performance with a 99.1% identification rate for the dup2 probe set, which has been reported to be the most challenging data set in the FERET testing protocol.

B. Evaluation of the CAS-PEAL-R1 Dataset

The CAS-PEAL-R1 DB includes 30,863 face images of 1,040 subjects depicting the various characteristics of poses, expressions, accessories, and lighting conditions (PEAL). In our experiments, we adopted the standard evaluation protocol [7] where five data sets, including training, gallery, expression, lighting, and accessory were used for FR experiments. The training set comprises 1,200 face images of 300 subjects, while the gallery set comprises 1,040 images of 1,040 different subjects. Additional details of the remaining three probe sets have been described in [7]. Figure 5 presents some aligned and cropped example images from three different probe sets used in our experiments. It must be noted that a training set was used to construct our GDCNN ensemble framework—each GDCNN member was created by fine-tuning pre-trained VGG-Face or Lightened CNN models using this training set—while the gallery and the other three probe sets were used for testing.

Table II presents the rank-1 recognition rate of our GDCNN ensemble method on the CAS-PEAL-R1 probe data sets, which has been compared with the state-of-the-art FR methods. As can be observed, our proposed method outperforms the best results among existing methods; specifically, the identification rate can be improved for the accessory and lighting probe sets, respectively. Moreover, the proposed GDCNN ensemble approach demonstrates even better FR performances compared with conventional DCNN-based FR approaches in which *RGB* color or grayscale face images are used as input representations (i.e., the so- called "baseline DCNN" approach in Table II). Compared with the baseline DCNN approach, in the case of the 'Lighting' probe set, the rank-1 recognition rate can be improved by 6.5% and 5.8% points for the 'VGG-Face' and 'Lightened CNN' architectures, respectively, using Gabor face input representations. This result confirms that integrating Gabor face representations with popular DCNN architectures is indeed a good choice for improving FR performance when considering DCNN-based FR algorithms.

C. Evaluation of the LFW Dataset

We compare our proposed GDCNN ensemble method with some of the state-of-the-art FR methods on Labeled Faces in the Wild (LFW) DB [8]. The LFW dataset contains

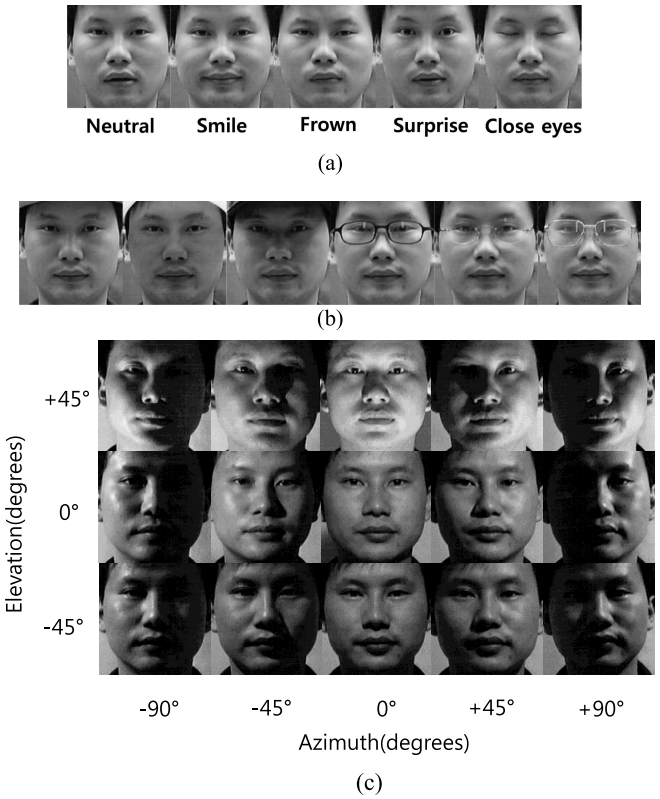


Fig. 5. Examples of cropped facial images depicting the (a) expression, (b) accessory, and (c) lighting probe sets, respectively.

more than 13,000 face images, downloaded from the Internet, of 5,749 different individuals such as celebrities, and public figures. Owing to unconstrained image acquisition environments, the LFW dataset includes significant variations in facial poses, illumination conditions, and expressions, and many of the face images are occluded [8].

For comparison, we adopted two standard identification protocols based on the LFW datasets proposed in [34]: (1) closed set and open set identification protocols. In closed set identification, the gallery set includes 4,249 identities, each with only a single example, and the probe set includes 3,143 faces that belong to the same set of identities. In open set identification, the gallery includes 596 identities, each with a single example, and the probe set includes 596 genuine probes and 9,491 imposter ones. We evaluate performance based on the Rank-1 detection and identification rate (DIR), which is a fraction of the genuine probes that are matched correctly at Rank-1 and at the false alarm rate (FAR) of the rejection process (the fraction of imposter probe images that have not been rejected). We report the 'Rank-1 identification rate' and 'DIR at 1% FAR' in the close and open set identification tasks for direct comparison with other state-of-the-art FR methods. This guarantees fair and reliable comparisons of our method against recently developed DCNN-based FR methods such as DeepFace [11].

Table III presents a comparison with the state-of-the-art methods under closed- and open-set identification tasks on the LFW dataset. As can be observed from Table III,

TABLE II

COMPARISON OF THE RANK-1 IDENTIFICATION RATE BETWEEN OUR GDCNN ENSEMBLE AND THE STATE-OF-THE-ART FR METHODS TESTED WITH THE STANDARD CAS-PEAL-R1 EVALUATION PROTOCOL

Method	Type of probe sets		
	Expression	Accessory	Lighting
LBP [3]	97.0	89.0	29.0
LGBP [39]	95.0	87.0	51.0
LVP [46]	96.0	86.0	29.0
HGGP [40]	96.0	92.0	62.0
LLGP [41]	96.0	90.0	52.0
DT-LBP [42]	98.0	92.0	41.0
DLBP [43]	99.0	92.0	41.0
DFD [47]	99.6	96.9	63.9
JFL [33]	99.7	97.2	67.4
Baseline DCNN (VGG-Face [15] + RGB color input)	**100**	97.5	69.2
Baseline DCNN (Lightened CNN [27]+ RGB color input)	**100**	96.2	67.5
Our GDCNN ensemble (VGG-Face [15]+ Gabor input)	**100**	**99.9**	**75.7**
Our GDCNN ensemble (Lightened CNN [27] + Gabor input)	**100**	**99.9**	72.3

The bold values denote the best result of FR methods in each probe set. The results of other methods are from the original paper.

TABLE III

COMPARISON WITH THE STATE-OF-THE-ART METHODS UNDER CLOSED- AND OPEN-SET IDENTIFICATION TASKS ON THE LFW DATASET. NOTE THAT IN THE PROPOSED METHOD, "VGG-FACE" [15] WAS USED TO CONSTRUCT AN ENSEMBLE OF GDCNNS

Method	Rank-1(%)	DIR @ 1% FAR(%)
COTS-s1 [34]	56.7%	25%
COTS-s1+s4 [34]	66.5%	35%
High Dimensional LBP [48]	18.1%	7.89%
DeepFace [11]	64.9%	44.5%
VGG-Face [15]	74.1%	52%
web-scale training Fusion [49]	82.5%	61.9%
DeepID2+ [13]	95.0%	80.7%
Noisy Softmax [50]	92.6%	78.4%
C2D-CNN [51]	91.9%	63.3%
Proposed GDCNN Ensemble	93.6%	77.1%

the proposed GDCNN ensemble achieves better or comparable FR performance compared with other deep learning-based FR methods. For the close- and open-set identification protocols, our method can achieve state-of-the-art results of 93.6% and 77.1%, respectively. These results indicate that our method outperforms other state-of-the-art methods, except for DeepID2+ [13]. It must be noted that DeepID2+ is a model ensemble consisting of a large number of deep networks trained with millions of images from private datasets, whereas we use only the CASIA dataset for training our GDCNN ensemble, which has less than 500K images. As such, direct comparisons with DeepID2+ might be unfair, owing to limitations in the computation of resources and the amount of training data.

TABLE IV

COMPARISON OF RANK-1 FACE IDENTIFICATION ACCURACY (WITH 1M DISTRACTORS) OF DIFFERENT METHODS UNDER MEGAFACE CHALLENGE 1 USING FACESCRUB AS THE PROBE SET. NOTE THAT IN OUR METHOD, "VGG-FACE" [15] WAS USED TO CONSTRUCT AN ENSEMBLE OF GDCNNS

Method	Rank-1(%)
Triplet [14]	64.79
Center Loss [58]	65.49
SphereFace [59]	72.729
CosFace [60]	77.11
AM-Softmax [61]	72.47
SphereFace+ [62]	73.03
ArcFace [63]	77.50
Our GDCNN Ensemble with Softmax	76.85
Our GDCNN Ensemble with CosFace	78.42
Our GDCNN Ensemble with ArcFace	80.09

The high performance results obtained using our proposed method are because of the following reasons: (1) the application of Gabor face representations (as input) can make the DCNN-based FR approach more robust to variations in poses, illumination conditions, expressions, and occlusion (please refer to Section VI for more details on the discussion that supports this argument), (2) multiple DCNNs learned with different Gabor face representations (i.e., GDCNN ensemble) provide different discriminant information and are mutually compensational in terms of improvements in FR performance, and (3) our GDCNN ensemble combination allows for complementary face identity determination via the fusion of "the number of votes " and the corresponding "confidence value" computed from an ensemble of GDCNNs.

D. Evaluation of the MegaFace Dataset

MegaFace [56] is a very challenging testing benchmark that was recently released for large-scale face identification and verification. The gallery set in Megaface includes more than 1 million face images and the probe set comprises 106,863 face images of 530 celebrities from Facescrub [57]. In our study, we evaluate the performance of our proposed method on the Megaface Challenge 1 [56] in which the gallery set incorporates more than 1 million images from 690K subjects. For a direct and fair comparison of the existing results that use small training datasets (defined as small if it contains less than 0.5M images of 20K subjects), we train our GDCNN ensemble on the publicly available CASIA-WebFace [29] dataset, which includes approximately 0.49M face images from 10,575 subjects.

In Table IV, our proposed GDCNN ensemble−trained with softmax loss function− achieves a comparable rank-1 identification rate (76.85%) with that of the recent best performances (77.50%) reported in [63]. To validate whether our GDCNN ensemble generalizes well with other popular loss functions, we performed additional experiments by incorporating CosFace [60] or ArcFace [63] loss into the construction of our GDCNN ensemble. It must be noted that compared to softmax loss, CosFace, and ArcFace are more beneficial for enforcing a more distant gap between the nearest (hard-to-classify) identity classes [60], [63]. As shown in Table IV, our GDCNN ensemble coupled with ArcFace is ranked first for challenge 1, attaining a new state-of-the-art result with a margin of 2.59% on rank-1 identification accuracy, which further demonstrates the effectiveness of our GDCNN ensemble.

E. Computational Time

We analyzed complexity in terms of computational time on the LFW dataset. Our hardware configuration comprises Intel Xeon E5-2620 v4 CPUs, 256-GB RAM with two NVidia Titan X Pascal GPUs for acceleration. In this analysis, we employed a VGG-Face [15] network to create our GDCNN ensemble members. When measuring the training time over a dataset of around 12,000 face images collected from the LFW dataset, the time required to generate 40 GDCNN networks as an ensemble (including the computation of Gabor face representations) is approximately 104 hours, while the training time needed for a single GDCNN network is approximately 2.46 hours. However, it must be noted that the average testing time for recognition when using our GDCNN ensemble is as low as 0.071 seconds (per face image). The training of our GDCNN ensemble is computationally time-consuming compared to the single DCNN-based approaches. However, for practical FR applications, training can be performed offline and testing must be executed in a real-time manner. Moreover, training time can be greatly reduced by up to factors of 10 to 50 using multiple GPU-accelerated implementations and fully parallel implementations of the DCNN algorithm [35], [36].

VI. DISCUSSION AND CONCLUSION

We performed intensive experiments to compare our proposed ensemble combination approach (described in Section 3.3) with the commonly used ensemble (classifier) combination schemes [23], [24] such as the product rule, sum rule, min rule, max rule, median rule, and majority voting [22]. It must be noted that the aforementioned combination rules were implemented based on posterior or class probabilities defined in Eq. (5)−each computed by the respective GDCNN ensemble member. Moreover, it must be noted that conventional majority voting [22], [23] is equivalent to the special case of using only $N_{\text{vote}}^{(n)}$ with an exclusion confidence $C_{\text{conf}}^{(n)}$ in our proposed combination approach (see Eq. (10)).

Figure 6 presents the FR accuracy of several ensemble combination approaches obtained using the FERET testing protocol based on the '**dup2**' probe set. It must be noted that in Figure 6, FR based on the best single GDCNN serves as a baseline method. For this, we first generated 40 individual GDCNNs (each trained with a particular Gabor face representation) and the best single GDCNN was then selected based on the testing performances obtained using all 40 individual GDCNNs. From Figure 6, we can observe the following: 1) all ensemble combination approaches exhibit much better FR performances than the best single GDCNN; this justifies the benefit of combining numerous DCNNs learned with different Gabor face representations in terms of offering additional discriminative power and 2) the proposed

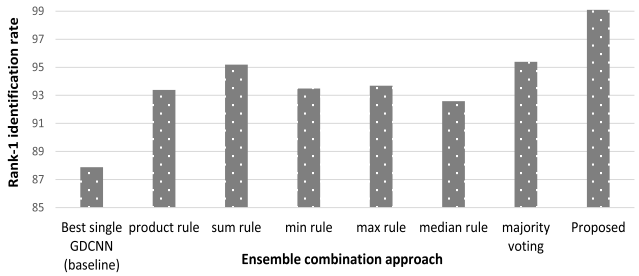


Fig. 6. Effectiveness of using our proposed GDCNN ensemble combination approach. Note that for all ensemble combination approaches, 40 GDCNN members were created using the "VGG-Face" [15] architecture coupled with 40 different Gabor representations.

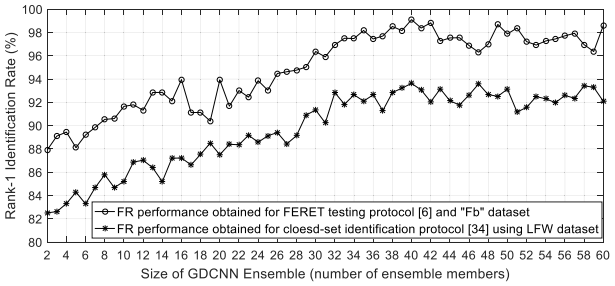


Fig. 7. Effect of the size of our GDCNN ensemble (i.e., the number of ensemble members) on FR performance.

combination approach outperforms all other ensemble combination strategies; in particular, compared to the conventional majority voting scheme, approximately 4% improvement in FR accuracy can be achieved by imposing associated confidences on "the number of votes" as proposed in our method.

Further, we examined the effect of altering the size of our GDCNN ensemble (i.e., the number of GDCNN ensemble members) on FR performance. The size of the ensemble was varied by using different combinations of orientation and scale parameters of Gabor face representation as in Eq. (1). In this study, we used ten orientation and six scale parameters, and therefore, the GDCNN ensemble size was varied in the range of [2, 60]. Figure 7 presents variations in FR performance with respect to changes in our GDCNN ensemble size. It can observed that FR performance is generally improved as the size of the ensemble increases until it saturates when the ensemble comprises approximately 40 members. It must be noted that variations in FR performance among ensembles with sizes of 40 and higher are not significant. Considering a balance between FR accuracy and the computational cost required for training, our GDCNN ensemble comprising of 40 members (i.e., eight orientations and five scales) is appropriated for FR (at least, in the data sets used in our experiments).

Prior to concluding this paper, we presented feature (activation) maps for examining the internal representations learned by our GDCNNs on FR tasks. This may be helpful to better understand why our proposed GDCNN can achieve improved FR performance compared to current DCNN-based FR approaches that rely on the use of grayscale or color input representations. Figure 8 visualizes numerous feature maps from our GDCNNs trained with the "VGG-Face" network configuration. For comparison purposes, feature maps from the VGG-Face that were trained with grayscale face images are also displayed. The conv2_1 layer of VGG-Face network [15] was chosen and then the strongest activations for this selected layer were displayed. From Figure 8, we can observe the following: (1) the activated neurons of the feature maps in our GDCNNs for the same identity are more similar than those in the baseline DCNN with grayscale input images (2) the feature maps from our GDCNNs exhibit invariances to a large extent under severe illumination and

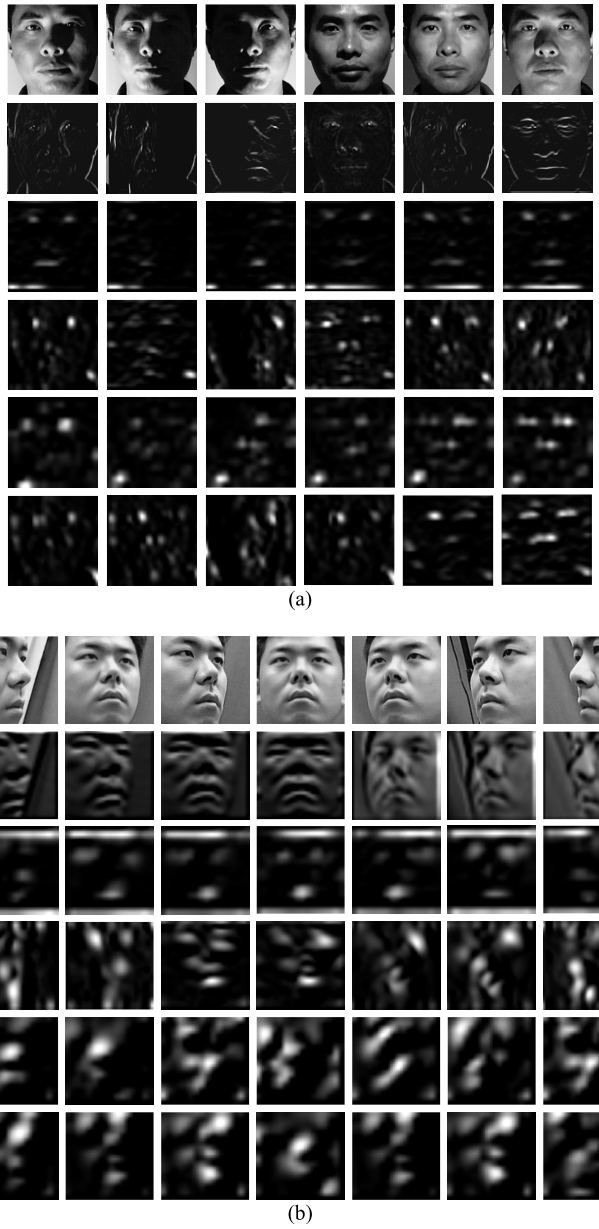


Fig. 8. Visualization of feature (activation) maps [37], [38] from a selected layer of our GDCNN trained on VGG-Face [15] network, including (a) illumination variation or (b) viewpoint (pose) variation. The strongest activations for the selected conv2_1 layer are displayed. Original face images and feature maps generated using grayscale input images (for comparison) are presented in the first and second rows, respectively, and feature maps of 5^{th}, 6^{th}, 11^{th}, 15^{th} Gabor face representations are presented in the third, fourth, fifth, sixth rows, respectively. It can be better viewed electronically when zoomed in.

pose variations. Based on these observations, the most plausible reason for superior performance of our GDCNN ensemble is that the feature maps from our GDCNNs include *very common activated neurons* even in the case where face images are captured under severe pose and illumination variations. Moreover, we determined that our GDCNNs captures ***complementary activation patterns*** from different Gabor face representations, and therefore they are likely to be *mutually compensational* for improving FR accuracy; this finding can justify the usefulness of our GDCNN ensemble approach in terms of combining a set of DCNNs learned with different Gabor face representations.

In this paper, we developed a new and novel FR solution that effectively incorporates different and multiple Gabor face representations as inputs into the popular DCNN architectures for improved FR performance. The key characteristics of our method include 1) construction of a so-called Gabor DCNN (GDCNN) ensemble whose members are learned with a particular type of Gabor face representation 2) combination of multiple FR outcomes obtained from an ensemble of GDCNNs in an effective and adaptive manner; this can be achieved by our proposed confidence-based majority voting algorithm that accounts for both the number of votes for an identity label and the corresponding confidence (belief) values during the fusion. Extensive and comparative experiments were conducted to test the proposed GDCNN ensemble on four public face DBs. Experimental results demonstrate that our method exhibits good generalization and provides a competitive face identification solution under both constrained and unconstrained FR circumstances. In addition, our method achieves state-of-the-art FR performances on standard benchmarks.

REFERENCES

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.

[2] C. Ding and D. Tao, "A comprehensive survey on pose-invariant face recognition," *ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 3, p. 37, 2016.

[3] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[4] J. Y. Choi, Y. M. Ro, and K. N. Plataniotis, "Color local texture features for color face recognition," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1366–1380, Mar. 2012.

[5] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 467–476, Apr. 2002.

[6] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.

[7] W. Gao *et al.*, "The CAS-PEAL large-scale Chinese face database and baseline evaluations," *IEEE Trans. Syst., Man, Cybern. A, Syst., Hum.*, vol. 38, no. 1, pp. 149–161, Jan. 2008.

[8] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," *Univ. Massachusetts, Amherst*, vol. 1, no. 2, pp. 7–49, Oct., 2007.

[9] P. Yang, S. Shan, W. Gao, S. Z. Li, and D. Zhang, "Face recognition using Ada-boosted Gabor features," in *Proc. 6th IEEE Int. Conf. Autom. Face Gesture Recognit.*, May 2004, pp. 356–361.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[11] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1701–1708.

[12] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1891–1898.

[13] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2892–2900.

[14] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 815–823.

[15] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. Brit. Mach. Vis. Conf.*, 2015, vol. 1. no. 3, p. 6.

[16] G. Levi and T. Hassner, "Emotion Recognition in the Wild via Convolutional Neural Networks and Mapped Binary Patterns," in *Proc. ACM Int. Conf. Multimodal Interact.*, 2015, pp. 503–510.

[17] B. Kwolek, "Face detection using convolutional neural networks and Gabor filters," in *Proc. Int. Conf. Artif. Neural Netw.*, Sep. 2005, pp. 551–556.

[18] S. Xie, S. Shan, X. Chen, and J. Chen, "Fusing local patterns of Gabor magnitude and phase for face recognition," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1349–1361, May 2010.

[19] J. Gu *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, May 2017.

[20] J. Wu, "Introduction to convolutional neural networks," Ph.D. dissertation, Nat. Key Lab Novel Softw. Technol., Nanjing Univ., Nanjing, China, 2017.

[21] J. Yosinski *et al.*, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3320–3328.

[22] J. Kittler, M. Hatef, R. P. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.

[23] Z. H. Zhou, *Ensemble Methods: Foundations and Algorithms*. Boca Raton, FL, USA: CRC Press, 2012.

[24] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Hoboken, NJ, USA: Wiley, 2004.

[25] G. Cheng, P. Zhou, and J. Han, "Rifd-cnn: Rotation-invariant and Fisher discriminative convolutional neural networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2884–2893.

[26] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[27] X. Wu, R. He, and Z. Sun, "A light CNN for deep face representation with noisy labels," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 11, pp. 2884–2896, Nov. 2018.

[28] A. Vedaldi and K. Lenc, "MatConvNet: Convolutional neural networks for MATLAB," in *Proc. 23rd ACM Int. Conf. Multimedia*, 2015, pp. 689–692.

[29] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," 2014, *arXiv:1411.7923*. [Online]. Available: https://arxiv.org/abs/1411.7923

[30] *CASIA WebFace Database*. Accessed: 2014. [Online]. Available: http://www.cbsr.ia.ac.cn/english/CASIA-WebFace-Database.html

[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

[32] P. J. Phillips *et al.*, "Overview of the face recognition grand challenge," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jun. 2005, pp. 947–954.

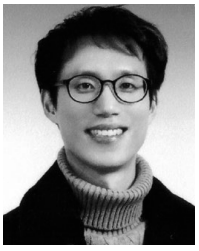
[33] J. Lu, V. E. Liong, G. Wang, and P. Moulin, "Joint Feature Learning for Face Recognition," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 7, pp. 1371–1383, Jun. 2015.

[34] L. Best-Rowden, H. Han, C. Otto, B. F. Klare, and A. K. Jain, "Unconstrained face recognition: Identifying a person of interest from a media collection," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 12, pp. 2144–2157, Dec. 2014.

[35] D. Strigl, K. Kofler, and S. Podlipnig, "Performance and scalability of GPU-based convolutional neural networks," in *Proc. 18th Euromicro Conf. Parallel, Distrib. Netw.-Based Process.*, Feb. 2010, pp. 317–324.

[36] C. Farabet, B. Martini, P. Akselrod, S. Talay, Y. LeCun, and E. Culurciello, "Hardware accelerated convolutional neural networks for synthetic vision systems," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2010, pp. 257–260.

[37] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Aug. 2014, pp. 818–833.
[38] *Visualize Activations of a Convolutional Neural Networks Using MATLAB*. Accessed: 2019. [Online]. Available: https://kr.mathworks.com/help/nnet/examples/visualize-activations-of-a-convolutional-neural-network.html
[39] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," in *Proc. 10th IEEE Int. Conf. Comput. Vis.*, vol. 1, Oct. 2005, pp. 786–791.
[40] B. Zhang, S. Shan, X. Chen, and W. Gao, "Histogram of Gabor phase patterns (HGPP): A novel object representation approach for face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 57–68, Jan. 2007.
[41] S. Xie, S. Shan, X. Chen, X. Meng, and W. Gao, "Learned local Gabor patterns for face representation and recognition," *Signal Process.*, vol. 89, no. 12, pp. 2333–2344, Dec. 2009.
[42] D. Maturana, D. Mery, and A. Soto, "Face recognition with decision tree-based local binary patterns," in *Proc. 10th Asian Conf. Comput. Vis.*, vol. 6495, Nov. 2010, pp. 618–629.
[43] D. Maturana, D. Mery, and A. Soto, "Learning discriminative local binary patterns for face recognition," in *Proc. FG*, Mar. 2011, pp. 470–475.
[44] N.-S. Vu and A. Caplier, "Enhanced patterns of oriented edge magnitudes for face recognition and image matching," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1352–1365, Mar. 2012.
[45] Z. Chai, Z. Sun, H. Méndez-Vázquez, R. He, and T. Tan, "Gabor ordinal measures for face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 1, pp. 14–26, Jan. 2014.
[46] X. Meng, S. Shan, X. Chen, and W. Gao, "Local visual primitives (LVP) for face modelling and recognition," *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, vol. 2, Aug. 2006, pp. 536–539.
[47] Z. Lei, M. Pietikäinen, and S. Z. Li, "Learning discriminant face descriptor," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 289–302, Feb. 2014.
[48] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 3025–3032.
[49] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Web-scale training for face identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2746–2754.
[50] B. Chen, W. Deng, and J. Du, "Noisy softmax: Improving the generalization ability of DCNN via postponing the early softmax saturation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5372–5381.
[51] J. Li, T. Qiu, C. Wen, K. Xie, and F. Q. Wen, "Robust face recognition using the deep C2D-CNN model based on decision-level fusion," *Sensors*, vol. 18, no. 7, p. 2080, Jun. 2018.
[52] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.
[53] A. Kinnikar, M. Husain, and S. M. Meena, "Face recognition using Gabor filter and convolutional neural network," in *Proc. ACM Int. Conf. Informat. Anal.*, Aug. 2016, p. 113.
[54] S. S. Sarwar, P. Panda, and K. Roy, "Gabor filter assisted energy efficient fast learning convolutional neural networks," in *Proc. IEEE/ACM Int. Symp. Low Power Electron. Design (ISLPED)*, Jul. 2017, pp. 1–6.
[55] M. Wang and W. Deng, "Deep face recognition: A survey," 2018, *arXiv:1804.06655*. [Online]. Available: https://arxiv.org/abs/1804.06655
[56] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, "The megaface benchmark: 1 Million faces for recognition at scale," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4873–4882.
[57] H.-W. Ng and S. Winkler, "A data-driven approach to cleaning large face datasets," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 343–347.
[58] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2016, pp. 499–515.
[59] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 212–220.
[60] H. Wang *et al.*, "Cosface: Large margin cosine loss for deep face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 5265–5274.
[61] F. Wang, J. Cheng, W. Liu, and H. Liu, "Additive margin softmax for face verification," *IEEE Signal Process. Lett.*, vol. 25, no. 7, pp. 926–930, Jul. 2018.
[62] W. Liu *et al.*, "Learning towards minimum hyperspherical energy," *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2018, pp. 6222–6233.
[63] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," 2018, *arXiv:1801.07698*. [Online]. Available: https://arxiv.org/abs/1801.07698



Jae Young Choi (M'09) received the M.S. and Ph.D. degrees from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2008 and 2011, respectively. In 2008, he was a Visiting Scholar with the University of Toronto. From 2011 to 2012, he was a Postdoctoral Researcher with the University of Toronto. He was a Postdoctoral Fellow with the University of Pennsylvania (UPenn) from 2012 to 2013. He was a Senior Engineer with Samsung Electronics from 2013 to 2014. He is currently an Associate Professor with the Division of Computer & Electronics Systems Engineering, Hankuk University of Foreign Studies - Global Campus. His research interests include pattern recognition, machine learning, deep learning, image processing, and computer vision. Especially, he has developed several pioneering algorithms for automatic face recognition using facial color information. He is the author or co-author of over 90 refereed research publications in the aforementioned research areas. Dr. Choi was a recipient of the Samsung HumanTech Thesis Prize in 2010.

Bumshik Lee (M'07) received the B.S. degree in electrical engineering from Korea University, Seoul, South Korea, and the M.S. and Ph.D. degrees in information and communications engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, respectively. He was a Research Professor with KAIST, South Korea, in 2014. He was a Postdoctoral Scholar with the University of California, San Diego (UCSD), CA, USA, from 2012 to 2013. He was a Principal Engineer with the Advanced Standard Research and Development Laboratory, LG Electronics, Seoul, from 2015 to 2016. He is currently an Assistant Professor with the Department of Information and Communication Engineering, Chosun University, South Korea. His research interests include pattern recognition, video compression and processing, video security, and medical image processing.