An analytic gabor feedforward network for single-sample and pose-invariant face recognition

Oh, Beom-Seok; Toh, Kar-Ann; Teoh, Andrew Beng Jin; Lin, Zhiping

2018

Oh, B.-S., Toh, K.-A., Teoh, A. B. J., & Lin, Z. (2018). An Analytic Gabor Feedforward Network for Single-Sample and Pose-Invariant Face Recognition. *IEEE Transactions on Image Processing*, 27(6), 2791-2805. doi:10.1109/TIP.2018.2809040

<https://hdl.handle.net/10356/86017>

<https://doi.org/10.1109/TIP.2018.2809040>

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at:
<https://doi.org/10.1109/TIP.2018.2809040>.

Downloaded on 26 Apr 2023 22:15:00 SGT

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/323371693>

An Analytic Gabor Feedforward Network for Single-Sample and Pose-Invariant Face Recognition

Article in IEEE Transactions on Image Processing · June 2018

DOI: 10.1109/TIP.2018.2809040

CITATIONS

2

READS

136

4 authors:

Beom-Seok Oh

Nanyang Technological University

33 PUBLICATIONS 305 CITATIONS

[SEE PROFILE](#)

K.A. Toh

Yonsei University

180 PUBLICATIONS 2,337 CITATIONS

[SEE PROFILE](#)

Andrew Beng Jin Teoh

Yonsei University

310 PUBLICATIONS 5,156 CITATIONS

[SEE PROFILE](#)

Zhiping Lin

Nanyang Technological University

258 PUBLICATIONS 3,134 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:

information theoretic learning [View project](#)

Collaboration with EEE, NTU [View project](#)

An Analytic Gabor Feedforward Network for Single-sample and Pose-invariant Face Recognition

Beom-Seok Oh, *Member, IEEE*, Kar-Ann Toh*, *Senior Member, IEEE*,
Andrew Beng Jin Teoh, *Senior Member, IEEE*, Zhiping Lin, *Senior Member, IEEE*,

Abstract—Gabor magnitude is known to be among the most discriminative representations for face images due to its space-frequency co-localization property. However, such property causes adverse effects even when the images are acquired under moderate head pose variations. To address this pose sensitivity issue as well as other moderate imaging variations, we propose an analytic Gabor feedforward network which can absorb such moderate changes. Essentially, the network works directly on the raw face images and produces directionally projected Gabor magnitude features at the hidden layer. Subsequently, several sets of magnitude features obtained from various orientations and scales are fused at the output layer for final classification decision. The network model is analytically trained using a single sample per identity. The obtained solution is globally optimal with respect to the classification total error rate. Our empirical experiments conducted on five face datasets (six subsets) from the public domain show encouraging results in terms of identification accuracy and computational efficiency.

Index Terms—Face Recognition Across Pose, Gabor Filtering, Single Hidden Layer Feedforward Network, Information Fusion

I. INTRODUCTION

DURING the past few decades, face recognition has been found to be among those active research topics in pattern recognition and computer vision. The main reason is that face recognition possesses huge potential in a wide variety of application domains including forensics, law enforcement, and entertainment. To gain a deep penetration into the deployment of such applications, achieving a certain level of recognition accuracy performance under various imaging conditions is mandatory. However, images acquired for such applications are often contaminated by imaging noises such as variation of pose, illumination, expression, etc. These noises often induce highly overlapping intra- and inter-identity distributions, and hence deteriorate the recognition accuracy. For those challenging security related applications such as passport identification [1]–[5], watch-list (e.g., a terrorist) screening [6], [7] and video surveillance [1], [8] under the near-infrared or infrared illumination [7] in which only a single face image per subject

is available for training, the difficulty towards an accurate recognition task is further aggravated [1]–[8].

Despite the several decades of research on face recognition, there remained much room for recognition performance enhancement. Particularly, as pointed out in [9], [10], face recognition across pose variation remains largely unsolved. This is partly due to self-occlusion and partly due to the drastic change in appearance from projecting a 3D rigid body onto a 2D surface under various illumination settings. A possible resolution by deep networks as seen in [11]–[14] takes advantage of abundant training samples to construct the corrupted information. Particularly, the deep networks, which are tuned on a set of pose-varied face images per subject, extract pose invariant features in an hierarchical (depth) manner. Although a pre-trained deep network model can be used for transfer learning, collecting a large enough number of labelled data for deep network tuning could still be costly. Moreover, a network which was pre-trained using the RGB images is not directly transferable to another visual spectrum (such as the infrared and the near-infrared images). A re-training of the network for such applications requires yet a significant amount of the already scarce data with ground truth. Furthermore, expert knowledge in network design and sophisticated parameter tuning effort may be needed for adapting the network to the new recognition paradigm [10].

Several methods which address the pose issue have been proposed in the literature (see [9], [10]). These methods can be categorized into three approaches [10]: i) face image synthesis based [13], [15]–[22], ii) multi-view subspace learning based [23], [24], and iii) pose-robust feature extraction based [11]–[14], [25]–[31] approaches. The main characteristics of these methods are summarized in Table I. It is observed from this table that each category possesses its own strength in different aspects. For example, approaches 1, 2 and deep learning based methods of approach 3 can work even for probe images with significant head pose variation. However, except for some recently proposed deep learning-based methods [1], [5]¹ which use single sample per person face recognition, many of these approaches require a large amount of data for training and are computationally demanding. Such requirements hinder them from deployment in real-world applications. On the other

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2015R1D1A1A09061316).

B.-S. Oh and Z. Lin are with the School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore.

K.-A. Toh and A. B. J. Teoh are with the School of Electrical and Electronic Engineering, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Republic of Korea. * Corresponding author: kato@yonsei.ac.kr

Most of this work was performed when Beom-Seok Oh was with Yonsei University.

¹The single sample per person domain adaptation network (SSPP-DAN) method proposed in [5] fine-tuned the pre-trained VGG-Face model [32] using a set of synthesized face images generated from the input sample to overcome the single training data constraint. Since the SSPP-DAN utilizes the 3D face information beyond the scope of our study, this work is excluded from our empirical comparison study.

TABLE I
A BRIEF SUMMARY OF EXISTING APPROACHES TO POSE-INVARIANT FACE RECOGNITION

Approaches	Pose range*	No. imgs/subject for training (Multi pose images?)	Computational efficiency	Note
1. Face synthesis based				
- 3D pose normalization [15], [16]	Full [‡]	Large (Needed)	Low	- May encounter image synthesis problems [§] - Some methods require face shape fitting
- 3D modeling by image reconstruction [17]				
- 2D pose normalization [18]–[20]	Wide ^b	1 – a few (Depends on method)	Medium-Low	- May encounter image synthesis problems [§] - Piece-, patch- or pixel-wise warping is required
- 2D regression model [13], [21], [22]				
2. Multi-view subspace learning based				
- Linear or nonlinear model [23], [24]	Wide	Large (Needed)	Low	- Its recognition performance is highly affected by the amount of training data
3. Pose-robust feature extraction based				
- Deep learning based features [1], [5], [11]–[14]	Wide	1 – very large (Depends on method)	Low	- Training DNN parameters is a non-trivial task - Some recent works SSPP-DAN [5], SSAE [1] used a single sample per person for training the deep learning network
- Engineered features [25]–[31]				
	Mode-rate [‡]	1 – a few (Depends on method)	Medium-High	- Landmark- or key-points detection is required

*: A coverable range of head pose in which a method can produce the reasonably good recognition accuracy performance (about 90% accuracy, for example).

[‡], ^b, [‡]: ‘Full’ (roughly about -90° to +90°), ‘Wide’ (roughly about -60° to +60°) and ‘Moderate’ (roughly about -30° to +30°).

[§]: Suffering from artifacts and blurring effects in a synthesized image, and losing fine facial textures (e.g., moles etc.).

Abbreviation: Deep neural network (DNN), Single sample per person domain adaptation network (SSPP-DAN), Stacked supervised auto-encoder (SSAE).

hand, the engineered features based methods of approach 3 (see the last row of Table I) are free from such requirements. Our proposed method belongs to this approach.

The main issue shared by engineered feature based methods is upon how to build a good semantic correspondence among the face images. A typical solution is to use either manually picked or automatically detected landmark points from the images [9], [10]. We note here that picking corresponding points manually is impractical. This is particularly true for the case with a huge number of images. Extracting the corresponding points automatically with a certain level of accuracy is another difficult task. Therefore, it is necessary and beneficial to extract discriminative and rich facial features without needing to establish the correspondences.

It is noted here that extracting and processing rich facial information raise two issues namely, high computational cost and feature redundancy. Moreover, any additional step such as fusion for performance enhancement further increases the computational load. Motivated by these observations, in this work, a single hidden layer [33] analytic Gabor feedforward network (AGFN) is proposed for single-sample and pose-invariant face recognition.

The proposed AGFN packs effective processing steps into a single network structure for efficient processing. Essentially, AGFN extracts Gabor features which were theoretically proven to be among the best representation for modeling human’s receptive field profiles of cortical simple cell (see [34] for details). Via an accumulation process on small image regions, AGFN achieves an invariance property to moderate translation. Such an accumulation process eventually enables the AGFN to tolerate moderate variation of pose as well as variation of illumination, expression and acquisition time. The input layer takes in raw images and the hidden layer extracts directionally projected Gabor features from local image patches. These features are then projected onto a whitened low-dimensional subspace. The outputs of hidden layer obtained based on various orientation and scale settings are finally fused at the output layer for identity classification.

The main contributions of this work are enumerated as follows:

- We propose a computationally efficient and yet effective single hidden layer feedforward network structure for face recognition under moderate pose tolerance. The network is also tolerant to a moderate amount of external imaging noises such as variation of illumination, expression and acquisition time. The network works directly on raw face images, and its model is analytically trained using a single image per subject.
- We have found that the directional accumulation of Gabor elements expands the size of Gabor response area. Such feature provides tolerance to moderate image translation, and thus plays an important role in mitigating the translational effects caused by moderate changes in head pose. The accumulation can be performed efficiently via the proposed simplification of Gabor filtering process.
- We provide extensive experiments on five public face datasets (six subsets) comparing AGFN with competing Gabor-based face recognition methods. Our results show that AGFN outperforms these compared methods in terms of accuracy and efficiency under moderate conditions of imaging noise.

The remaining parts of this paper are organized as follows. In Section II, we provide some background knowledge on Gabor filtering and a total error rate minimization based classifier for immediate reference. Section III presents the details of Gabor features extraction within the proposed AGFN model. Section IV presents an extensive empirical evaluation based on six public face datasets/subsets and state-of-the-art methods. Concluding remarks are given in Section V.

II. PRELIMINARY

In this section, we provide the background knowledge of two methods, namely the two-dimensional Gabor filtering for image representation and the total error rate minimization for multi-category classification, for immediate reference.

A. Two-dimensional (2D) Gabor filtering

Let $\mathbf{z} = [c, r]^T$ denote a pixel coordinate, where T indicates the vector/matrix transposition, and suppose $j = \sqrt{-1}$. According to [35], [36], a Gabor kernel is defined as follows:

$$\Psi_{\mu, v}(\mathbf{z}) = \frac{\|\mathbf{k}_{\mu, v}\|^2}{\sigma^2} e^{-\frac{\|\mathbf{k}_{\mu, v}\|^2 \|\mathbf{z}\|^2}{2\sigma^2}} \times \left[e^{j\mathbf{k}_{\mu, v}^T \mathbf{z}} - e^{-\frac{\sigma^2}{2}} \right], \quad (1)$$

where $\Psi_{\mu, v}(\mathbf{z}) \in \mathbb{C}$ is a complex Gabor element obtained using orientation μ and scale v at pixel coordinate $\mathbf{z}$. σ denotes the standard deviation and $\|\cdot\|$ indicates the L_2 -norm operator. The kernel matrix $\Psi_{\mu, v}$ is of size $h \times w$ pixels. The symbol $\mathbf{k}_{\mu, v} = (k_v \cos \varphi_\mu, k_v \sin \varphi_\mu)^T$ denotes a wave vector in which $k_v = k_{\max}/f^v$ is the frequency where $k_{\max}$ is the maximum frequency and f is the spacing between kernels in frequency domain. $\varphi_\mu = \mu\pi/8 \in [0, \pi)$ indicates the orientation.

The Gabor representation of an image matrix $\mathbf{X} \in \mathbb{R}^{p \times q}$ is the result of convolving $\mathbf{X}$ with a Gabor kernel $\Psi_{\mu, v}$. The convolution can be performed in both spatial and frequency domains. However, we shall focus on the spatial domain convolution² in which the proposed method is originated (see Section III for details). This convolution can be written as:

$$\mathbf{O}_{\mu, v}(x, y) = \sum_{c=1}^w \sum_{r=1}^h \mathbf{X}(x-c, y-r) \Psi_{\mu, v}(c, r), \quad (2)$$

where $\mathbf{O}_{\mu, v} \in \mathbb{C}^{p \times q}$, (x, y) is a pixel coordinate of image $\mathbf{X}$, (c, r) is a pixel coordinate of kernel $\Psi_{\mu, v} \in \mathbb{C}^{h \times w}$, $p > h$, $q > w$, and both h and w are odd numbers. To handle pixels falling on the image boundary, a zero padding can be adopted.

The next step is to compute magnitude and/or phase features from the $\mathbf{O}_{\mu, v}$ matrix [36], [37]. Except for some works [38]–[40] to list just a few, literature discarded the phase information due to its sensitivity to changes in rotation and translation while retaining the magnitude information. For similar reason, our proposed method adopts only the magnitude feature for recognition, which is defined as follows:

$$\mathbf{M}_{\mu, v}(x, y) = \sqrt{\text{Re}(\mathbf{O}_{\mu, v}(x, y))^2 + \text{Im}(\mathbf{O}_{\mu, v}(x, y))^2}, \quad (3)$$

where $\mathbf{M}_{\mu, v} \in \mathbb{R}^{p \times q}$, $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ respectively correspond to real and imaginary parts of $\mathbf{O}_{\mu, v}$.

B. TER minimization for multi-category classification

According to [41], the total error rate (TER) is defined as a summation of the false acceptance rate and the false rejection rate of a predictor's output at a certain decision threshold τ . Let N_C be the total number of classes where each class represents an identity. Then, samples belonging to the C^{th} -category, $C = 1, \dots, N_C$, are treated as positive class (denoted by a superscript '+') while all the other non- C^{th} -category samples are considered as negative class (denoted by a superscript '-'). With an appropriate normalization plus inclusion of an offset term η , minimization of TER with

²Due to its efficiency, most of the conventional Gabor face features (e.g., magnitude and phase) are computed in the frequency domain. However, as we shall show in Sections III-A2 and III-A3, the proposed features can also be extracted efficiently in the spatial domain.

respectively to a classifier which is linear in its parameters β_C can be solved in analytic form [41]:

$$\beta_C = (b\mathbf{I} + \mathbf{P}_C^T \mathbf{W}_C \mathbf{P}_C)^{-1} \mathbf{P}_C^T \mathbf{W}_C \mathbf{y}_C, \quad (4)$$

where b is a regularization factor, $\mathbf{P}_C = \begin{bmatrix} \mathbf{P}_C^- \in \mathbb{R}^{m_C^- \times d} \\ \mathbf{P}_C^+ \in \mathbb{R}^{m_C^+ \times d} \end{bmatrix} \in \mathbb{R}^{m \times d}$ is a regressor matrix, m_C^+ and m_C^- ($m = m_C^+ + m_C^-$) respectively indicate the positive class and negative class populations. $\mathbf{y}_C = \begin{bmatrix} \mathbf{y}_C^- \\ \mathbf{y}_C^+ \end{bmatrix} = \begin{bmatrix} (\tau - \eta)\mathbf{1}_C^- \in \mathbb{N}^{m_C^-} \\ (\tau + \eta)\mathbf{1}_C^+ \in \mathbb{N}^{m_C^+} \end{bmatrix} \in \mathbb{N}^m$ is a target vector, $\mathbf{1}_C^- = [1, \dots, 1]^T \in \mathbb{N}^{m_C^-}$ and $\mathbf{1}_C^+ = [1, \dots, 1]^T \in \mathbb{N}^{m_C^+}$. $\mathbf{W}_C = \mathbf{W}_C^- + \mathbf{W}_C^+ \in \mathbb{R}^{m \times m}$ is a diagonal weighting matrix defined for the C -th class in which $\mathbf{W}_C^- = \text{diag}(1/m_C^-, \dots, 1/m_C^-, 0, \dots, 0)$ and $\mathbf{W}_C^+ = \text{diag}(0, \dots, 0, 1/m_C^+, \dots, 1/m_C^+)$. Here $\mathbf{I}$ is an identity matrix with dimension matching that of $\mathbf{P}_C^T \mathbf{P}_C$.

III. PROPOSED METHOD

In this section, we propose a single hidden layer feedforward network, called analytic Gabor feedforward network (AGFN), for efficient extraction of Gabor features without needing the time consuming convolutional operation in full mode. Another important characteristics of AGFN is that no iteration is needed for network training and the solution is globally optimal with respect to classification total error rate.

By removing redundant computations while extracting Gabor variant features at the input and the hidden layers, we achieve both efficiency and effectiveness at the same time. The outputs from the hidden layer are subsequently fused to produce the final decision in analytic solution. The following subsections provide details of AGFN.

A. Gabor-based structured projection

Gabor magnitude is known to be among the most discriminative representations for face [34]–[36], [42]. However, the high computational cost due to image convolution hinders its real-time application potential. Another shortcoming of the representation is its sensitiveness to variation of illumination and head pose. To address these issues, the extracted Gabor features are treated as two directional projections with random accumulation of local regions such that moderate translations caused by head pose variation can be tolerated.

1) *Simplification of Gabor filtering process:* To achieve this goal, the matrix convolution operation discussed in (2) is revisited based on the vector multiplication view point. Consider an image matrix $\mathbf{X} \in \mathbb{R}^{p \times q}$ and a set of Gabor kernels $\Psi_{\mu, v} \in \mathbb{C}^{h \times w}$, $\mu \in \{0, \dots, \mu^{\max}\}$ and $v \in \{0, \dots, v^{\max}\}$ being constructed using (1). Starting around each image pixel, let $\mathbf{A}_{x, y} \in \mathbb{R}^{h \times w}$ be an image patch of $\mathbf{X}$ which is centered on a pixel coordinate (x, y) , and it has the same size with that of $\Psi_{\mu, v}$. Then, equation (2) can be re-written as follows:

$$\begin{aligned} \mathbf{O}_{\mu, v}(x, y) &= \sum_{c=1}^w \sum_{r=1}^h \mathbf{A}_{x, y}(c, r) \Psi_{\mu, v}(c, r) \\ &= \mathbf{a}_{x, y}^T \Psi_{\mu, v}, \end{aligned} \quad (5)$$

where $\mathbf{a}_{x, y} \in \mathbb{R}^{hw \times 1}$ and $\Psi_{\mu, v} \in \mathbb{C}^{hw \times 1}$ respectively denote vectorized $\mathbf{A}_{x, y}$ and $\Psi_{\mu, v}$.

Consider a set of $n = (\mu^{\max} + 1)(\nu^{\max} + 1)$ kernels being pre-generated and stacked to form a kernel matrix $\tilde{\Psi} = [\psi_{0,0}, \dots, \psi_{0,\nu^{\max}}, \psi_{1,0}, \dots, \psi_{\mu^{\max},\nu^{\max}}] \in \mathbb{C}^{hw \times n}$. Consequently, the matrix convolution (5) can be re-written as:

$$\tilde{\mathbf{O}}(x, y) = \mathbf{a}_{x,y}^T \tilde{\Psi}, \quad (6)$$

where $\tilde{\mathbf{O}}(x, y) \in \mathbb{C}^{1 \times n}$ and $\tilde{\mathbf{O}} \in \mathbb{C}^{p \times q \times n}$. As a result of this simplification, we are able to compute all the n Gabor responses by a single vector-matrix multiplication. This simplified representation enables us to design localized directional projections to extract locally accumulated Gabor features which will be described below. Such directional projections are tolerant to noises originated from variation of head pose, expression, and acquisition time.

2) Vertically projected local Gabor features, $\mathbf{g}_{ver}$ features:

The proposed $\mathbf{g}_{ver}$ features consist of magnitudes of locally accumulated Gabor responses along the y -axis direction. Let $s_y \in \{1, \dots, p-l\}$ be an index of the first cell where the projection (i.e., accumulation) starts, and l be an extraction size which controls the number of cells to be accumulated within each image columns. With these two parameters, we define vertically projected Gabor responses $\mathbf{O}_{\mu,v}(x, \Sigma_{\{s_y, l\}})$ as:

$$\mathbf{O}_{\mu,v}(x, \Sigma_{\{s_y, l\}}) = \sum_{i=s_y}^{s_y+l-1} (\mathbf{a}_{x,i}^T \boldsymbol{\psi}_{\mu,v}) = \mathbf{v}_x^T \boldsymbol{\psi}_{\mu,v}, \quad (7)$$

where $\mathbf{O}_{\mu,v}(x, \Sigma_{\{s_y, l\}}) \in \mathbb{C}^{1 \times q}$, and $\mathbf{v}_x = \sum_{i=s_y}^{s_y+l-1} \mathbf{a}_{x,i} \in \mathbb{R}^{hw \times 1}$ is a vertically summed image vector along the y -direction at the x -th column. See equation (5) for details of $\mathbf{a}_{x,y}$ and $\boldsymbol{\psi}_{\mu,v}$.

It is worth emphasizing that the above projection discards those Gabor responses that do not belong to a local region of interest (i.e., corresponding to pixel coordinates $(x, y < s_y)$ and $(x, y > s_y + l)$). In this work, we randomly set the location s_y while l is an adjustable parameter. Because of this localized strategy, we achieve algorithmic efficiency which other Gabor-based algorithms do not possess. Fig.1 (a) illustrates the process of computing $\mathbf{O}_{\mu,v}(x, \Sigma_{\{s_y, l\}})$.

Thus far, we have worked on a single accumulated image vector $\mathbf{v}_x$ and a single Gabor kernel $\boldsymbol{\psi}_{\mu,v}$ per operation. By considering multiple number of image vectors and kernels as in (6), equation (7) can be represented compactly as:

$$\tilde{\mathbf{O}}_{ver} = \mathbf{V}^T \tilde{\Psi}, \quad (8)$$

where $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_q] \in \mathbb{R}^{hw \times q}$ is a stacked set of image vectors $\mathbf{v}_x$, and $\tilde{\mathbf{O}}_{ver} \in \mathbb{C}^{q \times n}$ denotes the projected Gabor responses over the entire range of μ and ν . With this generalization, extraction of the vertically projected Gabor responses from an image matrix over n number of Gabor kernels can be completed by a single matrix-matrix multiplication. Subsequently, the magnitude of $\tilde{\mathbf{O}}_{ver}$ is computed using (3) and then being converted into a feature vector $\mathbf{g}_{ver} \in \mathbb{R}^{qn \times 1}$.

3) *Horizontally projected local Gabor features, $\mathbf{g}_{hor}$ features:* Similar to $\mathbf{g}_{ver}$ features, horizontally projected Gabor features, which is called $\mathbf{g}_{hor}$ features, are proposed. Similar to equation (7), here Gabor responses of a local region are accumulated along the direction of x -axis as follows:

$$\mathbf{O}_{\mu,v}(\Sigma_{\{s_x, l\}}, y) = \sum_{i=s_x}^{s_x+l-1} (\mathbf{a}_{i,y}^T \boldsymbol{\psi}_{\mu,v}) = \mathbf{h}_y^T \boldsymbol{\psi}_{\mu,v}, \quad (9)$$

Fig. 1. An illustration for the extraction of $\mathbf{g}_{ver}$ and $\mathbf{g}_{hor}$ features. In this illustration, l number of complex Gabor responses are accumulated within (a) column vectors for $\mathbf{g}_{ver}$ and (b) row vectors for $\mathbf{g}_{hor}$, respectively. Note that a complex number is illustrated as a vector in the Cartesian coordinate.

where $\mathbf{h}_y = \sum_{i=s_x}^{s_x+l-1} \mathbf{a}_{i,y} \in \mathbb{R}^{hw}$ is a summed image vector along the horizontal direction. $\mathbf{O}_{\mu,v}(\Sigma_{\{s_x, l\}}, y) \in \mathbb{C}^{p \times 1}$ indicates the horizontally summed Gabor responses, and $s_x \in \{1, \dots, q-l\}$ denotes randomly determined index of the first cell in which the accumulation starts along each image rows (see Fig.1 (b)).

Similar to the matrix-based formulation in (8), equation (9) can be compactly represented using matrix notation as:

$$\tilde{\mathbf{O}}_{hor} = \mathbf{H}^T \tilde{\Psi}, \quad (10)$$

where $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_p] \in \mathbb{R}^{hw \times p}$ represents a stacked set of $\mathbf{h}_y$, and $\tilde{\mathbf{O}}_{hor} \in \mathbb{C}^{p \times n}$ is the horizontally extracted Gabor responses. Similar to $\mathbf{g}_{ver} \in \mathbb{R}^{qn \times 1}$ features, $\mathbf{g}_{hor} \in \mathbb{R}^{pn \times 1}$ features contain the vectorized magnitudes of $\tilde{\mathbf{O}}_{hor}$.

4) *Controlling the number of extracted features by projection size k :* We have introduced two adjustable parameters, namely the extraction size l which controls the number of Gabor responses to be accumulated, and the cell index s_y or s_x where the accumulation starts. Both $s_{[y,x]}$ and l determine upon which part of a face image and how much of local facial information will be extracted. We note here that such locality in the feature extraction stage gives rise to a concern that some important facial information might be missed. For example, $\mathbf{g}_{ver}$ which extracted features at $s_y = 1$ and $l = 3$, may contain non-facial information such as hair and/or background.

In order to alleviate such lacking of facial information, a new parameter, projection size k , is introduced to both the horizontal and vertical projections. Suppose each computation of $\mathbf{g}_{ver}$ or $\mathbf{g}_{hor}$ features per image constitutes a single round of feature extraction. Then, k controls the number of rounds per image. Particularly, k number of $s_{[y,x]}$ are randomly generated and utilized: $[s_y^1, \dots, s_y^k]$ for $\mathbf{g}_{ver}$ features and $[s_x^1, \dots, s_x^k]$ for $\mathbf{g}_{hor}$ features. Since k is directly related to the amount of extracted information, it can tradeoff between accuracy and efficiency. Our focus in this work is to observe under what setting of k is appropriate in balancing between accuracy and efficiency. With inclusion of k projection rounds, the

size of resulted feature vector becomes $\mathbf{g}_{ver} \in \mathbb{R}^{kqn \times 1}$ and $\mathbf{g}_{hor} \in \mathbb{R}^{kpn \times 1}$, respectively.

5) *Empirical analysis regarding the physical meaning of Gabor responses accumulation:* The proposed $\mathbf{g}_{ver}$ and $\mathbf{g}_{hor}$ features consist of magnitudes of directionally accumulated Gabor responses. Here, we observe the underlying mechanism of the accumulation and its effects on translation invariance corresponding to moderate face pose variation. Essentially, our analysis shows that the output region of Gabor responses is directionally expanded as a result of the accumulation. Particularly, such region expansion increases the chance of matching identical facial components which appear at different image locations caused by pose variation.

Let $\mathbf{r}_{\mu,v} = \text{Re}(\Psi_{\mu,v})$ and $\mathbf{i}_{\mu,v} = \text{Im}(\Psi_{\mu,v})$, then equation (3) for any given image coordinate (x,y) can be re-written as:

$$\mathbf{M}_{\mu,v}(x,y) = \sqrt{(\mathbf{r}_{\mu,v}^T \mathbf{a}_{x,y})^2 + (\mathbf{i}_{\mu,v}^T \mathbf{a}_{x,y})^2} = \sqrt{\mathbf{a}_{x,y}^T \mathbf{S} \mathbf{a}_{x,y}}, \quad (11)$$

where $\mathbf{S} = \mathbf{r}_{\mu,v} \mathbf{r}_{\mu,v}^T + \mathbf{i}_{\mu,v} \mathbf{i}_{\mu,v}^T \in \mathbb{R}^{hw \times hw}$. Based on (11), both $\mathbf{g}_{ver}$ feature at coordinate x (7) and $\mathbf{g}_{hor}$ feature at coordinate y (9) can now respectively be represented as the square root of directionally accumulated $\mathbf{a}_{x,y}^T \mathbf{S} \mathbf{a}_{x,y} \in \mathbb{R}$ as follows:

$$\mathbf{M}_{\mu,v}(x, \Sigma_{\{s_y, l\}}) = \sqrt{\Sigma_{i=s_y}^{s_y+l-1} (\mathbf{a}_{x,i}^T \mathbf{S} \mathbf{a}_{x,i})} = \sqrt{\mathbf{v}^T \mathbf{S} \mathbf{v}}, \quad (12)$$

$$\mathbf{M}_{\mu,v}(\Sigma_{\{s_x, l\}}, y) = \sqrt{\Sigma_{i=s_x}^{s_x+l-1} (\mathbf{a}_{i,y}^T \mathbf{S} \mathbf{a}_{i,y})} = \sqrt{\mathbf{h}^T \mathbf{S} \mathbf{h}}, \quad (13)$$

where $\mathbf{v} = \Sigma_{i=s_y}^{s_y+l-1} \mathbf{a}_{x,i}$ and $\mathbf{h} = \Sigma_{i=s_x}^{s_x+l-1} \mathbf{a}_{i,y}$.

Equations (12) and (13) show that matrix $\mathbf{S}$ remains unchanged for l number of image patches (a) accumulation (as in $\mathbf{v}$ and $\mathbf{h}$). Particularly, only those image patches which contain corresponding information to the adopted Gabor kernel contribute to $\mathbf{g}_{ver}$ and $\mathbf{g}_{hor}$ features computation. To explore the relationship between such kernel dependent summation of patches and recognition accuracy, two synthetic image-based case studies: Case study 1 and 2, are conducted as follows.

a) *Case study 1 (Fig. 2 (a)–(d)):* We observe the similarity among the Gabor magnitude features by means of the utilized image patches and a particular Gabor kernel. Consider an image shape with its translated version in Fig. 2 (a). Such geometrical translation of objects (e.g., facial components) on a 2D image plane represents an effect caused by changes in head pose. Then, the extracted features are shown in Fig. 2 (b) where only diagonal image edges are being extracted due to the alignment of Gabor kernel. We note that this empirical analysis holds for other settings of μ and v as well.

Fig. 2 (c) shows image patches $\mathbf{a}_{x,y}$ taken from two synthetic images $\mathbf{X}_1$ and $\mathbf{X}_2$ (see Fig. 2 (a)) respectively. Here, $\mathbf{X}_2$ is three pixels shifted from $\mathbf{X}_1$ along the x -direction. The conventional approach compares magnitude values computed from each patch of the same pixel coordinate. As shown in Fig. 2 (c), this approach may produce a low similarity between $\mathbf{X}_1$ and $\mathbf{X}_2$. The main reason is that no patch contains the targeted edge at the same or similar pixel coordinates.

On the other hand, as shown in Fig. 2 (d), the horizontally accumulated image patch (i.e., the vector $\mathbf{h}$ in equation (13)) appeared to be similar to each other at $l = 6$. This shows

Fig. 2. An illustrative example for Gabor elements accumulation.

that $\mathbf{g}_{hor}$ features can be less sensitive to differences of object location caused by horizontal pose variation than that of the conventional Gabor-based approach. In a similar manner, it can be inferred that $\mathbf{g}_{ver}$ features can be less sensitive to vertical pose variation than that of the conventional approach.

b) *Case study 2 (Fig. 2 (e)–(h)):* The above empirical analysis is now extended to synthetic face images. The first row of Fig. 2 (f) shows the real part of Gabor kernels (generated at $\mu = \{0, \dots, 7\}$ and $v = 0$) which are treated as images. Magnitude features resulted from the conventional approach and the $\mathbf{g}_{hor}$ features at $l = 5$ and $l = 9$ are respectively shown in the second to fourth row of Fig. 2 (f). By comparing the three magnitude results, it is observed that the size of magnitude response area becomes wider as a result of accumulation except for $\mu = 0$ and $\mu = 4$.

To support the above observation, we now consider three pose-varied synthetic face images (see Fig. 2 (g)): frontal, right pose and left pose (rotated 20° from the frontal). The conventional Gabor magnitude features and the proposed $\mathbf{g}_{hor}$ features at $l = \{3, 5, 7, 9\}$ are extracted from each of the images, respectively. The Euclidean distance is then computed over pairs of different poses: frontal-left, frontal-right and left-right, per method per l setting. Fig. 2 (h) shows that $\mathbf{g}_{hor}$ features extracted at higher extraction size $l \in \{7, 9\}$ produced lower Euclidean distance values than that of low extraction size at $l \in \{3, 5\}$, and than that of the conventional one which is equivalent to $l = 1$.

Fig. 3. An illustration of the proposed AGFN model. $\star \in \{ver, hor\}$, $d = kqn$ if $\star = ver$ and $d = kpn$ if $\star = hor$, and $n = (\mu^{\max} + 1)(\nu^{\max} + 1)$.

B. Analytic Gabor Feedforward Network

The extracted local Gabor features: $\mathbf{g}_{ver} \in \mathbb{R}^{kpn \times 1}$ and $\mathbf{g}_{hor} \in \mathbb{R}^{kqn \times 1}$, would be of high-dimension. This is particularly true if input images are of high resolution and when a high value of k is adopted. Moreover, due to the nature of Gabor representation resulted from eight orientations and five scales, the feature vector would be highly correlated and redundant.

The proposed AGFN is designed to handle these high-dimensional and redundancy problems while increasing the discriminativity by information fusion. After performing feature extraction at the hidden layer (see Fig. 3), the information are fused and classified at the output layer. By integrating these steps into a simple network, we could achieve algorithmic simplicity and efficiency while enjoying good accuracy performance. Due to an approximated counting formulation, AGFN is free from iterative parameter learning where the output weights are analytically estimated.

1) *The proposed network architecture:* Suppose we have m training sample pairs $(\mathbf{X}_i, y_i)$, $i = 1, \dots, m$, where $\mathbf{X}_i \in \mathbb{R}^{p \times q}$ and $y_i \in \mathbb{R}$ respectively indicate the i -th raw image matrix and its corresponding class label (i.e., user identity). The proposed AGFN model is defined as follows (see Fig. 3):

$$y_i = \beta_C^T (\mathbf{w}^T \phi(\mathbf{g}_*^i)), C = 1, \dots, N_C \quad (14)$$

where $\mathbf{g}_*^i$ denotes the proposed Gabor features (either $\mathbf{g}_{ver}^i \in \mathbb{R}^{kpn \times 1}$ or $\mathbf{g}_{hor}^i \in \mathbb{R}^{kqn \times 1}$, see Section III-A) extracted from $\mathbf{X}_i$, and $\phi(\cdot)$ indicates the sigmoid activation function. $\mathbf{w} \in \mathbb{R}^{d \times \tilde{d}}$ indicates the set of $\tilde{d}$ projection basis vectors learned by a whitened principal component analysis (WPCA), where $d \gg \tilde{d}$, $d = kpn$ if $\star = ver$ otherwise $d = kqn$, and $\beta_C = [\beta_{C1}, \dots, \beta_{C\tilde{d}}]^T \in \mathbb{R}^{\tilde{d}}$ denotes the output weight vector learned for the C -th identity. The unsupervised WPCA subspace learning is adopted because the supervised learning method (e.g., linear discriminant analysis [43]) requires more than a single image per subject for training.

As shown in equation (14) and Fig. 3, the n number of hidden nodes constitute the hidden layer. Each hidden node corresponds to a combination of μ and ν , and produces a directionally projected Gabor feature vector. The outputs from the hidden nodes are subsequently activated and projected onto $\tilde{d}$ -dimensional whitened subspace for data compaction (see Section III-B2 for details). The low dimensional feature vector is finally incorporated into the output weight vector β_C to produce a final decision (see Section III-B4 for details). According to the feature type being extracted, we call the proposed model as AGFN $_{\star}$ where $\star = ver$ refers to AGFN $_{ver}$ and $\star = hor$ refers to AGFN $_{hor}$ hereafter for convenience. AGFN $_{fusion}$ indicates the AGFN which adopts both $\mathbf{g}_{ver}^i$ and $\mathbf{g}_{hor}^i$ features being concatenated in feature level.

From the feature extraction view point, the hidden nodes of AGFN extract the proposed local Gabor features. By having multiple hidden nodes with different values for μ and ν , different Gabor facial information are extracted simultaneously. From the information fusion view point, the output layer of AGFN fuses the extracted Gabor features over the entire range of μ and ν by a simple weighted sum rule. In a nutshell, both feature extraction and information fusion stages are integrated and absorbed into the AGFN model.

2) *Learning the subspace $\mathbf{w}$ by a Whitened PCA:* In order to reduce the dimension of hidden layer outputs as well as to reduce their correlation, we learn a projection subspace using WPCA. Let $\mathbf{P}_{PCA} = [\phi(\mathbf{g}_*^1), \dots, \phi(\mathbf{g}_*^m)]_{d \times m}$ be a matrix of hidden layer outputs in which each column vector corresponds to a training image. The whitened subspace matrix is defined as $\mathbf{w} = \mathbf{E}\mathbf{\Lambda}^{-\frac{1}{2}}$, where $\mathbf{E} = [\mathbf{e}_1, \dots, \mathbf{e}_{\tilde{d}}] \in \mathbb{R}^{d \times \tilde{d}}$ and $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_{\tilde{d}}\}$ respectively denote the leading $\tilde{d}$ eigenvectors and eigenvalues (i.e., $\lambda_i > \lambda_{i+1}$) computed from the covariance matrix of $\mathbf{P}_{PCA}$.

3) *Learning the output weights β_C by TER:* Suppose the given training data consists of N_C subjects. Our task is to identify a subject from the given test face image among the N_C subjects. To this end, we learn an output weight vector $\hat{\beta}_C \in \mathbb{R}^{\tilde{d}}$ per subject which is indexed by $C = 1, \dots, N_C$. Within the training set, those samples corresponding to each of the C -th subject are considered as the positive class while the remaining non- C -th samples constitute the negative class (see Section II-B for details). Correspondingly, an indicator target label [44] can be constructed to train these multiple sets of binary targets.

With a regression matrix $\mathbf{P}_C = [\mathbf{P}_C^-, \mathbf{P}_C^+]^T \in \mathbb{R}^{m \times \tilde{d}}$, where $\mathbf{P}_C^- = \left[\mathbf{w}^T \phi(\mathbf{g}_*^1), \dots, \mathbf{w}^T \phi(\mathbf{g}_*^{m_C^-}) \right] \in \mathbb{R}^{\tilde{d} \times m_C^-}$, and $\mathbf{P}_C^+ = \left[\mathbf{w}^T \phi(\mathbf{g}_*^1), \dots, \mathbf{w}^T \phi(\mathbf{g}_*^{m_C^+}) \right] \in \mathbb{R}^{\tilde{d} \times m_C^+}$ computed using the training data, the output weight vector $\hat{\beta}_C$ is computed using Equation (4) [45]. Following other Gabor-based algorithms [36], we normalize each column of $\mathbf{P}_C$ to follow zero mean and unit variance. The procedure for training AGFN model is summarized in Algorithm 1.

4) *Prediction of unseen test data:* Let $\mathbf{X}_t \in \mathbb{R}^{p \times q}$ represent a test image which does not belong to the training samples. Its class (i.e., identity) label can be predicted using the estimated

Algorithm 1 AGFN: Training phase

1: **Given:** m training sample pairs $(\mathbf{X}_i \in \mathbb{R}^{p \times q}, y_i \in \mathbb{R}), i = 1, \dots, m$, projection size k , extraction size l , Sigmoid function $\phi(\cdot)$, bias $b = 10^{-4}$, threshold $\tau = 0$, offset $\eta = 1, \star \in \{ver, hor\}$

2: $\# \leftarrow p$ if $\star = ver$ or q if $\star = hor$

3: $[s^1, \dots, s^k] \leftarrow \lfloor rand(1, k) \times (\# - l) \rfloor + 1$

4: $\Psi \leftarrow n$ number of vectorized Gabor kernels (1)

5: **for** $i \leftarrow 1$ **to** m **do** ▷ Extract $\mathbf{g}_\star^i$ features

6: $\tilde{\mathbf{O}}_\star^i \in \mathbb{C}^{k\# \times n} \leftarrow \mathbf{V}_i^T \Psi$ if $\star = ver$ or $\mathbf{H}_i^T \Psi$ if $\star = hor$

7: $\mathbf{M}_\star^i \in \mathbb{R}^{k\# \times n} \leftarrow$ Magnitude of $\tilde{\mathbf{O}}_\star^i$

8: $\mathbf{g}_\star^i \in \mathbb{R}^{k\# \times 1} \leftarrow$ Vector form of $\mathbf{M}_\star^i$

9: **end for**

10: Normalize $[\mathbf{g}_\star^1, \dots, \mathbf{g}_\star^m]$ to a zero mean and unit variance

11: $\mathbf{P}_{PCA} \in \mathbb{R}^{d(=k\#n) \times m} \leftarrow \phi([\mathbf{g}_\star^1, \dots, \mathbf{g}_\star^m])$ ▷ Learn $\mathcal{W}$

12: $\mathbf{E} \in \mathbb{R}^{d \times \tilde{d}}, \mathbf{A} \in \mathbb{R}^{\tilde{d} \times \tilde{d}} \leftarrow$ Top $\tilde{d}$ (95% of the eigenenergy) eigenvectors and eigenvalues obtained from $Cov(\mathbf{P}_{PCA}) \in \mathbb{R}^{m \times m}$

13: $\mathcal{W} \in \mathbb{R}^{d \times \tilde{d}} \leftarrow \mathbf{E} \mathbf{A}^{-\frac{1}{2}}$

14: $\mathbf{P} \in \mathbb{R}^{d \times m} \leftarrow \mathcal{W}^T \phi([\mathbf{g}_\star^1, \dots, \mathbf{g}_\star^m])$

15: **for** $C \leftarrow 1$ **to** N_C **do** ▷ Learn $\hat{\mathbf{P}}_C$

16: Construct a diagonal weighting matrix $\mathbf{W} \in \mathbb{R}^{m \times m}$

17: $\mathbf{P}_C^+ \in \mathbb{R}^{d \times m_C^+} \leftarrow$ Vectors of $\mathbf{P}$ corresponding to C th class

18: $\mathbf{P}_C^- \in \mathbb{R}^{d \times m_C^-} \leftarrow$ Vectors of $\mathbf{P}$ corresponding to non- C th class

19: $\mathbf{y}_C^+ \in \mathbb{N}^{m_C^+ \times 1} \leftarrow (\tau + \eta) \mathbf{1}^+$ and $\mathbf{y}_C^- \in \mathbb{N}^{m_C^- \times 1} \leftarrow (\tau - \eta) \mathbf{1}^-$

20: $\mathbf{P}_C \in \mathbb{R}^{m \times \tilde{d}} \leftarrow [\mathbf{P}_C^+, \mathbf{P}_C^-]^T$ and $\mathbf{y}_C \in \mathbb{R}^{m \times 1} \leftarrow [\mathbf{y}_C^+; \mathbf{y}_C^-]$

21: $\hat{\mathbf{P}}_C \in \mathbb{R}^{\tilde{d} \times 1} \leftarrow (\mathbf{b} \mathbf{I} + \mathbf{P}_C^T \mathbf{W} \mathbf{P}_C)^{-1} \mathbf{P}_C^T \mathbf{W} \mathbf{y}_C$

22: **end for**

Note: 'A $\leftarrow$ B' indicates that 'B' is assigned to 'A'.

output weights obtained in (4) as: $\hat{\mathbf{y}}_t = \mathbf{p}_t^T [\hat{\mathbf{P}}_1, \hat{\mathbf{P}}_2, \dots, \hat{\mathbf{P}}_{N_C}]$ where $\hat{\mathbf{y}}_t = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_{N_C}]$ denotes the estimated model outputs over N_C classes, and $\mathbf{p}_t = \mathcal{W}^T \phi(\mathbf{g}_\star^t) \in \mathbb{R}^{d \times 1}$ is the regression vector constructed using the test sample. The class label can then be predicted using the one-versus-all technique: $cls(\mathbf{X}_t) = \arg \max_C \hat{y}_t$. In other words, no prior class information is needed for this test sample prediction.

5) *Comparison with convolutional neural network (CNN):* In this subsection, the proposed AGFN is compared with the well-known CNN [46]–[48], in terms of architectural and learning properties. The upper panel of Fig. 4 illustrates the architectural difference in the hidden layer(s) between the two networks. The table shown in the lower part of Fig. 4 provides a brief summary of both networks for comparison.

Generally, the AGFN can be considered as a layer of the CNN. However, as shown in Fig. 4, the architectures of the hidden layer(s) of both networks have clear difference. The AGFN consists of a single hidden layer and its input weights (i.e., filter masks) are pre-fixed to Gabor kernels. Due to the fixed nature of input weights, only the output weights are to be estimated. Our approach for learning the output weights is to adopt an analytic solution based method so that no iterative search is required.

On the other hand, training the CNN is no simple task as it consists of $N > 2$ hidden layers with nonlinearity within each layer. Filter masks for local convolution should be learned by using a large amount of image data (either face images or images of other objects) [49]. In addition to the filters, fully connected weights should also be learned for the final classification task. Moreover, sophisticated parameter tuning

Fig. 4. Comparison between AGFN and CNN in terms of their architectural similarities and differences.

is often required to achieve outstanding recognition accuracy performance [50].

Despite these differences in architecture and learning mechanism, AGFN shares some functional similarities with that of CNN. As shown in Fig. 4, the AGFN hidden layer performs a Gabor based convolution, accumulated local Gabor features extraction, and feature activation using the sigmoid function. Particularly, the Gabor convolution step (denoted by 'a' in Fig. 4) is similar with the convolution step (d) in CNN. Similarly, extracting accumulated Gabor features followed by the sigmoid activation (b) is similar to taking a non-linear operation (Sigmoid, Rectified linear units [51] etc.) followed by a mean pooling (e).

IV. EXPERIMENTS

In this section, the one-to-N identification performance of the proposed AGFN is evaluated via extensive empirical experiments. Firstly, we analyze the effects of each AGFN parameter towards the identification performance empirically. Next the identification capability of AGFN is evaluated under head pose changes as well as several external imaging conditions. The following subsections describe our experimental design, protocols and results in detail.

A. Datasets, Preprocessing and Protocols

Six public face datasets/subsets (see Fig. 5): FERET b-series [52], CMU-PIE [53], Multi-PIE [54], FERET f-series [52], LFW-a [55] and COX [56], are utilized in this experimental study. The first three subsets are selected to evaluate the tolerance of AGFN to head pose changes. The last three are chosen to evaluate the tolerance of AGFN to variation of illumination, acquisition time and the wild condition. Note that the COX dataset is a recently released large-scale face dataset, and is provided with the standard evaluation protocol for the single-sample face recognition. Except for LFW10, subjects

Database name	A*	B*	Crop. size	Imaging noise(s)	Data protocol		Remarks	Sample images geometrically normalized and/or cropped (Rotated pose angle towards his left (+) or right (-) direction from the frontal pose)									
	C*	D*	Refs.		For training	For evaluation											
FERET b-series	200	1	130×90	Head pose only	Images of partition <i>ba</i>	Images of partitions <i>bb, bc, bd, be, bf, bg, bh, bi</i>	Images are geometrically normalized based on fiducial points released by the authors of [25].	<i>ba</i> (0°) <i>bb</i> (+60°) <i>bc</i> (+40°) <i>bd</i> (+25°) <i>be</i> (+15°) <i>bf</i> (-15°) <i>bg</i> (-25°) <i>bh</i> (-40°) <i>bi</i> (-60°)									
	200	1,600	[52]														
CMU-PIE	68	1	160×140	Head pose only	Images of partition <i>c27</i>	Images of partitions <i>c22, c02, c37, c05, c29, c11, c14, c34</i>	Images are geometrically normalized based on manually detected fiducial points.	<i>c27</i> (0°) <i>c22</i> (+60°) <i>c02</i> (+44°) <i>c37</i> (+31°) <i>c05</i> (+16°) <i>c29</i> (-17°) <i>c11</i> (-32°) <i>c14</i> (-46°) <i>c34</i> (-66°)									
	68	544	[53]														
Multi-PIE	249	1	160×140, 32×32	Head pose plus illumination	Images of partition <i>05_1</i> with illum. ID07	Images of partitions <i>14_0, 05_0 & 05_1</i> (other illum.)	Subjects of <i>Session 1</i> are used. Images are with neutral expression, 19 illuminations, 3 poses. Similar protocols with that of [22].	<i>05_1</i> (0°) <i>14_0</i> (-15°) <i>05_0</i> (+15°)									
	249	14,193	[54,22]														
FERET f-series	1,196	1	150×130	Expression, acquisition time	Images of partition <i>fa</i>	Images of partitions <i>fb, fc, Dup1, Dup2</i>	Images are aligned using Transform-Invariant PCA [57], and processed using Processing sequences (PS) normalization [59].
	1,196	2,345	[52,57,58]														
LFW10 (LFW-a)	158	5	56×58	Wild condition	Randomly selected 1 st half	Randomly selected 2 nd half	Identification protocol [60]. Only subs. ≥10 imgs. are used. Img. cropping according to [61].
	790	790	[55,60,61]														
COX	700	1	60×48	Wild cond., low resolution	All still images	All video frames of Camcorder 3 (V3)	Video-to-still face recognition (V3-S). Following the standard protocol [56]. Each video frame is regarded as an image. PS [59]	Still image Face images extracted from a video captured by Camcorder 3									
	700	12K	[56]														

* (A) Number of subjects utilized in our experiments, (B) Number of training samples (i.e., images) per subject, Total number of samples utilized for (C) training, and (D) test performance evaluation

Fig. 5. Geometrically aligned and/or cropped sample images of the data sets.

of the datasets possess only one image for training. Fig. 5 provides details of the datasets (e.g., data statistics, protocols, etc.) utilized in our experimental study.

B. Empirical analysis on AGFN parameters

This experiment aims to study the effects of AGFN parameters namely, the projection size k and the extraction size l , on the accuracy of face identification. Since k is closely related to the dimension of extracted features, both recognition accuracy and training CPU time are measured. Next, we study the role of l in face recognition across pose variation. The FERET b-series, which contains variation of head pose, is chosen for these studies. The adopted parameters for Gabor kernel construction [36] are: $\mu \in \{0, \dots, 7\}$, $v \in \{0, \dots, 4\}$, $k_{max} = \pi/2$, $\sigma = 2\pi$, $f = \sqrt{2}$, and $h = w = 21$. The reduced dimension $\tilde{d}$ by WPCA is calculated and set at a number corresponding to 95% of eigenenergy (see line 12 of Algorithm 1). Following [45], we set threshold $\tau = 0$ and offset $\eta = 1$. For statistical evidence over the randomness in $s_{[x,y]}$, the AGFN is iterated 10 times where the average performance is recorded.

1) *Effect of the projection size k* : The test recognition accuracies of AGFNs at various projection sizes $k \in \{1, 2, \dots, 20\}$ are shown in Fig. 6 (a). Images of partition *ba* are used for training while partition *bf* images are used for testing. Here, l is fixed at 3 in order to focus only on observing the effect of k . From this sub-figure, AGFNs show lower accuracy at low k values (e.g., $k < 7$) than that at high k values (e.g., $k > 10$). It can be seen that the accuracy values become stable at $k > 8$ and then get saturated after $k = 15$. Here we note that a similar trend is observed for experiments on all other datasets/subsets. Due to space constraint, these results are not presented here.

The CPU time elapsed for i) extracting the $\mathbf{g}_{ver}$ and/or $\mathbf{g}_{hor}$ features; and ii) estimating the output weights $\hat{\beta}$, are shown in Fig. 6 (b). This sub-figure reveals that all AGFNs spent

Fig. 6. Accuracy and CPU time performances of AGFN resulted at $k \in \{1, \dots, 20\}$ and $l = 3$ on FERET b-series

similar amount of time to estimate $\hat{\beta}$ regardless of k settings. This is because WPCA projects $\mathbf{g}_{ver}$ and $\mathbf{g}_{hor}$ features onto a low dimensional space. On the other hand, the elapsed time for feature extraction is observed to increase linearly with respect to k . Since AGFN_{fusion} computes both $\mathbf{g}_{ver}$ and $\mathbf{g}_{hor}$ features, extracting this fused feature takes longer time than these of individual ones. To balance between accuracy performance and computational efficiency, we fix k at 15 for the rest of our experiments.

2) *Effect of the extraction size l* : Fig. 7 shows the test recognition accuracy at various extraction sizes $l \in \{1, 2, \dots, 9, 10, 20, 30, \dots, 80, 90\}$ at $k = 15$ using the FERET b-series subset. The frontal face images of *ba* partition (single image per subject) are used for training while non-frontal face images of partitions *bb, bc, bd* and *be* are respectively used for test performance evaluation across pose variations.

It is observed from Fig. 7 that both AGFN_{ver} and AGFN_{hor} can almost perfectly identify face images from partition *be* at low extraction size $l < 5$. They also work reasonably well for images from partition *bd* at $l < 5$. Their accuracy values,

Fig. 7. Test recognition accuracies, which are plotted over extraction sizes $l \in \{1, 2, \dots, 9, 10, 20, 30, \dots, 80, 90\}$, obtained at $k = 15$ using FERET b-series.

however, drop (significantly for $AGFN_{ver}$) at high extraction sizes (e.g., $l > 10$). We note here that the images from both $be(+15^\circ)$ and $bd(+25^\circ)$ contain moderate pose variation.

Different from the results of partition be and bd discussed above, $AGFN_{hor}$ has produced higher accuracy values at $l > 10$ than those at $l \leq 10$ for partitions $bb(+40^\circ)$ and $bc(+60^\circ)$. In other words, $AGFN_{hor}$ shows better recognition performance at higher l setting when probe images contain large pose variation. $AGFN_{ver}$, again, shows low accuracy due to images from partitions bb and bc contain only horizontal pose variation.

In summary, the obtained results are observed as follows:

- The projection size k is closely related to the recognition accuracy as well as computational efficiency. In order to balance between them, we fix $k = 15$ where the accuracy values started to get saturated.
- The extraction size l plays an important role in mitigating difficulties caused by pose changes. This observation is congruent to our analysis on synthetic images (see Section III-A5 and Fig. 2).

C. Evaluation of AGFN performances under pose variation

The main goal for conducting this experiment is two-folded. The first is to investigate under what level of head pose as well as external imaging conditions can AGFN tolerate in terms of recognition accuracy. The second is to position AGFN among those competing state-of-the-art face recognition methods by comparing their recognition accuracy. To achieve this goal, AGFN and several competing methods are evaluated under three scenarios as shown in Table II. Under Scenarios I and II, the tolerances of AGFN to head pose variation only and head pose together with illumination changes are respectively evaluated. The tolerance of AGFN to other external imaging conditions is then evaluated under Scenario III.

Apart from the evaluation scenarios, Table II shows five groups of algorithms for performance comparison under the three scenarios mentioned above. Group 1 consists of competing Gabor-based face recognition methods being implemented in our experiments. These methods are listed in Table III together with their parameter settings. Group 2 consists of state-of-the-art pose-invariant face recognition methods (see the second column of Table IV and V). These Group 2 methods are 2D based and published recently with evaluation results on FERET b-series and/or CMU-PIE subsets. Group 3 consists of recently published deep learning based methods (see Table VI) which are evaluated under a single-sample face recognition protocol. Groups 4 and 5 consist of state-of-the-art

TABLE II
THREE SCENARIOS FOR PERFORMANCE EVALUATION AND COMPARISON

Scenario	Target external imaging conditions	Adopted dataset/subset	Performance Comparison with
I	Head pose, Single-sample face recognition	FERET b-series, CMU-PIE	Algo. Groups 1, 2, VGG-Face [32]
II	Head pose+illumination, Single-samp. face recog.	Multi-PIE	Algo. Group 3, VGG-Face [32]
III	Expression, Acquisition time, Wild condition,	FERET f-series, LFW10, COX	Algo. Groups 1, 4, 5, VGG-Face [32]

Algo. Group 1: State-of-the-art Gabor-based face recognition methods. Accuracy performances of this group are obtained from our experiments.
 Algo. Group 2[§]: State-of-the-art pose-invariant face recognition methods.
 Algo. Group 3[§]: Recently published deep learning based methods for single-sample face recognition.
 Algo. Group 4[§]: State-of-the-art face recognition methods published with evaluation results on FERET f-series subset.
 Algo. Group 5[§]: State-of-the-art face recognition methods published with evaluation results on COX subset.
 §: Accuracy performances of groups 2 to 5 are directly taken from literature. The accuracy performance of the proposed method has been observed under similar experimental protocols except for image resolution and adopted normalization techniques.

TABLE III
GROUP 1 METHODS WITH THEIR PARAMETER SETTINGS

Methods	Parameter settings
GaborKPCA [†] [63]	Polynomial kernel; Mahalanobis distance; $d \in \{40, 80, \dots, 400\}$ for all except for CMU-PIE; $d \in \{5, 10, \dots, 65\}$ for CMU-PIE
LGBPHS [64]	Block size 4×8 ; Number of bins 16; Histogram intersection
GRCM [65]	AIRM similarity [66]
MBC-O [‡] [67]	Num. of blocks $M_b = 4$ and $M_r = 2$; Num. of scale = 3; BFLD dim. $d = 200$
MOST [‡] [42]	Gaussian kernel (with $t = 30$); Euclidean distance
All methods listed above	$\mu \in \{0, 1, \dots, 7\}$; $v \in \{0, 1, \dots, 4\}$; $k_{max} = \pi/2$; $\sigma = 2\pi$; $f = \sqrt{2}$; $h = w = 21$; Filtering in the frequency domain

Abbreviations: Gabor-based Kernel PCA (GaborKPCA); Gabor-based region covariance matrix (GRCM); Local Gabor binary patterns with histogram sequence (LGBPHS); Monogenic binary coding w/ orientation (MBC-O); Multiple orientation and scale transform (MOST).

b: Code source: http://luks.fe.uni-lj.si/sl/osebje/vitimir/face_tools/PhDface/

§: The authors released source code for ‘MBC-O’rientation, ‘MBC-A’mplitude and ‘MBC-P’hase. We selected MBC-O due to its better performance than the other two [67]. Code source: <http://www.vision.ee.ethz.ch/~yangme/paper/TIFS/MBc.zip>

‡: Code source: http://mcs.sysu.edu.cn/user/renchx/ArtList/Cate_8533/P1/E

Note: LGBPHS and GRCM are implemented by ourselves.

face recognition methods published recently with evaluation results on FERET f-series and COX subset, respectively. In addition to these five groups of competing methods, a CNN-based method namely the VGG-Face [32], is also included in all of our comparison studies (see Table II). Particularly, the pre-trained VGG-Face model [62] has been used as an extractor of deep features (of 4,096 dimension) followed by the TER classifier for class label prediction. For convenience, we shall use ‘VGG-Face + TER’ and ‘VGG-Face’ interchangeably hereafter. The publicly available source code for VGG-Face [62] has been adopted in our experimental study.

1) *Scenario I*: Under this scenario, we evaluate the tolerance of AGFNs and competing methods to horizontal head pose changes on both FERET b-series and CMU-PIE subsets. The best test results obtained from the various parameter settings using these two subsets are respectively tabulated in Table IV and V. Note that these tables consist of four groups of methods: i) three AGFNs (denoted as Group 0 in these tables), ii) Group 1 methods, iii) Group 2 methods, and iv) VGG-Face method. Within Group 0, apart from single-image training, we have included a subgroup for two-image training to demonstrate the advantage of having more images for training. In the rest of this subsection, we shall discuss

TABLE IV

THE BEST TEST RECOGNITION ACCURACY (%) ON FERET B-SERIES. THE ACCURACY VALUES OF THE FIRST TWO GROUPS ARE OBTAINED FROM OUR EXPERIMENTS. THE ACCURACY VALUES OF THE THIRD GROUP ARE DIRECTLY TAKEN FROM LITERATURE. IMAGES OF $ba(0^\circ)$ ARE USED FOR TRAINING

Algo. Group	Methods		Image notation / Rotated degree of head from a frontal pose								Landmark detection (No. of pts)	Pose estimation / compensation	Multiple non-frontal faces in training?
			<i>bb</i> +60°	<i>bc</i> +40°	<i>bd</i> +25°	<i>be</i> +15°	<i>bf</i> -15°	<i>bg</i> -25°	<i>bh</i> -40°	<i>bi</i> -60°			
0	Single-image training	AGFN _{ver}	11.7 (<i>l</i> =2)	34 (<i>l</i> =1)	90.9 (<i>l</i> =1)	97.7 (<i>l</i> =1)	99.2 (<i>l</i> =2)	95.5 (<i>l</i> =1)	68.6 (<i>l</i> =4)	16.2 (<i>l</i> =3)	Manual (3)	APK ¹ / FA ²	No
		AGFN _{hor}	45.4 (<i>l</i> =85)	60.2 (<i>l</i> =35)	93.6 (<i>l</i> =2)	97.6 (<i>l</i> =6)	99.3 (<i>l</i> =1)	96.7 (<i>l</i> =2)	78.2 (<i>l</i> =9)	52.4 (<i>l</i> =55)			
		AGFN _{fusion}	31.1 (<i>l</i> =60)	47.8 (<i>l</i> =25)	92.6 (<i>l</i> =1)	97.4 (<i>l</i> =2)	99.4 (<i>l</i> =1)	95.9 (<i>l</i> =3)	77.5 (<i>l</i> =1)	40.6 (<i>l</i> =50)			
	Two-image tr.	AGFN _{hor} (TI ³ : <i>be</i> and <i>bf</i>)	70.6 (<i>l</i> =50)	93.5 (<i>l</i> =15)	100 (<i>l</i> =1)	TI -	TI -	100 (<i>l</i> =1)	93.6 (<i>l</i> =25)	75 (<i>l</i> =90)	Manual (3)	APK / FA	Yes
AGFN _{hor} (TI: <i>bd</i> and <i>bg</i>)		76.4 (<i>l</i> =25)	97.6 (<i>l</i> =8)	TI -	100 (<i>l</i> =1)	100 (<i>l</i> =1)	TI -	95.3 (<i>l</i> =2)	77.3 (<i>l</i> =85)				
1	GaborKPCA (<i>d</i>)		5 (320)	5 (240)	40 (400)	48 (360)	86 (400)	64.5 (400)	22 (400)	7 (120)	Manual (3)	APK / FA	No
	LGBPHS		4	18	81.5	88	95	85.5	52	6			
	GRCM		5	12	24	61.5	69	28.5	28	21			
	MBC-O		0.5	10.5	80	95.5	99.5	90.5	31.5	2.5			
	MOST		8.5	23.5	85	91.5	98	92.5	56	10			
2	PAN [68]		44	81.5	93	97	98.5	91.5	78.5	52.5	Manual (58)	N.A. / PN ⁴ , SF ⁵ , PW ⁶	Yes
	EGFC-S1 [19]		93	99	100	100	100	100	100	94	Manual (5)	Auto / FA, IS ⁷ , 3D ⁸	Yes
	ADMCLS [24]		70	82	94	95	96	94	85	79	Manual (4)	APK / FA	Yes
	RidgeGab [22]		82.5	~95 [†]	97.5	~96	~95	~95	~90	~78	Manual (5)	APK / FA, IS	Yes
	PAF [31]		93.7	98	98.5	99.2	99.2	98.5	98	93.7	Auto (34)	Auto / FA, SF, 3D	Yes
	SPAE [13]		77	96	98	99	99	99	95	77	Manual (5)	N.A. / FA, DL ⁹	Yes
	MVV-AE [20]		45.4	88.1	98	99.8	99.5	98.2	84.5	42.5	SIFT ¹⁰ (123)	APK / PW, IS	Yes
	Cament et al. [29]		89	99.5	100	100	100	100	98.1	84.6	Auto (66)	Auto / FA, SF	Yes
	Haghighat et al. [2]		91.5	96	100	100	100	100	99	93	Auto (68)	Auto / FA, SF	No
VGG-Face [32] + TER			55.5	92	99	99	99	99.5	89.5	69.5	N.A.	N.A. / DL	Yes

Abbreviations - 1: Assumed pose in known (APK), 2: Face alignment (FA), 3: Training image (TI), 4: Pose normalization (PN), 5: Shape fitting (SF), 6: Patch warping (PW), 7: Image synthesis (IS), 8: Some steps require 3D facial information (3D), 9: Deep learning (DL), 10: Scale invariant feature transform (SIFT).

†: '~' indicates the value is approximated from a graph.

about the results per group to provide an overview.

The Groups 0 and 1 in Tables IV and V show that AGFNs outperform those Group 1 methods in test recognition accuracy. Particularly, single-image AGFN_{hor} (see the upper panel of Group 0) produces the best test accuracies over various partitions of the two subsets. When the probe images contain moderate head pose variation, MBC-O, MOST (for partitions from bd to bg of FERET b-series, see Table IV), and GRCM, MBC-O (for partition $c05$ of CMU-PIE, see Table V) achieve about 89–98% of accuracy. However, their performances drop significantly when the probe images contain significant pose variation (e.g., partitions bb , bc , bh and bi of Table IV, and partitions $c22$, $c02$, $c14$ and $c34$ of Table V).

Different from those AGFNs and Group 1 methods, VGG-Face and Group 2 methods generally work well under moderately rotated poses (e.g., $\pm 30^\circ$). For images with moderate pose variation (e.g., partitions from bd to bg of Table IV, and partitions from $c37$ to $c11$ of Table V), the AGFNs' accuracies appear to be comparable or slightly worse than those of the compared VGG-Face and Group 2 methods. Except for PAN, ADMCLS, MVV-AE and Sarfraz et al., VGG-Face and Group 2 methods generally show good recognition performances regardless of the rotation degree of the head. For example, the reported results of EGFC-S1, RidgeGab and Cament et al., obtained using partition $bb(+60^\circ)$ are $> 82\%$. The best accuracy achieved by AGFN for the partition is 45.4%.

Here, it is noted that the proposed AGFN is among the

simplest algorithms within the compared pose-invariant face recognition methods. Particularly, AGFN uses only a single-image for training and does not require any sophisticated processing for pose compensation. Approaches to pose compensation include, but not limited to, image synthesis, image warping and face shape fitting. As shown in the last second column of Tables IV and V, the Group 2 methods require pose compensation to reduce differences caused by pose changes. For example, PAF requires several complicated processing steps (e.g., face shape fitting, 3D face modeling and filter transformation) to reduce the pose gap. Finally, most of those Group 2 methods require multiple non-frontal face images (see the last column of the tables) for training. In contrast, the proposed AGFN does not have such requirement.

To observe the merits of using multiple number of non-frontal face images in training, a case study is conducted here. In this extended study, the accuracy performance of AGFN_{hor} is observed when two non-frontal face images per subject is used for training. Interestingly, as shown in the lower panel of Group 0 of Tables IV and V, a large gap of accuracy improvement is observed. For example, about 22–30% of accuracy increment on $bb(+60^\circ)$ and $bi(-60^\circ)$ partitions is observed from Table IV. These results are either comparable or slightly worse than the VGG-Face and Group2 methods.

2) *Scenario II*: Different from the two subsets (i.e., FERET b-series and CMU-PIE) of Scenario I, the Multi-PIE subset used in this scenario consists of images with moderate ($\pm 15^\circ$)

TABLE V

THE BEST TEST RECOGNITION ACCURACY (%) ON CMU-PIE. THE ACCURACY VALUES OF THE FIRST TWO GROUPS ARE OBTAINED FROM OUR EXPERIMENTS. THE ACCURACY VALUES OF THE THIRD GROUP ARE DIRECTLY TAKEN FROM LITERATURE. IMAGES OF $c27(0^\circ)$ ARE USED FOR TRAINING

Algo. Group	Methods		Image notation / Rotated degree of head from a frontal pose								Landmark detection (No. of pts)	Pose estimation / compensation	Multiple non-frontal faces in training?
			c22 +60°	c02 +44°	c37 +31°	c05 +16°	c29 −17°	c11 −32°	c14 −46°	c34 −66°			
0	Single-image training	AGFN _{ver}	16.2 (<i>l</i> =4)	76.5 (<i>l</i> =2)	48.8 (<i>l</i> =4)	97.7 (<i>l</i> =2)	60 (<i>l</i> =1)	34.1 (<i>l</i> =4)	13.2 (<i>l</i> =1)	10.6 (<i>l</i> =6)	Manual (3)	APK / FA	No
		AGFN _{hor}	37.9 (<i>l</i> =89)	91.8 (<i>l</i> =85)	93.8 (<i>l</i> =75)	99.4 (<i>l</i> =70)	98.2 (<i>l</i> =65)	93.5 (<i>l</i> =90)	69.4 (<i>l</i> =90)	26.2 (<i>l</i> =90)			
		AGFN _{fusion}	27.7 (<i>l</i> =65)	85 (<i>l</i> =55)	84.4 (<i>l</i> =80)	98.8 (<i>l</i> =70)	93.5 (<i>l</i> =35)	78.8 (<i>l</i> =85)	57.1 (<i>l</i> =90)	21.8 (<i>l</i> =90)			
	Two-image tr.	AGFN _{hor} (TI: c05 and c29)	48.2 (<i>l</i> =85)	96.5 (<i>l</i> =8)	98 (<i>l</i> =35)	TI -	TI -	98.2 (<i>l</i> =25)	86.5 (<i>l</i> =85)	35.3 (<i>l</i> =85)	Manual (3)	APK / FA	Yes
		AGFN _{hor} (TI: c37 and c11)	74.1 (<i>l</i> =25)	96.2 (<i>l</i> =45)	TI -	98.5 (<i>l</i> =40)	98.2 (<i>l</i> =15)	TI -	90.3 (<i>l</i> =45)	41.2 (<i>l</i> =85)			
1	GaborKPCA (<i>d</i>)		8.8 (60)	33.8 (65)	20.6 (55)	54.4 (50)	20.1 (50)	17.7 (35)	7.4 (25)	5.9 (50)	Manual (3)	APK / FA	No
	LGBPHS		2.9	20.6	44.1	58.8	10.3	4.4	2.9	2.9			
	GRCM		14.7	20.6	39.7	89.7	51.5	54.4	11.8	4.4			
	MBC-O		1.5	25	8.8	98.5	11.8	5.9	2.9	0			
	MOST		4.4	30.9	17.7	35.3	10.3	8.8	1.5	2.9			
2	Sarfraz et al. [69]		~40 [†]	~78	~85	~83	~88	~83	~75	~37	N.A.	APK / IS	Yes
	RidgeGab [22]		~73	100	100	100	100	100	~83	~73	Manual (5)	APK / FA, IS	Yes
	ADMCLS [24]		79	88	100	100	100	100	97	85	Manual (4)	APK / FA	Yes
	PAF [31]		72.9	97.9	100	100	100	100	98.9	75	Auto (34)	Auto / FA, SF, 3D	Yes
	Arashloo et al. [28]		92	100	100	100	100	100	100	95	Auto (dense)	N.A. / PW, IS	No
	Cament et al. [29]		61.2	90.9	92.9	100	100	100	74.3	47.5	Auto (66)	Auto / FA, SF	Yes
	Haghighat et al. [2]		-	-	100	100	100	100	-	-	Auto (68)	Auto / FA, SF	No
	VGG-Face [32] + TER		52.9	89.7	98.5	98.5	98.5	98.5	77.9	29.4	N.A.	N.A. / DL	Yes

See the note of Table IV for abbreviations (e.g., APK, FA, IS, PW, SF, DL) / †: '~' indicates the value is approximated from a graph.

horizontal pose variation plus 19 illumination changes. Moreover, the proposed AGFN_{hor} (selected as the representative among AGFNs based on the results of Tables IV and V) is compared with the deep learning-based Group 3 methods and VGG-Face in terms of their recognition accuracies.

When the low resolution images (of 32×32 pixels) are utilized (see the second, third, fifth and sixth rows of Table VI), VGG-Face outperforms all compared methods except for SSAE under the 05_I partition in terms of test recognition accuracy. The AGFN_{hor} produces higher accuracy values than that of DAE and MDAE. However, as shown in the fourth row of Table VI, the accuracy values of AGFN_{hor} increase significantly when the images are photometrically normalized using processing sequences (PS) [59] and of high resolution (of 160×140 pixels). It is worth noting here that different from the deep learning methods, the efficiency of AGFNs is not much affected by the resolution of adopted images.

3) *Scenario III*: Under this scenario, AGFNs and the compared VGG-Face, Groups 1 and 4 methods are evaluated under two different external imaging conditions: variation of expression and difference in acquisition time using FERET f-series. AGFNs, VGG-Face, Groups 1 and 5 methods are subsequently evaluated under the wild condition using both LFW10 and COX subsets.

The test recognition accuracies on FERET f-series are tabulated in Table VII. To improve the readability of results over the four subsets, only the best performances of GaborKPCA and AGFNs are tabulated along the used parameter setting as in [38]. It is observed from Table VII that the proposed AGFNs produce comparable recognition performances with competing

TABLE VI
COMPARISON OF TEST RECOGNITION ACCURACY (%) WITH DEEP LEARNING METHODS UNDER POSE-INVARIANT AND SINGLE-SAMPLE FACE RECOGNITION PROTOCOL (EVALUATED ON MULTI-PIE SUBSET)

Algo. Group	Methods	Notation / Rotatd degree			Image resolution (Normalization)
		05_I 0°	14_0 -15°	05_0 +15°	
0	AGFN _{hor} at $h = w^{\dagger} = 13$	76.10 (<i>l</i> = 1)	38.65 (<i>l</i> = 3)	37.78 (<i>l</i> = 2)	32×32 pixel (No normali.)
	AGFN _{hor} at $h = w = 9$	89.11 (<i>l</i> = 1)	49.87 (<i>l</i> = 1)	48.33 (<i>l</i> = 2)	32×32 pixel (With PS [59])
	AGFN _{hor} at $h = w = 21$	100 (<i>l</i> = 2)	97.78 (<i>l</i> = 4)	98.74 (<i>l</i> = 6)	160×140 pixel (With PS [59])
3*	DAE [70]	59.06	19.1	22.49	32×32 pixel (<i>l</i> ₂ -norm.)
	MDAE [1]	71.34	20.46	24.12	
	SSAE [1]	97.93	67.19	63.25	
VGG-Face [32] + TER		92.79	87.38	85.12	32×32 pixel (<i>l</i> ₂ -norm.)
		98.88	96.62	96.36	160×140 pixel (<i>l</i> ₂ -norm.)

Abbreviations - Denoising auto-encoder (DAE), Modified DAE (MDAE), Stacked supervised auto-encoder (SSAE), Processing sequences (PS) normalization.

*The accuracy values of Group 3 methods are directly taken from [1].

[†] $h = w$ indicates the Gabor kernel size in pixels (see the first paragraph of Section IV-B).

state-of-the-arts for all the four probe sets. Particularly, AGFNs achieve around 99% recognition accuracy on the first two probe sets, and around 93–94% recognition accuracy on the last two probe sets.

Fig. 8 and Table VIII respectively show the test recognition accuracies of the evaluated methods on the LFW10 and COX subsets. VGG-Face significantly outperforms the AGFNs and Group 1 methods in terms of the test accuracy on LFW10 subset (see Fig. 8). Among the Gabor-based meth-

TABLE VII
TEST RECOGNITION ACCURACY (%) ON FERET F-SERIES

Methods	f_b	f_c	$Dup1$	$Dup2$	Avg.	Year
GaborKPCA (d)	86.8	69.6	61.6	67.5	71.4	2004
LGBPHS	93.4	93.3	80.3	82.1	87.3	2005
Weighted HGPP [38] [†]	97.5	99.5	79.5	77.8	88.6	2007
GRCM	81.0	32.5	36.4	43.2	48.3	2008
HEC [37] [‡]	99.0	99.0	92.0	88.0	94.5	2009
S[LGBP+LGXP] [39] [‡]	99.9	99.9	94.0	93.0	96.7	2010
Tan&Triggs [59] [‡]	98.0	98.0	90.0	85.0	92.8	2010
E-GV-LBP-M [71] [‡]	98.4	98.9	81.9	81.6	90.2	2011
LMG+EJS [72] [‡]	99.8	100	88.0	84.2	93.0	2011
POEM+WPCA [73] [‡]	99.6	99.5	88.8	85.0	93.2	2012
MBC-F [67] [‡]	99.7	99.5	93.6	91.5	96.1	2012
G-LQP [74] [‡]	99.9	100	93.2	91.0	96.0	2012
s-POEM+WPCA [75] [‡]	99.4	100	91.7	90.2	95.3	2013
GOM [58] [‡]	99.9	100	95.7	93.1	97.1	2014
MOST [‡]	94.8	73.7	79.5	79.5	81.9	2014
H-Groupwise MRF [76] [‡]	99.7	99.2	94.7	93.6	96.8	2014
DFD [77] [‡]	99.4	100	91.8	92.3	95.9	2014
CBFD+WPCA [78] [‡]	99.8	100	93.5	93.2	96.6	2015
VGG-Face [32] + TER	98.83	95.36	90.17	93.16	94.38	2015
KED [79] [‡]	99.5	100	91.5	86.3	94.4	2016
AGFN _{ver}	99.0	99.0	93.3	93.0	96.1	Ours
AGFN _{hor}	($l=7$)	($l=1$)	($l=1$)	($l=4$)	96.4	
AGFN _{fusion}	($l=3$)	($l=2$)	($l=3$)	($l=4$)	96.7	

[†]: The accuracies are taken from the original papers. Except for image resolution and adopted normalization techniques, the data protocols of this subset are standardized [52].
[‡]: The source codes released by the authors contain only essential functions for MOST algorithm. Components for accuracy performance computation are not provided. Thus, the results shown in this table may not fully exploit its capability.

ods, AGFN_{fusion} achieves the best test recognition accuracy performance. However, for the COX subset, the VGG-Face fails to maintain such a good accuracy performance (see Table VIII). This could be due to the large difference between the COX images (see the last row of Fig. 5) and those VGG images [32] used in training the VGG-Face model. According to [32], the VGG images were acquired from the web similar to that of the LFW images. Moreover, the authors of [80] have pointed out that a CNN trained only on still images under the video-to-still face recognition protocol may not produce a good recognition accuracy performance for the similar reason discussed above. Since the VGG-Face is a very deep network [32], fine-tuning the model on COX dataset is not a trivial task unless we have very powerful computing resources. Among the compared methods, the best accuracy performance of the point-to-set matching-based Group 5 methods appear to be about 10% higher than that of the AGFNs. Considering the point-to-point matching is a more difficult protocol than that of the point-to-set matching, it can be seen that the proposed method works reasonably well.

D. CPU processing time performance

AGFNs, VGG-Face and the Group 1 methods are now evaluated in terms of their CPU processing time elapsed for extracting Gabor or deep features and training a model to evaluate 100 test samples using the LFW10 subset (see Table IX). To evaluate the efficiency of the proposed feature extraction, the CPU time elapsed for extracting the conventional Gabor features (denoted as ‘Conventional Gabor’) is also included in the table. The CPU times for GaborKPCA and AGFNs are

Fig. 8. Comparison of test recognition accuracy (%) on LFW 10 subset. Since LGBPHS, GRCM, MBC-O, MOST and VGG-Face are not influenced by l and d , their results are repeated to from a line for better visibility.

TABLE VIII
TEST RECOGNITION ACCURACY (%) ON COX DATASET. SINCE THE VIDEOS RECORDED USING CAMCORDER 3 (V3) ARE LONGER THAN CAMCORDERS 1 (V1) AND 2 (V2) VIDEOS, ONLY THE V3 IS UTILIZED IN THIS STUDY

Methods	V3-S	Remarks
PSDML [81]	25.43	Rank-1 accuracy, Point-to-set matching (a video is regarded as a set of images), Group 5 methods
LERM-EG [82]	47.87	
PSCL-EG [56]	48.43	
LERM-EA [82]	59.76	
PSCL-EA [56]	47.74	
LERM-ES [82]	58.37	One-to-N classification accuracy, Point-to-point matching, (each frame of a video is regarded as an image)
PSCL-ES [56]	53.26	
VGG-Face	21.65	
AGFN _{ver}	43.30 ($l=5$)	
AGFN _{hor}	45.13 ($l=3$)	
AGFN _{fusion}	49.58 ($l=10$)	

Abbreviations: Point-to-set distance metric learning (PSDML), Point-to-set correlation learning (PSCL), Learning Euclidean-to-Riemannian Metric (LERM), Euclidean-to-AffineGrassmannian (EA), Euclidean-to-Grassmannian (EG), Euclidean-to-SPD (ES), Symmetric positive definite (SPD)

respectively measured at $d = 400$, and $l = 2$ and $k = 15$ which showed the best test accuracy performance. The CPU time was measured using a PC of 3.4 GHz CPU with 16 GByte RAM and Matlab [83].

As shown in Table IX, AGFNs take less processing time than VGG-Face and all other compared Gabor-based face recognition methods. Individual AGFN_{ver} or AGFN_{hor} is about 2–11 times faster, and AGFN_{fusion} is about 1.4–8.8 times faster than most compared methods in terms of training speed. The AGFNs also outperform all compared methods in terms of testing time (about 1.3–22 times faster). The table also shows that extraction of individual $\mathbf{g}_{ver}$ and $\mathbf{g}_{hor}$ features is about 7 times faster than that for the conventional Gabor magnitude features. Among those Group 1 methods, the algorithm MOST takes the lowest processing time for both training and testing while MBC-O takes the highest processing time.

E. Summary of results and observations

The evaluation results and observations are summarized as follows:

- The proposed AGFNs have produced about 90–96% of accuracy when the probe images contain moderate pose variation (about -30° to $+30^\circ$). For similar probe images, the best accuracy performances achieved by the

TABLE IX
A COMPARISON OF CPU TIME (SEC.) PERFORMANCE ON LFW10 SUBSET

Methods	Time for extracting features ($\diamond$)	Time for training ($\blacklozenge$)	Total time ($\diamond + \blacklozenge$)	Time for evaluating 100 test samples
GaborKPCA ($d = 400$)	82.34	-	82.34	12.00
LGBPHS	310.37	-	310.37	49.33
GRCM	87.99	-	87.99	13.46
MBC-O [†]	-	-	449.29	71.14
MOST	67.89	8.32	76.21	8.79
VGG-Face + TER	327.37	32.75	360.12	40.11
Conventional Gabor (Gabor+WPCA+TERRM)	90.51	26.10	116.61	21.19
AGFN _{ver} ($l = 1$)	12.87	25.72	38.59	3.21
AGFN _{hor} ($l = 1$)	12.98	25.06	38.04	3.27
AGFN _{fusion} ($l = 1$)	25.81	25.83	51.64	6.57

$\diamond$: Time consumed for Gabor filtering and feature extraction from the Gabor responses, or extracting deep features

$\blacklozenge$: Time consumed for training a classifier (i.e., estimating the classifier coefficients)

[†]: Different from other methods, MBC-O computes Gabor features from image patches. We thus measure total time consumed to produce the final features.

Group 1 methods were about 80–92%. The performances of AGFNs are seen to be comparable or slightly lower than the VGG-Face and Group 2 methods.

- For those significantly rotated (e.g., $\pm 60^\circ$) images, all evaluated AGFNs and Group 1 methods failed to produce satisfactory performance. Particularly, AGFN_{hor} achieved about 26–52% accuracy, while the best performance of Group 1 methods was about 0.5–10%. When two non-frontal face images are available for training, the accuracy performance of AGFN_{hor} was observed to be either comparable or slightly worse than that of VGG-Face and Group2 methods which used multiple images for training.
- When only low resolution images were available, AGFN_{hor} appeared to be better than Group 3 methods except for SSAE and VGG-Face in terms of accuracy, under the single-sample face recognition protocol. However, it was empirically observed that the recognition accuracy of AGFN_{hor} could be significantly increased when photometrically normalized (using PS method) high resolution images were available.
- The proposed AGFNs have outperformed the Group 1 methods under variation of expression and acquisition time (on FERET f-series), and under the wild condition (on LFW10). Finally, AGFNs produced similar or slightly lower accuracy performances than those of point-to-set matching methods under the video-to-still face recognition problem (on COX).
- Since AGFNs extract localized Gabor features by discarding redundant Gabor convolutions, it is computationally efficient (e.g., training individual AGFN_{ver} or AGFN_{hor} was about 2–11 times faster than competing Gabor-based face recognition methods).

V. CONCLUSION

In this paper, a single hidden layer analytic Gabor feedforward network was proposed for face recognition. Essentially, the input layer took in raw images and propagated geometrically localized sub-image patches to the hidden layer. At the hidden layer, vertically and horizontally accumulated Gabor magnitude features were extracted over several orientation

and scale settings. After a whitening process with dimension reduction, the extracted features were finally fused at the output layer for classification decision based on the total error rate minimization. Our empirical results on CMU-PIE, Multi-PIE, FERET, LFW and COX datasets showed comparable or better accuracy and CPU processing time performances than state-of-the-art and/or Gabor-based face recognition methods. Although the shallow networks are computationally efficient, their generalization ability for complex data could be limited. The proposed method trades off between such weakness and computational efficiency for applications in which only a single training sample per person is available.

ACKNOWLEDGEMENT

The authors are thankful to the associate editor and the anonymous reviewers for their constructive and insightful comments to improve the manuscript. In addition, we would like to thank Dr. Cheng-Yaw Low for sharing his experience in deep learning.

REFERENCES

- [1] S. Gao, Y. Zhang, K. Jia, J. Lu, and Y. Zhang, "Single Sample Face Recognition via Learning Deep Supervised Autoencoders," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 10, pp. 2108–2118, 2015.
- [2] M. Haghghat, M. Abdel-Mottaleb, and W. Alhalabi, "Fully automatic face normalization and single sample face recognition in unconstrained environments," *Expert Syst. Appl.*, vol. 47, pp. 23–34, 2016.
- [3] M. Yang, X. Wang, G. Zeng, and L. Shen, "Joint and Collaborative Representation with Local Adaptive Convolution Feature for Face Recognition with Single Sample Per Person," *Pattern Recogn.*, vol. 66, pp. 117–128, 2017.
- [4] J.-Y. Zhu, W.-S. Zheng, F. Lu, and J.-H. Lai, "Illumination Invariant Single Face Image Recognition under Heterogeneous Lighting Condition," *Pattern Recogn.*, vol. 66, pp. 313–327, 2017.
- [5] S. Hong, W. Im, J. Ryu, and H. S. Yang, "SSPP-DAN: Deep Domain Adaptation Network for Face Recognition with Single Sample Per Person," *arXiv preprint arXiv:1702.04069*, 2017.
- [6] Y. Gao, J. Ma, and A. L. Yuille, "Semi-Supervised Sparse Representation Based Classification For Face Recognition With Insufficient Labeled Samples," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2545–2560, 2017.
- [7] F. Mokhayeri, E. Granger, and G.-A. Bilodeau, "Domain-Specific Face Synthesis for Video Face Recognition from a Single Sample Per Person," *arXiv preprint arXiv:1801.01974*, 2018.
- [8] L. Zhuang, T.-H. Chan, A. Y. Yang, S. S. Sastry, and Y. Ma, "Sparse Illumination Learning and Transfer for Single-Sample Face Recognition with Image Corruption and Misalignment," *Int. J. Comput. Vision*, vol. 114, no. 2–3, pp. 272–287, 2015.
- [9] X. Zhang and Y. Gao, "Face recognition across pose: A review," *Pattern Recogn.*, vol. 42, no. 11, pp. 2876–2896, 2009.
- [10] C. Ding and D. Tao, "A Comprehensive Survey on Pose-Invariant Face Recognition," *ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 3, pp. 37:1–37:42, 2016.
- [11] Z. Zhu, P. Luo, X. Wang, and X. Tang, "Deep learning identity-preserving face space," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 113–120.
- [12] Y. Zhang, M. Shao, E. K. Wong, and Y. Fu, "Random Faces Guided Sparse Many-to-one Encoder for Pose-invariant Face Recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2416–2423.
- [13] M. Kan, S. Shan, H. Chang, and X. Chen, "Stacked Progressive Auto-Encoders (SPA) for Face Recognition Across Poses," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, 2014, pp. 4321–4328.
- [14] J. Yim, H. Jung, B. I. Yoo, C. Choi, D. Park, and J. Kim, "Rotating Your Face Using Multi-task Deep Neural Network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, 2015, pp. 676–684.
- [15] J. Heo and M. Savvides, "Gender and Ethnicity Specific Generic Elastic Models from a Single 2D Image for Novel 2D Pose Face Synthesis and Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 12, pp. 2341–2350, 2012.

- [16] H. Han and A. K. Jain, "3D Face Texture Modeling from Uncalibrated Frontal and Profile Images," in *Proc. IEEE Int. Conf. Biometrics, Theory, Appl. Syst.*, 2012, pp. 223–230.
- [17] V. Blanz and T. Vetter, "Face Recognition Based on Fitting a 3D Morphable Model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1063–1074, 2003.
- [18] H. T. Ho and R. Chellappa, "Pose-Invariant Face Recognition using Markov Random Fields," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1573–1584, 2013.
- [19] S. Li, X. Liu, X. Chai, H. Zhang, S. Lao, and S. Shan, "Morphable Displacement Field Based Image Matching for Face Recognition across Pose," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 102–115.
- [20] Y. Gao and H. J. Lee, "Cross-pose face recognition based on multiple virtual views and alignment error," *Pattern Recogn. Lett.*, vol. 65, pp. 170–176, 2015.
- [21] X. Chai, S. Shan, X. Chen, and W. Gao, "Locally Linear Regression for Pose-invariant Face Recognition," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1716–1725, 2007.
- [22] A. Li, S. Shan, and W. Gao, "Coupled Bias-Variance Tradeoff for Cross-Pose Face Recognition," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 305–315, 2012.
- [23] A. Li, S. Shan, X. Chen, and W. Gao, "Maximizing Intra-individual Correlations for Face Recognition Across Pose Differences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, 2009, pp. 605–611.
- [24] A. Sharma, M. A. Haj, J. Choi, L. S. Davis, and D. W. Jacobs, "Robust pose invariant face recognition using coupled latent space discriminant analysis," *Comput. Vis. Image Underst.*, vol. 116, no. 11, pp. 1095–1110, 2012.
- [25] L. Wiskott, J.-M. Fellous, N. Kruger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 775–779, 1997.
- [26] C. D. Castillo and D. W. Jacobs, "Using stereo matching with general epipolar geometry for 2d face recognition across pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2298–2304, 2009.
- [27] 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 Recogn.*, 2013, pp. 3025–3032.
- [28] S. R. Arashloo and J. Kittler, "Fast pose invariant face recognition using super coupled multiresolution Markov Random Fields on a GPU," *Pattern Recogn. Lett.*, vol. 48, pp. 49–59, 2014.
- [29] L. A. Cament, F. J. Galdames, K. W. Bowyer, and C. A. Perez, "Face recognition under pose variation with local Gabor features enhanced by Active Shape and Statistical Models," *Pattern Recogn.*, vol. 48, no. 11, pp. 3371–3384, 2015.
- [30] C. Ding, C. Xu, and D. Tao, "Multi-task Pose-invariant Face Recognition," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 980–993, 2015.
- [31] D. Yi, Z. Lei, and S. Z. Li, "Towards Pose Robust Face Recognition," in *Proc. IEEE Int. Conf. Pattern Recogn.*, 2013, pp. 3539–3545.
- [32] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," in *Proc. the British Mach. Vision Conf.*, vol. 1, no. 3, 2015, p. 6.
- [33] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. John Wiley & Sons, 2012.
- [34] F. Anselmi, J. Z. Leibo, L. Rosasco, J. Mutch, A. Tacchetti, and T. Poggio, "Magic materials: a theory of deep hierarchical architectures for learning sensory representations," *CBCL paper*, 2013.
- [35] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Am. A*, vol. 2, no. 7, pp. 1160–1169, 1985.
- [36] 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, 2002.
- [37] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical Ensemble of Global and Local Classifiers for Face Recognition," *IEEE Trans. Image Process.*, vol. 18, no. 8, pp. 1885–1896, 2009.
- [38] 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, 2007.
- [39] 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, 2010.
- [40] C. Wang, Z. Chai, and Z. Sun, "Face Recognition Using Histogram of Co-occurrence Gabor Phase Patterns," in *Proc. IEEE Int. Conf. Image Process.*, 2013, pp. 2777–2781.
- [41] K.-A. Toh and H.-L. Eng, "Between Classification-Error Approximation and Weighted Least-Squares Learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 4, pp. 658–669, April 2008.
- [42] C.-X. Ren, D.-Q. Dai, X.-X. Li, and Z.-R. Lai, "Band-reweighted Gabor Kernel Embedding for Face Image Representation and Recognition," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 725–740, 2014.
- [43] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, 1997.
- [44] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*. Springer-Verlag New York, 2001, vol. 1.
- [45] K.-A. Toh, "Deterministic neural classification," *Neural Computat.*, vol. 20, no. 6, pp. 1565–1595, 2008.
- [46] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological cybernetics*, vol. 36, no. 4, pp. 193–202, 1980.
- [47] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *P. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [48] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [49] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep Learning Face Representation by Joint Identification-Verification," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 1988–1996.
- [50] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: A simple deep learning baseline for image classification?" *IEEE Trans. Image Process.*, no. 12, pp. 5017–5032, 2015.
- [51] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
- [52] 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, 2000.
- [53] T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) database," in *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognit.*, 2002, pp. 46–51.
- [54] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image and Vision Comput.*, vol. 28, no. 5, pp. 807–813, 2010.
- [55] L. Wolf, T. Hassner, and Y. Taigman, "Effective unconstrained face recognition by combining multiple descriptors and learned background statistics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 1978–1990, 2011.
- [56] Z. Huang, S. Shan, R. Wang, H. Zhang, S. Lao, A. Kuerban, and X. Chen, "A Benchmark and Comparative Study of Video-Based Face Recognition on COX Face Database," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5967–5981, 2015.
- [57] W. Deng, J. Hu, J. Lu, and J. Guo, "Transform-Invariant PCA: A Unified Approach to Fully Automatic FaceAlignment, Representation, and Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 6, pp. 1275–1284, 2014.
- [58] Z. Chai, Z. Sun, H. Mendez-Vazquez, R. He, and T. Tan, "Gabor Ordinal Measures for Face Recognition," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 1, pp. 14–26, 2014.
- [59] X. Tan and B. Triggs, "Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, 2010.
- [60] K.-C. Fan and T.-Y. Hung, "A Novel Local Pattern Descriptor-Local Vector Pattern in High-Order Derivative Space for Face Recognition," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 2877–2891, 2014.
- [61] M. Tistarelli, Y. Sun, and N. Poh, "On the Use of Discriminative Cohort Score Normalization for Unconstrained Face Recognition," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 12, pp. 2063–2075, 2014.
- [62] A. Vedaldi and K. Lenc, "MatConvNet – Convolutional Neural Networks for MATLAB," in *Proceeding of the ACM Int. Conf. on Multimedia*, 2015.
- [63] C. Liu, "Gabor-based Kernel PCA with Fractional Power Polynomial Models for Face Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 572–581, 2004.
- [64] 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. IEEE Int. Conf. Comput. Vis.*, 2005, pp. 786–791.
- [65] Y. Pang, Y. Yuan, and X. Li, "Gabor-Based Region Covariance Matrices for Face Recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 7, pp. 989–993, 2008.
- [66] R. Bhatia, *Positive definite matrices*. Princeton University Press, 2009.

- [67] M. Yang, L. Zhang, S. C.-K. Shiu, and D. Zhang, "Monogenic Binary Coding: An Efficient Local Feature Extraction Approach to Face Recognition," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 6, pp. 1738–1751, 2012.
- [68] H. Gao, H. K. Ekenel, and R. Stiefelhagen, "Pose Normalization for Local Appearance-Based Face Recognition," in *Proc. Int. Conf. Adv. Biometrics*, 2009, pp. 32–41.
- [69] M. S. Sarfraz and O. Hellwich, "Probabilistic learning for fully automatic face recognition across pose," *Image Vis. Comput.*, vol. 28, no. 5, pp. 744–753, 2010.
- [70] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, pp. 3371–3408, 2010.
- [71] Z. Lei, S. Liao, M. Pietikainen, and S. Z. Li, "Face Recognition by Exploring Information Jointly in Space, Scale and Orientation," *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 247–256, 2011.
- [72] C. A. Perez, L. A. Cament, and L. E. Castillo, "Methodological improvement on local Gabor face recognition based on feature selection and enhanced Borda count," *Pattern Recogn.*, vol. 44, no. 4, pp. 951–963, 2011.
- [73] 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, 2012.
- [74] S. U. Hussain, T. Napoléon, and F. Jurie, "Face Recognition using Local Quantized Patterns," in *Proc. Brit. Mach. Vis. Conf.*, 2012, pp. 1–11.
- [75] N.-S. Vu, "Exploring patterns of gradient orientations and magnitudes for face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 2, pp. 295–304, 2013.
- [76] S. Liao, D. Shen, and A. C. S. Chung, "A Markov Random Field Groupwise Registration Framework for Face Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 4, pp. 657–669, 2014.
- [77] Z. Lei, M. Pietikainen, and S. Z. Li, "Learning Discriminant Face Descriptor," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 289–302, 2014.
- [78] J. Lu, V. E. Liong, X. Zhou, and J. Zhou, "Learning Compact Binary Face Descriptor for Face Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 10, pp. 2041–2056, 2015.
- [79] K.-K. Huang, D.-Q. Dai, C.-X. Ren, and Z.-R. Lai, "Learning Kernel Extended Dictionary for Face Recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 5, pp. 1082–1094, 2017.
- [80] C. Ding and D. Tao, "Trunk-Branch Ensemble Convolutional Neural Networks for Video-based Face Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017.
- [81] P. Zhu, L. Zhang, W. Zuo, and D. Zhang, "From point to set: Extend the learning of distance metrics," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2664–2671.
- [82] Z. Huang, R. Wang, S. Shan, and X. Chen, "Learning euclidean-to-riemannian metric for point-to-set classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, June 2014, pp. 1677–1684.
- [83] The MathWorks, "MATLAB," <http://www.mathworks.com/>.