Intraspectrum Discrimination and Interspectrum Correlation Analysis Deep Network for Multispectral Face Recognition

Fei Wu, Xiao-Yuan Jing, Xiwei Dong, Ruimin Hu, *Senior Member, IEEE*, Dong Yue, Lina Wang, Yi-Mu Ji, Ruchuan Wang, and Guoliang Chen

***Abstract*—Multispectral images contain rich recognition information since the multispectral camera can reveal information that is not visible to the human eye or to the conventional RGB camera. Due to this characteristic of multispectral images, multispectral face recognition has attracted lots of research interest. Although some multispectral face recognition methods have been presented in the last decade, how to fully and effectively explore the intraspectrum discriminant information and the useful interspectrum correlation information in multispectral face images for recognition has not been well studied. To boost the performance of multispectral face recognition, we propose an intraspectrum discrimination and interspectrum correlation analysis deep network (IDICN) approach. Multiple spectra are divided into several spectrum-sets, with each containing a group of spectra within a small spectral range. The IDICN network contains a set of spectrum-set-specific deep convolutional neural networks attempting to extract spectrum-set-specific features, followed by a spectrum pooling layer, whose target is to select a group of spectra with favorable discriminative abilities adaptively. IDICN jointly learns the nonlinear representations of the selected spectra, such that the intraspectrum Fisher loss and the interspectrum discriminant correlation are minimized. Experiments on the well-known Hong Kong Polytechnic University, Carnegie Mellon University, and the University of Western Australia multispectral face datasets demonstrate the superior performance of the proposed approach over several state-of-the-art methods.**

***Index Terms*—Deep convolutional neural networks (DCNNs), intraspectrum discriminant information exploration, multispectral face recognition, spectra selection, useful interspectrum correlation information exploration.**

Manuscript received January 28, 2018; revised May 15, 2018 and July 29, 2018; accepted October 2, 2018. Date of publication November 6, 2018; date of current version January 21, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61702280, in part by the National Post-Doctoral Program for Innovative Talents under Grant BX20180146, in part by the NSFC-Key Project of General Technology Fundamental Research United Fund under Grant U1736211, in part by the National Key Research and Development Program of China under Grant 2017YFB1401302, Grant 2017YFB0202204, and Grant 2017YFB1401301, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20170900, in part by the Key Research and Development Program of Jiangsu Province under Grant BE2017166, in part by the Natural Science Foundation Outstanding Youth Fund of Jiangsu Province under Grant BK20170100, in part by the Natural Science Fund for Colleges and Universities in Jiangsu Province under Grant 17KJB520025, in part by the Scientific Research Staring Foundation for Introduced Talents in NJUPT (NUPTSF) under Grant NY217009, and in part by 1311 Project of NJUPT. This paper was recommended by Associate Editor H. Yin. *(Corresponding author: Xiao-Yuan Jing.)*

F. Wu, X. Dong, and D. Yue are with the College of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (e-mail: wufei_8888@126.com).

X.-Y. Jing is with the School of Computer, Wuhan University, Wuhan 430072, China, and also with the College of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (e-mail: jingxy_2000@126.com).

R. Hu and L. Wang are with the School of Computer, Wuhan University, Wuhan 430072, China.

Y.-M. Ji, R. Wang, and G. Chen are with the College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2018.2876591

I. Introduction

AMONG various recognition tasks in the computer vision and pattern classification communities, face recognition is one of the most extensively studied fields [1]–[3]. Recent studies have found that face recognition based on multispectral images has the following advantages, where the multispectral image is a data cube consisting of two spatial dimensions and a spectral dimension.

1) The multispectral images can detect distinct patterns contained in human faces where such discriminative patterns may not be captured by trichromatic (RGB) color or monochromatic (grayscale) cameras [4].
2) Multispectral imaging allows the spectral distribution of an imaged object to be different from the others, which makes the multispectral images invariant to illumination changes [5].
3) The person-to-person spectral variability for different tissue types is also helpful for recognition [6]. Therefore, multispectral face recognition has attracted a lot of research interest [7]–[9].

Most of existing multispectral face recognition methods focus on spectrum information fusion (including image level fusion [10], [11]; feature level fusion [12]; decision level fusion [5], [13], [14]; and hierarchical fusion [15]). Some other works take the following issues into consideration, including spectral band selection algorithm [11], [16], [17]; task-specific feature learning algorithm [18]–[20]; cross-spectral recognition application [21]–[26]; multispectral face recognition system [27]; etc.

A. Motivation and Contribution

Multispectral images contain rich recognition information since the multispectral camera can reveal information that is not visible to the conventional RGB camera. And multispectral

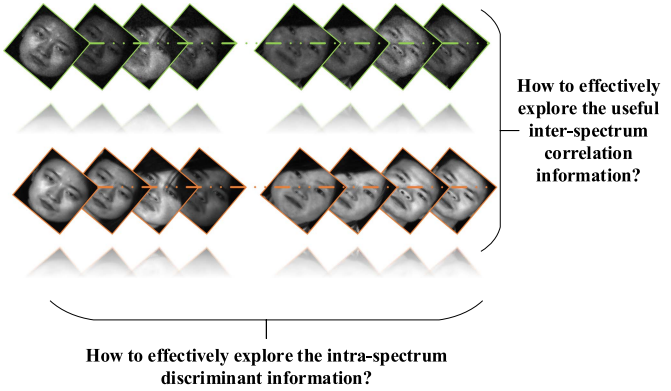


Fig. 1. Current challenges in multispectrum face recognition.

imaging can alleviate some challenges of face recognition, for example, illumination variation [28], [29]. Thus, multispectral imaging itself is beneficial to face recognition, and the study of multispectral images-based face recognition has practical significance. Multispectral face recognition is an important research branch of face recognition.

Discriminant feature learning from multiple spectra is the key of the success of multispectral face recognition. Besides, as discussed in [20], the similarity between different spectra of multispectral images is large. Hence, how to explore the useful correlation and reduce the adverse redundancy among multiple spectra is a significant factor that is worthy of attention. Focusing on existing representative multispectral face recognition methods, we found that there exists room to improve their performance. Theoretically, how to effectively explore the intraspectrum discriminant information and the useful interspectrum correlation information in multispectral face images simultaneously has not been well studied, which is an important research topic. This challenge of multispectrum face recognition is illustrated in Fig. 1.

In this paper, we make effort to cope with this challenge. We summarize the contributions of this paper as follows.

1) We design a supervised deep neural network model for multispectral face recognition, which can effectively explore the intraspectrum discriminant information and interspectrum correlation information.
2) The designed model contains a set of spectrum-set-specific deep convolutional neural networks (DCNNs) attempting to extract spectrum-set-specific features, followed by a spectrum pooling layer and the output layer. The spectrum pooling layer can dynamically and adaptively select a group of spectra with favorable discriminative abilities in the process of feature learning.
3) We design the objective function on the output layer, which aims to jointly learn the nonlinear representations of the selected spectra, such that the intraspectrum Fisher loss and the interspectrum discriminant correlation are minimized.
4) The proposed approach for multispectral face recognition is named intraspectrum discrimination and interspectrum correlation analysis deep network (IDICN). It is evaluated on three widely used hyper-spectral face datasets, that is, Hong Kong Polytechnic University (HK PolyU) [16], Carnegie Mellon University (CMU) [30], and the University of Western Australia (UWA) [11]. The experimental results show that it can significantly outperform the state-of-the-art related methods.

B. Organization

The rest of this paper is organized as follows. In Section II, we briefly introduce the related work. In Section III, we describe our IDICN approach. The optimization process is provided in Section IV. Experimental results are reported in Section V. Finally, the conclusions are drawn in Section VI.

II. Related Work

A. Multispectral Face Recognition Methods

Over the past several years, we have witnessed some methods developed for multispectral face recognition [31]. One of the earliest work was done by Pan *et al.* [6], which examines the utility of using near-infrared hyperspectral images for face recognition with an algorithm based on the spectral comparison of combinations of tissue types. Existing multispectral face recognition methods can be generally categorized into three types.

1) Methods Focusing on Spectrum Information Fusion or Cross-Spectral Recognition: Spectrum information fusion methods exploit synergistic integration of the information obtained from different spectra or from multiple classifiers to improve the overall recognition performance. Omri *et al.* [10] presented a multispectral pixel level fusion algorithm, which first decomposes visible and near-infrared images into different bands using discrete wavelet transform, and then performs the fusion process through singular value decomposition or principal component analysis that can highlight the most salient features needed to be preserved for the final fused image. Uzair *et al.* [11] designed a multispectral recognition algorithm (named "Band fusion+PLS"), which uses a spatio-spectral covariance for band fusion and partial least square regression for classification. In band fusion stage, Band fusion+PLS incorporates local spatial information and removes noise by averaging both the spectral and spatial dimensions. Based on particle swarm optimization, Raghavendra *et al.* [12] presented two image fusion schemes (a weighting scheme and a selection scheme) for combining visible and near-infrared face images. Arandjelović *et al.* [13] provided a decision level fusion algorithm for visual and infrared spectra. Chang *et al.* [5] provided multiple image fusion methods, including physics-based weighted fusion, illumination adjustment, and rank-based decision level fusion (RDLF), for multispectral face images. Singh *et al.* [15] presented a two-level hierarchical fusion scheme. A discrete wavelet transform-based image fusion algorithm is used to fuse the multispectral images at image level. At feature level, the amplitude and phase features are extracted from the fused images, and an SVM learning algorithm is employed to select either the amplitude or phase features to generate the fused feature set.

Cross-spectral recognition makes efforts to bring images of different spectra to a same platform such that comparison and

recognition can be done [25], [32]. Nicolo and Schmid [23] developed a cross-spectral recognition method, which encodes magnitude and phase of multispectral face images filtered with a bank of Gabor filters, and compares the encoded images by using the symmetric I-divergence distance. Klare and Jain [24] presented a heterogeneous face recognition method, which uses kernel prototype similarities to recognize face images of different modalities. Yi *et al.* [27] developed a multispectral face recognition system working in visible and near-infrared spectra, which is robust to various spoofing attacks and is user cooperation free.

2) Methods Focusing on Feature Learning: Most of existing multispectral face recognition methods employ traditional feature learning techniques to extract features for classification. Di *et al.* [16] used the bidirectional 2-D principal component analysis [$(2\text{-D})^2$PCA] method for multispectral face recognition. Uzair *et al.* [33] developed a spatio-spectral feature extraction algorithm by employing the 3-D discrete cosine transform. Based on [33], Bianco [34] further investigated the performance improvement when 1-D projections of the whole spectral data along the spectral dimension are used. Sharma and Van Gool [35] regarded each band of multispectral image as a separate image, and extracted features on these images by using the traditional feature descriptors, that is, histograms of oriented gradients, local binary patterns (LBPs), and scale-invariant feature transform.

Recently, focusing on the characteristics of multispectral face recognition task, some task-specific feature learning methods have been developed. Zhao and Sun [19] represented a multispectral image as a third-order tensor and presented a sparse tensor embedding (STE) method, which learns transformation matrices through preserving the sparse information of the third-order tensors. Aiming to exploit the spatial and spectral features in multispectral images, Liang *et al.* [36] designed a 3-D high-order texture pattern descriptor based on local derivative pattern. Jing *et al.* [20] presented a multispectral low-rank structured dictionary learning (MLSDL) method, which learns multiple low-rank spectrum-specific dictionaries and a low-rank spectrum-common dictionary, and uses a collaborative representation-based classifier for classification.

Existing multispectral face recognition methods make use of the characteristics of the multispectral image to improve the face recognition performance. However, as analyzed in Section I-A, how to effectively explore the intraspectrum discriminant information and the interspectrum correlation information in multispectral face images simultaneously has not been well studied.

3) Methods Focusing on Spectral Band Selection: Discriminative spectral band selection is an important research topic. By using spectral band selection, accurate recognition can be realized with only a small number of discriminative spectra, which is beneficial to the real-time multispectral face recognition application. Chang *et al.* [37] introduced a distribution separation measure and then designed an illumination-specific spectral range selection algorithm that selects a minimal set of narrowband spectral ranges by ranking the separation values. According to the absorption characteristics of hemoglobin, Di *et al.* [16] selected two disjoint band subsets centered at 540 and 580 nm. Bouchech *et al.* [38] assumed that there exist two probability density functions (PDFs), which are computed using a finite mixture of Gaussian model. Based on the PDFs, they decided whether the given spectral band is good enough to be selected for recognition purposes or not. In [39], multispectral images are organized as a 3-mode tensor, and multilinear sparse decomposition technique is utilized to select optimal spectral bands. Uzair *et al.* [11] provided a band selection algorithm, which drops the least discriminative bands from the hyperspectral cube until only a discriminative combination is left.

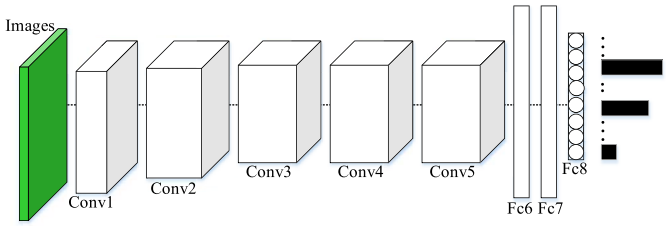


Fig. 2. Illustration of the architecture of CaffeNet.

Current spectral band selection methods usually regard the feature learning and spectral band selection processes as two serial and independent steps. Different from these methods, our IDICN approach can perform feature learning and spectral band selection by using a unified deep learning-based model.

B. Deep Learning Methods

1) Convolutional Neural Network-Based Methods: In recent years, deep learning is gaining more and more attention [40], and convolutional neural networks (CNNs) [41] have demonstrated their outstanding representation power in a wide range of computer vision applications, such as face recognition [42], object classification [43], visual tracking [44], etc. By training a DCNN that consists of five convolutional layers with the largescale ImageNet dataset, Krizhevsky *et al.* [45] developed the AlexNet, which brings significant performance improvement in image classification. Based on the Caffe framework [46], Jia *et al.* provided a CaffeNet implementation of AlexNet. CaffeNet is a slightly modified version of AlexNet, with some normalizing and pooling layers being switched. Fig. 2 shows the architecture of the CaffeNet, which contains five convolutional layers and three fully connected layers. The output of the last fully connected layer (Fc8) is fed into a softmax that produces a distribution over the class labels.

2) Multidataset-Based Deep Learning Methods: In order to analyze the relationship or extract the features among given two or multiple datasets, the following representative algorithms have been presented. Andrew *et al.* [47] presented a deep canonical correlation analysis (DCCA) method, which is a nonlinear extension of canonical correlation analysis [48], to learn complex nonlinear transformations of two views of data such that the resulting representations are highly linearly correlated. By imposing regularizations in the deep neural network, Wu *et al.* [49] tried to explore the interfeature (visual and audio features) and interclass relationships for video classification.

Wang *et al.* [50] presented a CNNs-based multimodal method for RGB-D object recognition, which learns transformation matrices for two modalities with a large margin criterion and a maximal cross-modality correlation criterion. Focusing on the RGB-D scene recognition task, Zhu *et al.* [51] considered the inter- and intra-modality correlation for samples and meanwhile regularizing the learned features to be discriminative and compact. Tao *et al.* [52] presented a scheme called deep multiview feature learning, which exploits the collaboration between handcrafted and deep learning features, for person reidentification [53]. Kan *et al.* [54] developed a multiview deep network (MvDN) for cross-view classification, which seeks for a view-invariant representation shared between multiple views by maximizing the between-class variations and minimizing the within-class variations of feature representations from both intraview and interview aspects.

The works in [47] and [49]–[52] were designed for the application including two data sets (e.g., two views of data, visual and audio features, or RGB and depth images), and cannot be directly used for multispectral face recognition. The MvDN network model [54] does not employ CNN, and it only focuses on discriminant feature learning without considering the useful correlation information exploration.

3) CNN-Based Hyperspectral Image Classification Methods: Deep learning-based methods have drawn much attention in remote sensing image analysis, and some CNN-based methods have been designed for hyperspectral image classification [55], [56]. Based on CNN, Chen *et al.* [57] designed a regularized deep feature extraction method (e.g., the L2 regularization) to extract the spectral, spatial, and spectral–spatial features of hyperspectral images. Li *et al.* [58] presented a DCNN with pixel-pair features (CNN-PPFs) algorithm, which utilizes the DCNN to learn pixel-pair features. For each testing pixel, CNN-PPF classifies the constructed neighboring pixel-pairs, and determines the label via a voting strategy based on joint classification results. Maggiori *et al.* [59] devised a fully convolutional architecture, which only involves a series of convolution and deconvolution operations to produce the output classification maps, and used a two-step training strategy for the pixelwise classification of satellite imagery.

There exist differences between these CNN-based hyperspectral image classification methods and our approach. This kind of methods mainly focus on pixel-level classification, while our proposed IDICN approach focuses on the image-level recognition, that is exploiting the higher-level information of image and assigning a class label for the whole image.

III. Our Approach

A. Overview

Assume that $X = [X_1, \ldots, X_M]$ denotes the face image set of M spectra, $X_k = [x_1^k, \ldots, x_N^k] \in \mathbb{R}^{d_x \times N}$ is sample set of the kth spectrum, and $\{l_1, \ldots, l_N\}$ is the label set corresponding to samples in X_k. For the input images of multiple spectra, the spectrum-specific features can be obtained jointly by passing through the spectrum-specific DCNNs. However, it consumes large computational cost when the number of spectra is large. For example, if we want to jointly extract spectrum-specific features for face samples in the CMU [30] dataset that includes 65 spectra covering the spectral range of 450–1090 nm, we need to train a huge network containing 65 spectrum-specific DCNNs, which requires large computational cost and is not necessary. Furthermore, the deep learning technique usually requires a large number of samples for the training process, while image samples per person of a certain spectrum are usually very unique. Considering that there exists similarity between spectra with neighboring ranges of wavelength, in this paper, we divide the spectra into several sets, with each containing a set of spectra within a small spectral range (e.g., from 400 to 450 nm). And each spectrum-set has approximatively equal number of spectra. We regard samples of multiple spectra in a spectrum-set as the input for DCNN. Thus, we can obtain the spectrum-set-specific features for image samples of the specific spectrum-set by passing through the spectrum-set-specific DCNN.

Suppose that the spectra are divided into H spectrum-sets, then for any sample x_i^k of the kth spectrum in the hth spectrum-set, its feature representation z_i^k in the second fully connected layer (Fc7) of the spectrum-set-specific DCNN is formulated as $z_i^k = w_h(x_i^k; \varsigma_h) \in \mathbb{R}^{d_z}$. Here, ς_h denotes all parameters in DCNN of the hth spectrum-set (including the weight matrix O_r^h and bias vector b_r^h of the rth layer, $r = 1, \ldots, R$ and $R = 7$ in a DCNN). We initialize the spectrum-set-specific DCNNs by using the parameters of the CaffeNet [46] trained on the ImageNet dataset.

Considering that large correlation exists between multiple spectra and also in order to achieve the computational efficiency in the testing procedure, we design a spectrum pooling layer to select spectra with strong discriminative abilities for subsequent network training. As reported in [11], similar recognition results can be achieved by using a small number of discriminative spectra as the results using all spectra. According to the Fisher criterion [60], we calculate the Fisher score of each spectrum in spectrum-set by

$$F(Z_k) = \frac{\mathrm{tr}\left(G_B^k\right)}{\mathrm{tr}\left(G_W^k\right)} \tag{1}$$

where $\mathrm{tr}(\cdot)$ is the trace of a matrix, G_B^k and G_W^k are separately the intraspectrum between-class scatter matrix and the intraspectrum within-class scatter matrix calculated on the feature samples $Z_k = [z_1^k, \ldots, z_N^k]$ of the kth spectrum. Specifically,

$$G_B^k = \frac{1}{N}\sum_{c=1}^{C} N_c\left(m_c^k - m^k\right)\left(m_c^k - m^k\right)^T \tag{2}$$

$$G_W^k = \frac{1}{N}\sum_{c=1}^{C}\sum_{i=1}^{N_c}\left(z_{ci}^k - m_c^k\right)\left(z_{ci}^k - m_c^k\right)^T \tag{3}$$

where C denotes the number of classes, N_c denotes the number of samples in the cth class of each spectrum, z_{ci}^k is the ith feature sample of the cth class in the kth spectrum, $m_c^k = (1/N_c)\sum_{i=1}^{N_c} z_{ci}^k$ is the mean of the cth class in the kth

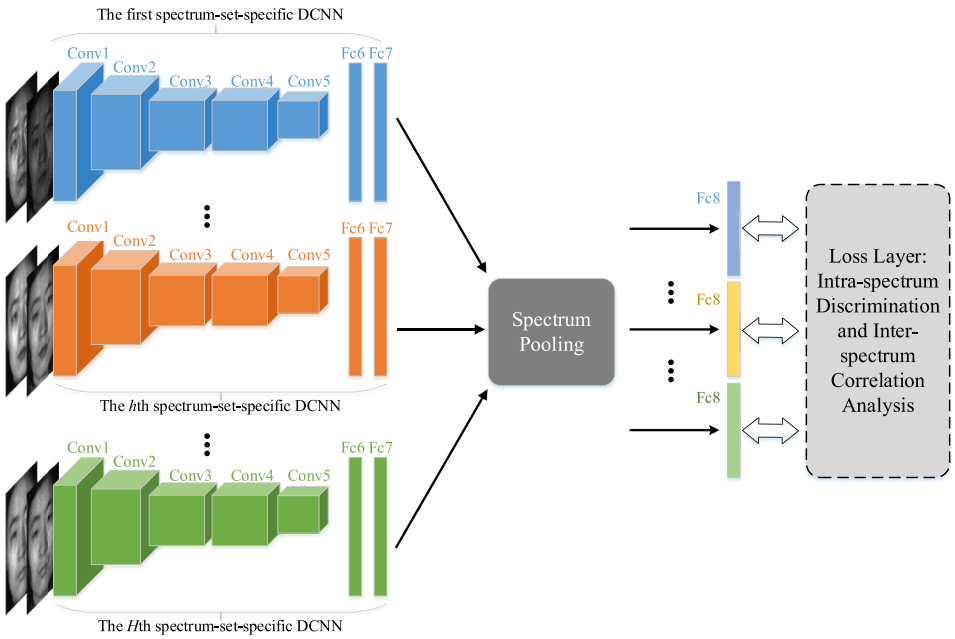


Fig. 3. Overview of the proposed IDICN network. The input of the network is the 2-D images of all spectra. The DCNNs [46] (containing five convolutional layers Conv1–Conv5, and two fully connected layers Fc6 and Fc7) for multiple spectrum-sets are trained separately, and the features of the second fully connected layers (Fc7) are fed into the spectrum pooling layer. The spectrum pooling layer chooses a group of spectra with strong discriminative abilities. The features of the selected spectra are further transformed with a fully connected layer (the output layer Fc8) for each selected spectrum. The representations of the output layers are required to minimize the objective of intraspectrum discrimination and interspectrum correlation analysis.

spectrum, and $m^k = (1/N)\sum_{i=1}^{N} z_i^k$ is the mean of all feature samples in the kth spectrum. We rank the Fisher scores of different spectra and select Q spectra with largest Fisher scores. The spectra selected in the training stage will be used for recognition.

The features of the selected spectra will experience further transformation with a fully connected layer, such that the feature representations in the final fully connected layer (Fc8, the output layer) will minimize the objective of intraspectrum discrimination and interspectrum correlation analysis, which will be elaborated in detail in the following sections. Finally, the learned features of samples corresponding to the selected spectra are fed into the simple and widely used nearest neighbor classifier for recognition. The architecture of our IDICN network is illustrated in Fig. 3.

B. Intraspectrum Discriminant Analysis

Assume that $Y_q = [y_1^q, \ldots, y_N^q] \in \mathbb{R}^{d_y \times N}$ is the feature representation of the top layer (output layer Fc8) corresponding to the qth spectrum in the selected Q spectra, and the qth selected spectrum corresponds to the kth spectrum of all spectra, that is, $Y_q = f_q(Z_k; \theta_q)$, where θ_q is a vector of parameters containing the weight matrix parameter O^q and bias vector parameter b^q. In order to ensure that Y_q is discriminative, the following Fisher loss should be minimized:

$$\sum_{q=1}^{Q} \mathrm{Disc}(Y_q) = \sum_{q=1}^{Q} \left(\mathrm{tr}(S_W^q) - \mathrm{tr}(S_B^q)\right) \tag{4}$$

where S_W^q and S_B^q are the intraspectrum within-class scatter matrix and the intraspectrum between-class scatter matrix calculated on Y_q, respectively.

The within-class scatter matrix can be further formulated as

$$S_W^q = \frac{1}{N} Y_q \left(I - \sum_{c=1}^{C} \frac{1}{N_c} e_c e_c^T \right) Y_q^T = \frac{1}{N} Y_q \mathcal{A}_q Y_q^T \tag{5}$$

and the total scatter matrix can be defined as

$$S_K^q = \frac{1}{N} Y_q \left(I - \frac{1}{N} e e^T \right) Y_q^T = \frac{1}{N} Y_q \mathcal{B}_q Y_q^T \tag{6}$$

then the between-class scatter matrix can be formulated as

$$S_B^q = S_K^q - S_W^q = \frac{1}{N} Y_q (\mathcal{B}_q - \mathcal{A}_q) Y_q^T = \frac{1}{N} Y_q \mathcal{H}_q Y_q^T \tag{7}$$

where I is an identity matrix, $e \in \mathbb{R}^N$ is a vector with all elements as 1, and $e_c \in \mathbb{R}^N$ is a vector with $e_c(i) = 1$ if the ith sample is from the cth class.

Thus, we should minimize the objective of

$$\sum_{q=1}^{Q} \mathrm{tr}\left(\frac{1}{N} Y_q (\mathcal{A}_q - \mathcal{H}_q) Y_q^T \right). \tag{8}$$

C. Interspectrum Correlation Analysis

According to [47] and [61], the total correlation of data matrices Ψ and Φ can be defined as the trace norm of L

$$\mathrm{Corr}(\Psi, \Phi) = \|L\|_{\mathrm{tr}} = \mathrm{tr}\left(L^T L\right)^{1/2} \tag{9}$$

where $L = \Sigma_{uu}^{-1/2} \Sigma_{uv} \Sigma_{vv}^{-1/2}$. Here, Σ_{uu}, Σ_{vv}, and Σ_{uv}, which are represented by centered data matrices $\bar{\Psi}$ and $\bar{\Phi}$, are separately the covariance of Ψ, the covariance of Φ, and the cross covariance between Ψ and Φ.

Assume that $\bar{Y}_q = Y_q - (1/N) Y_q e e^T = [\bar{y}_1^q, \ldots, \bar{y}_N^q]$ and $\bar{Y}_g$ denote the centered data matrices of the qth and the gth

spectra, respectively. To leverage the useful correlation among multiple spectra, we define the interspectrum within-class cross covariance as

$$\Sigma_{qg}^{W} = \frac{1}{C}\sum_{c=1}^{C}\Sigma_{qcgc} = \frac{1}{C}\sum_{c=1}^{C}\frac{1}{N_c-1}\bar{Y}_{qc}\bar{Y}_{gc}^{T} \tag{10}$$

where $\bar{Y}_{qc}$ is a feature sample subset in $\bar{Y}_q$ containing feature samples from the cth class.

Considering the interspectrum total cross covariance is given as $\Sigma_{qg}^{P} = [1/(N-1)]\bar{Y}_q\bar{Y}_g^T$, the interspectrum between-class cross covariance can be defined as

$$\Sigma_{qg}^{B} = \Sigma_{qg}^{P} - \Sigma_{qg}^{W} = \frac{1}{N-1}\bar{Y}_q\bar{Y}_g^T - \frac{1}{C}\sum_{c=1}^{C}\frac{1}{N_c-1}\bar{Y}_{qc}\bar{Y}_{qc}^T \tag{11}$$

Σ_{qg}^{W} can be reformulated as

$$\begin{aligned}\Sigma_{qg}^{W} &= \frac{1}{C}\sum_{c=1}^{C}\frac{1}{N_c-1}\bar{Y}_{qc}\bar{Y}_{gc}^T \\ &= \frac{1}{C}\bar{Y}_q\begin{bmatrix}\frac{1}{N_1-1} & & \\ & \ddots & \\ & & \frac{1}{N_C-1}\end{bmatrix}\bar{Y}_g^T \\ &= \frac{1}{C}\bar{Y}_q E\bar{Y}_g^T\end{aligned} \tag{12}$$

then

$$\begin{aligned}\Sigma_{qg}^{B} &= \Sigma_{qg}^{P} - \Sigma_{qg}^{W} \\ &= \frac{1}{N-1}\bar{Y}_q\bar{Y}_g^T - \frac{1}{C}\bar{Y}_q E\bar{Y}_g^T \\ &= \bar{Y}_q\left(\frac{1}{N-1}I - \frac{1}{C}E\right)\bar{Y}_g^T.\end{aligned} \tag{13}$$

To make full use of the favorable cross-spectral correlation (i.e., the interspectrum within-class correlation) and reduce the adverse redundancy among different spectra, we should minimize the following multispectral discriminant correlation:

$$\begin{aligned}&\frac{1}{Q(Q-1)}\sum_{q=1}^{Q}\sum_{\substack{g=1\\g\neq q}}^{Q}\mathrm{DCorr}(Y_q, Y_g) \\ &= \frac{1}{Q(Q-1)}\sum_{q=1}^{Q}\sum_{\substack{g=1\\g\neq q}}^{Q}\left\|L^{qg}\right\|_{\mathrm{tr}} \\ &= \frac{1}{Q(Q-1)}\sum_{q=1}^{Q}\sum_{\substack{g=1\\g\neq q}}^{Q}\mathrm{tr}\left(L^{qgT}L^{qg}\right)^{\frac{1}{2}}\end{aligned} \tag{14}$$

where $L^{qg} = \Sigma_{qq}^{-1/2}(\Sigma_{qg}^{B} - \Sigma_{qg}^{W})\Sigma_{gg}^{-1/2}$, $\Sigma_{qq} = (1/[N-1])\bar{Y}_q\bar{Y}_q^T + \gamma I$, and γ is a positive regularization parameter.

D. Unified Objective Function

By combining the objective of intraspectrum discriminant analysis in (8) and the objective of interspectrum correlation analysis in (14), we can obtain the unified objective function of our IDICN approach

$$\arg\min_{\substack{\varsigma_1,\ldots,\varsigma_h,\ldots,\varsigma_H\\ \theta_1,\ldots,\theta_q,\ldots,\theta_Q}} J \tag{15}$$

where

$$J = \frac{1}{2}\sum_{q=1}^{Q}\mathrm{Disc}(Y_q) + \frac{\alpha}{2Q(Q-1)}\sum_{q=1}^{Q}\sum_{\substack{g=1\\g\neq q}}^{Q}\mathrm{DCorr}(Y_q, Y_g) \tag{16}$$

α is a balance factor for balancing two terms.

IV. Optimization

Like most of existing deep learning works, we employ the gradient descent algorithm to optimize the designed deep network. Based on the gradient descent, our IDICN is optimized by following the chain rule, that is, feeding forward and calculating the loss of the unified objective in (16), and then back propagating the loss to each layer, such that the whole network can be updated.

We first calculate the gradient of the loss J with respect to Y_q. According to [47], assume that the singular value decomposition of L is $L = UDV^T$, then

$$\begin{aligned}\frac{\partial\mathrm{Corr}(\Psi,\Phi)}{\partial\Psi} &= \frac{\partial\mathrm{tr}\left(L^TL\right)^{1/2}}{\partial\Psi} \\ &= \frac{1}{p-1}\left(2\nabla_{uu}\bar{\Psi} + \nabla_{uv}\bar{\Phi}\right)\end{aligned} \tag{17}$$

where $\nabla_{uu} = -(1/2)\Sigma_{uu}^{-1/2}UDU^T\Sigma_{uu}^{-1/2}$, $\nabla_{uv} = \Sigma_{uu}^{-1/2}UV^T\Sigma_{vv}^{-1/2}$, and p denotes the number of columns in Ψ and Φ. Therefore,

$$\begin{aligned}\frac{\partial J}{\partial Y_q} &= \frac{1}{N}Y_q\left(\mathcal{A}_q - \mathcal{H}_q\right) \\ &\quad + \frac{\alpha}{Q(Q-1)}\left(\frac{1}{N-1}\sum_{\substack{g=1\\g\neq q}}^{Q}\left(2\nabla_{qq}\bar{Y}_q + \nabla_{qg}\bar{Y}_g\right)\right)\end{aligned} \tag{18}$$

where $\nabla_{qq} = -(1/2)\Sigma_{qq}^{-1/2}U^{qg}D^{qg}(U^{qg})^T\Sigma_{qq}^{-1/2}$ and $\nabla_{qg} = \Sigma_{qq}^{-1/2}U^{qg}(V^{qg})^T\Sigma_{gg}^{-1/2}$. Here, U^{qg}, D^{qg}, and V^{qg} are the matrices decomposed by L^{qg} with singular value decomposition as $L^{qg} = U^{qg}D^{qg}(V^{qg})^T$. The other gradient calculations, like $(\partial Y_q/\partial Z_k)$, $(\partial Y_q/\partial\theta_q)$, and $(\partial Z_k/\partial\varsigma_h)$, can be easily performed according to the "UFLDL Tutorial" website [62].

Assume that $\vartheta = [\varsigma_1,\ldots,\varsigma_H,\theta_1,\ldots,\theta_Q]$ denotes the parameter of the whole IDICN network. At the $(t+1)$th iteration, these parameters are updated with full-batch optimization using the limited-memory BFGS algorithm [63] with the gradient calculated as follows:

$$\begin{aligned}\Delta\vartheta^t &= \left[\Delta\varsigma_1^t,\ldots,\Delta\varsigma_H^t,\Delta\theta_1^t,\ldots,\Delta\theta_Q^t\right] \\ \Delta\theta_q^t &= \frac{\partial J}{\partial Y_q}\cdot\frac{\partial Y_q}{\partial\theta_q},\ q=1,\ldots,Q \\ \Delta\varsigma_h^t &= \frac{\partial J}{\partial Y_q}\cdot\frac{\partial Y_q}{\partial Z_k}\cdot\frac{\partial Z_k}{\partial\varsigma_h},\ h=1,\ldots,H.\end{aligned} \tag{19}$$

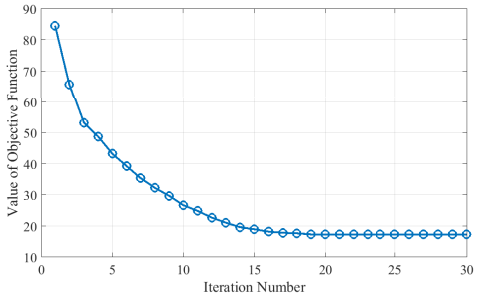


Fig. 4. Convergence curve of the IDICN network on the HK PolyU dataset.

Algorithm 1 IDICN for Multispectral Face Recognition

1) **Input**
Training data $X = \left[X_1, \cdots, X_M\right]$ of M spectra, number of required spectra, i.e., Q, and the multi-spectral query sample ρ.
2) **Discriminant feature learning and spectrum selection**
For $t = 1$ to ξ (the iteration number)
   a) **Forward propagation**
      i) Obtain the spectrum-set-specific feature representation $Z_k = \left[z_1^k, \cdots, z_N^k\right]$ for the $k^{th}(k = 1, \cdots, M)$ spectrum.
      ii) Select Q discriminative spectra according to the Fisher score of each spectrum, which is calculated with Eq. (1).
      iii) Further transform the features of the selected spectra with a fully connected layer (Fc8).
   b) **Back propagation**
   Update parameters $\vartheta = \left[\varsigma_1, \cdots, \varsigma_H, \theta_1, \cdots, \theta_Q\right]$ by using Eqs. (18) and (19).
End
3) **Multi-spectral face recognition**
Obtain the output feature representations of query sample and training samples, i.e., $\rho^q_{transform}\Big|_{q=1}^{Q}$ and $Y_q = \left[y_1^q, \cdots, y_N^q\right]\Big|_{q=1}^{Q}$, by using the learned parameters and the selected spectra. Concatenate these feature representations of Q spectra into vectors and utilize the nearest neighbor classifier for classification.

The procedure of using IDICN for multispectral face recognition is summarized in Algorithm 1. For fusing the output feature representations of Q selected spectra of training/query samples, we tried multiple fusion strategies, and found that the concatenation rule, which concatenates feature representations of all selected spectra of one sample into one vector, can achieve stable and better performances. We thus concatenate feature representations of selected spectra for classification. Fig. 4 shows the convergence curve of the IDICN network on the HK PolyU dataset [16]. One can see that the energy converges quickly and well. In the CMU and UWA datasets, the IDICN network can also converge with about 20 iterations.

V. Experiments

A. Dataset

In our experiments, we employ three widely used hyper-spectral face datasets, namely HK PolyU [16], CMU [30], and UWA [11] datasets, to evaluate our IDICN approach.

HK PolyU dataset includes hyper-spectral image cubes from 48 subjects, which are acquired by using the CRIs VariSpec liquid crystal tuneable filter (LCTF) under a halogen light. The spectral range is from 400 to 720 nm with a step length of 10 nm, producing 33 bands in all. This dataset contains appearance variations of the subjects, for example, changes of hair style and skin condition. Each of the first 25 subjects has four to seven cubes while the remaining 23 subjects have one cube per subject. Following [11] and [16], we use the first 25 subjects, comprising 113 hyper-spectral image cubes for experiment. The first six and the last three bands have high noise. As recommended by [16], we remove these bands in experiment. Fig. 5 shows demo face images from 12 spectra of one subject in the HK PolyU dataset.

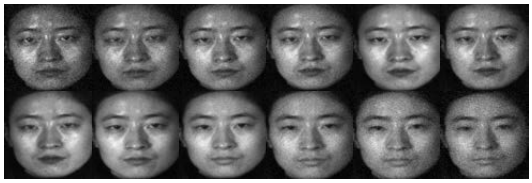


Fig. 5. Multispectral demo face images of one subject in HK PolyU.

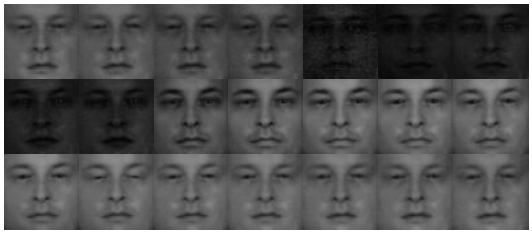


Fig. 6. Multispectral demo face images of one subject in the CMU dataset.

CMU dataset contains 54 diverse faces, which is obtained by using the prototype CMU-developed spectro-polarimetric camera. The hyper-spectral range is from 450 to 1090 nm with a step length of 10 nm (65 bands in all). Multiple sessions (up to five) were taken on most of the subjects over a period of several weeks. Each session consists of four different illumination conditions that include all lights on, center light on, left light on, and right light on. Like in [11], we only use the cubes acquired with all lights turned on, and the experimental data consists of 147 hyper-spectral cubes of 48 subjects where each subject has 1–5 cubes. Fig. 6 shows demo face images from 21 spectra of one subject in the CMU dataset.

UWA dataset consists of 120 hyper-spectral image cubes of 70 subjects, with each subject having one or two cubes. The hyper-spectral images are captured by using the CRIs VariSpec LCTF integrated with a photon focus camera. Each hyperspectral image cube contains 33 bands covering the spectral range of 400–720 nm with a 10-nm step. Fig. 7 shows demo face images from 12 spectra of one subject in the UWA dataset. Table I gives a summary about three used datasets.

B. Compared Methods and Experimental Settings

We compare our approach with three types of related methods: 1) five multispectral face recognition methods, that is, RDLF [5], 2-D^{2}PCA [16], STE [19], band fusion and partial least square regression (Band fusion+PLS) [11],

TABLE I
DETAILS OF THREE DATASETS

Dataset	Number of subjects	Number of spectra	Spectral range	Number of total cubes	Size of training set	Size of testing set
HK PolyU	25	33 (24 used)	400-720nm	113	50 cubes	63 cubes
CMU	48	65	450-1090nm	147	48 cubes	99 cubes
UWA	70	33	400-720nm	120	70 cubes	50 cubes

Fig. 7. Multispectral demo face images of one subject in the UWA dataset.

TABLE II
SELECTED SPECTRAL BANDS ON THREE DATASETS

Dataset	Spectral range	Selected subsets of spectral bands
HK PolyU	460-690nm	{490,540,550,580,630}nm
CMU	450-1090nm	{570,640,720,930,1010}nm
UWA	400-720nm	{530,540,610,630,690}nm

and MLSDL [20]; 2) one deep learning method, that is, MvDN [54]; and 3) two traditional face recognition methods, that is, LBP [64] and 2-D whitening reconstruction with whitened horizontally centered PCA (TWR+PCA II) [65]. LBP and TWR+PCA II are grayscale image-based face recognition algorithms. In order to get the grayscale images corresponding to multispectral images, like in [11], we transform multispectral images into RGB images using CIE 2006 tristimulus color matching functions [66], and then convert the RGB images to grayscale. We report the experimental results of compared methods with our own implementation. Since the codes of these compared methods (except for LBP) have not been publicly available, we reimplemented these methods and carefully optimized the hyper-parameters as recommended by the original authors. We were able to successfully reimplement the competing methods, matching the accuracy on published results.

In HK PolyU dataset, we randomly select two cubes per person for training, and use the remaining cubes for testing. For CMU dataset, like in [11], we randomly select one cube per subject to construct the training set and use the remaining cubes for testing. For the UWA dataset, following [11] and [20], we randomly select one cube for each of the 70 subjects to construct the training set, and use the remaining 50 cubes to construct the testing set. Training and testing set division on these three datasets is summarized in Table I. On these datasets, we run all compared methods 20 times to report the average recognition results. To leverage the parameters of the CaffeNet pretrained on the ImageNet dataset, we resize each spectral image to 227×227 in three datasets.

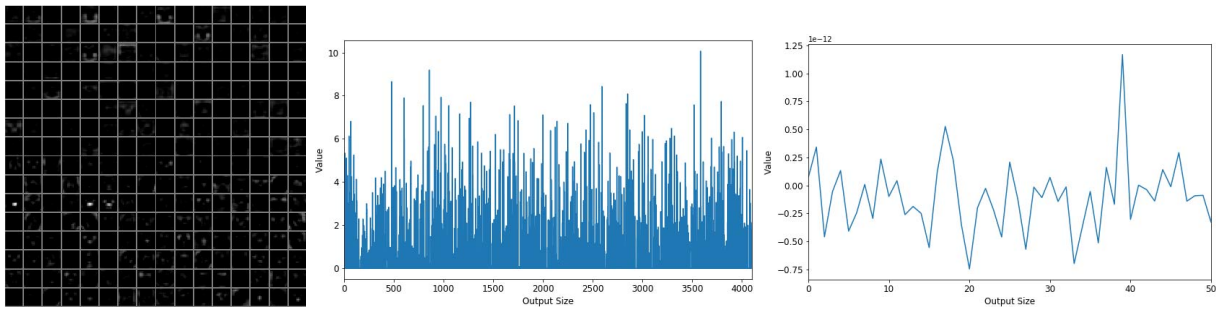
There are five hyper-parameters in our approach, namely the number of divided spectrum-sets (H), the dimension of feature representation of the top layer (d_y), the number of selected spectra (Q), the regularization parameter γ, and the balance factor α in (16). On HK PolyU dataset, we set H, d_y, γ, and α by twofold cross validation on training set. Through cube division according to the spectral bands, fivefold and threefold cross validations on training set are separately conducted on CMU and UWA for parameter tuning. Specifically, they are set as $H = 3$, $\gamma = 10^{-2}$, and $\alpha = 10^{-3}$ on HK PolyU; $H = 6$, $\gamma = 10^{-3}$, and $\alpha = 10^{-2}$ on CMU; and $H = 4$, $\gamma = 10^{-3}$, and $\alpha = 10^{-3}$ on UWA. Parameter d_y is tuned over the range $[50, 500]$ with step length of 50. We found that changing this parameter would not affect the results that much and we set it as 50 on three datasets.

Considering that previous band selection works choose about five spectra for training recognition model [11], [16], [20], we set parameter Q as 5 on three datasets. And the selected spectral bands are summarized in Table II. For network depth, we tried ResNet-50 [67] that includes 50 layers instead of CaffeNet to learn spectrum-set-specific features, and found that comparable results can be obtained by using more layers. We also extracted LBP features on raw images and tried to perform our IDICN on LBP images. And we found that the performances of IDICN on LBP images are inferior to those of IDICN on raw images.

C. IDICN Implementation Details

The common data augmentation practice, that is, horizontal flipping, is adopted to avoid overfitting. The image of a certain spectrum is input into the corresponding spectrum-set-specific DCNN. The details of each layer in IDICN network are tabulated in Table III. In the spectrum-pooling layer, Q spectra with largest Fisher scores are determined, and the output features of Fc7 corresponding to the selected spectra are fed to the fully connected layers Fc8, which has d_y neurons. Finally, the features of size $d_y \times 1$ corresponding to the input images of the selected spectra can be obtained. Fig. 8(a)–(c) separately shows the images of output features of the Conv5, Fc7, and Fc8 layers of IDICN on the HK PolyU dataset.

We initialize spectrum-set-specific DCNNs with the weights and biases of the first seven layers (discarding the softmax layer) from the pretrained CaffeNet (BVLC reference CaffeNet), which is provided in the Caffe library. For the fully connected layer Fc8, we initialize the weights from a zero-mean Gaussian distribution with standard deviation 0.01, and initialize the biases with the constant 1. IDICN is trained with full-batch optimization. We use an equal learning rate for all layers. The learning rate is initialized at 0.0001 and is divided by 10 for every 1000 iterations. The training is repeated until the network is converged or the predefined number of iterations is reached. We deem the network converged if the values of objective function in adjacent iterations are close enough. And the predefined maximum number of iterations is 5000.



(a) (b) (c)

Fig. 8. Output feature images of the (a) Conv5, (b) Fc7, and (c) Fc8 layers on the HK PolyU dataset.

TABLE III
DETAILS OF EACH LAYER IN IDICN NETWORK

Name	Type	Input size	Filter number /size /stride /pad	With ReLU	With normalization
Conv1 Pool1	Convolution Max pool	227×227 55×55	96 /11×11 /4 /0 N/A /3×3 /2 /0	yes no	no yes
Conv2 Pool2	Convolution Max pool	27×27 27×27	256 /5×5 /1 /2 N/A /3×3 /2 /0	yes no	no yes
Conv3	Convolution	13×13	384 /3×3 /1 /1	yes	no
Conv4	Convolution	13×13	384 /3×3 /1 /1	yes	no
Conv5 Pool5	Convolution Max pool	13×13 13×13	256 /3×3 /1 /1 N/A /3×3 /2 /0	yes no	no no
Fc6 Drop6	Fc Dropout	6×6 4096×1	N/A N/A	yes N/A	no N/A
Fc7 Drop7	Fc Dropout	4096×1 4096×1	N/A N/A	yes N/A	no N/A
Pool	Spectrum-pooling	4096×1	N/A	no	no
Fc8	Fc	4096×1	N/A	no	no

We use the deep learning framework/tool Caffe [46] that is developed by the Berkeley Vision and Learning Center (BVLC) and community contributors. It was downloaded at https://github.com/BVLC/caffe.git. Caffe is installed on the Ubuntu 16.04 64bit operating system with the hardware configuration comprising Intel i7-6800K CPU, NVIDIA GeForce GTX TITAN X GPU, 32 GB DDR3 memory, and SSD hard drive. And we use OpenCV 3.2, CUDA 8.0, and cuDNN v5.1. We train the network on a single NVIDIA TITAN X GPU.

D. Comparison With the Baselines

Table IV lists the average recognition results on three datasets. We can see that our IDICN approach improves the average recognition rate at least by 3.43% = (99.76%−96.33%) on HK PolyU, 0.87% = (100.00%−99.13%) on CMU, and 1.34% = (99.85%−98.51%) on UWA. The reasons for this improvement lie in that our approach is based on the powerful DCNNs, which can help to fully extract the spectrum-set-specific features. Besides, the spectra with favorable discriminative abilities are effectively selected. Furthermore, the intraspectrum discrimination and interspectrum discriminant correlation is fully considered to learn favorable features.

To statistically analyze the recognition rates in Table IV, we also conduct a statistical test, that is, Mcnemar's test [68]. This test can provide statistical significance between our IDICN approach and other compared methods. Here, the Mcnemar's test uses a significance level of 0.05, that is, if the p-value is below 0.05, the performance difference between two compared methods is regarded as statistically significant. Table V shows the p-values between IDICN and other compared methods on three datasets. According to the table, our approach makes a statistically significant difference in comparison with related methods.

TABLE IV
AVERAGE RECOGNITION RATES (%, ±STANDARD DEVIATION) OF COMPARED METHODS ON THREE DATASETS

Method	HK PolyU	CMU	UWA
LBP	90.05±2.42	96.61±1.67	96.20±1.05
TWR+PCA II	88.46±2.15	89.13±2.03	85.60±2.10
RDLF	86.26±2.84	88.25±2.32	88.01±2.61
$(2D)^2$PCA	71.11±3.16	72.10±5.41	83.85±2.42
STE	88.17±2.35	90.83±2.01	89.85±2.84
Band fusion+PLS	95.20±1.60	99.13±0.57	98.26±1.21
MLSDL	96.33±1.42	99.04±0.66	98.51±1.04
MvDN	94.28±1.69	95.24±1.45	96.34±1.40
IDICN	**99.76**±0.43	**100.00**±0.00	**99.85**±0.25

In addition, we employ the cumulative match characteristic (CMC) [69] curve as our evaluation metric. A CMC curve plots the recognition rate against the returned 1 : n candidate list size. The faster the CMC curve approaches 100%, the better the recognition method is. Fig. 9(a)–(c) shows the rank-n average recognition rate of compared methods over 20 random running on three datasets. We can observe that our IDICN can always approach 100% quickly. On the CMU dataset, IDICN can even reach 100% in rank-1 case.

E. Discussion

1) Effects of the Intraspectrum Discriminant Analysis and the Interspectrum Correlation Analysis: Intraspectrum discriminant analysis and interspectrum correlation analysis are the two most important components in our IDICN approach. In this section, we evaluate the effects of these two components. We conduct IDICN without the intraspectrum discriminant analysis part and/or without the interspectrum correlation analysis part. We call the version of IDICN without the intraspectrum discriminant analysis part as ICN, and call the version of IDICN without the interspectrum correlation analysis part as IDN. We call the version without both the

TABLE V
p-VALUES BETWEEN IDICN AND OTHER COMPARED METHODS ON THREE DATASETS

Dataset	IDICN							
	LBP	TWR+PCA II	RDLF	$(2D)^2$PCA	STE	Band fusion+PLS	MLSDL	MvDN
HK PolyU	4.20×10^{-4}	3.06×10^{-4}	4.56×10^{-5}	1.92×10^{-6}	3.21×10^{-4}	4.01×10^{-3}	1.07×10^{-3}	7.94×10^{-3}
CMU	8.84×10^{-3}	2.81×10^{-5}	1.77×10^{-6}	2.48×10^{-8}	5.24×10^{-4}	1.96×10^{-2}	1.05×10^{-2}	8.27×10^{-3}
UWA	7.37×10^{-4}	1.49×10^{-5}	3.55×10^{-5}	9.04×10^{-5}	1.08×10^{-5}	8.49×10^{-3}	1.62×10^{-2}	4.88×10^{-3}

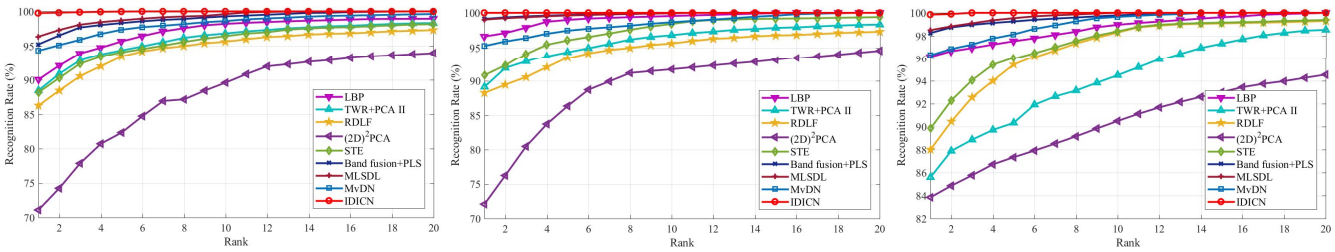


Fig. 9. Performance comparison using CMC curves on (a) HK PolyU, (b) CMU, and (c) UWA datasets.

TABLE VI
AVERAGE RECOGNITION RATES (%, ±STANDARD DEVIATION) OF DCNNS+SP, ICN, IDN, IDICN$_{TC}$, AND IDICN

Method	HK PolyU	CMU	UWA
DCNNs+sp	88.20±2.35	91.65±2.11	91.03±2.05
ICN	92.38±1.44	94.70±1.27	93.47±1.40
IDN	94.05±1.33	95.16±1.15	95.62±1.06
IDICN$_{TC}$	98.84±0.62	99.28±0.10	98.91±0.47
IDICN	99.76±0.43	100.00±0.00	99.85±0.25

intraspectrum discriminant and interspectrum correlation analysis parts as DCNNs+sp, which only uses the pretrained CaffeNet to extract features and employs the spectrum pooling layer to select discriminant spectra. Besides, in order to evaluate the discrimination consideration in the interspectrum correlation analysis part, we also experiment with the version of IDICN that minimizes the interspectrum total cross covariance and the intraspectrum Fisher loss. This version is called IDICN$_{TC}$. Table VI shows the comparison of recognition results on three datasets.

From the table, the recognition performances of DCNNs+sp, ICN, and IDN are obviously inferior to that of IDICN, and when discrimination is not considered in interspectrum correlation analysis, IDICN experiences a slight performance degradation. This result means that the designed intraspectrum discriminant analysis and interspectrum correlation analysis components can help to learn features with favorable discriminative capability.

In addition, there exist large differences between ICN and DCCA [47]. Specifically, the differences mainly lie in the following aspects.

1) ICN can be used in the scenario containing multiple datasets/spectra, while DCCA can only deal with the applications including two data sets.
2) ICN treats the interspectrum within-class and between-class correlation differently, such that the favorable cross-spectral correlation can be fully used and the adverse redundancy among spectra is reduced, while DCCA only focuses on the interview total correlation.
3) ICN uses the spectrum pooling layer to dynamically and adaptively select discriminative spectra, while DCCA does not provide the view selection scheme.

2) Effect of the Designed Spectrum Pooling Layer: We design a spectrum pooling layer to select spectra with strong discriminative abilities for network training. To evaluate the effectiveness of the layer in IDICN, we compare the recognition performance of the complete set of spectra with the selected spectra. Table VII summarizes the results. In this table, we also report the training and testing time (the testing time for the total testing set) to observe the computational cost.

From the table, we can see that our IDICN approach can obtain the same or even better recognition results using 5 selected spectra out of 24 spectra on HK PolyU, 5 out of 65 spectra on CMU, and 5 out of 33 spectra on UWA, as compared with the results using the complete set of spectra. As to the running time, although performing the spectrum selection practice costs a little more training time, it significantly reduces the testing time. These two aspects indicate that the spectrum pooling layer is effective.

3) Parameter Analysis: Next, we provide a discussion about the sensitivity of IDICN to different choices of parameters H, γ, α, and Q. We take the HK PolyU dataset as an example and conduct experiments by changing values of H, γ, α, and Q. When evaluating one parameter, the others are fixed to the values introduced in Section V-B. Fig. 10(a)–(d) separately shows the recognition rates versus different values of H, γ, α, and Q. From the figure, IDICN is not sensitive to the choice of H in the range [3, 5], γ in the range $[10^{-3}, 10^{-1}]$, and α in the range $[10^{-4}, 10^{-2}]$. In addition, nearly stable recognition results can be obtained when Q is in the range [5, 10]. For simplicity, we set $H = 3$, $\gamma = 10^{-2}$, $\alpha = 10^{-3}$, and $Q = 5$ for HK PolyU in experiments. A similar phenomenon also exists on the other two datasets.

4) How Does Our Approach Perform on Larger Dataset: To investigate whether our approach can perform well on larger dataset, in this section, we design an experiment for

TABLE VII
AVERAGE RECOGNITION RESULTS (%) AND EXECUTION TIME (SECOND) OF OUR APPROACH WHEN THE SPECTRUM POOLING LAYER IS USED OR NOT

Dataset	Using complete spectra			Using selected spectra		
	Recognition rate	Training time	Testing time	Recognition rate	Training time	Testing time
HK PolyU	99.70	2980	147	99.76	3205	18
CMU	100.00	3864	310	100.00	4259	25
UWA	99.82	3144	152	99.85	3373	19

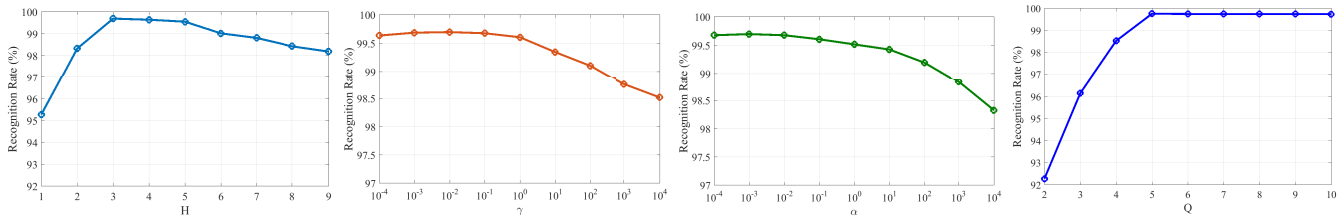


Fig. 10. Recognition results on the HK PolyU dataset versus different values of (a) H, (b) γ, (c) α, and (d) Q.

evaluation. We combine the HK PolyU and UWA datasets, both of which have the spectral range of 400–720 nm, such that the combined dataset contains 233 cubes of 95 subjects. For simplicity, all the spectral bands including the first 6 and the last 3 bands in HK PolyU are used for the experiment. We call the combination of HK PolyU and UWA datasets as H-U dataset. We randomly select two cubes per person for the 25 subjects in the HK PolyU part and one cube per person for the 70 subjects in the UWA part for training, and use the remaining cubes for testing. For parameters H, γ, and α in our IDICN, they are separately set as $H = 3$, $\gamma = 10^{-2}$, and $\alpha = 10^{-3}$. In experiment, the spectral bands of 490, 540, 580, 630, and 690 nm are selected in IDICN.

Table VIII reports the average recognition results over 20 random running of compared methods. Comparing the results in the table with those reported in Table IV, we can see that the recognition performance experiences a significant decline when larger dataset with more classes is used. The reasons for the performance drop may lie in the following two aspects.

1) The combination of two datasets may make the classification task more difficult, since more subjects should be accurately recognized. The HK PolyU dataset and the UWA dataset are collected in different scenarios, which may lead to the dataset characteristics difference.
2) Different from the experimental setting on the HK PolyU dataset in Table IV, all the spectral bands including the first 6 and the last 3 bands in HK PolyU are used in experiment on H-U dataset, while these nine bands have high noise [16].

In Table VIII, IDICN$_{CS}$ denotes the version of IDICN that uses the complete spectra. From the table, we can see that better recognition result can be achieved by IDICN as compared with IDICN$_{CS}$, and the result improvement reaches $0.44\% = (84.89\% - 84.45\%)$. In addition, it is obvious that IDICN obtains better recognition results than the related multispectral face recognition methods and the compared deep learning method, and the performance improvement is at least $5.48\% = (84.89\% - 79.41\%)$. According to these results, although our IDICN will experience performance decline on a larger dataset, it can still achieve good recognition result as compared with the related state-of-the-art methods.

TABLE VIII
AVERAGE RECOGNITION RATES (%, ±STANDARD DEVIATION) OF COMPARED METHODS ON THE H-U DATASET

Method	H-U
RDLF	70.98±3.19
(2D)^{2}PCA	62.72±4.73
STE	72.36±3.25
Band fusion+PLS	78.84±2.96
MLSDL	79.24±2.60
MvDN	79.41±2.54
IDICN	84.89±2.05
IDICN$_{CS}$	84.45±2.07

VI. CONCLUSION

In this paper, we propose a novel multispectral face recognition approach named IDICN. It provides a supervised deep neural network model containing multiple spectrum-set-specific DCNNs that can effectively learn the spectrum-set-specific features and a spectrum pooling layer that can adaptively select spectra with favorable discriminative abilities. IDICN also gives a novel loss function, which can fully explore the intraspectrum discriminant information and the useful interspectrum correlation information in multispectral face image data simultaneously.

We evaluate our approach on three widely used datasets. The experimental results demonstrate that our approach is superior to the related methods. The results also validate the effectiveness of the intraspectrum discriminant analysis part, interspectrum correlation analysis part, and the designed spectrum pooling layer in our approach. In addition, the experimental results show the efficiency of the IDICN network, which converges quickly with a small amount of running time.

For the future work, we will collect multispectral face images in uncontrolled environment. We will then further evaluate the proposed approach on these multispectral face images. We will also dedicate to addressing more challenging problems

in practical multispectral face recognition for promoting the deployment of multispectral face recognition.

ACKNOWLEDGMENT

The authors would like to thank the Editors and anonymous reviewers for their constructive comments and suggestions.

REFERENCES

[1] A. K. Jain and S. Z. Li, *Handbook of Face Recognition*. London, U.K.: Springer, 2011.
[2] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surveys*, vol. 35, no. 4, pp. 399–458, 2003.
[3] D. Tao, Y. Guo, Y. Li, and X. Gao, "Tensor rank preserving discriminant analysis for facial recognition," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 325–334, Jan. 2018.
[4] W. Cho, "Hyperspectral data acquisition and its application for face recognition," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Univ. Tennessee, Knoxville, TN, USA, 2015.
[5] H. Chang, A. Koschan, B. Abidi, and M. Abidi, "Fusing continuous spectral images for face recognition under indoor and outdoor illuminants," *Mach. Vis. Appl.*, vol. 21, no. 2, pp. 201–215, 2010.
[6] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, "Face recognition in hyperspectral images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1552–1560, Dec. 2003.
[7] H. Wang, T. C. Bau, and G. E. Healey, "Expression-invariant face recognition in hyperspectral images," *Opt. Eng.*, vol. 53, no. 10, 2014, Art. no. 103102.
[8] D. Zhang, Z. Guo, and Y. Gong, "Multispectral biometrics systems," in *Multispectral Biometrics*. Cham, Switzerland: Springer, 2016, pp. 23–35.
[9] N. T. Vetrekar, R. Raghavendra, A. A. Gaonkar, G. M. Naik, and R. S. Gad, "Extended multi-spectral face recognition across two different age groups: An empirical study," in *Proc. 10th Indian Conf. Comput. Vis. Graph. Image Process.*, 2016, Art. no. 78.
[10] F. Omri, S. Foufou, and M. Abidi, "NIR and visible image fusion for improving face recognition at long distance," in *Proc. 6th Int. Conf. Image Signal Process.*, 2014, pp. 549–557.
[11] M. Uzair, A. Mahmood, and A. Mian, "Hyperspectral face recognition with spatiospectral information fusion and PLS regression," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1127–1137, Mar. 2015.
[12] R. Raghavendra, B. Dorizzi, A. Rao, and G. H. Kumar, "Particle swarm optimization based fusion of near infrared and visible images for improved face verification," *Pattern Recognit.*, vol. 44, no. 2, pp. 401–411, 2011.
[13] O. Arandjelović, R. Hammoud, and R. Cipolla, "Thermal and reflectance based personal identification methodology under variable illumination," *Pattern Recognit.*, vol. 43, no. 5, pp. 1801–1813, 2010.
[14] Y. Zheng, "Orientation-based face recognition using multispectral imagery and score fusion," *Opt. Eng.*, vol. 50, no. 11, 2011, Art. no. 117202.
[15] R. Singh, M. Vatsa, and A. Noore, "Hierarchical fusion of multi-spectral face images for improved recognition performance," *Inf. Fusion*, vol. 9, no. 2, pp. 200–210, 2008.
[16] W. Di, L. Zhang, D. Zhang, and Q. Pan, "Studies on hyperspectral face recognition in visible spectrum with feature band selection," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 40, no. 6, pp. 1354–1361, Nov. 2010.
[17] A. Koschan, Y. Yao, H. Chang, and M. Abidi, "Multispectral face imaging and analysis," in *Handbook of Face Recognition*. London, U.K.: Springer, 2011, pp. 401–428.
[18] M. A. Akhloufi and A. Bendada, "Locally adaptive texture features for multispectral face recognition," in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, 2010, pp. 3308–3314.
[19] H. Zhao and S. Sun, "Sparse tensor embedding based multispectral face recognition," *Neurocomputing*, vol. 133, pp. 427–436, Jun. 2014.
[20] X.-Y. Jing *et al.*, "Multi-spectral low-rank structured dictionary learning for face recognition," *Pattern Recognit.*, vol. 59, pp. 14–25, Nov. 2016.
[21] J. Chen *et al.*, "Learning mappings for face synthesis from near infrared to visual light images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 156–163.
[22] B. Klare and A. K. Jain, "Heterogeneous face recognition: Matching NIR to visible light images," in *Proc. IEEE 20th Int. Conf. Pattern Recognit.*, 2010, pp. 1513–1516.
[23] F. Nicolo and N. A. Schmid, "Long range cross-spectral face recognition: Matching SWIR against visible light images," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 6, pp. 1717–1726, Dec. 2012.
[24] B. F. Klare and A. K. Jain, "Heterogeneous face recognition using kernel prototype similarities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1410–1422, Jun. 2013.
[25] F. Juefei-Xu, D. K. Pal, and M. Savvides, "NIR-VIS heterogeneous face recognition via cross-spectral joint dictionary learning and reconstruction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2015, pp. 141–150.
[26] C. Chen and A. Ross, "Matching thermal to visible face images using hidden factor analysis in a cascaded subspace learning framework," *Pattern Recognit. Lett.*, vol. 72, pp. 25–32, Mar. 2016.
[27] D. Yi, Z. Lei, Z. Zhang, and S. Z. Li, "Face anti-spoofing: Multispectral approach," in *Handbook of Biometric Anti-Spoofing*. London, U.K.: Springer, 2014, pp. 83–102.
[28] A. K. Jain, P. Flynn, and A. A. Ross, *Handbook of Biometrics*. Boston, MA, USA: Springer, 2007.
[29] D. Zhang, Z. Guo, and Y. Gong, *Multispectral Biometrics: Systems and Applications*. Cham, Switzerland: Springer, 2016.
[30] L. J. Denes, P. Metes, and Y. Liu, "Hyperspectral face database," Robot. Inst. Carnegie Mellon Univ., Pittsburgh, PA, USA, Rep. CMU-RI-TR-02-25, 2002.
[31] W. Cho, J. Jang, A. Koschan, M. A. Abidi, and J. Paik, "Hyperspectral face recognition using improved inter-channel alignment based on qualitative prediction models," *Opt. Exp.*, vol. 24, no. 24, pp. 27637–27662, 2016.
[32] S. Hu, N. J. Short, P. K. Gurram, K. P. Gurton, and C. Reale, "MWIR-to-visible and LWIR-to-visible face recognition using partial least squares and dictionary learning," in *Face Recognition Across the Imaging Spectrum*. Cham, Switzerland: Springer, 2016, pp. 69–90.
[33] M. Uzair, A. Mahmood, and A. S. Mian, "Hyperspectral face recognition using 3D-DCT and partial least squares," in *Proc. 24th Brit. Mach. Vis. Conf. (BMVA)*, 2013, pp. 1–9.
[34] S. Bianco, "On the usefulness of hyperspectral imaging for face recognition," *J. Electron. Imag.*, vol. 25, no. 6, 2016, Art. no. 63020.
[35] V. Sharma and L. Van Gool, "Image-level classification in hyperspectral images using feature descriptors, with application to face recognition," *CoRR*, vol. abs/1605.03428, 2016. [Online]. Available: http://arxiv.org/abs/1605.03428
[36] J. Liang, J. Zhou, and Y. Gao, "3D local derivative pattern for hyperspectral face recognition," in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit.*, 2015, pp. 1–6.
[37] H. Chang, Y. Yao, A. Koschan, B. Abidi, and M. Abidi, "Improving face recognition via narrowband spectral range selection using Jeffrey divergence," *IEEE Trans. Inf. Forensics Security*, vol. 4, no. 1, pp. 111–122, Mar. 2009.
[38] H. J. Bouchech, S. Foufou, and M. Abidi, "Dynamic best spectral bands selection for face recognition," in *Proc. IEEE 48th Annu. Conf. Inf. Sci. Syst.*, 2014, pp. 1–6.
[39] H. J. Bouchech, S. Foufou, and M. Abidi, "Multilinear sparse decomposition for best spectral bands selection," in *Proc. 6th Int. Conf. Image Signal Process.*, 2014, pp. 384–391.
[40] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
[41] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[42] Y. Sun, X. Wang, and X. Tang, "Sparsifying neural network connections for face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4856–4864.
[43] C. R. Qi *et al.*, "Volumetric and multi-view CNNs for object classification on 3D data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 5648–5656.
[44] H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4293–4302.
[45] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 26th Annu. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[46] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 675–678.

[47] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep canonical correlation analysis," in *Proc. 30th Int. Conf. Mach. Learn.*, 2013, pp. 1247–1255.
[48] D. R. Hardoon, S. R. Szedmak, and J. R. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Comput.*, vol. 16, no. 12, pp. 2639–2664, 2004.
[49] Z. Wu, Y.-G. Jiang, J. Wang, J. Pu, and X. Xue, "Exploring inter-feature and inter-class relationships with deep neural networks for video classification," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 167–176.
[50] A. Wang, J. Lu, J. Cai, T.-J. Cham, and G. Wang, "Large-margin multi-modal deep learning for RGB-D object recognition," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 1887–1898, Nov. 2015.
[51] H. Zhu, J.-B. Weibel, and S. Lu, "Discriminative multi-modal feature fusion for RGBD indoor scene recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2969–2976.
[52] D. Tao, Y. Guo, B. Yu, J. Pang, and Z. Yu, "Deep multi-view feature learning for person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 10, pp. 2657–2666, Oct. 2018.
[53] D. Tao *et al.*, "Person re-identification by dual-regularized KISS metric learning," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2726–2738, Jun. 2016.
[54] M. Kan, S. Shan, and X. Chen, "Multi-view deep network for cross-view classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4847–4855.
[55] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016.
[56] Y. Li, W. Xie, and H. Li, "Hyperspectral image reconstruction by deep convolutional neural network for classification," *Pattern Recognit.*, vol. 63, pp. 371–383, Mar. 2017.
[57] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.
[58] W. Li, G. Wu, F. Zhang, and Q. Du, "Hyperspectral image classification using deep pixel-pair features," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 844–853, Feb. 2017.
[59] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 645–657, Feb. 2017.
[60] 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, Jul. 1997.
[61] F. Yan and K. Mikolajczyk, "Deep correlation for matching images and text," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3441–3450.
[62] A. Ng, J. Ngiam, C. Y. Foo, Y. Mai, and C. Suen. (2013). *Unsupervised Feature Learning and Deep Learning Tutorial*. [Online]. Available: http://deeplearning. stanford. edu/tutorial
[63] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Math. Program.*, vol. 45, nos. 1–3, pp. 503–528, 1989.
[64] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.
[65] X. Shi *et al.*, "Two-dimensional whitening reconstruction for enhancing robustness of principal component analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 10, pp. 2130–2136, Oct. 2016.
[66] CIE-CMFs. *CIE 2006 Color Matching Functions*. Accessed: Oct. 5, 2017. [Online]. Available: http://cvrl.ioo.ucl.ac.uk/cmfs.htm
[67] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[68] B. A. Draper, W. S. Yambor, and J. R. Beveridge, "Analyzing PCA-based face recognition algorithms: Eigenvector selection and distance measures," *Empirical Evaluation Methods in Computer Vision*. River Edge, NJ, USA: World Sci., 2002, pp. 1–15.
[69] D. Tao, L. Jin, Y. Wang, and X. Li, "Person reidentification by minimum classification error-based KISS metric learning," *IEEE Trans. Cybern.*, vol. 45, no. 2, pp. 242–252, Feb. 2015.

Fei Wu received the Ph.D. degree in computer science from the Nanjing University of Posts and Telecommunications (NJUPT), Nanjing, China, in 2016.

He is currently with the College of Automation, NJUPT. He has authored over 30 scientific papers. His current research interests include pattern recognition, artificial intelligence, and computer vision.

Xiao-Yuan Jing received the Doctoral degree in pattern recognition and intelligent system from the Nanjing University of Science and Technology, Nanjing, China, in 1998.

He is currently a Professor with the School of Computer, Wuhan University, Wuhan, China, and the College of Automation, Nanjing University of Posts and Telecommunications, Nanjing. His current research interests include artificial intelligence and pattern recognition.

Xiwei Dong is currently pursuing the Ph.D. degree in computer science with the Nanjing University of Posts and Telecommunications, Nanjing, China.

His current research interests include machine learning, pattern recognition, and computer vision.

Ruimin Hu (SM'10) was born in 1964. He received the Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, China, in 1994.

He is currently a Professor and the Dean of the School of Computer Science, Wuhan University, Wuhan, where he is also the Director of the National Engineering Research Center for Multimedia Software. His current research interests include multimedia network communication, security and surveillance technology, and big data analytics.

Dong Yue received the Ph.D. degree from the South China University of Technology, Guangzhou, China, in 1995.

He is currently the Dean of the College of Automation, Nanjing University of Posts and Telecommunications, Nanjing, China, and also a Changjiang Professor. His current research interests include analysis and synthesis of networked control systems, and multiagent systems.

Dr. Yue is currently an Associate Editor of the IEEE Control Systems Society Conference Editorial Board and the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS.

Lina Wang received the Ph.D. degree in computer science and technology from Northeastern University, Shenyang, China, in 2001.

She is a Professor with Wuhan University, Wuhan, China. Her current research interests include system security, multimedia security, and cloud computing.

Yi-Mu Ji received the Ph.D. degree in computer science from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2006.

He is a Professor with the Nanjing University of Posts and Telecommunications, Nanjing, China. His current research interests include intelligent driving and data processing.

Ruchuan Wang received the B.S. degree in computational mathematics from the PLA Information Engineering University, Zhengzhou, China, in 1968.

He was a Visiting Scholar with Bremen University, Bremen, Germany, Munich University, Munich, Germany, and Max-Planck Institute, Germany, from 1984 to 1992. He is a Professor with the Nanjing University of Posts and Telecommunications, Nanjing, China. His current research interests include wireless sensor networks and information security.

Guoliang Chen received the B.S. degree from Xi'an Jiaotong University, Xi'an, China, in 1961.

He is an Academician of the Chinese Academy of Sciences, Beijing, China. He is currently the Director of the National High Performance Computing Center, Hefei, China. He has published nine books and over 200 research papers. His current research interests include parallel algorithms, computer architectures, computer networks, and computational intelligence.