

Received February 2, 2020, accepted February 8, 2020, date of publication February 12, 2020, date of current version March 3, 2020. *Digital Object Identifier* 10.1109/ACCESS.2020.2973472

Local Feature Hashing With Binary Auto-Encoder for Face Recognition

JING CHEN[®] AND YUNXIAO ZU[®]

School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China Corresponding author: Yunxiao Zu (zuyx@bupt.edu.cn)

This work was supported by the Beijing Key Laboratory of Work Safety Intelligent Monitoring, School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing, China.

ABSTRACT The learning-based hashing has recently made encouraging progress in face recognition. However, most existing hashing methods disregard the discrete constraint during optimization, inducing the accumulated quantization errors. In this work, we develop an effective learning-based hashing model, namely local feature hashing with binary auto-encoder (LFH-BAE), to directly learn local binary descriptors in the Hamming space. It attempts to exploit structure factors to well reconstruct the face image from binary codes. Specifically, we first introduce a binary auto-encoder to learn a hashing function to project each face region into high-quality binary codes. Since the original problem is a tricky combinational function, we then present a softened version to decompose it into separate tractable sub-problems. Next, we propose an effective alternating algorithm based on the augmented Lagrange method (ALM) to solve these sub-problems, which helps to generate strong discriminative and excellent robust binary codes. Moreover, we utilize the discrete cyclic coordinate descent (DCC) method to optimize binary codes to reduce the loss of useful information. Lastly, we cluster and pool the obtained binary codes, and construct a histogram feature as the final face representation for each image. Extensive experimental results on four public datasets including FERET, CAS-PEAL-R1, LFW and PaSC show that our LFH-BAE is superior to most state-of-the-art face recognition algorithms.

INDEX TERMS Binary auto-encoder, binary feature, discrete optimization, face recognition, feature hashing.

I. INTRODUCTION

Face recognition has been a popular research topic in computer vision due to its potential applications in various real-world scenarios. Over the past thirty decades, a variety of face recognition algorithms have been proposed in the literatures [1]–[11]. These algorithms have achieved encoura-ging recognition performance but often only in the controlled scenarios. With the development of the social network technology and image capturing devices, there has been grown interest in face recognition under the uncontrolled scenarios. However, face recognition under those conditions is even more difficult due to the unpredictability of numerous variations such as pose, occlusion, aging, illumination, expression and resolution. According to the existing works, face representation is the most important component of a face recognition system. Therefore, the key challenge for practical face recognition system is to obtain robust face representa-tions that are invariant to the real-world variations.

The associate editor coordinating the review of this manuscript and approving it for publication was Juan Wang^(b).

In recent years, local features have been studied extensively due to their stableness and robustness to local changes. Numerous local feature algorithms have sprung up in the field of face recognition. Based on the designing methodology, existing local features can be classified into two categories: hand-crafted features [1], [2], [4] and learning-based features [3], [5], [7]. Typical hand-crafted features are local binary pattern (LBP) [1], scale-invariant feature transform (SIFT) [2] and Gabor wavelets [4]. The hand-crafted features are data-independent and usually require strong prior knowledge to elaborately engineer them, which only work well against the variations of the controlled scenarios. Unlike the hand-crafted features, the learning-based features are extracted from the raw face data in a data-driven way and they can obtain satisfying performance in the unconstrained scenarios. Various learning-based features have been proposed such as local quantized pattern (LQP) [12], discriminant face descriptor (DFD) [3], compact binary feature descriptor (CBFD) [5] and context-aware local binary feature learning (CA-LBFL) [7]. Among them, the binary codes obtained by the learning-based hashing methods



FIGURE 1. The framework of our proposed LFH-BAE model for face recognition. In the training procedure, we first extract patch-wise pixel difference vectors (PDVs) from each face image and learn an encoding matrix R_1 and an encoding bias δ_1 to project the PDVs into high-quality binary codes, which are produced in the Hamming space in a closed form. Then, an unsupervised clustering is performed on the binary codes to obtain a codebook D for feature pooling. In the testing procedure, we first extract the PDVs for the testing image and then utilize the learned encoding matrix and bias to project the PDVs into discriminative binary codes. Finally, we use the learned codebook to pool the binary codes into histogram feature representations.

(i.e., CBFD and CA-LBFL) show stronger discriminative power due to their compact nature, energy saving and even distribution.

Due to the binary constraint imposed on the binary codes, the discrete hashing optimization is a NP-hard problem. To simplify the solving process, the above learning-based hashing methods follow the two-step "relaxing+thresholding" optimization procedure to learn binary codes. Specifical-ly, they first relax the binary constraint to obtain continuous solutions, and then binarize the resulting values to binary codes by thresholding. However, such traditional two-step optimization disregards the binary nature of the hashing codes, inducing the accumulated quantization errors and leading to the loss of useful information. Therefore, how to learn high-quality binary codes to eliminate the accumulated quantization errors remains a central problem in face recognition.

In this paper, we introduce a simple unsupervised binary hashing model, dubbed local feature hashing with binary auto-encoder (LFH-BAE), to directly learn binary codes in the Hamming space for face representation. The proposed LFH-BAE differs from previous binary hashing methods for face recognition in two aspects. For one thing, the proposed LFH-BAE combines the binary encoder with decoder, so the encoder can receive the feedback of the decoder in the feature learning stage. Thus, the binary auto-encoder can exploit more salient information for subsequent feature pooling. For another, it respects the binary nature of the problem when implementing the discrete hashing optimization, so that the accumulated quantization errors can be somehow eliminated. Fig. 1 illustrates the framework of our proposed LFH-BAE

vectors (PDVs) from each face image and cascade them into a large pixel difference matrix (PDM). Inspired by the fact that local binary codes are robust to intra-class variations and auto-encoder can capture the manifold structure, we then introduce the binary auto-encoder to learn a hashing function to project PDM into high-quality binary codes. It is tricky to solve the original objective function of the auto-encoder due to the three hard constraints imposed on the output of the encoder. To address this, we propose a softened version of the binary auto-encoder by introducing an auxiliary variable to minimize the binary quantization loss. By doing so, we can break the combinational complexity of the original problem and convert it into separate feasible sub-problems. We put forward an effective alternating optimization algorithm based on the augmented Lagrange method (ALM) [13] to solve these sub-problems, which not only can explicitly cope with the discrete constraint, but also consider the bit-independent constraint and bit-balanced constraint together. To deal with the most pivotal sub-problem, i.e., discrete optimization for binary codes, we utilize the discrete cyclic coordinate descent (DCC) method to directly learn binary codes in the Hamming space. The ALM-based alternating optimization algorithm enables LFH-BAE to generate strong discriminative and excellent robust binary codes, which are very favorable for face recognition. Finally, we construct a codebook by clustering on the obtained binary codes and pool them into a high-dimensional histogram feature as the final representation for each face image. We perform extensive experiments on FERET, CAS-PEAL-R1, LFW and PaSC datasets and experimental results show that our proposed model outperforms most existing state-of-the-art methods.

model. Firstly, we extract the patch-wise pixel difference

The main contributions of this work are summarized as follows:

- 1) We propose an unsupervised binary auto-encoder to capture the salient structure inherent in the raw data and learn high-quality binary features for face representa-tion. With the well-trained binary autoencoder, the learned binary codes can best approximate the global data space and deliver more identity information.
- 2) To reduce the combinational complexity of the original problem, we transform the binary auto-encoder with three constraints to a softened version by minimizing the binary quantization loss, and then decompose the original problem into separate tractable sub-problems.
- 3) Moreover, we propose an efficient alternating optimiza-tion algorithm based on the ALM to directly compute the discrete solution in the Hamming space. Our method can not only address the binary constraint, but also take the bit-balanced and bit-independent constraints into consideration. Thus, the accumulated quantization errors caused by the "relaxing+thresholding" optimization method can be effectively eliminated and high-quality binary codes can be obtained.

4) We conduct extensive face recognition experiments on four public datasets consisting of FERET, CAS-PEAL-R1, LFW and PaSC to evaluate the effectiveness and generalization ability of our LFH-BAE. The results on these datasets demonstrate that our model is superior to most of state-of-the-art face recognition algorithms.

II. BACKGROUND AND RELATED WORKS

Since our proposed hashing model is a learning-based binary auto-encoder for face recognition, in this section, we concen-trate on the following three related topics: 1) face feature learning, 2) feature representation based on auto-encoders, and 3) binary hashing.

A. FACE FEATURE LEARNING

Recently, feature learning has made abundant achievements in the field of face recognition [3], [14]–[19]. Lei et al. [3] combine the image filter with the optimal soft sampling to obtain LBP-like features. Chang et al. [14] incorporate the sparse constraints into the low-rank regression optimization to analysis the importance of each feature under the PCA criteria. Aslan et al. [15] propose a multi-channel multi-model method that uses various techniques to extract features from multiple regions. Yi et al. [16] utilize the Restricted Boltzmann Machines (RBMs) to learn the shared representations from the extracted Gabor features to eliminate the heterogeneity locally. He et al. [17] propose a dynamic feature matching (DFM) method, which deals with partial face images by combining the fully convolutional network (FCN) with sparse representation classification (SRC). Liu et al. [18] propose a deep hypersphere embedding method to learn face features with angular margin. Liu et al. [19] present a graph-based manifold learning method to extract sparse feature representations for face recognition.

Since the learning-based algorithms can exploit discriminative information from the face data in a data-driven way, these methods, especially the convolutional neural networks (CNNs), perform better than the hand-crafted methods in various face recognition tasks. However, there are two drawbacks of CNNs in practical applications. Firstly, the CNNs are required to estimate extensive parameters from a large amount of labeled samples, which leads to time consuming in the training process. Secondly, due to privacy protection, it is very difficult to collect the labeled data in real applications. Therefore, we propose an unsupervised low-level feature learning algorithm for face recognition.

B. FEATURE REPRESENTATION BASED ON AUTO-ENCODES

As a typical learning method, auto-encoder can extract structure factors inherent in the input data automatically. It has shown its effectiveness and efficiency in various compu-ter vision tasks [20]–[24]. Kan *et al.* [20] use the stacked progressive auto-encoder to extract the pose-robust features for face recognition. Cheng *et al.* [21] propose a double channel stacked sparse denoising auto-encoder (DC-SSDA) to eliminate or reduce the influence of the occlusions. Gao *et al.* [22] first associate the label information with the auto-encoder to form a deep supervised auto-encoder. Then, they introduce the similarity preservation term into the supervised auto-encoder to extract robust representations for single sample per person face recognition. The work [23] employs the stacked auto-encoder to pre-train the weights of the deep CNN and improves the performance of the facial emotion recognition. Xu *et al.* [24] use two shallow neural networks to connect two auto-encoders to deal with the age-invariant face recognition and retrieval problems.

The objective functions of these auto-encoder methods are similar to that of ours. Different from these methods, on the one hand, we only employ one layer auto-encoder to learn projections effectively and efficiently. The reason for choosing this strategy is that binary codes show strong robustness to local intra-class variants. On the other hand, there are three challenging constraints imposed on the output of the binary encoder, so it is more difficult to optimize the objective function than those of existing auto-encoders.

C. BINARY HASHING

Binary codes has gained increasing attention due to their high storage efficiency, fast matching speed and strong robustness in computer vision. In recent years, a lot of works on binary hashing have emerged in the area of computer vision [25]-[34]. Locality Sensitive Hashing (LSH) [25] is a classic hashing method in existing hashing strategies, which obtains hashing functions according to random projections. As an extension of the LSH, density sensitive hashing (DSH) [26] exploits the geometric information of the data to generate hashing codes consistent with the distribution of data points. The work [27] introduces an effective sparse regularizer in the objective function to improve the computational efficiency and solve the overfitting problem. Zhu et al. [28] propose a discrete semantic transfer hashing (DSTH) method to exploit the auxiliary contextual patterns to enhance the semantics of hashing codes for image retrieval. He et al. [29] present a bidirectional discrete matrix factorization hashing (BDMFH) model to force the hashing codes to inherit the latent structure of the raw data. Supervised discrete multi-view hashing (SDHM) [30] applies the integrated multi-view feature mapping and latent binary coding to the image data to learn discriminative hashing codes. Robust discrete code modeling (RDCM) [31] uses the $\ell_{2,p}$ norm to restrain the impact of the unreliable discrete codes and noise labels. The collective reconstructive embeddings (CRE) [32] method employs specific hashing modalities for varying modes to reduce the loss of detailed information. Xu et al. [33] develop a discrete cross-modal hashing (DCH) method to exploit the discriminative information of the class labels to learn unified hashing codes for cross-modal retrieval. Han et al. [34] employ the residual net architecture into the hashing model to address the uneven

TABLE 1. Notations and their descriptions in this paper.

Notation	Description
X	$\mathbf{X} = {\{\mathbf{x}_i\}}_{i=1}^N \in \mathbb{R}^{d \times N}$: training set of N PDV samples
R	the radius of sampling patch
Κ	binary code length
D	codebook size
$\mathbf{R}_1, \mathbf{\delta}_1$	$\mathbf{R}_1 \in \mathbb{R}^{K \times d}$, $\mathbf{\delta}_1 \in \mathbb{R}^K$: encoding matrix and bias of encoder
$\mathbf{R}_2, \mathbf{\delta}_2$	$\mathbf{R}_2 \in \mathbb{R}^{d \times K}$, $\mathbf{\delta}_2 \in \mathbb{R}^d$: decoding matrix and bias of decoder
1	column vector with all ones
Ι	identity matrix

distribution problem. However, these available hashing methods are proposed for visual search.

Some learning-based hashing methods are also put forward for face recognition. For example, Lu et al. [5] propose a compact binary feature descriptor (CBFD) method to learn compact binary features for face representation in an unsupervised way. They then develop a simultaneous local binary feature learning and encoding (SLBFLE) [6] approach to jointly learn the projection matrix and the codebook (also known as dictionary) for each face patch to obtain discriminative face representations. The work [7] presents a context-aware local binary feature learning (CA-LBFL) method, an improvement of CBFD, which exploits the contextual information within each binary feature to extract more robust information. However, these algorithms ignore the discrete constraint in the binary hashing optimization, causing the loss of useful information and obtaining inferior hashing functions. In contrast, our proposed LFH-BAE respects the discrete nature of the binary codes and constrains the hashing optimization in the Hamming space to obtain high-quality hashing functions.

III. LOCAL FEATURE HASHING WITH BINARY AUTO-ENCODER

In this section, we first describe how to extract the PDM from a face image patch. Then, we present the formulation and optimization process of LFH-BAE in details. Finally, we show the face representation process based on the LFH-BAE. The notifications of this work are summarized in table 1.

A. THE EXTRACTION PROCESS OF PDM

Since PDV measures the difference between the central pixel and the neighboring ones of the same patch, it can encode important visual patters [5], and we utilize the PDVs to learn the face feature representation. Let $\Omega = {\Omega_1, \Omega_2, \dots, \Omega_n}$ be the training set including *n* original face samples, where Ω_i is the *i*th face image and $1 \le i \le n$. We extract the patch-wise PDVs from each face image and cascade them into a PDM. All *N* PDV samples of the training set are grouped into a large PDM $\mathbf{X} = {\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N}$, where $\mathbf{x}_j \in \mathbb{R}^d$ is the *j* – thPDV and $1 \le j \le N$. Fig. 2 illustrates the extraction approach of the PDM from an original face image. The neighbors of each pixel of the original image are considered within a range



FIGURE 2. An illustration of extracting a PDM from an original face image. Given a face patch with the size of (2R + 1)(2R + 1), where *R* is the radius of the neighborhood size. We first compute the differences between the central patch with that of the neighboring patches, and then cascade all the obtained vectors into a PDM. The column vector of the PDM is denoted as the PDV. In order to make the figure concise, *R* is set to be 1 and there are 8 neighboring patches selected, so the PDM is cascaded by eight 8-dimensinal PDVs.

of (2R + 1)(2R + 1), so the dimensionality of each PDV can be computed as d = (2R + 1)(2R + 1) - 1.

B. FORMULATION

The binary hashing codes can well reflect the visual informa-tion of local image region, and encode important visual patterns (such as edges and lines) of the face image, thereby effectively alleviating the influence of noise factors. In order to obtain binary codes with three properties of effective reconstruction ability, excellent robustness and strong discriminative power, our proposed model should follow the following three restrictions. 1) The objective function should comply with a whole reconstruction criterion to exploit structure information to best approximate the entire face data space. 2) The binary bits should be independent and balanced to reduce redundancy factors while delivering more discrimi-native information. 3) The binary constraint of the problem should be taken into consideration in the hashing optimiza-tion to avoid the accumulated quantization errors to obtain high-quality binary codes.

By integrating the above first two restrictions, we formulate the following objective function

$$\begin{split} \min_{\{\mathbf{R}_{i},\delta_{i}\}_{i=1}^{2}} (1/2) \left\| \mathbf{X} - (\mathbf{R}_{2}(\mathbf{R}_{1}\mathbf{X} + \delta_{1}\mathbf{1}^{\mathrm{T}}) + \delta_{2}\mathbf{1}^{\mathrm{T}}) \right\|_{F}^{2} \\ &+ (\beta/2)(\|\mathbf{R}_{1}\|_{F}^{2} + \|\mathbf{R}_{2}\|_{F}^{2}) \\ \text{s.t.} (\mathbf{R}_{1}\mathbf{X} + \delta_{1}\mathbf{1}^{\mathrm{T}}) \in \{-1, 1\}^{K \times N}, \\ (\mathbf{R}_{1}\mathbf{X} + \delta_{1}\mathbf{1}^{\mathrm{T}})\mathbf{1} = \mathbf{0}, \\ (\mathbf{R}_{1}\mathbf{X} + \delta_{1}\mathbf{1}^{\mathrm{T}})(\mathbf{R}_{1}\mathbf{X} + \delta_{1}\mathbf{1}^{\mathrm{T}})^{\mathrm{T}} = N\mathbf{I}_{K}. \end{split}$$
(1)

where β is a turning parameter to adjust the balance of different terms. The encoding matrix $\mathbf{R}_1 \in \mathbb{R}^{K \times d}$ and bias

 $\delta_1 \in \mathbb{R}^{K \times 1}$ map each PDV $\mathbf{x}_j \in \mathbb{R}^d$ into a low-dimensional binary vector $(\mathbf{R}_1 \mathbf{x}_i + \delta_1) \in \{-1, 1\}^{K \times 1}$, while $\mathbf{R}_2 \in \mathbb{R}^{d \times K}$ and $\delta_2 \in \mathbb{R}^{d \times 1}$ in turn map the binary code back to \mathbb{R}^d in an attempt to reconstruct the PDV. Note that the weight-tying constraint $(\mathbf{R}_1 = (\mathbf{R}_2)^T)$ is not involved in our binary auto-encoder.

The first term of problem (1) ensures that the reconstructtion loss between the input and the output of the binary auto-encoder is minimized, which enables (dis)similar input data to map to (dis)similar binary codes. The second term is added as the regularization on \mathbf{R}_1 and \mathbf{R}_2 to prevent overfitting. The discrete constraint of (1) is to ensure that the output of the encoder are binary codes. The second bit-balanced constraint guarantees that each binary bit has equal chance to occur (almost half of them are encoded to -1, and others are 1), so that the information entropy is maximized. The last bit-independent constraint is to make sure that the learned bits are orthogonal to remove the redundant information.

The three constraints imposed on the hashing codes, which have not been involved in existing auto-encoders, bring new challenges to the optimization of the objective function. In this work, we propose a softened version of the binary auto-encoder to overcome the barriers of combinational complexi-ty of the original problem. Specifically, we introduce a new auxiliary variable **B** and reformulate (1) as

$$\min_{\mathbf{B}, \{\mathbf{R}_{i}, \delta_{i}\}_{i=1}^{2}} (1/2) \left\| \mathbf{X} - (\mathbf{R}_{2}\mathbf{B} + \delta_{2}\mathbf{1}^{\mathrm{T}}) \right\|_{F}^{2} + (\alpha/2) \left\| \mathbf{B} - (\mathbf{R}_{1}\mathbf{X} + \delta_{1}\mathbf{1}^{\mathrm{T}}) \right\|_{F}^{2} + (\beta/2)(\|\mathbf{R}_{1}\|_{F}^{2} + \|\mathbf{R}_{2}\|_{F}^{2}) \text{s.t. } \mathbf{B} \in \{-1, 1\}^{K \times N}, \quad \mathbf{B}\mathbf{1} = \mathbf{0}, \ \mathbf{B}\mathbf{B}^{\mathrm{T}} = N\mathbf{I}_{K}.$$
(2)

where $\alpha(\alpha > 0)$ adjusts the balance between terms. Such a softened version is encouraging due to the absence of the sgn(·) function in the encoder. The second term of (2) minimizes the quantization loss by compelling the output of encoder close to binary hashing values, and meanwhile the first term still guarantees a good reconstruction of the global input space. The solution of (2) will be closer to that of the original problem (1) by setting the parameter α reasonably large enough.

By introducing the auxiliary variable **B**, the awkward original problem can be decomposed into separate tractable subproblems. We solve them via alternating optimization, which will be discussed in details in the next section.

C. OPTIMIZATION

The objective function (2) is a NP-hard problem due to the three constraints imposed on the binary codes. To address this, we utilize the alternating approach based on the ALM [13] to optimize (2), i.e., optimizing the objective function for one variable each time while the others are fixed. We optimize the variables by iteratively solving the following sub-problems.

1) FIX δ_1 , δ_2 AND B, OPTIMIZE \mathbf{R}_1 , \mathbf{R}_2

When δ_1 , δ_2 and **B** are fixed, the objective functions for **R**₁ and **R**₂ can be rewritten as

$$\min_{\mathbf{R}_{1}}(\alpha/2) \left\| \mathbf{B} - (\mathbf{R}_{1}\mathbf{X} + \boldsymbol{\delta}_{1}\mathbf{1}^{\mathrm{T}}) \right\|_{F}^{2} + (\beta/2) \left\| \mathbf{R}_{1} \right\|_{F}^{2} \quad (3)$$

$$\min_{\mathbf{R}_{2}}(1/2) \left\| \mathbf{X} - (\mathbf{R}_{2}\mathbf{B} + \boldsymbol{\delta}_{2}\mathbf{1}^{\mathrm{T}}) \right\|_{F}^{2} + (\beta/2) \left\| \mathbf{R}_{2} \right\|_{F}^{2}$$
(4)

The objective functions (3) and (4) are both regularized least squares problems. By computing the derivatives of the functions for $\mathbf{R_1}$ and $\mathbf{R_2}$, and then set them to 0, respectively, we get the closed form solutions for $\mathbf{R_1}$ and $\mathbf{R_2}$ as follows

$$\mathbf{R}_{\mathbf{1}} = \alpha (\mathbf{B} - \boldsymbol{\delta}_{\mathbf{1}} \mathbf{1}^{\mathrm{T}}) \mathbf{X}^{\mathrm{T}} (\alpha \mathbf{X} \mathbf{X}^{\mathrm{T}} + \beta \mathbf{I})^{-1}$$
(5)

$$\mathbf{R}_2 = (\mathbf{X} - \boldsymbol{\delta}_2 \mathbf{1}^{\mathrm{T}}) \mathbf{B}^{\mathrm{T}} (\mathbf{B} \mathbf{B}^{\mathrm{T}} + \beta \mathbf{I})^{-1}$$
(6)

2) FIX \mathbf{R}_1 , \mathbf{R}_2 AND \mathbf{B}_1 , OPTIMIZE δ_1 , δ_2

When \mathbf{R}_1 , \mathbf{R}_2 and \mathbf{B} are fixed, the objective functions for δ_1 and δ_2 are reformulated as follows

$$\min_{\boldsymbol{\delta}_1} \left\| \mathbf{B} - (\mathbf{R}_1 \mathbf{X} + \boldsymbol{\delta}_1 \mathbf{1}^{\mathrm{T}}) \right\|_F^2$$
(7)

$$\min_{\boldsymbol{\delta}_2} \left\| \mathbf{X} - (\mathbf{R}_2 \mathbf{B} + \boldsymbol{\delta}_2 \mathbf{1}^{\mathrm{T}}) \right\|_F^2$$
(8)

The closed solutions for (7) and (8) with respect to δ_1 and δ_2 can be denoted as

$$\boldsymbol{\delta}_1 = (1/N)(\mathbf{B} - \mathbf{R}_1 \mathbf{X})\mathbf{1}$$
(9)

$$\boldsymbol{\delta}_2 = (1/N)(\mathbf{X} - \mathbf{R}_2 \mathbf{B})\mathbf{1}$$
(10)

3) FIX \mathbf{R}_1 , \mathbf{R}_2 , δ_1 AND δ_2 , OPTIMIZE \mathbf{B}

We first fix \mathbf{R}_1 , δ_1 , \mathbf{R}_2 and δ_2 , and then discard the constant terms, problem (2) for **B** can be rewritten as

$$\min_{\mathbf{B}}(1/2) \|\mathbf{R}_{2}\mathbf{B}\|_{F}^{2} - \operatorname{Tr}(\mathbf{B}^{\mathrm{T}}\mathbf{R}_{2}(\mathbf{X} - \boldsymbol{\delta}_{2}\mathbf{1}^{\mathrm{T}})) - \alpha \operatorname{Tr}(\mathbf{B}^{\mathrm{T}}(\mathbf{R}_{1}\mathbf{X} + \boldsymbol{\delta}_{1}\mathbf{1}^{\mathrm{T}}))$$

s.t. $\mathbf{B} \in \{-1, 1\}^{K \times N}, \quad \mathbf{B}\mathbf{1} = \mathbf{0}, \ \mathbf{B}\mathbf{B}^{\mathrm{T}} = N\mathbf{I}_{K}$ (11)

where $Tr(\cdot)$ denotes the trace norm.

According to the above third restriction (see section III B) of our LFH-BAE, the discrete optimization should be constrained in the Hamming space to reduce the information loss caused by the traditional two-step optimization procedure. In this work, we employ ALM [13] to optimize (11) and adopt the DCC method to explicitly address the discrete constraint with one step. Our main idea is to separate the three challenging constraints by introducing another auxiliary variable **Z** and impose the balanced and independent constraints on it, and meanwhile apply quadratic penalty between **B** and **Z**. Thereby, the objective function (11) is transformed into an augmented Lagrangian (AL) function

$$\min_{\mathbf{B},\mathbf{Z}} (1/2) \|\mathbf{R}_{2}\mathbf{B}\|_{F}^{2} - \operatorname{Tr}(\mathbf{B}^{\mathrm{T}}\mathbf{R}_{2}(\mathbf{X} - \boldsymbol{\delta}_{2}\mathbf{1}^{\mathrm{T}})) - \alpha \operatorname{Tr}(\mathbf{B}^{\mathrm{T}}(\mathbf{R}_{1}\mathbf{X} + \boldsymbol{\delta}_{1}\mathbf{1}^{\mathrm{T}})) + (\gamma/2) \|\mathbf{B} - \mathbf{Z} + \boldsymbol{\delta}_{\gamma}/\gamma\|_{F}^{2}$$

s.t.
$$\mathbf{B} \in \{-1, 1\}^{K \times N}, \quad \mathbf{Z}\mathbf{1} = \mathbf{0}, \ \mathbf{Z}\mathbf{Z}^{\mathrm{T}} = N\mathbf{I}_{K}.$$
 (12)

where Δ_{γ} is defined as the difference between the target variable **B** and auxiliary one **Z**, and γ is defined as the AL penalty parameter.

When \mathbf{R}_1 , δ_1 , \mathbf{R}_2 , δ_1 , \mathbf{Z} and $\boldsymbol{\Delta}_{\gamma}$ are fixed, problem (12) for **B** can be rewritten as

$$\min_{\mathbf{B}}(1/2) \|\mathbf{R}_{2}\mathbf{B}\|_{F}^{2} - \operatorname{Tr}(\mathbf{B}^{\mathrm{T}}\mathbf{R}_{2}(\mathbf{X} - \boldsymbol{\delta}_{2}\mathbf{1}^{\mathrm{T}}))
- \alpha \operatorname{Tr}(\mathbf{B}^{\mathrm{T}}(\mathbf{R}_{1}\mathbf{X} + \boldsymbol{\delta}_{1}\mathbf{1}^{\mathrm{T}}))
- \gamma \operatorname{Tr}(\mathbf{B}^{\mathrm{T}}(\mathbf{Z} + \boldsymbol{\Delta}_{\gamma}/\gamma))
\text{s.t. } \mathbf{B} \in \{-1, 1\}^{K \times N}$$
(13)

We directly optimize **B** in the Hamming space with respect-ting the binary constraint $\mathbf{B} \in \{-1, 1\}^{K \times N}$. Let $\tilde{\mathbf{X}} =$ $\mathbf{X} - \delta_2 \mathbf{1}^T$, $\mathbf{O} = \mathbf{R}_1 \mathbf{X} + \delta_1 \mathbf{1}^T$, $\mathbf{P} = \mathbf{Z} - \mathbf{\Delta}_{\gamma} / \gamma$, with a series of mathema-tical operations, (13) can be equivalently transformed into the following problem

$$\min_{\mathbf{B}} \|\mathbf{R}_{2}\mathbf{B}\|_{F}^{2} - 2\mathrm{Tr}(\mathbf{B}^{\mathrm{T}}\mathbf{R}_{2}^{\mathrm{T}}\tilde{\mathbf{X}} + \alpha\mathbf{B}^{\mathrm{T}}\mathbf{O} + \gamma\mathbf{B}^{\mathrm{T}}\mathbf{P})$$

s.t. $\mathbf{B} \in \{-1, 1\}^{K \times N}$ (14)

For convenience, problem (14) can be simply rewritten as

$$\min_{\mathbf{B}} \|\mathbf{R}_2 \mathbf{B}\|_F^2 - 2\text{Tr}(\mathbf{B}^{\mathsf{T}} \mathbf{Q}) \quad \text{s.t. } \mathbf{B} \in \{-1, 1\}^{K \times N} \quad (15)$$

where $\mathbf{Q} = \mathbf{R}_{2}^{\mathrm{T}} \tilde{\mathbf{X}} + \alpha \mathbf{O} + \gamma \mathbf{P}$. Similar to the optimization idea of [35] and [36], we adopt the DCC method to learn the binary codes **B** bit-by-bit. Let \mathbf{b}^{T} , \mathbf{q}^{T} and \mathbf{r}^{T} be the *i*th row of **B**, **Q** and $\mathbf{R}_2^{\mathrm{T}}$ respectively, where $i = 1, 2, \cdots, K$, we utilize **B**, **Q** and \bar{R} to denote the remaining matrix of **B**, **Q** and **R**₂^T, respectively. From the derivation in [37], problem (15) can be converted into the following form

$$\min_{\mathbf{b}} (\mathbf{r}^{\mathrm{T}} \bar{\mathbf{B}}^{\mathrm{T}} \bar{\mathbf{B}} - \mathbf{q}^{\mathrm{T}}) \mathbf{b} \quad \text{s.t. } \mathbf{b} \in \{-1, 1\}^{N}$$
(16)

The optimal solution for **b** can be denoted as

$$\mathbf{b} = \operatorname{sgn}(\mathbf{q} - \bar{\boldsymbol{B}}^{\mathrm{T}} \bar{\boldsymbol{R}}^{\mathrm{T}} \mathbf{r})$$
(17)

4) FIX \mathbf{R}_1 , \mathbf{R}_2 , δ_1 , δ_2 , B AND Δ_{γ} , OPTIMIZE Z

When \mathbf{R}_1 , δ_1 , \mathbf{R}_2 , δ_2 , \mathbf{B} and $\boldsymbol{\Delta}_{\gamma}$ are fixed, problem (12) for \mathbf{Z} can be reformulated as

$$\min_{\mathbf{Z}} \|\mathbf{B} - \mathbf{Z} + \mathbf{\Delta}_{\gamma} / \gamma \|_{F}^{2} \quad \text{s.t. } \mathbf{Z}\mathbf{1} = \mathbf{0}, \quad \mathbf{Z}\mathbf{Z}^{\mathrm{T}} = N\mathbf{I}_{K}$$
(18)

The above objective function is simplified to the following maximization problem

$$\max_{\mathbf{Z}} \operatorname{Tr}(\mathbf{Z}^{\mathrm{T}}\mathbf{S}) \quad \text{s.t. } \mathbf{Z}\mathbf{1} = \mathbf{0}, \mathbf{Z}\mathbf{Z}^{\mathrm{T}} = N\mathbf{I}_{K}$$
(19)

where $\mathbf{S} = \mathbf{B} + \mathbf{\Delta}_{\gamma} / \gamma$. By performing singular value decomposition (SVD), S can be formulated as $S = U\Lambda V^T =$ $\sum_{k=1}^{r} \sigma_k \mathbf{u}_k \mathbf{v}_k^{\mathrm{T}}$, where r(r < K) is the rank of **S**, $\boldsymbol{\Lambda}$ is a rectangular diagonal matrix and its diagonal elements $\sigma_1, \sigma_2, \cdots, \sigma_r$ are non-negative singular values of S. U = $[\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r]$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r]$ are left- and right-singular vectors of S, respectively. Based on the above an- alyses, it is easy to derive that $\mathbf{V}^{\mathrm{T}}\mathbf{Z}^{\mathrm{T}}\mathbf{U}\mathbf{U}^{\mathrm{T}}\mathbf{Z}\mathbf{V} = N\mathbf{I}$.

Input: training data: **X**, binary code length: *K*, iteration number: T, parameters: α , β , γ , η .

Output: encoding matrix: \mathbf{R}_1 , encoding bias δ_1 , decoding matrix: \mathbf{R}_2 , decoding bias: δ_2 .

1: Initialization: Initialize B using ITQ, initialize Z randomly, initialize $\delta_1 = 0$, $\delta_2 = 0$ and $\Delta_r = 0$. 2: repeat

 $(\mathbf{R}_1, \mathbf{R}_2) - \mathbf{step}$: Optimize \mathbf{R}_1 and \mathbf{R}_2 according 3: to (5) and (6), respectively.

4:
$$(\delta_1, \delta_2) - \text{step}$$
: Optimize δ_1 and δ_2 according to (9) and (10), respectively.

repeat 5:

- $\mathbf{B} \mathbf{step}$: Optimize \mathbf{B} utilizing DCC method 6: with (17) by iteratively learning $\{\mathbf{b}_i\}_{i=1}^K$ bit-by-bit.
- $\mathbf{Z} \mathbf{step}$: Optimize \mathbf{Z} according to (21). 7:
- $(\mathbf{\Delta}_r, \gamma)$ **step** : Update $\mathbf{\Delta}_r$ and γ according 8: to (22). until convergences

9:

- 10: until convergences
- 11: **Return:** \mathbf{R}_1 , $\boldsymbol{\delta}_1$, \mathbf{R}_2 and $\boldsymbol{\delta}_2$.

On the basis of the Theorem 3 in [38] and [39], the optimal value of Z can be obtained if and only if $V^{T}Z^{T}U =$ $diag(\sqrt{N})$. The solution of **Z** can be written as

$$\mathbf{Z} = \sqrt{N} \mathbf{U} \mathbf{V}^{\mathrm{T}}$$
(20)

In addition, in order to meet the balanced constraint Z1 = 0, the Gram-Schmidt process is used to construct the matrices $\bar{U} \in \mathbb{R}^{K \times (K-r)}$ and $\bar{V} \in \mathbb{R}^{N \times (K-r)}$, which satisfy the following four conditions: $\bar{\boldsymbol{U}}^{\mathrm{T}}\bar{\boldsymbol{U}} = \mathbf{I}_{K-r}$, $[\mathbf{U}, \mathbf{1}]^{\mathrm{T}}\bar{\boldsymbol{U}} = \mathbf{0}$, $\bar{\boldsymbol{V}}^{\mathrm{T}}\bar{\boldsymbol{V}} = \mathbf{I}_{K-r}$, $\mathbf{V}\bar{\mathbf{V}}^{\mathrm{T}} = \mathbf{0}$. So the ultimate optimal solution for **Z** is

$$\mathbf{Z} = \sqrt{N} [\mathbf{U}, \bar{\boldsymbol{U}}] [\mathbf{V}, \bar{\boldsymbol{V}}]^{\mathrm{T}}$$
(21)

5) FIX \mathbf{R}_1 , \mathbf{R}_2 , δ_1 , δ_2 , B AND Z, OPTIMIZE Δ_{γ}

When **B** and **Z** are fixed, Δ_{γ} and γ are updated according to the following rules

$$\mathbf{\Delta}_{\gamma} = \mathbf{\Delta}_{\gamma} + \gamma (\mathbf{B} - \mathbf{Z}), \quad \gamma = \eta \gamma \tag{22}$$

The learning rate $\eta(\eta > 0)$ controls the convergence of the objective function.

We repeat the above steps until the objective function converges. The detailed procedure of our proposed method is summarized in Algorithm 1. After obtaining $(\mathbf{R}_1, \boldsymbol{\delta}_1)$, the hash function can be denoted as $\mathbf{H}(\mathbf{X}) = sgn(\mathbf{R}_1\mathbf{X} + \boldsymbol{\delta}_1\mathbf{1}^{\mathrm{T}}).$

D. FACE REPRESENTATION BASED ON LFH-BAE

Having obtained the hashing function equipped with the encoding matrix \mathbf{R}_1 and the encoding bias δ_1 , each PDV is first projected into a low-dimensional hashing vector. Since the traditional K-means is an effective and simple



FIGURE 3. The process of face representation based on the LFH-BAE. Firstly, each training face image is divided into several non-overlapping regions, which are used to learn the encoding matrix R_1 , the encoding bias δ_1 and the codebook D. Then, they are applied on corresponding regions to extract histogram features, which are cascaded into a longer feature vector for face representation. Finally, the classifier with cosine similarity is used for face matching.



FIGURE 4. The aligned and cropped face samples with 128 \times 128 pixels from the FERET.

unsupervised clustering method, we apply it to the training set to construct the codebook **D**. Then, we pool each hashing code as a bin with the codebook, and use histogram feature to represent the hashing codes of the same face region. At last, the histogram features of all the regions are cascaded into a high-dimensional feature, which is considered as the final representation for the whole face image. Fig. 3 illustrates the approach of face representation based on LFH-BAE.

IV. EXPERIMENTS

We conduct extensive experiments on four public face datasets to evaluate the performance of our LFH-BAE model. To be more specific, we first use two constrained datasets (FERET and CAS-PEAL-R1) to show the effectiveness of our LFH-BAE model. Then, we employ two other unconstrained datasets (LFW and PaSC) to demonstrate the robustness and discriminative power of the LFH-BAE model. At last, we perform cross-dataset evaluation to estimate the generalization ability of our proposed model.

A. PERFORMANCE ON FERET

The FERET is a widely used large-scale public database, which contains 13539 human faces of 1565 subjects, varying in age, gender and race. Following the standard FERET evolution protocol [40], we conduct experiments on six disjoint sets: one *training* set, one *gallery* set (*fa*), and four probe sets (*fb*, *fc*, *dup1* and *dup2*). All images are aligned according to the provided eye coordinates and cropped into 128×128 pixels. Some cropped images from the FERET are shown in Fig. 4.

TABLE 2. The rank-one and average (avr) recognition rates (%) of LFH-BAE on FERET versus various values of α , β and γ .

Pa	arameter	'S		c	1 1	1 0	
α	β	γ	fb	tc	dupl	dup2	Ave.
0.1	0.1	1	99.8	100.0	94.5	93.6	97.0
0.01	0.1	1	99.5	100.0	95.3	94.0	97.2
0.01	0.01	1	99.5	100.0	94.7	93.2	96.9
0.001	0.1	1	99.6	100.0	94.9	93.6	97.0
0.001	0.01	1	99.6	100.0	94.5	92.3	96.6
0.01	0.1	0.1	99.5	100.0	95.0	94.9	97.3
0.01	0.1	10	99.6	100.0	94.3	92.7	96.7
0.01	0.1	100	99.8	100.0	95.8	96.2	97.9
0.01	0.1	1000	99.7	100.0	95.4	94.4	97.4
0.001	0.01	100	99.5	100.0	94.7	93.6	97.0

In our experiments, we divide each face image into 8×8 non-overlapping regions, and evaluate our model with two different neighborhood radius sizes (R = 3 and R = 5), yielding 48- and 120-dimensional PDV for each pixel, respectively. The *training* set is utilized for feature learning, and the learned hashing function and codebook are performed on the other five sets for feature extraction. Whitened PCA (WPCA) is applied on the high-dimensional histogram representation to generate a compressed 1195-dimentional feature vector to reduce redundancy. The compressed feature vector is fed to the nearest neighbor classifier for face recognition.

1) PARAMETER ANALYSIS

In this section, we test the impact of different parameters in our LFH-BAE model on the FERET. We first study the effects of one parameter while the others are fixed. Then, we apply the determined parameters to the subsequent experiments.

We fix R = 5 to explore the influence of the coefficients α , β , γ on recognition rates. The learning rate η , code length *K* and codebook size are set to 1.5, 25 and 600, respectively, and the impact of them will be studied in the following experiments. Table 2 shows the recognitions rates on the FERET for different empirical values. It can be observed that the best performance is obtained when α , β , γ are chosen to be 0.01, 0.1 and 100, respectively.

Then, the influence of the code length is examined when the codebook size is fixed at 600. The code length should be set at 25 for the highest recognition rate, as depicted in Fig. 5(a). With the fixed parameters in the above experiments, varying dictionary sizes are also studied and Fig. 5(b) shows that the best results are obtained when the codebook size is 600.

Next, we study the effects of the learning rate η in our LFH-BAE model. The other parameters are the same as those determined in the above experiments. The values of learning rate are empirically varied from 1 to 9. Fig. 6 shows the encoding-decoding time and the average recognition rates of our LFH-BAE versus different values of η . We can obseve that the encoding-decoding time decreases non-linearly as the learning rate increases, and the average recognition rate reaches a maximum at $\eta = 1.5$. This demenstrates that a small learning rate may result in a poor convergence of the multiplier iteration, while a large learning rate may force



FIGURE 5. The rank-one and average recognition rates (%) of LFH-BAE on FERET versus different values of (a) code length and (b) dictionary size.



FIGURE 6. Encoding-decoding time and average rates of LFH-BAE on four probe sets of FERET versus different values of learning rate.

much ill-conditioning upon the objective function too early. The key is to find a right balance point where the LFH-BAE not only ensures fast encoding and decoding but also achieves high recognition rate. Thus, η is set to 1.5 as a moderate rate in the following experiments.

Lastly, we investigate the impact of iterations T on LFH-BAE. The property of the proposed alternating optimization process and ALM optimization theory make Algorithm 1 converge, and enable that the objective function value non-incremental and bounded. Fig. 7 shows the convergence of the objective function of our LFH-BAE. We observe that the objective function converges in 3 iterations.

2) COMPARISON WITH EXISTING HASHING ALGRITHMS

Although most existing hashing algorithms are presented for visual retrieval [25], [26], [41]–[44], they can still be implemented and remain effective in face recognition tasks. In this section, we compare our LFH-BAE with the following seven representative hashing methods: One Layer Anchor Graphs Hashing (AGH1) [41], Two Layer Anchor



FIGURE 7. Objective function value of LFH-BAE versus different iteration numbers on FERET.

Graphs Hashing (AGH2) [41], Locality-Sensitive Hashing (LSH) [25], Spherical Hashing (SPH) [42], PCA-iterative Quantization (PCA-ITQ) [43], Density Sensitive Hashing (DSH) [26], and Hashing with Angular Reconstructive Embeddings (ARE) [44]. The first six algorithms are all conducted using the source codes and determined parameters of the original authors, except that ARE is implemented by ourselves. To ensure a fair comparison, we only replace the objective function of LFH-BAE (i.e., problem (1)) with those of other hashing methods, while the other steps remain the same. In this experiment, the neighborhood radius size is 5 and other parameters of those hashing algorithms are consistent with LFH-BAE. WPCA is employed to project the high-dimensional feature into a 1195-dimensional representation vector. Table 3 shows the rank-one recognition rates of different algorithms on the FERET.

It can be obviously seen that our LFH-BAE model outperforms those existing hashing algorithms. On one hand, the above first six hashing methods ignore the discrete

TABLE 3. The rank-one and average recognition rates (%) of our LFH-BAE and other existing binary hashing methods on the FERET.

Algorithm	fb	fc	dup1	dup2	Ave.
AGH1 [41]	98.7	99.0	84.2	83.3	91.3
AGH2 [41]	99.1	98.5	84.2	83.8	91.4
LSH [25]	99.5	100.0	93.6	91.0	96.0
SPH [42]	99.7	100.0	95.2	93.6	97.1
PCA-ITQ [43]	99.3	100.0	94.5	94.0	97.0
DSH [26]	99.4	100.0	92.7	92.9	96.3
ARE [45]	99.7	100.0	95.6	94.9	97.6
LFH-BAE (R=5)	99.8	100.0	95.8	96.2	97.9

constraint in the optimization process, thereby inducing the accumulated quantization errors. On the other hand, ARE is dedicated to preserving cosine similarity, it may lack the ability of extracting structure information from the input face data. The reconstruction term in LFH-BAE helps to exploit the inter-class structure factors that can approximate the entire image data space, and the direct discrete optimization can reduce the loss of useful information. Therefore, our proposed hashing model achieves the best performance on the face recognition task.

3) COMPARISON WITH STATE-OF-THE-ART FACE ALGORITHMS

In this section, we evaluate the performance of our LFH-BAE model on the FERET with the standard evaluation protocol. We compare our proposed LFH-BAE model with the hand-crafted local descriptors (such as LGBP, DT-LBP, LGFV//LN//SNP, FCPRF, LDENP) and learning-based descriptors (such as shallow descriptors: DFD, CBFD, SLBFLE, SPMBD, and deep descriptors: 2-FFC_{Gabor-PCA}, BISF+K-DAN, PCANet+_{SF}). Table 4 lists the rank-one recognition rates of our LFH-BAE and other approaches on the FERET.

We can observe that our LFH-BAE model obtains much competitive performance on the four probe sets of the FERET. As a tangible description, our LFH-BAE model obtains much higher accuracies on all probe sets than the hand-crafted descriptors, because it doesn't require prior knowledge and can adaptively extract more discriminative information from the original data. Compared with the learning-based hashing descriptors, LFH-BAE achieves the highest average rate, which further proves the direct discrete optimization can facilitate binary features retaining more useful information. Moreover, the recognition performance of LFH-BAE is also better than the real-valued descriptors (such as shallow descriptors LGFV//LN//SNP, FCPRF and deep descriptors 2-FFC_{Gabor-PCA}, PCANet+_{SF}), which indicates that the quantized binary features are more robust to intra-class variations in face images. It is worth noting that LFH-BAE obtains excellent performance when two-scale PDVs with different radii are cascaded.

4) RECONSTRUCTION ERROR

In this section, we compare the reconstruction ability of the learning-based hashing methods, including PCA-ITQ,
 TABLE 4.
 The rank-one and average recognition rates (%) of our LFH-BAE and state-of-the-art algorithms on the FERET.

Algorithm	fb	fc	dup1	dup2	Ave.
$LBP [1]^{a}$	97.0	79.0	66.0	64.0	76.5
LGBP [9] ^a	94.0	97.0	68.0	53.0	78.0
DT-LBP [45] ^a	99.0	100.0	84.0	80.0	90.8
LDP [8] ^a	94.0	83.0	62.0	53.0	73.0
DLBP [46] ^a	99.0	99.0	86.0	85.0	92.3
I-LQP [12] ^a	99.2	69.6	65.8	48.3	70.7
LGFV//LN//SNP [47] ^a	99.2	100.0	91.7	89.3	95.0
FCPRF [48] ^a	99.7	98.5	89.1	85.5	93.2
LDENP [49] ^a	97.8	90.0	83.0	86.8	89.4
WCBP+FLD+CD [50] ^a	99.0	99.5	94.7	90.2	96.9
CFV [51] ^a	99.0	100.0	95.0	95.3	97.3
LBPNet [52] ^a	99.6	100.0	94.2	93.6	96.9
DFD [3] ^a	99.4	100.0	91.8	92.3	95.9
CBFD [5] ^a	99.8	100.0	93.5	93.2	96.6
CA-LBMFL [7] ^a	99.8	100.0	95.3	95.3	97.6
SLBFLE [6] ^a	99.9	100.0	95.2	92.7	97.0
SPMBD [53] ^a	99.7	100.0	95.7	97.0	98.1
2-FFC _{Gabor-PCA} [54] ^a	99.5	100.0	96.3	94.9	97.7
BISF + K-DAN [55] ^a	99.6	100.0	93.5	91.0	96.0
PCANet+ _{SF} [56] ^a	99.4	100.0	96.4	95.7	97.9
LFH-BAE (R=3)	99.8	100.0	95.3	95.3	97.6
LFH-BAE (R=5)	99.8	100.0	95.8	96.2	97.9
LFH-BAE (R=3+5)	99.8	100.0	96.4	96.6	98.2

^aThe results of other approaches are cited from the original papers.

 TABLE 5. The mean reconstruction error and average recognition rates

 (%) of our LFH-BAE and other binary learning algorithms on four probe sets of FERET.

Algorithm	Reconstruction Error	Ave.
CBFD [5]	1.58×10^{7}	96.7
PCA-ITQ [43]	1.44×10^{7}	97.1
CA-LBFL [7]	1.43×10^{7}	97.3
LFH-FBAE (R=5)	1.15×10^{7}	97.9

CBFD, CA-LBFL and our LFH-BAE. Each face image of the FERET is divided into $8 \times 8 = 64$ non-overlapping regions, and the neighborhood radius is set to 5. The parameters of the other three hashing methods are consistent with those of our LFH-BAE, so that all methods can be compared fairly. We compute the mean reconstruction errors on the FERET.

For PCA-ITQ, CBFD and CA-LBFL, the decoder is unavailable in their implementation, but it is taken into account in the design of LFH-BAE and can be obtained in the learning process. Therefore, following [35], we compute the optimal linear decoders for the other three methods. Specifically, we use the learned binary codes of the training data to compute the optimal linear decoder (\mathbf{R}_2 , δ_2). Let **Y** be the binary codes of the testing data, the reconstructed testing data can be calculated as $\hat{\mathbf{X}} = \mathbf{R}_2\mathbf{Y} + \delta_2\mathbf{1}^T$.

Table 5 tabulates the mean reconstruction errors and average recognition rates of different hashing methods. It indicates that the reconstruction abilities of PCA-ITQ and CA-LBFL are comparable, and our LFH-BAE is optimum in terms of reconstruction ability and recognition performance.

This experimental configuration confirms that our LFH-BAE can extract the structure information inherent in the input face data, which can improve the discriminative power of the feature representations.

 TABLE 6. The rank-one and average recognition rates (%) of various learning strategies on the FERET.

Algorithm	fb	fc	dup1	dup2	Ave.
LFTH-BAE	99.7	100.0	95.3	94.9	97.4
LFH-BE	99.7	100.0	95.6	95.3	97.7
LFL-RAE	99.8	100.0	93.1	92.3	96.3
LFH-BAE	99.8	100.0	95.8	96.2	97.9

5) VARIOUS LEARNING STRATEGIES

In order to explore the contributions of discrete optimization and binary auto-encoder of the proposed LFH-BAE model, we develop three learning strategies to study the importance of these components:

- 1) TLFH-BAE: Traditional local feature hashing with binary auto-encoder.
- 2) LFH-BE: Local feature hashing with only binary encoder.
- 3) LFL-RAE: Local feature learning with real-valued auto-encoder.

The parameter settings in this part are consistent with those in Part 4). TLFH-BAE uses the traditional two-step "relaxing+thresholding" procedure to learn binary codes. LFH-BE conduct the iterative optimization without considering the reconstruction term in problem (2) (i.e., disregarding the first term). LFL-RAE learns real-valued features by discarding the discrete constraint. The remaining stages of these three learning strategies are the same as those of LFH-BAE, and each face sample is projected into a 1195-dimensinal representation vector with WPCA. The rank-one recognition rates of the three strategies and our LFH-BAE on the FERET are shown in Table 6. We can summarize the following four conclusions:

- 1) Discrete optimization can eliminate the accumulated quantization errors, thereby reducing the loss of useful information.
- Binary auto-encoder can exploit structure factors from the face image and extract excellent discriminative features.
- 3) Compared with the real-valued codes, binary codes show stronger robustness to intra-class changes of the face image.
- When these three components are combined together, significantly high recognition performance can be obtained.

6) COMPUTATIONAL TIME

At last, we conduct experiment on the FERET to estimate the computational cost of the real-valued descriptors and LBP-like features. The parameter settings in this experiment are also the same as those in above Part 4), and they are applied to other methods to ensure a fair comparison. For the CBFD, CA-LBFL and our LFH-BAE, a 600-dimentional feature vector is extracted from each region, so the dimension of the cascaded feature vector for each image is 38400 ($600 \times$ 64 = 38400). Note that all methods are implemented with the matlab software and carried out on a PC with a 2.70 GHz

TABLE 7. The average recognition rates (%), feature dimension and computational time (ms) comparison of the proposed LFH-BAE and other LBP-like feature extraction methods.

Algorithm	Ave.	Feature Dimension	Time(ms)
FCPRF [48]	93.6	35400	299.3
2-FFC _{Gabor-PCA} [54]	97.5	262144	129.7
LBP [1]	82.6	3776	41.6
LGBP [9]	93.5	655360	1235.2
DFD [3]	94.9	65536	680.4
CBFD [5]	96.7	38400	132.1
CA-LBFL [7]	97.3	38400	140.2
LFH-BAE (R=5)	97.9	38400	139.3



FIGURE 8. The aligned and cropped face samples with 150 \times 130 pixels from the CAS-PEAL-R1.

and a 24G RAM configurations. Table 7 shows the feature dimensions and computational time of different methods.

We can observe that feature descriptors with high dimensions can improve the recognition performance to a certain extent, but the computational cost increases accordingly. Compared with the real-valued 2-FFC_{Gabor-PCA}, our LFH-BAE is slightly inferior in the computational cost, but it is excellent in terms of storage efficiency and recognition performance. Moreover, although there are encoding biases in our LFH-BAE, it's computational cost is almost the same as CBFD and CA-LBFL, but lower than that of DFD. This is because only one scale PDV for each pixel is extracted in LFH-BAE, while multiple scale PDVs for each pixel are extracted in DFD. Therefore, the computational cost of our LFH-BAE is less than that of DFD.

B. PERFORMANCE ON CAS-PEAL-R1

The CAS-PEAL-R1 dataset contains more than 9000 face images from 1040 individuals with various variabilities, such as accessory, pose, lighting and expression. Following the standard evaluation protocol [57], we conduct the experiment on five subsets for face recognition, i.e., *training*, *gallery*, *accessory*, *lighting* and *expression*. The *training* set contains 1200 face images of 300 individuals (4 images per subject). The *gallery* set consists of 1040 subjects with 1040 face images, which are captured under normal conditions. The probe sets are collections of *accessory*, *expression* and *lighting*, consisting of 2285, 2243, 1570 images, respectively. All face images of the above five sets are aligned and cropped into 150×130 pixels according to the provided eye coordinates. Fig. 8 shows the aligned and cropped examples of the CAS-PEAL-R1.

TABLE 8. The ran	k-one recognition rates	(%) of our LFH-BAE and
state-of-the-art a	Igorithms on the CAS-PE	AL-R1.

Algorithm	expression	accessory	lighting
LGBP [9] ^a	95.0	87.0	51.0
LLGP [58] ^a	96.0	90.0	52.0
DT-LBP [45] ^a	98.0	92.0	41.0
DLBP [46] ^a	99.0	92.0	41.0
WCBP+FLD+CD [50] ^a	98.0	93.9	69.0
DFD [3] ^a	99.6	96.9	63.9
CBFD [5] ^a	99.7	97.2	67.4
IFL [59] ^a	99.3	96.5	64.3
JFL [59] ^a	99.7	97.2	67.4
SPMBD [53] ^a	99.7	97.3	77.0
FCPRF [48]	99.2	96.2	55.6
LFH-BAE (R=3)	99.5	97.4	80.1
LFH-BAE (R=5)	99.8	97.4	81.8
LFH-BAE (R=3+5)	99.8	97.6	82.3

^aThe results of other approaches are cited from the original papers.

Firstly, we learn the LFH-BAE on the *training* set and then use the learned projections to extract features from the *gallery* and the other three probe sets. Parameters of the LFH-BAE model are kept the same as those used on the FERET dataset. Finally, WPCA is used to reduce the high dimension to 1039 and the cosine metric¹ is applied for face matching. Table 8 tabulates the rank-one recognition rates of our LFH-BAE and other different face descriptors on the CAS-PEAL-R1. Compared with state-of-the-art methods, our proposed LFH-BAE model obtains the best recognition rates on all probe sets. In particular, the proposed LFH-BAE obtains higher accuracy than the previous best SPMBD with the gain of 5.3% on the *lighting* set. It is demonstrated that our LFH-BAE shows strong robustness to expression, occlusion, especially in the face of lighting.

C. PERFORMANCE ON LFW

The LFW dataset [60] contains more than 13,000 face images of 5,749 subjects and it is designed to study face recognition of unconstrained scenarios. As all face images are collected in the wild conditions, they suffer from large intra-class variations, such as varying poses, expressions, illuminations and backgrounds. In this experiment, we use the unsupervised setting [61] to evaluate the verification performance of LFH-BAE on LFW. According to the standard evolution protocol [61], there are 3000 matched pairs and 3000 mismatched pairs in the "View 2" dataset. It is divided into 10 folds, and each fold consists of 300 matched image pairs and 300 mismatched pairs, which are randomly selected from the original dataset. We use a conventional 2D affine transformation to align each face image and crop it into a region with the size of 128×128 pixels. The aligned and cropped face samples are shown in Fig. 9.



FIGURE 9. The aligned and cropped face samples wit 128 \times 128 pixels from the LFW.

TABLE 9. The AUC (%) comparisons with other approaches on LFW under the unsupervised setting.

Algorithm	AUC	Algorithm	AUC
SIFT [2] ^a	54.07	SLBFLE [6] ^a	92.00
LBP [1] ^a	75.47	SPMBD [53] ^a	91.16
LARK [62] ^a	78.30	MS-LZM [63] ^a	95.15
LHS [64] ^a	81.07	PCLWT [65] ^a	96.88
LQP [12] ^a	87.00	LBPNet [52] ^a	94.04
MRF-MLBP [66] ^a	89.94	2-FFC _{Gabor-PCA} [54] ^a	95.54
DFD [3] ^a	83.70	PCANet+ _{SF} [56] ^a	94.36
JFL [59] ^a	91.03	LFH-BAE (R=3)	89.33
CBFD [5] ^a	88.65	LFH-BAE (R=5)	90.72
CA-LBFL [7] ^a	91.66	LFH-BAE (R=3+5)	91.95

^aThe results of other approaches are cited from the original papers.

In this experiment, parameters of the proposed LFH-BAE are consistent with those used on the above FERET dataset. We perform feature learning on the training set, and apply WPCA to the high-dimensional feature to obtain a 700-dimensional representation vector. The nearest neighbor classifier with cosine similarity is used for face verification. Table 9 and Fig. 10 show the excellent average AUC and ROC curves of the proposed LFH-BAE and state-of-the-art strategies for the unsupervised setting, respectively. Note that we only plot the ROC curves for the algorithms that are released on the LFW website.²

It can be observed that our LFH-BAE model achieves competitive performance with the existing learning-based hashing methods such as CA-LBFL, SLBFLE and SPMBD, and outperforms hand-crafted feature descriptors such as LHS, LARK and MRF-MLB. This again demonstrated that LFH-BAE can not only exploit structure factors from face images data-adaptively, but also eliminate the accumulated quantization errors, therefore more discriminative features are extracted for face verification.

Although the deep LBPNet [52] show better performance than our LFH-BAE, it requires prior knowledge to design a deep funneled network containing multiple types LBP filters and PCA filters. The deep network PCANet+_{SF} [56] utilizes various filters to extract various PCA-based features, increasing the computational cost. The methods in [63], [66] and the deep real-values descriptor 2-FFC_{Gabor-PCA} achieve excellent results on the LFW dataset with the unsupervised setting, they all use the high-fidelity pose and expression

¹As described in section III D, the statistical histogram feature is used to represent the learned hashing codes, the cosine similarity of two faces can be denoted as $sim(i, j) = \mathbf{y}_i^T \cdot \mathbf{y}_j / \|\mathbf{y}_i\|_2 \cdot \|\mathbf{y}_j\|_2$, where \mathbf{y}_i and \mathbf{y}_j are the feature representations of the *i* – th and *j* – th face images, respectively.

²The results of the other algorithms are available at http://viswww.cs.umass.edu/lfw/results.html#Unsupervised



FIGURE 10. The ROC curves of different approaches on LFW under the unsupervised setting.

normalization (HPEN) [68] to process the face images and extract features at multiple scales, which is time or storage consuming in the recognition stage. On the contrary, we only use a simple 2D affine transformation to process the face images, and use a simple learning-based hashing strategy to extract only two-scale features to achieve encouraging performance.

D. PERFORMANCE ON PASC

The combination of point-and-shoot camera technology and social network technology increases the difficulty of face recognition tasks in real scenarios. In this section, we employ a recently released challenging face dataset, namely PaSC, to estimate the robustness of our LFH-BAE model. The still image portion of the PaSC includes 9376 images from 293 subjects. The balance of alternative sensors, distance to cameras, varying views and different positions are taken into account in image collection. This results in rare complications in controlled and less controlled scenarios, including poor focusing, over and under exposure, blur, and poor lighting. As the constraints on imaging conditions are relaxed, the challenges of face recognition increase dramatically. The standard evaluation protocol [69] for the still portion of PaSC is to compare all images of the query set to those of target set. Both target set and query set are allocated the same number of images (4688 images). We align each image with the provided eye coordinates³ and crop it into the size of 128×128 pixels. Fig. 11 shows some aligned and cropped sample images of the PaSC.

We utilize the target set to train the proposed LFH-BAE model, and extract feature representations from the target set and query set. The parameters are kept as the same as those used on the FERET dataset and each face image sample is projected into a 500-dimensional representation space

TABLE 10. The verifica	ation rates (%) at FAR	= 1% of different algorith
on the all and frontal	settings of the PaSC.	-

Algorithm	Frontal	All
LBP [1]	22.3	10.9
LRPCA [68] ^a	19.0	10.0
CohortLDA [68] ^a	22.0	8.0
PittPatt [68] ^a	55.0	-
LPQ [69] ^a	23.1	13.2
$BSIF[70]^{a}$	24.9	14.3
CBFD [5]	34.7	18.9
CA-LBFL [7]	45.4	25.7
SPMBD [53] ^a	47.5	26.9
LFH-BAE (R=3)	38.8	21.2
LFH-BAE (R=5)	48.7	27.5
LFH-BAE (R=3+5)	50.7	29.0

^aThe results of other approaches are cited from the original paper.

with WPCA. By matching images in the target set to those in the query set, we compute the cosine distance for all pairs and obtain a 4688×4688 similarity matrix. The resulting similarity matrix is the basis for plotting the ROC curves for the frontal and all (frontal and non-frontal) images. In addition to comparing with the algorithms published on the PaSC website (one commercial PittPatt algorithm and two baseline real-valued descriptors: LRPCA, CohortLDA), we also compare the LFH-BAE model with other existing popular methods, such as LBP, LPQ, BSIF, CBFD and CA-LBFL. Table 10 and Fig. 12 show the verification rates at 1.0% FAR and ROC curves for the frontal images and all images, respectively. In order to provide a concise figure, we only plot the ROC curves of some representative algorithms.

We can see that the proposed LFH-BAE model is significantly superior to the existing popular methods and baseline algorithms, except for the commercial PittPatt. The minimal gains of verification rates for all images and frontal images are 2.1% and 3.2%, respectively. Moreover, feature representations that learned from two-scale cascaded PDVs can further improve the verification performance of LFH-BAE.

The results demonstrate that, even for the point-and-shoot data, our LFH-BAE model is capable of extracting robust and discriminative feature representations. Binary auto-encoder not only promotes the learned binary features to inherit salient factors inherent in the raw data, but also make them show strong robustness to intra-class variations. Moreover, the ALM-based optimization strategy avoids the accumulated quantization errors and reduces the loss of discriminative information. Therefore, our LFH-BAE model can be employed in the real-world scenarios to handle the challenging face recognition tasks.

E. PERFORMANCE ON CROSS-DATASET

In practical applications, the appearances of the training set and testing set are apparently different due to the varying scenarios. In this section, we conduct a cross-dataset experiments to further investigate the generalization capability of our proposed strategy. Firstly, we utilize the constraint

³The eye coordinates of all still images are available at http://www.cs.colostate.edu/~ vision/pasc/index.php



(a) Frontal images

(b) non-Frontal images





FIGURE 12. The ROC curves of different algorithms on the (a) all setting and (b) frontal setting of the PaSC.

TABLE 11. Cross-dataset evaluation of our proposed LFE-SBAE on different training sets.

Training Set	LFW	PaSC	
	AUC	All	Frontal
FERET	89.72	26.2	46.3
LFW	91.95	27.8	49.3
PaSC	90.13	29.0	50.7

FERET dataset, the unconstraint LFW⁴ and PaSC datasets to train our LFH-BAE model to obtain the encoding matrix \mathbf{R}_1 , the encoding bias δ_1 and the codebook **D**. Then, we evaluate the performance of the learned LFH-BAE models on LFW⁵ and PaSC. In these experiments, we only use the two-scale PDVs and the other parameters are consistent with the above sections.

Table 11 shows the AUC of LFW and verification rates (All and Frontal) at 1.0% FAR of PaSC for the cross- and same-dataset evaluations. We can observe that our LFH-BAE

 5 In the cross-dataset evaluation, the evaluation protocol setting of LFW is consistent with that of part C in this section.

achieves excellent performance on the same-dataset evaluation, but in the case of cross-dataset evaluation, its performance decreases slightly. This is because the images of the training set and testing set are captured under different environments. Nevertheless, the performance of the cross-dataset evaluation is still comparable with that of state-of-the-art strategies based on the same-dataset evalu-ation (see the above sections C and D), which demonstrates the effectiveness of the LFH-BAE under different conditions. Moreover, the performance of the unconstrained cross-dataset (unconstrained training set and unconstrained testing set) evaluation is slightly better than that of the semi-unconstrained cross-dataset (constrained training set and unconstrained testing set) evaluation, which can be attributed to more diverse intra-class variations of the unconstrained training dataset. In a word, the LFH-BAE has a good generalization ability and can be competent for face recognition tasks under different conditions.

V. CONCLUSION

In this work, we propose a simple and effective unsupervised learning-based hashing model for face recognition.

⁴In the cross-dataset evaluation, the "View 1" subset of LFW is utilized as the training set.

Unlike previous hashing methods, we respect the binary nature of the problem and enforce the binary constraint in the discrete optimization to learn high-quality binary codes. To exploit the salient semantics inherent in the raw data, we utilize the binary auto-encoder that seeks to well reconstruct a face image from the learned robust binary codes. When optimizing the awkward original objective function, we simplify it into separate manageable sub-problems by introducing an auxiliary variable, and then solve them via the ALM-based alternating optimization. Such an alternating optimization not only respects the discrete constraint, but also takes the bit-independent and bit-balanced constraints into consideration. When solving the critical sub-problem, i.e., discrete optimization, we adopt the DCC method to learn binary codes in a closed form in the Hamming space. Experiments are conducted on four public face datasets and the results demonstrate that our proposed method achieves superior or very competitive performance compared with state-ofthe-art algorithms. Our LFH-BAE concentrates on modeling the entire data space to extract the salient inter-class factors. However, the local geometric structure of the face image is still important and should be taken into consideration in the future to improve the performance our model. Moreover, we also intend to apply our model to other computer vision tasks such as video surveillance and multi-media search.

REFERENCES

- T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.
- [2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [3] Z. Lei, M. Pietikäinen, and S. Z. Li, "Learning discriminant face descriptor," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 289–302, Feb. 2014.
- [4] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 467–476, Apr. 2002.
- [5] 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, Oct. 2015.
- [6] J. Lu, V. E. Liong, and J. Zhou, "Simultaneous local binary feature learning and encoding for homogeneous and heterogeneous face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 8, pp. 1979–1993, Aug. 2018.
- [7] Y. Duan, J. Lu, J. Feng, and J. Zhou, "Context-aware local binary feature learning for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 5, pp. 1139–1153, May 2018.
- [8] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 533–544, Feb. 2010.
- [9] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1, Oct. 2005, pp. 786–791.
- [10] Z. Liu, J. Wang, G. Liu, and L. Zhang, "Discriminative low-rank preserving projection for dimensionality reduction," *Appl. Soft Comput.*, vol. 85, Dec. 2019, Art. no. 105768.
- [11] Z. Lai, W. K. Wong, Y. Xu, J. Yang, and D. Zhang, "Approximate orthogonal sparse embedding for dimensionality reduction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 4, pp. 723–735, Apr. 2016.
- [12] S. U. Hussain, T. Napoléon, and F. Jurie, "Face recognition using local quantized patterns," in *Proc. Brit. Mach. Vis. Conf.*, Sep. 2012, pp. 52–61.

- [13] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, Mar. 2011.
- [14] X. Chang, F. Nie, Y. Yang, C. Zhang, and H. Huang, "Convex sparse PCA for unsupervised feature learning," ACM Trans. Knowl. Discovery Data, vol. 11, no. 1, pp. 1–16, Jul. 2016.
- [15] M. S. Aslan, Z. Hailat, T. K. Alafif, and X.-W. Chen, "Multi-channel multi-model feature learning for face recognition," *Pattern Recognit. Lett.*, vol. 85, pp. 79–83, Jan. 2017.
- [16] D. Yi, Z. Lei, and S. Z. Li, "Shared representation learning for heterogenous face recognition," in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, May 2015, pp. 1–7.
- [17] L. He, H. Li, Q. Zhang, and Z. Sun, "Dynamic feature learning for partial face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7054–7063.
- [18] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "SphereFace: Deep hypersphere embedding for face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 212–220.
- [19] Z. Liu, Z. Lai, W. Ou, K. Zhang, and R. Zheng, "Structured optimal graph based sparse feature extraction for semi-supervised learning," *Signal Process.*, vol. 170, May 2020, Art. no. 107456, doi: 10.1016/j.sigpro.2020.107456.
- [20] M. Kan, S. Shan, H. Chang, and X. Chen, "Stacked progressive autoencoders (SPAE) for face recognition across poses," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1883–1890.
- [21] L. Cheng, J. Wang, Y. Gong, and Q. Hou, "Robust deep auto-encoder for occluded face recognition," in *Proc. 23rd ACM Int. Conf. Multimedia MM*, 2015, pp. 1099–1102.
- [22] S. Gao, Y. Zhang, K. Jia, J. Lu, and Y. Zhang, "Single sample face recognition via learning deep supervised autoencoders," *IEEE Trans. Inf. Foren*sics Secur., vol. 10, no. 10, pp. 2108–2118, Oct. 2015.
- [23] A. Ruiz-Garcia, M. Elshaw, A. Altahhan, and V. Palade, "Stacked deep convolutional auto-encoders for emotion recognition from facial expressions," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 1586–1593.
- [24] C. Xu, Q. Liu, and M. Ye, "Age invariant face recognition and retrieval by coupled auto-encoder networks," *Neurocomputing*, vol. 222, pp. 62–71, Jan. 2017.
- [25] L. Paulevé, H. Jégou, and L. Amsaleg, "Locality sensitive hashing: A comparison of hash function types and querying mechanisms," *Pattern Recognit. Lett.*, vol. 31, no. 11, pp. 1348–1358, Aug. 2010.
- [26] Z. Jin, C. Li, Y. Lin, and D. Cai, "Density sensitive hashing," *IEEE Trans. Cybern.*, vol. 44, no. 8, pp. 1362–1371, Aug. 2014.
- [27] Y. Xia, K. He, P. Kohli, and J. Sun, "Sparse projections for highdimensional binary codes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3332–3339.
- [28] L. Zhu, Z. Huang, Z. Li, L. Xie, and H. T. Shen, "Exploring auxiliary context: Discrete semantic transfer hashing for scalable image retrieval," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5264–5276, Feb. 2018.
- [29] S. He, B. Wang, Z. Wang, Y. Yang, F. Shen, Z. Huang, and H. T. Shen, "Bidirectional discrete matrix factorization hashing for image search," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2019.2941284.
- [30] L. Xu, L. Zhu, J. Li, H. Zhang, and H. T. Shen, "Efficient supervised discrete multi-view hashing for large-scale multimedia search," *IEEE Trans. Multimedia*, to be published, doi: 10.1109/TMM.2019.2947358.
- [31] Y. Luo, Y. Yang, F. Shen, Z. Huang, P. Zhou, and H. T. Shen, "Robust discrete code modeling for supervised hashing," *Pattern Recognit.*, vol. 75, pp. 128–135, Mar. 2018.
- [32] M. Hu, Y. Yang, F. Shen, N. Xie, R. Hong, and H. T. Shen, "Collective reconstructive embeddings for cross-modal hashing," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 2770–2784, Jun. 2019.
- [33] X. Xu, F. Shen, Y. Yang, H. T. Shen, and X. Li, "Learning discriminative binary codes for large-scale cross-modal retrieval," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2494–2507, May 2017.
- [34] L. Han, P. Li, X. Bai, C. Grecos, X. Zhang, and P. Ren, "Cohesion intensive deep hashing for remote sensing image retrieval," *Remote Sens.*, vol. 12, no. 1, p. 101, Dec. 2019.
- [35] T.-T. Do, D.-K.-L. Tan, T. T. Pham, and N.-M. Cheung, "Simultaneous feature aggregating and hashing for large-scale image search," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6618–6627.

- [36] X. Shen, F. Shen, L. Liu, Y.-H. Yuan, W. Liu, and Q.-S. Sun, "Multiview discrete hashing for scalable multimedia search," ACM Trans. Intell. Syst. Technol., vol. 9, no. 5, pp. 1–21, Jun. 2018.
- [37] F. Shen, C. Shen, W. Liu, and H. T. Shen, "Supervised discrete hashing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 37–45.
- [38] L. Zhu, Z. Huang, X. Chang, J. Song, and H. T. Shen, "Exploring consistent preferences: Discrete hashing with pair-exemplar for scala-ble landmark search," in *Proc. ACM Int. Conf. Multimedia*, Oct. 2017, pp. 726–734.
- [39] L. Zhu, Z. Huang, X. Liu, X. He, J. Sun, and X. Zhou, "Discrete multimodal hashing with canonical views for robust mobile landmark search," *IEEE Trans. Multimedia*, vol. 19, no. 9, pp. 2066–2079, Sep. 2017.
- [40] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 1997, pp. 137–143.
- [41] W. Liu, J. Wang, S. Kumar, and S.-F. Chang, "Hashing with graphs," in Proc. Int. Conf. Mach. Learn., Jul. 2011, pp. 1–8.
- [42] J.-P. Heo, Y. Lee, J. He, S.-F. Chang, and S.-E. Yoon, "Spherical hashing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2957–2964.
- [43] Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin, "Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 2916–2929, Dec. 2013.
- [44] M. Hu, Y. Yang, F. Shen, N. Xie, and H. T. Shen, "Hashing with angular reconstructive embeddings," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 545–555, Feb. 2018.
- [45] D. Maturana, D. Mery, and A. Soto, "Face recognition with decision treebased local binary patterns," in *Proc. Asian. Conf. Comput. Vis.*, Nov. 2010, pp. 618–629.
- [46] D. Maturana, D. Mery, and A. Soto, "Learning discriminative local binary patterns for face recognition," in *Proc. Face Gesture*, Mar. 2011, pp. 470–475.
- [47] F. Kamaruzaman and A. A. Shafie, "Recognizing faces with normalized local Gabor features and spiking neuron patterns," *Pattern Recognit.*, vol. 53, pp. 102–115, May 2016.
- [48] F. Shen, C. Shen, X. Zhou, Y. Yang, and H. T. Shen, "Face image classification by pooling raw features," *Pattern Recognit.*, vol. 54, pp. 94–103, Jun. 2016.
- [49] A. Pillai, R. Soundrapandiyan, S. Satapathy, S. C. Satapathy, K.-H. Jung, and R. Krishnan, "Local diagonal extrema number pattern: A new feature descriptor for face recognition," *Future Gener. Comput. Syst.*, vol. 81, pp. 297–306, Apr. 2018.
- [50] W. Li, Y. Wang, Z. Xu, Y. Jiang, Z. Lu, and Q. Liao, "Weighted contourlet binary patterns and image-based Fisher linear discriminant for face recognition," *Neurocomputing*, vol. 267, pp. 436–446, Dec. 2017.
- [51] H. Wang, J. Hu, and W. Deng, "Compressing Fisher vector for robust face recognition," *IEEE Access*, vol. 5, pp. 23157–23165, Sep. 2017.
- [52] M. Xi, L. Chen, D. Polajnar, and W. Tong, "Local binary pattern network: A deep learning approach for face recognition," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2016, pp. 3224–3228.
- [53] L. Tian, Y. Ming, X. Hong, G. Zhao, and M. Pietikäinen, "Sparse projections matrix binary descriptors for face recognition," *Neurocomputing*, vol. 297, pp. 8–21, Jul. 2018.
- [54] C.-Y. Low, A. B.-J. Teoh, and C.-J. Ng, "Multi-fold Gabor, PCA, and ICA filter convolution descriptor for face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 1, pp. 115–129, Jan. 2019.
- [55] C.-Y. Low, J. Park, and A. B.-J. Teoh, "Stacking-based deep neural network: Deep analytic network for pattern classification," *IEEE Trans. Cybern.*, to be published.
- [56] C. Y. Low, A. B. J. Teoh, and K. A. Toh, "Stacking PCANet+: An overly simplified convnets baseline for face recognition," *IEEE Signal Process. Lett.*, vol. 24, no. 11, pp. 1581–1585, Nov. 2017.
- [57] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao, "The CAS-PEAL large-scale chinese face database and baseline evaluations," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 38, no. 1, pp. 149–161, Jan. 2008.
- [58] S. Xie, S. Shan, X. Chen, X. Meng, and W. Gao, "Learned local Gabor patterns for face representation and recognition," *Signal Process.*, vol. 89, no. 12, pp. 2333–2344, Dec. 2009.

- [59] J. Lu, V. E. Liong, G. Wang, and P. Moulin, "Joint feature learning for face recognition," *IEEE Trans. Inf. Forensics Secur.*, vol. 10, no. 7, pp. 1371–1383, Jul. 2015.
- [60] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Vis. Lab, Univ. Massachusetts, Amherst, Amherst, MA, USA, Tech. Rep. 07-49, Oct. 2007.
- [61] G. B. Huang and E. Learned-Miller, "Labeled faces in the wild: Updates and new reporting procedures," Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep. 14-003, May 2014.
- [62] H. J. Seo and P. Milanfar, "Face verification using the lark representation," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 4, pp. 1275–1286, Dec. 2011.
- [63] U. Hussain and F. Jurie, "Local higher-order statistics (LHS) for texture categorization and facial analysis," in *Proc. Comput. Vis. ECCV*, Oct. 2012, pp. 1–12.
- [64] E. Basaran, M. Gökmen, and M. Kamasak, "An efficient multiscale scheme using local zernike moments for face recognition," *Appl. Sci.*, vol. 8, no. 5, p. 827, May 2018.
- [65] M. Uzun-Per and M. Gökmen, "Face recognition with patch-based local Walsh transform," *Signal Process., Image Commun.*, vol. 61, pp. 85–96, Feb. 2018.
- [66] S. R. Arashloo and J. Kittler, "Efficient processing of MRFs for unconstrained-pose face recognition," in *Proc. IEEE 6th Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS)*, Sep. 2013, pp. 1–8.
- [67] X. Zhu, Z. Lei, J. Yan, D. Yi, and S. Z. Li, "High-fidelity pose and expression normalization for face recognition in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 787–976.
- [68] J. R. Beveridge, K. W. Bowyer, P. J. Flynn, S. Cheng, P. J. Phillips, D. S. Bolme, B. A. Draper, G. H. Givens, Y. M. Lui, M. N. Teli, H. Zhang, and W. T. Scruggs, "The challenge of face recognition from digital pointand-shoot cameras," in *Proc. IEEE 6th Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS)*, Sep. 2013, pp. 1–8.
- [69] E. Rahtu, J. Heikkilä, V. Ojansivu, and T. Ahonen, "Local phase quantization for blur-insensitive image analysis," *Image Vis. Comput.*, vol. 30, no. 8, pp. 501–512, Aug. 2012.
- [70] J. Ylioinas, A. Hadid, J. Kannala, and M. Pietikäinen, "An in-depth examination of local binary descriptors in unconstrained face recognition," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 4471–4476.



JING CHEN received the M.S. degree from the the Department of Electronic Engineering, Beijing University of Posts and Telecommunications, China, in 2016, where she is currently pursuing the Ph.D. degree with the Beijing Key Laboratory of Work Safety Intelligent Monitoring. Her research interests include computer vision, pattern recognition, and machine learning.



YUNXIAO ZU received the Ph.D. degree in advanced electrotechnology form Tianjin University, Tianjin, China, in 2000. She was a Post-doctoral Fellow with the Institute of Advanced Electrotechnology, Tsinghua University, from 2001 to 2003. She was also a Visiting Scholar with the Department of Electronic and Electrical Engineering, Imperial College London, London, from 2006 to 2007. She is currently a Professor and the Director of the Signals, Circuits and Systems

Center, School of Electronic Engineering, Beijing University of Posts and Telecommunications. Her major research interests include wireless communication theory, cognitive radio, and intelligent information processing.