Sparse Projections Matrix Binary Descriptors for Face Recognition

Chunxiao Fan[a], Lei Tian[a,b], Yue Ming[a,], Xiaopeng Hong[b], Guoying Zhao[b], Matti Pietikäinen[b]

[a]*School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing, P.R.China, 100876*
[b]*The Center for Machine Vision and Signal Analysis, Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland, 90014*

Abstract

In recent years, the binary feature descriptor has achieved great success in face recognition (FR) field, such as local binary pattern (LBP). It is well known that the high-dimensional feature representations can contain more discriminative information, therefore, it is natural for us to construct the high-dimensional binary feature for FR task. However, the high-dimensional representations would lead to high computational cost and overfitting. Therefore, an effective sparsity regularizer is necessary. In this paper, we introduce the sparsity constraint into the objective function of general binary codes learning framework, so that the problem of high computational cost and overfitting can be somehow solved. There are three main requirements in our objective function: First, we require that the high-dimensional binary codes have the minimized quantization loss compared with centered original data. Second, we require the projection matrices are sparse, so that the projection process would not take lots of computational resource even faced with high-dimensional original data. Third, for a mapping (hashing) function, the bit-independence and bit-balance are two excellent properties for generating discriminative binary codes. We also empirically show that the high-dimensional binary codes can obtain

Email addresses: fcxg100@163.com (Chunxiao Fan), Lei.Tian@oulu.fi, tianlei189@sina.com (Lei Tian), myname35875235@126.com (Yue Ming), Xiaopeng.Hong@oulu.fi (Xiaopeng Hong), guoying.zhao@oulu.fi (Guoying Zhao), Matti.Pietikainen@oulu.fi (Matti Pietikäinen)

more discriminative ability by pooling process with an unsupervised clustering method. Therefore, a discriminative and low-cost Sparse Projection Matrix Binary Descriptors (SPMBD) is learned by the data-driven way. Extensive experimental results on four public datasets show that our SPMBD descriptor outperforms other existing face recognition algorithms and demonstrate the effectiveness and robustness of the proposed methods.

Keywords: Local Descriptor, Feature Learning, Binary Coding, Sparse Projection Matrix, Face Recognition.

---

1. Introduction

Face recognition (FR) has achieved great success in over three decades, but it attracts researchers' attention due to there are still many problems unsolved in real applications, such as the extreme intraclass variations in unconstrained scenario and large number of subject classes in crowded video surveillance scenario [1, 2, 3, 4]. Therefore, in this paper, we mainly focus on the FR problem in constrained and unconstrained scenarios.

There are two main reasons why current FR system can not work well in real life. On the one hand, the face images (especially, captured in unconstrained environment) consist of a various of variabilities, such as pose, expression, illumination, blur and so on. These intra-class variations dramatically reduce the recognition accuracy of FR system. On the other hand, with the development of Internet and image capture device, the recent FR system is conducted on large-scale datasets. But the large-scale samples inevitably lead to high computational cost and can not directly be applied on many real scenarios, such as mobile phones and wearable devices.

According to the existing works, feature representation is a key breakthrough point to achieve excellent performance for FR task. Over decades, a various of feature representations based FR methods are proposed, such as SIFT [5] and LBP [6]. However, their success mainly comes from designing features manually and elaborately, these methods can obtain excellent performance only when

strong task-specific prior knowledge is provided (**Problem 1**). What is more, many high-dimensional features are proposed in recent years, such as high-dim LBP [3] and MDML-DCPs [4]. They have empirically found that a high-dimensional feature representation is necessary to achieve good performance. However, when the original feature's dimensionality d is large, the projection matrix could contain millions parameters, it can easily tend to overfitting (**Problem 2**). The computational cost in terms of both time and memory is also a non-negligible problem for high-dimensional data (**Problem 3**). Therefore, how to learn a data-oriented and non-overfitting feature representation is a key concern for FR task.

1.1. Motivation

The work [2, 7, 8] has proved that the combination of binary descriptors and histogram representations are insensitive to local variabilities and the work in [3, 4] has proved that high-dimensional representations can achieve better performance in FR task, respectively. However, there are few methods which integrate these two excellent properties into a general framework. Therefore, we would like to learn discriminative high-dimensional binary descriptors and histogram-based representations from data itself.

According to the above analysis, there are three necessary optimization terms need to be introduced into our objective function to overcome the above problems. They are **quantization loss term**, **sparsity regularizer term** and **bit-balance and bit-independence term**. The quantization loss term makes the quantization error between real-valued PDVs and high-dimensional binary codes is minimized, so that the identity information of face image is preserved as much as possible. The sparsity regularizer term makes the projection matrices of binary coding process are properly sparse, it is expected to restrict non-zero elements' number in the projection matrix and somehow solve the **Problem 2** and **Problem 3** together. The bit-balance term leads to each bit have the same probability to be 0 and 1, the bit-independence term leads to the learned different bits are independent of each other, so that the learned binary codes

can contain more information than those bit-dependent codes when bit length is fixed. Moreover, the whole optimization process is conducted based on data-driven way, so that the **Problem 1** can be solved.

Figure 1 shows the work-flow of the proposed SPMBD method. Firstly, the patch-wise pixel difference vectors (PDVs) are extracted, which is able to implicitly describe the basic visual patterns of face images. Secondly, we optimize the objective function of our method by iterative scheme. Lastly, inspired by the work [9], the binary codes learned from the above step are clustered as a dictionary and are pooled into a set of statistical features to make the final feature representations have more discriminative power.

1.2. Contributions

The contributions of our proposed can be listed as follows:

1. We propose a sparse projections matrix binary descriptors for high-dimensional feature representations in FR task. Since there is the quantization loss term in our objective function, our SPMBD feature keeps the identity information as much as possible.
2. The sparsity property effectively reduces the number of non-zero elements in the training process, so that the possibility of overfitting is substantially decreased in trained model. The sparsity makes our trained model lie on the balance point where the model is neither underfitting nor overfitting. The sparsity term also dramatically reduces the computational cost.
3. Our proposed method are extensively evaluated on four public datasets: FERET, CAS-PEAL-R1, PaSC and LFW, and our method obtains better performance than other recent works in both constrained and unconstrained FR scenarios.

In the following section, we review some recent related works in the Section 2. Then, in the Section 3, we detail our proposed SPMBD learning method. Specially, we analyze the existing problems of recent binary codes based FR method and formulate the objective function of our SPMBD in subsection 3.2.

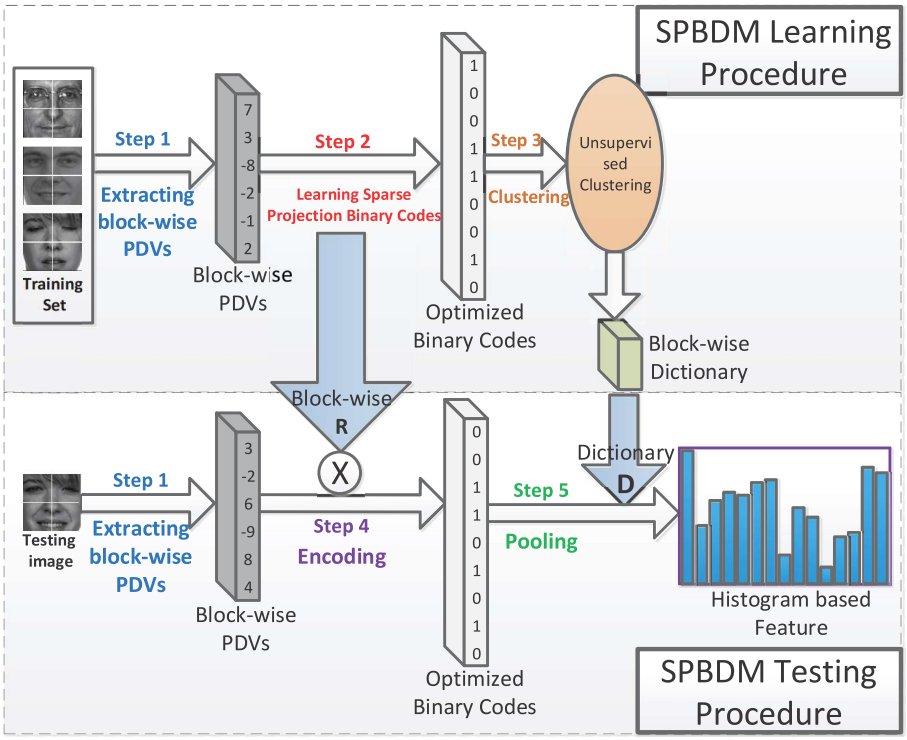


Figure 1: The illustration of our proposed SPMBD method. In **Step 1**, we extract PDVs from each local region for training and testing images. In **Step 2**, we learn a sparse feature mapping from the training images, so that the extracted PDVs are projected as discriminative and high-dimensional binary codes. In **Step 3**, the optimized binary codes in **Step 2** are clustered by unsupervised clustering method, the region-wise dictionaries are obtained. In **Step 4**, having learned sparse projection matrix R in **Step 2**, we encode the PDVs into binary codes. In **Step 5**, the learned binary codes are pooled as histogram-based features by using the learned dictionary D in **Step 3**.

The overall optimization process is detailly described in Subsection 3.3. In the section 4, we discuss the relationship between our porposed method and previous methods. At last, we provide the performance comparison of our SPMBD methods and other existing methods on public face databases in Section 5 and summarize our work in the Section 6.

2. Related Work

Since our proposed method is a learning-based sparsity-encouraging binary coding method for face representation, in this section, we put our attention on these three most relevant topics: **face feature learning**, **face sparse representation** and **binary coding learning**.

2.1. Face Feature Learning

In recent years, the learning-based methods with low-level features have achieved great success for face recognition [9, 10, 11]. Since different neighboring patches have different contributions to face representation, Lei *et al.* [9] propose the optimal soft sampling to assign the different weights for each neighboring patch in a supervised way. Zhang *et al.* [10] employ the Weber local descriptors to extract multi-scale features from local patches centered at the facial landmarks and randomly selects a subset of the learned local features. The work in [11] extracts different feature representations for different modalities by exploiting different complexity deep model, and the outputs of different network are concatenated and compressed by stacked auto-encoders. It achieves excellent performance by using publicly available training set. Because of learning-based methods self-learn discriminative information from the data and do not need to provide the prior knowledge, they obtain better performance than those hand-crafted methods.

2.2. Face Sparse Representation

Wright *et al.* [12] first introduce sparse representation codes (SRC) into FR task, but the small sample size problem is a severe problem in this work. This problem is solved in the work [13] by constructing an auxiliary intra-class dictionary which contains all possible variation between the training samples and testing sample. The work [14] first constructs the intraclass and interclass scatter matrix by weighted elastic net and the weighted sparse preserving embedding technology is employed to find a subspace which makes the ratio of the interclass scatter to the intraclass scatter is maximized. In work [15],

the sparse neighborhood graph is first constructed, then the low-dimensional embedding is learned based on the intrinsic geometry metric. The work [16] considers the axis-symmetrical nature of faces and produce approximately axis-symmetrical virtual dictionary for face sparse representation.

2.3. Binary Code Learning

Lots of binary codes learning algorithms are also proposed in recent decades. Li *et al.* [17] introduce a number of labeled facial attributes into binary codes encoding process for face image retrieval and facial attributes prediction, just like killing two birds with one stone. The BitHash method [18] proposes an unbiased estimate of pairwise Jaccard similarity and employs one bit per hash value to represent data sample. Gao *et al.* [19] propose the Batch-Orthogonal Locality Sensitive Hashing (BOLSH) for face recognition in movie videos, which can be understood as the special case of locality-sensitive hashing. The BOLSH learns these orthogonal projections from the part original data and group their binary representations as a batch.

What is more, some recent works for high-dimensional binary codes are proposed. Yu *et al.* [20] learn binary codes by mapping the data with circulant matrix. Since circulant structure can be effectively computed by Fast Fourier Transformation, the projection process for high-dimensional data can be accelerated. Bilinear Projections [21] is also designed for high-dimensional binary code, it exploits natural two-dimensional structure of existing descriptors and decomposes a large projection matrix into two small matrices.

3. Sparse Projections Matrix Binary Descriptors Learning

In this section, we first show the extraction process of PDV, then we describe the formulation and optimization process of SPMBD descriptor. At last, we detail how to encode the binary codes into histogram-based features.

It is well-known that the identity-bearing information for each face region is different. In order to exploit these position-specific information, we intend

to learn region-wise SPMBD descriptor. The (training/testing) face images are divided into a number of local regions. These local SPMBD features are learned from each region of face images and these region-wise features are concatenated as the output feature. Moreover, we use the Whitening PCA (WPCA) to normalize the variance of our learned features and further make our representation efficiently. At last, the corresponding features obtained from WPCA are fed into 1-NN classifier to measure the similarity of face images.

3.1. The Extraction of PDVs

This section corresponds the Step 1 in Fig 1. Given training set $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n]$, where $\mathbf{a}_i$ denotes the ith training face image. We extract PDVs from a input face image at patch level and concatenate them as Pixel Difference Matrix (PDM) $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N]$, where $\mathbf{x}_i \in \mathbb{R}^d$ denotes the ith pixel difference vector and N is the number of PDVs. In order to clearly explain how we extract pixel difference matrix from the center patch and its neighboring patch, we show the extraction process of PDV in Fig. 2. It can be observed that the dimensionality d of a PDV is $[(2L+1) \times (2L+1) - 1]$ and L denotes the sampling radius. In order to exploit the multiple scales pixel-wise information, we can extract a set of PDVs with multiple radii and concatenate them into a long PDV. Moreover, we need to subtract the mean from input data before performing our proposed method, so that the learned binary codes can better represent variations of the data, i.e.,

$$\sum_{i=1}^{N} \mathbf{x}_i = 0, \tag{1}$$

and we will detail the reason why we remove the mean from input data in the next subsection.

3.2. Formulation

According the motivations of our method (i.e., Section 1.1), there are three requirements for the objective function of our method. First, the learned binary codes should similar to the original data as closely as possible. Second,

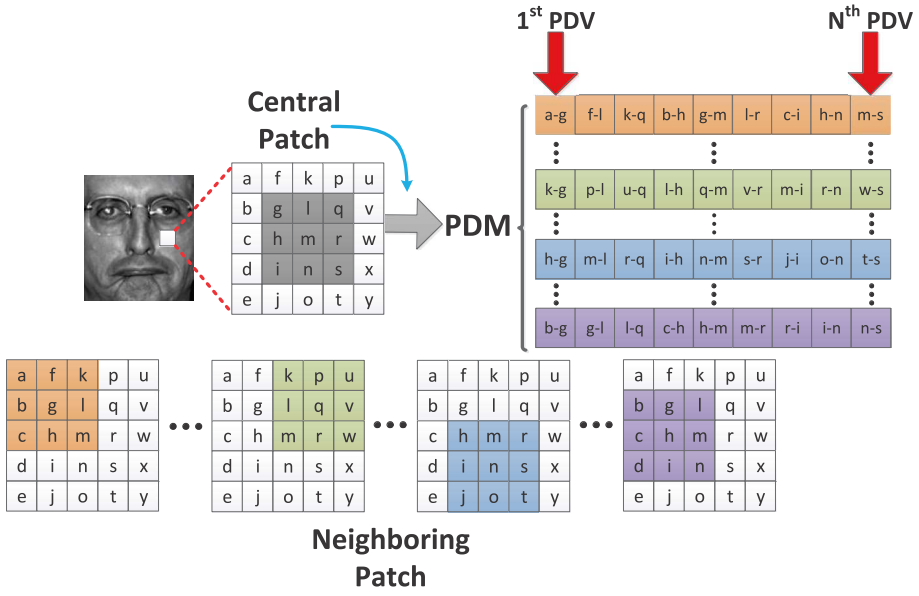


Figure 2: The example of extracting a patch-wise PDM and PDV from a face image. In order to make the figure concise, we set the sampling radius R to be 1, so the neighboring pixels in $(2R+1)\times(2R+1)$ space are concerned. Then we compare the pixels of each neighboring patches with that of central patch, and concatenate all difference vectors as the row vector of PDM. The PDV can be denoted as the column vector of the PDM. This figure is best viewed in color version.

the projection matrices should be properly sparse, so that the overfitting and computational cost during the projection process can be reduced. Third, each bit should have 50% to be 0 or 1 (i.e., bit-balance property), and different hashing (i.e., projection) functions should be independence (i.e., bit-independence property), so that the learned bits can carry more effective information with fixed bit length.

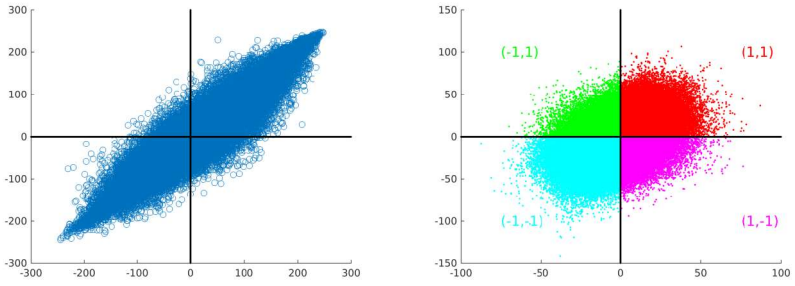
According to the above requirements, the objective function can be denoted as follow:

$$\begin{aligned} &\min_{\mathbf{R},\mathbf{B}} \|\mathbf{R}\mathbf{X}-\mathbf{B}\|_F^2 \\ &\text{s.t.} \quad |\mathbf{R}|_0 \le m, \quad ,\mathbf{R}^T\mathbf{R}=\mathbf{I} \end{aligned} \tag{2}$$

where sparse projection matrix $\mathbf{R} \in \mathbb{R}^{K\times d}$ to map each PDV $\mathbf{x}_i \in \mathbb{R}^{d\times 1}$ into K-bits binary codes $\mathbf{b}_i = [\mathbf{b}_{i,1}, \mathbf{b}_{i,2}, \cdots, \mathbf{b}_{i,K}]^T \in \{-1,1\}^{K\times 1}$ by the

equation $\mathbf{B} = \text{sign}(\mathbf{RX}) \in \mathbb{R}^{K \times N}$, $|\cdot|_0$ denotes non-zero elements' number of the projection matrix. There are three terms in the objective function of our SPMBD: quantization loss term, the sparsity term and bit-independence term. The main part of Eq. 2 corresponds to the minimization problem of quantization error (i.e., the first term). $|\mathbf{R}|_0 \leq m$ and $\mathbf{R}^T\mathbf{R} = \mathbf{I}$ corresponds to the sparsity term (i.e., the second term) and the bit-independence term (i.e., the third term), respectively. One may wondering that why there is no bit-balance term in the objective function. Because we have found that the learned bianry codes always satisfy the bit-balance property for the centered input data, and it is demonstrated in Figure 3. So this term is not necessary for our objective function as long as we pre-process the input data by removing its mean.

Compared with some SR-like methods, we control the sparsity of projection process by ℓ_0-regularization instead of ℓ_1-regularization. Since ℓ_0-regularization can directly determine the non-zero elements' number, so that the complexity of time or memory is also under our control. While it is not easy for ℓ_1-regularization, therefore, we choose ℓ_0-regularization to constrain the projection matrix.



(a) original PDVs (b) projected data

Figure 3: The first two-dimensional distribution examples of the original data and projected data, which come from the FERET training dataset. (a) The centered original PDV data and (b) the mapping data projected by the sparse projection matrix. This 2-D example demonstrates the projected data have the bit-balance property naturally.

3.3. Optimization

The objective function Eq. (2) is the m-sparsity problem [22], which is not convex due to the orthogonal and ℓ_0-regularization. Thanks to the development of the variable-splitting and penalty techniques [23, 24], we can adopt them in optimization process. The auxiliary variable $\mathbf{R}_2$ is introduced, and the sparse constraint is put on $\mathbf{R}_1$ and the orthogonality constraint on $\mathbf{R}_2$. In order to minimize the variable-splitting error, we also penalize the difference between $\mathbf{R}_1\mathbf{X}$ and $\mathbf{R}_2\mathbf{X}$. Therefore, we can rewrite the objective function (i.e., Eq. (2)) as following:

$$\begin{aligned} &\min_{\mathbf{R},\mathbf{B}} \|\mathbf{R}_2\mathbf{X} - \mathbf{B}\|_F^2 + \alpha \|\mathbf{R}_2\mathbf{X} - \mathbf{R}_1\mathbf{X}\|_F^2 \\ &\text{s.t.} \quad \mathbf{R}_2^T\mathbf{R}_2 = \mathbf{I}, \quad |\mathbf{R}_1|_0 \leq m \end{aligned} \tag{3}$$

and α is the penalty parameter to balance the contribution of two terms. The first term denotes the surrogate quantization error and the second term denotes the variable-splitting penalty error. It is well-known that the solution of Eq. 3 converges to that of Eq. 2 when $\alpha \to +\infty$. The original problem is convex to one of variable when other two are fixed.

3.3.1. Step 1: Updating $\mathbf{R}_1$ with fixed $\mathbf{R}_2$ and $\mathbf{B}$

This sub-problem can be rewritten as:

$$\begin{aligned} &\min_{\mathbf{R}} \|\mathbf{C}_1 - \mathbf{R}_1\mathbf{X}\|_F^2 \\ &\text{s.t.} \quad |\mathbf{R}_1|_0 \leq m \end{aligned} \tag{4}$$

where $\mathbf{C}_1 = \mathbf{R}_2\mathbf{X}$ is fixed. Similar to the optimization idea of the work [24], we firstly ignore the sparsity constraint and expand the Eq. 4 as following:

$$J = \operatorname{tr}\left(\mathbf{C}_1\mathbf{C}_1^T - 2\mathbf{C}_1^T\mathbf{R}_1\mathbf{X} + \mathbf{R}_1\mathbf{X}\mathbf{X}^T\mathbf{R}_1^T\right). \tag{5}$$

The above objective function can be optimized by gradient descent (GD) scheme. The sparsity constraint is then satisfied by directly thresholding the GD based solution by keeping m largest elements' magnitude and set that of the rest to be 0 as following:

$$\mathbf{R}_1^{t+1} = \phi_m\left(\mathbf{R}_1^t - \gamma\left(\mathbf{R}_2 - \mathbf{R}_1^t\right)\mathbf{X}\mathbf{X}^T\right), \tag{6}$$

where t is the number of iteration and $\gamma = -1/\|\mathbf{X}\|_F^2$ denotes the learning rate of GD scheme. The thresholding function ϕ_m can be denoted as following:

$$\phi_m(x) = \begin{cases} x, & x \geq x_m \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

where x_m is the mth largest element in the matrix $\mathbf{X}$. And in our implement, we find that the solution of Eq. 6 is almost converged in 5 iterations.

3.3.2. Step 2: Updating $\mathbf{R}_2$ with fixed $\mathbf{R}_1$ and $\mathbf{B}$

when $\mathbf{R}_1$ and $\mathbf{B}$ are fixed, the objective function can be denoted as following:

$$\begin{aligned} &\min_{\mathbf{R}_2} \|\mathbf{R}_2\mathbf{X} - \mathbf{C}_2\|_F^2 \\ &\text{s.t. } \mathbf{R}_2^T\mathbf{R}_2 = \mathbf{I} \end{aligned}, \tag{8}$$

where $\mathbf{C}_2 = (\mathbf{B} + \alpha\mathbf{R}_1\mathbf{X}) / (1 + \alpha)$ is fixed. This sub-problem can be considered as the classical orthogonal procrustes problem. The work [24] has proved that the procrustes problem is solvable when $K > d$. Readers can refer [24] for detailed derivation. When $K > d$, we first compute the Singular Value Decomposition (SVD) of the matrix $\mathbf{X}\mathbf{C}_2^T$ as $\mathbf{X}\mathbf{C}_2^T = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathbf{T}$, then set $\mathbf{R}_2$ as $\mathbf{V}\mathbf{U}^\mathbf{T}$. In order to solve the case of $K < d$, we project $\mathbf{X}$ to K-dimensional space by $\bar{\mathbf{X}} = \mathbf{W}\mathbf{X}$ and $\mathbf{W}$ is a largest eigenvalues based PCA projection matrix of size $K \times d$. Then $\mathbf{X}$ is substituted as $\bar{\mathbf{X}}$ in the sub-problem Eq. 8 and $\bar{\mathbf{R}}_2$ can be solved by the same case as $K > d$. At last, the orthogonal-constraint projection matrix $\mathbf{R}_2$ can be given by $\bar{\mathbf{R}}_2\mathbf{W}$.

3.3.3. Step 3: Updating $\mathbf{B}$ with fixed $\mathbf{R}_1$ and $\mathbf{R}_2$

This sub-problem can be rewritten as

$$\min_{\mathbf{R}_2} \|\mathbf{B} - \mathbf{C}_3\|_F^2, \tag{9}$$

where $\mathbf{C}_3 = \mathbf{R}_2\mathbf{X}$. The Eq. 9 is equivalent to $\max_{\mathbf{B}} \sum_{i,j} (\mathbf{C}_3)_{ij}\mathbf{B}_{ij}$, where i, j denotes the matrix elements' indexes. In order to maximize this equation, we need to set $\mathbf{B}_{ij} = 1$ when $(\mathbf{C}_3)_{ij} \geq 0$ and $\mathbf{B}_{ij} = -1$ otherwise. Therefore, $\mathbf{B}_{ij} = \text{sign}\left((\mathbf{C}_3)_{ij}\right) = \text{sign}\left((\mathbf{R}_2\mathbf{X})_{ij}\right)$ with element form, or simply $\mathbf{B} = \text{sign}(\mathbf{R}_2\mathbf{X})$

with matrix form. The pseudo code of our proposed SPMBD method is listed in *Algorithm 1*.

Algorithm 1 Sparse Projections Matrix Binary Descriptor learning algorithm.

Input: The training set $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n]$, the length of K, the penalty parameter α, the iteration number T and the sampling radius L.

Output: Sparse projection matrix $\mathbf{R}$ and binary codes $\mathbf{B}$.

Step 1 (Extraction of PDVs): Extracting the PDVs from $\mathbf{A}$ with sampling radius L and obtaining PDM $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N]$;

Step 2 (Pre-Processing):

  Subtract the mean $\bar{\mathbf{X}}$ from input data $\mathbf{X}$.

Step 3 (Optimization):

  3.1. Initialization:

    Initialize $\mathbf{R}_1^0 = \text{randn}\,(K, d)$.

    Initialize $\mathbf{B}^0 = \text{sign}\left(\mathbf{R}_1^0 \mathbf{X}\right)$.

  3.2. Iterative Optimization:

  for $t = 1, \cdots, T$ **do**

    3.2.1 Update $\mathbf{R}_1$ by solving Eq. 6.

    3.2.2 Update $\mathbf{R}_2$ by solving Eq. 8.

    3.2.3 Update B by solving Eq. 9.

  end for

Return: Sparse projection matrix $\mathbf{R}_1^T$ and corresponding binary codes $\mathbf{B}^T$.

3.4. Clustering and Pooling

This section corresponds to the Steps 3 and 5 in Fig. 1. Having learned the feature mapping matrix $\mathbf{R}_1$ by the above sparsity iterative scheme, so the PDVs are mapped into the high-dimensional feature vectors. In order to learn more data-adaptive binary codes, we first cluster learned features (see the step 3 in Fig. 1) by applying an unsupervised clustering method. We investigate different clustering methods, such as K-Means, Gaussian Mixture Model, mean-shift and spectral clustering, but there is no significant difference on performance in our

implementation. We think the learned binary codes already have elementary clustering structure after sparse projection, therefore, different methods have negligible effects on clustering. We choose the K-means in this paper for its simplicity and efficiency.

In the testing phase, in order to describe the statistical property of sparse projection binary code, each binary code is pooled as the bins of histogram by using the learned dictionary (see step 5 in Fig. 1). At last, we consider the histogram features as the output features.

4. Relation To Previous Methods

The SPMBD is a binary face descriptor which is designed to solve the overfitting and high computational cost problems of the high-dimensional representation. In our objective funciton, the quantization loss and bit-independence terms are used to learn the less-error and discriminative binary feature, and the sparsity term is introduced to reduce the overfitting of the existing descriptors.

There is a recent works, i.e., CBFD [2], which learns face representations by optimizing three terms in the objective function: quantization loss term, bit-balance term and variance-maximum term. Same to the motivation of our method, the quantization loss term of CBFD is to ensure the learned binary code lose identity information compared with original data as little as possible. The bit-balance term makes the binary codes centered, so that the output features can better describe the variances of the original data. The variance-maximum term makes the variance of the learned binary codes maximized, so that the output feature are more compact.

There are three main differences between ours and [2]: First, in order to handle the high-dimensional data, we introduce the ℓ_0 sparsity constraint into the objective function, while the CBFD can not handle the case of $K > d$. Second, we found that the output binary codes always satisfy the bit-balance property for the centered input data (refers to section 3.2), so this term is not

necessary as long as we pre-process the input data by removing its mean. In contrast, the CBFD preserves the bit-balance term in their objective function, what may lead to complicated optimization process. Third, the CBFD has variance-maximum term which make the CBFD features compact. However, the ℓ_0 constraint and discrete property of binary codes make our objective function is an NP-hard problem. It is challenging for us to solve the objective function with the variance-maximum term. Therefore, we separately append this property by using Whitened PCA in the next step, so that our final output features are also compact.

5. Experiments

We evaluate our SPMBD descriptor on constrained and unconstrained face recognition datasets. For the constrained FR scenario, we use FERET [25] and CAS-PEAL-R1 [26] to verify the discriminative ability of the SPMBD method. For the unconstrained FR scenario, we use the challenging PaSC [27] dataset to show the robustness of the SPMBD method. We also use the LFW [28] dataset to demonstrate the generalization of SPMBD method. At last, the performance analysis between our SPMBD method and state-of-the-art methods is described on the last subsection.

5.1. Testing on FERET

The FERET is a widely used public face dataset, which consists of over 13,000 face images of over 1,500 subjects. We follow the standard protocol of FERET dataset. In our experiment, all images are aligned by provided eye coordinates and cropped into 128×128 pixels, as shown in Fig. 4.

The training process is conducted on the *training* set, and the testing process is conducted on the other five subsets, i.e., *gallery*, *fb* (expression variation), *fc* (illumination variation) and *dup1, dup2* (aging variations). In our experiments, the size of each region's dictionary is set to be 1,000, and a face image is divided into 8×8 local regions. Therefore, a feature vector with 64,000 ($8 \times 8 \times 1000$)

dimensionality is learned by using SPMBD method for a face image. We apply the WPCA to compress features' dimensionality into 1,000. At last, the reduced features are fed into the Nearest Neighbor (NN) classifier with cosine metric.

Figure 4: The aligned and cropped face examples 128×128 from the FERET dataset.

5.1.1. Parameter Analysis

First, we evaluate the influence of different parameters of SPMBD method on FERET dataset, and further determine the parameter values which will be used in the next sections.

Binary Codes Length. We explore the influence of binary code's length K. To make the learned binary codes have the property of scale robustness, we set sampling radius r to be 3 and 5 to extract multi-scales original features. The penalty parameter α and the percentage of non-zero elements[1] p in sparse projections matrix $\mathbf{R}_1$ are empirically set to 10 and 0.9, and we will further discuss the impact of these parameters in next parts. We test the binary codes length K from 2^4 to 2^{12}. Fig. 5 shows the average recognition rate of different codes length on the FERET dataset.

According to experimental results, we can observe that high-dimensional binary codes usually have better performance. It has been proved that the

[1]The number of non-zero elements m can be denoted as $m = ceil\left(numel\left(\mathbf{R}_2\right) \times (1-p)\right)$ in our implementation.

fundamental assumption of this paper is correct. Especially, our SPMBD method obtained state-of-the-art performances when the length of binary codes K is set to between 256 and 1024, and the best performance is obtained when K is set to be 1024. Therefore, the binary codes length K of SPMBD is set to be 1024 in all following experiments.

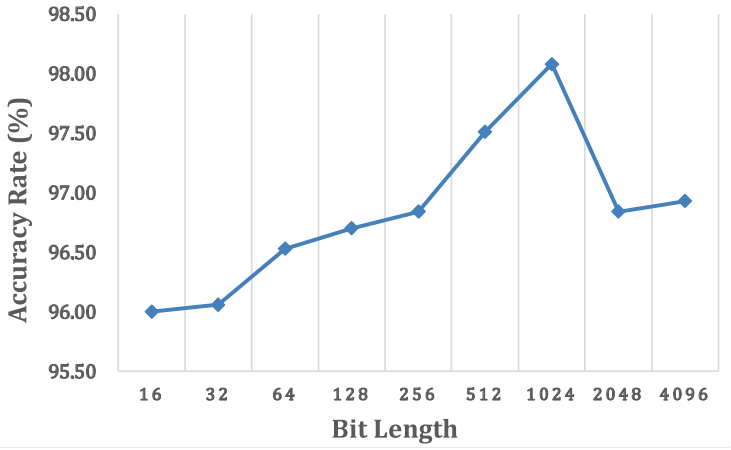


Figure 5: The average accuracy rate of SPMBD on four FERET probe sets when binary codes length is set to the different values.

The Percentage of Sparsity. Next, we study the impact of sparsity term in our SPMBD method in this part. The length of binary codes K is set to be 1024. And the penalty parameter α is empirically set to 10. Fig. 6 shows the accuracy and encoding time of our SPMBD versus different percentages of sparsity. We can observe that with the decrement of sparsity percentage, the encoding time of our SPMBD method increases linearly, and our SPMBD method can achieve best performance when percentage of sparsity p is set to be 0.9. It demonstrates that a dense projection matrix may lead to overfitting of the training model, and the excessive sparse projection matrix may lead to underfitting of the training model. Therefore, we set the percentage of sparsity p to be 0.9 as a trade off in subsequent experiments.

Impact of the penalty parameter α. Next, we study the impact of penalty parameter α of SPMBD in this part. The other parameters of SPMBD are the

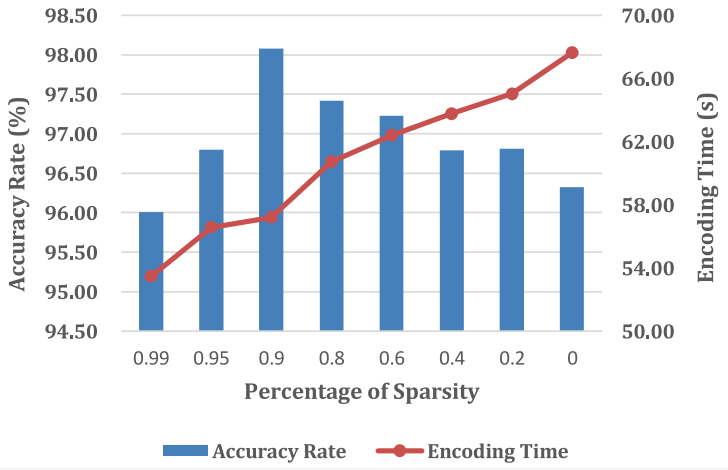


Figure 6: The average accuracy rate and encoding time of SPMBD on four FERET probe sets when the percentage of sparsity is set to be different values.

same as those used in the above experiment, and the percentage of sparsity p is set to 0.9.

The results are shown in Fig. 7. We can see that the values of the penalty parameter α have non-negligible effects on FERET datasets, and $\alpha = 10$ successfully balances the weight of surrogate quantization error and splitting penalty error.

Impact of the number of iterations. At last, we examine the impact of convergence for our SPMBD method. The parameters of SPMBD are the same as those used in the above experiments and the penalty parameter α is set to 10. The value of objective function is shown in Fig. 8. We vary the number of iterations from 1 to 60 and find that our SPMBD converges in about 5 iterations.

5.1.2. Binary Codes Learning Strategy

To demonstrate the effectiveness of sparse projection scheme, we compare our scheme with some existing binary code learning methods, such as, Locality-Sensitive Hashing (LSH) [29], Two Layer Anchor Graphs Hashing (AGH-2) [30], optimized version Circulant Binary Embedding (CBE-opt) and randomized

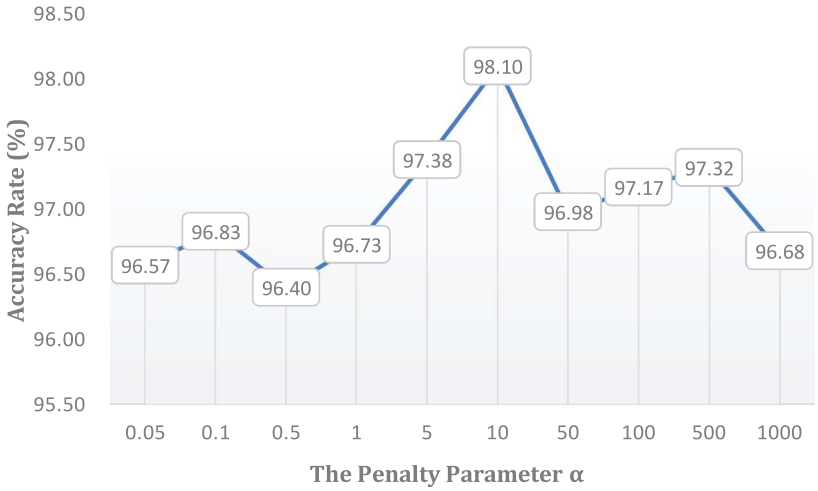


Figure 7: The average accuracy rate (%) of SPMBD on on four FERET probe sets when the penalty parameter α is set to be different values.

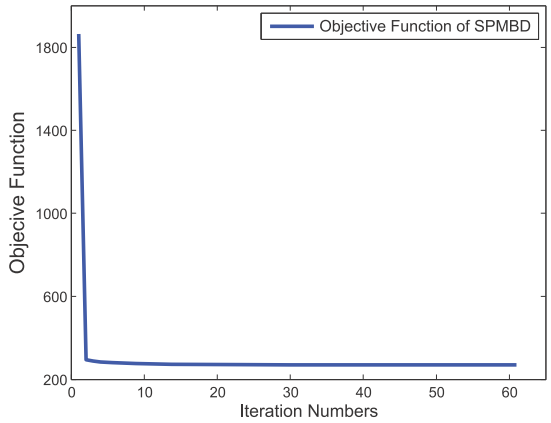


Figure 8: The value of objective function versus different iteration numbers on the FERET dataset.

version Circulant Binary Embedding (CBE-rand) [20]. Both LSH and AGH are existing dense based binary coding algorithms. CBE is designed to accelerate projection process for high-dimensional code[2]. We use the source codes provided

[2]Bilinear Projections is also designed for high-dimensional binary code, but this method require the input data are structured and the input PDVs don't meet this requirement.

by the authors to implement other four methods. We also remove the sparsity constraint from our objective function, denoted as SPMBD *($p = 0$)*, to further demonstrate the necessity of sparsity constraint. We just replace the objective function of the SPMBD with that of other binary learning algorithms, and keep other steps the same as **Algorithm 1**. It ensures the comparison experiment is totally fair. The parameters of our method are the same as those used in the above experiments. Table 1 lists the comparison results with WPCA on FERET dataset.

It can be observed that our sparse projections scheme outperforms other binary codes methods. On the one hand, some binary codes strategies (i.e., LSH and AGH-2) lack the description ability compared with our method. On the other hand, some binary codes strategies (i.e., CBE-rand and CBE-opt) overfit the original data due to their dense structure of projection matrix. Our SPMBD method introduces a sparsity regularizer which can effectively reduce the possible overfitting. Therefore, our method achieves better performance than others.

Table 1: The Rank-1 Recognition Rates (%) of our proposed SPMBD and other binary codes learning methods with WPCA on FERET Dataset.

Methods	*fb*	*fc*	*dup1*	*dup2*	*avg.*
LSH [29]	**99.7**	**100.0**	94.3	93.6	96.9
AGH-2 [30]	**99.7**	99.5	89.2	89.3	94.6
CBE-opt [20]	99.1	99.5	88.5	87.2	93.6
CBE-rand [20]	99.6	99.0	90.3	89.7	94.7
SPMBD ($p = 0$)	99.3	**100.0**	93.6	92.3	96.3
SPMBD	**99.7**	**100.0**	**95.7**	**97.0**	**98.1**

5.1.3. Recognition Results

We test the SPMBD method's performance on FERET probe sets with the standard evaluation protocol. Since our SPMBD is **local** learning based FR

method, therefore, we just compare our proposed descriptors with popular local face descriptors, such as LQP [31], DFD [9], CBFD [2] and IQBC [32]. Table 2 lists the accuracy of SPMBD and other approaches.

The proposed SPMBD method obtains the best performance on *fc*, *dup1* and *dup2*. Especially, our SPMBD achieves **95.7%** and **97.0%** accuracies on *dup1* and *dup2* sets. Our method improves recent DFD averagely 4.3% on *dup1* and *dup2* sets and improves recent CBFD averagely 3.0% these two sets. As far as we know, **it is a huge performance improvement for recent FR methods on these two sets, and it is even comparable with the DL-based methods.** On the *fb* set, only four images are misclassified by our method, but two of which are misclassified due to the mislabeling of dataset itself. The results demonstrate the superiority of our SPMBD method over previous methods.

5.2. Testing on CAS-PEAL-R1

The CAS-PEAL-R1 dataset consists of over 9,000 face images of 1,040 subjects with different variabilities. Similar to the above section, this database is also applied to show the description ability of our SPMBD to handle FR problems under controlled scenario. We follow the provided standard protocols and use five subsets, i.e., *accessory*, *training*, *expression*, *gallery*, and *lighting*. All face images are aligned by provided eye coordinates and cropped into 150×130 pixels, as shown in Fig. 9. What is more, we reduce the feature's dimensionality into 1,039 by WPCA and use the same parameters and classifier as those used in the FERET dataset for our SPMBD model. Table 3 tabulates accuracy on this database.

Clearly, compared with other existing face descriptors, our proposed SPMBD method obtains the best recognition rate on **all** testing sets. Especially, our SPMBD achieves **77.0%** accuracy on *lighting* set. **It is also a huge performance improvement for existing FR methods on this set.**

Table 2: The Rank-1 Recognition Rate (%) Comparison between state-of-the-art face methods with the standard FERET Evaluation Protocol.

Methods	*fb*	*fc*	*dup1*	*dup2*	year
LBP [6]*	97.0	79.0	66.0	64.0	2006
HGGP [33]*	97.6	98.9	77.7	76.1	2007
DT-LBP [34]*	99.0	100.0	84.0	80.0	2009
LDP [35]*	94.0	83.0	62.0	53.0	2010
GV-LBP-TOP [36]*	98.4	99.0	86.0	85.0	2011
DLBP [37]*	99.0	99.0	86.0	85.0	2011
I-LQP [31]*	99.2	69.6	65.8	48.3	2012
PEOM [38]*	97.0	95.0	77.6	76.2	2012
DFD [9]*	99.2	98.5	85.0	82.9	2014
CBFD [2]*	98.2	100.0	86.1	85.5	2015
IQBC [32]	98.2	100.0	85.5	85.9	2016
SPMBD	97.5	100.0	84.5	83.8	-
LBP+WPCA [6]*	98.5	84.0	79.4	70.0	2006
I-LQP+WPCA [31]*	99.8	94.3	85.5	78.6	2012
POEM+WPCA [38]*	99.6	99.5	88.8	85.0	2012
DFD+WPCA [9]*	99.4	**100.0**	91.8	92.3	2014
CBFD+WPCA [2]*	99.8	**100.0**	93.5	93.2	2015
SLBFLE+WPCA [7]*	**99.9**	**100.0**	95.2	92.7	2015
IQBC+WPCA [32]	99.7	**100.0**	94.9	95.3	2016
SPMBD+WPCA	99.7	**100.0**	95.7	**97.0**	-

* The results of other methods are directly cited from the original papers.

5.3. Testing on PaSC

Different from the above two sections, we investigate the accuracy of SPMBD method for uncontrolled FR scenarios in this section. The PaSC dataset is a new released challenging uncontrolled face dataset. This dataset releases 9,376 still images of 293 subjects and contains a large number of uncontrolled variabilities, such as poor lighting, large pose, occlusion, motion blur and poor focus. According to the standard evaluation protocol, both *target* set and *query*

Table 3: The accuracy (%) of our SPMBD and the state-of-the-art face recognition methods on CAS-PEAL-R1.

Methods	*Expression*	*Accessory*	*Lighting*	year
LGBP [39]*	95.0	87.0	51.0	2005
HGGP [33]*	96.0	92.0	62.0	2007
DT-LBP [34]*	98.0	92.0	41.0	2011
DLBP [37]*	99.0	92.0	41.0	2011
DFD [9]*	99.3	94.4	59.0	2014
CBFD [2]*	99.4	94.8	59.5	2015
IQBC [32]	99.5	95.1	70.4	2016
SPMBD	99.2	93.4	68.4	-
DFD+WPCA [9]*	99.6	96.9	63.9	2014
CBFD+WPCA [2]*	**99.7**	97.2	67.4	2015
IFL+WPCA [8]*	99.3	96.5	64.3	2015
JFL+WPCA [8]*	**99.7**	97.2	67.4	2015
IQBC+WPCA [32]	**99.7**	97.2	75.7	2016
SPMBD+WPCA	**99.7**	**97.3**	**77.0**	-

* The results of other methods are directly cited from the original papers.

Figure 9: The aligned and cropped examples with size 150×130 from the CAS-PEAL-R1 dataset.

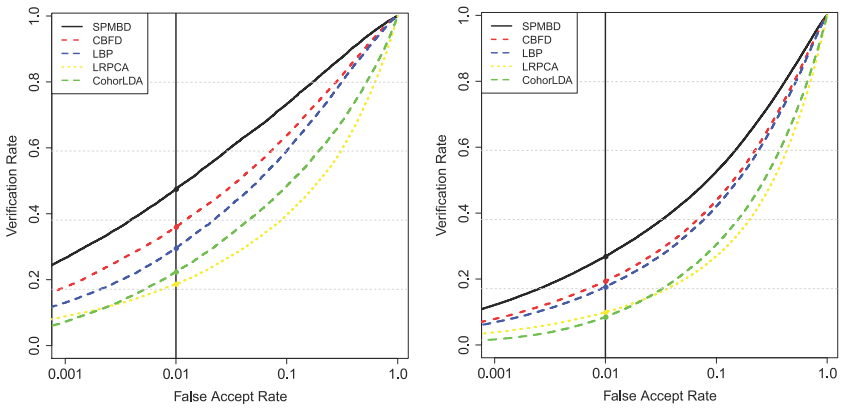
set contain 4,688 images. There are two evaluation scenarios in the standard protocol: all image and near-frontal evaluation protocol. All face images in the two subsets are aligned by provided eye coordinates[3] and cropped into 128×128 pixels, as shown in Fig. 10.

(a) frontal (b) non-frontal

Figure 10: The aligned and cropped face examples from the PaSC dataset. (a) frontal (or near-frontal) face images and (b) non-frontal face images. These examples shown some of the specially complicated intra-class variabilities in point-and-shoot scenario, such as large pose, poor focus and motion blur.

420 We compare our SPMBD method with the existing learning based FR method, such as CBFD, DFD, IQBC, LPQ, BSIF [40]. What is more, the

[3]The eye coordinates of each face image can be found at http://www.cs.colostate.edu/ vision/pasc/index.php



(a) frontal (b) non-frontal

Figure 11: The corresponding ROC curves of different methods on the PaSC dataset for (a). frontal and (b). all images scenarios respectively.

subspace learning methods (i.e., LRPCA, CohortLDA), and conventional LBP are evaluated as baseline methods. The parameters of our SPMBD method are the same as those used on the FERET and CAS-PEAL-R1 experiments. We use the *target* set to train our SPMBD model and reduce the 64,000-dimensional SPMBD feature to 500-dimensional by WPCA. Table 4 and Fig. 11 show the verification rate (VR) at FAR=0.01 and ROC curves of different descriptors for the above two testing sets, respectively.

We can observe that the proposed SPMBD method significantly outperforms the other methods for these two protocols. Specially, it improves state-of-the-art face descriptor CBFD by about 7.5% and 11.5% VR on all-images and frontal-images query set, respectively. This result again demonstrates that our SPMBD model can effectively learn discriminative and data-adaptive feature representations even faced with point-and-shoot scenario. It shows that our method can solve the face tag recommendation problem based on social networks. The quantization loss term and the bit-independence term make our SPMDB descriptor more discriminative, and the sparsity term avoid the overfitting problem and improve the robustness of our SPMDB feature. It is

the reason why SPMBD obtains the better results compared to other methods. This experiment also indicates our SPMBD method is capable of solving FR problem in real scenarios as long as our model can be trained with large enough training samples.

Table 4: Verification Rate (%) at FAR=0.01 on PaSC dataset for all images and frontal images scenarios

method	all	frontal	year
LBP [6]	17.6	29.6	2006
LRPCA [41]*	10.0	19.0	2011
CohortLDA [42]*	8.0	22.0	2012
LPQ [43] *	13.2	23.1	2012
BSIF [40]*	14.3	24.9	2014
DFD [9]	21.5	36.1	2014
CBFD [2]	19.4	36.0	2015
IQBC [32]	21.2	38.8	2016
SPMBD	**26.9**	**47.5**	-

* The results of other methods are directly cited from the original papers.

5.4. Cross-dataset Evaluation

In this section, we intend to investigate the generalization ability of our method with cross-dataset evaluation, i.e., we learn the SPMBD model from other datasets and test them on LFW dataset.

The LFW dataset is released for investigating the performance of uncontrolled FR scenario. This dataset contains over 13,000 face images and all of subjects are celebrity. Though these are some complicated variabilities, such as illumination, expression, pose, occlusion and aging, they are filmed by professional cameraman. Therefore, there are not poor focus and motion blur

problems in this dataset. According to the standard evaluation protocol, there are 300 positive face pairs and 300 negative face pairs in each fold and 10 folds are used on the *view 2* dataset. Though LFW dataset has six evaluation settings, we choose the *unsupervised* setting in this paper. Because this setting does not rely upon classifier training or metric learning, it may be the best choice to evaluate the discriminative ability of various descriptors. To demonstrate the generalization of SPMBD methods, we train the SPMBD model on the constrained datasets (i.e., FERET and PEAL) and the unconstrained dataset (i.e., PaSC), then test SPMBD model on the LFW dataset.

In this experiment, we employ the LFW-a dataset [44] and crop them into 150×130, as shown in Fig. 12. Due to that the images of LFW-a have been aligned, we do not perform the operation of aligning in this database. We keep the parameters of SPMBD method used in the above sections and employ the WPCA to compress the dimensionality of learned SPMBD features into 1,000. Table 5 tabulates the Area Under Curve (AUC) percent of different algorithms under the unsupervised setting, where *SPMBD-FERET* indicates the result obtained by using the *SPMBD* model which learned from *FERET* dataset. The rest legends can be understood by the same manner. We can observe that the SPMBD method obtains competitive results with some existing methods, such as CBFD, MRF-MLBP and IQBC, and achieves improvement than LHS, LQP, DFD and LMA. Since the intra-class variations in PaSC dataset are richer than the FERET and PEAL datasets, it is nature that the PaSC-learned SPMBD model slightly outperforms other SPMBD models.

Our experimental configuration do not strictly comply with the standard evaluation protocol of *unsupervised setting*, because of the outside data are used for training model. **But this experimental configuration makes it more convincing that our SPMBD method has excellent effectiveness and generalization ability. Since the SPMBD model is trained by different datasets which are collected under different scenarios, our SPMBD still obtains the comparable performance with the recently existing methods based on LFW database learning. Although FERET and**

PEAL datasets are controlled datasets and their appearances (i.e., variabilities) are clearly different from those of LFW database, the SPMBD model trained from these databases still perform very well on the LFW dataset. This result demonstrates that our SPMBD method has enough generalization ability to solve the FR task under different scenarios.

Figure 12: The aligned and cropped face examples with size 150×130 from the LFW-a dataset.

5.5. Experimental Discussion

In this subsection, we discuss the experimental results about the above four datasets.

1. The learning-based methods, such as our SPMBD and DFD, generally outperform the hand-crafted methods, such as SIFT and LBP. It is hard to take account of all possible variations in the real FR scenarios during the designing process of hand-crafted methods, but those learning-based methods can learn data-oriented features during the training process. That is why our method performs better than hand-crafted methods.
2. The learning-based binary feature methods, including our SPMBD, IQBC and CBFD, outperform learning-based real-value feature methods (i.e., LQP, DFD and so on). This is because binary methods are more insensitive to local variabilities than real-valued methods, therefore, the impact of some kinds of intra-class variabilities can be reduced.
3. Our SPMBD method outperforms those state-of-the-art dense-based binary feature methods, including CBFD and IQBC. First, the quantization

Table 5: The AUC (%) Comparisons with the Other Methods on LFW dataset under the Unsupervised Setting

Methods	AUC	year
SIFT [45]*	54.07	2002
LBP [6]*	75.47	2006
LARK [46]*	78.30	2011
LHS [47]*	81.07	2012
LQP [31]*	87.00	2012
MRF-MLBP [48]*	89.94	2013
DFD [9]*	83.70	2014
CBFD [2]*	88.89	2015
JFL [8]*	91.03	2015
LMA [49]*	83.04	2016
IQBC [32]	87.15	2016
SPMBD-FERET	89.53	-
SHBC-PEAL	87.94	-
SHBC-PaSC	89.83	-

* The results of other methods are directly cited from the original papers.

loss term in our objective function seeks to minimize the quantization error between the learned binary codes and original data. Second, the sparsity term reduces the possible overfitting during the encoding process, so that the learned feature also performs well on the testing set. Third, the bit-independence term ensures our feature to carry more identity information than other methods with fixed bit length.

4. The SPMBD method with WPCA obtains the best accuracy than existing methods on almost all testing datasets. Especially, our SPMBD achieves

95.7% and 97.0% accuracies on aging variation subsets of FERET dataset, 77.0% accuracy on *lighting* subset of CAS-PEAL-R1 dataset, and 26.9% and 47.5% VR at FAR=0.01 on *all* and *near-frontal* configurations of the PaSC dataset.

6. Conclusions And Future Work

In this paper, a sparse projections matrix binary descriptor is proposed, and it can be applied in various FR scenarios. There are three terms in our SPMBD objective function: the quantization loss term, the sparsity term and the bit-independence term. The experimental results demonstrate that our proposed SPMBD can obtain more discriminative information from raw data by introducing the quantization loss term and the bit-independence term, and can improve the robustness by employing the sparsity term.

According to the properties of our SPMBD method, we propose two interesting extensions of our SPMBD work:

1. Because of SPMBD method has enough generalized ability under different scenarios, we intend to further adjust the proposed SPMBD method and apply it to solve the current problems of surveillance based FR task.
2. The proposed SPMBD is an universal learning-based feature representation methods. Due to that its essence is still the description and representation for facial image, therefore, we can apply SPMBD into other face-related tasks, such as facial expression recognition.

Acknowledgment

This work was supported by the BUPT Excellent Ph.D. Students Foundation, the National Natural Science Foundation of China (Grants No. NSFC-61402046), Fund for Beijing University of Posts and Telecommunications (No.2013XZ10, 2013XD-04), Fund for the Doctoral Program of Higher Education of China (Grants No.20120005110002), the Academy of Finland,

Infotech Oulu, and Tekes Fidipro Program. This work is also partly supported by the Natural Science Foundation of China under the contract No. 61572205. The authors also want to acknowledge the supports of NVIDIA Corporation with the donation of the Tesla K40 and K80 GPU used for this research.

References

[1] H. Li, L. Zhang, B. Huang, X. Zhou, Sequential three-way decision and granulation for cost-sensitive face recognition, Knowledge-Based Systems 91 (2016) 241–251.

[2] J. Lu, V. Liong, X. Zhou, J. Zhou, Learning compact binary face descriptor for face recognition, Pattern Analysis and Machine Intelligence, IEEE Transactions on PP (99) (2015) 1–1. doi:10.1109/TPAMI.2015.2408359.

[3] D. Chen, X. Cao, F. Wen, J. Sun, Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3025–3032.

[4] C. Ding, J. Choi, D. Tao, L. S. Davis, Multi-directional multi-level dual-cross patterns for robust face recognition, IEEE transactions on pattern analysis and machine intelligence 38 (3) (2016) 518–531.

[5] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International journal of computer vision 60 (2) (2004) 91–110.

[6] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: Application to face recognition, Pattern Analysis and Machine Intelligence, IEEE Transactions on 28 (12) (2006) 2037–2041. doi:10.1109/TPAMI.2006.244.

[7] J. Lu, V. Erin Liong, J. Zhou, Simultaneous local binary feature learning and encoding for face recognition, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3721–3729.

[8] J. Lu, V. Liong, G. Wang, P. Moulin, Joint feature learning for face recognition, Information Forensics and Security, IEEE Transactions on 10 (7) (2015) 1371–1383. doi:10.1109/TIFS.2015.2408431.

[9] Z. Lei, M. Pietikainen, S. Z. Li, Learning discriminant face descriptor, Pattern Analysis and Machine Intelligence, IEEE Transactions on 36 (2) (2014) 289–302.

[10] Z. Zhang, L. Wang, Q. Zhu, S.-K. Chen, Y. Chen, Pose-invariant face recognition using facial landmarks and weber local descriptor, Knowledge-Based Systems 84 (2015) 78–88.

[11] C. Ding, D. Tao, Robust face recognition via multimodal deep face representation, IEEE Transactions on Multimedia 17 (11) (2015) 2049–2058. doi:10.1109/TMM.2015.2477042.

[12] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, Robust face recognition via sparse representation, Pattern Analysis and Machine Intelligence, IEEE Transactions on 31 (2) (2009) 210–227.

[13] W. Deng, J. Hu, J. Guo, In defense of sparsity based face recognition, in: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, 2013, pp. 399–406. doi:10.1109/CVPR.2013.58.

[14] L. Wei, F. Xu, A. Wu, Weighted discriminative sparsity preserving embedding for face recognition, Knowledge-Based Systems 57 (2014) 136–145.

[15] Y.-K. Lei, H. Han, X. Hao, Discriminant sparse local spline embedding with application to face recognition, Knowledge-Based Systems 89 (2015) 47 – 55. doi:http://dx.doi.org/10.1016/j.knosys.2015.06.016.
URL http://www.sciencedirect.com/science/article/pii/S0950705115002348

[16] Y. Xu, Z. Zhang, G. Lu, J. Yang, Approximately symmetrical face images for image preprocessing in face recognition and sparse representation based classification, Pattern Recognition 54 (2016) 68–82.

[17] Y. Li, R. Wang, H. Liu, H. Jiang, S. Shan, X. Chen, Two birds, one stone: Jointly learning binary code for large-scale face image retrieval and attributes prediction, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3819–3827.

[18] W. Zhang, J. Ji, J. Zhu, J. Li, H. Xu, B. Zhang, Bithash: An efficient bitwise locality sensitive hashing method with applications, Knowledge-Based Systems 97 (2016) 40–47.

[19] G. Gao, C. H. Liu, M. Chen, S. Guo, K. K. Leung, Cloud-based actor identification with batch-orthogonal local-sensitive hashing and sparse representation, IEEE Transactions on Multimedia 18 (9) (2016) 1749–1761. doi:10.1109/TMM.2016.2579305.

[20] F. X. Yu, S. Kumar, Y. Gong, S.-F. Chang, Circulant binary embedding, arXiv preprint arXiv:1405.3162.

[21] Y. Gong, S. Kumar, H. A. Rowley, S. Lazebnik, Learning binary codes for high-dimensional data using bilinear projections, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 484–491.

[22] T. Blumensath, M. E. Davies, Iterative thresholding for sparse approximations, Journal of Fourier Analysis and Applications 14 (5-6) (2008) 629–654.

[23] Y. Wang, J. Yang, W. Yin, Y. Zhang, A new alternating minimization algorithm for total variation image reconstruction, SIAM Journal on Imaging Sciences 1 (3) (2008) 248–272.

[24] Y. Xia, K. He, P. Kohli, J. Sun, Sparse projections for high-dimensional binary codes, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3332–3339.

[25] P. J. Phillips, H. Moon, S. Rizvi, P. J. Rauss, et al., The feret evaluation methodology for face-recognition algorithms, Pattern Analysis and Machine Intelligence, IEEE Transactions on 22 (10) (2000) 1090–1104.

[26] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, D. Zhao, The cas-peal large-scale chinese face database and baseline evaluations, Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 38 (1) (2008) 149–161.

[27] J. R. Beveridge, J. Phillips, D. S. Bolme, B. Draper, G. H. Givens, Y. M. Lui, M. N. Teli, H. Zhang, W. T. Scruggs, K. W. Bowyer, et al., The challenge of face recognition from digital point-and-shoot cameras, in: Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on, IEEE, 2013, pp. 1–8.

[28] G. B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled faces in the wild: A database for studying face recognition in unconstrained environments, Tech. rep., Technical Report 07-49, University of Massachusetts, Amherst (2007).

[29] A. Andoni, P. Indyk, Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions, Commun. ACM 51 (1) (2008) 117–122. doi:10.1145/1327452.1327494.

[30] W. Liu, J. Wang, S. Kumar, S.-F. Chang, Hashing with graphs, in: Proceedings of the 28th international conference on machine learning (ICML-11), 2011, pp. 1–8.

[31] S. U. Hussain, T. Napoléon, F. Jurie, Face recognition using local quantized patterns, in: British Machive Vision Conference, 2012, pp. 11–pages.

[32] L. Tian, C. Fan, Y. Ming, Learning iterative quantization binary codes for face recognition, Neurocomputing.

[33] B. Zhang, S. Shan, X. Chen, W. Gao, Histogram of gabor phase patterns (hgpp): A novel object representation approach for face recognition, Image Processing, IEEE Transactions on 16 (1) (2007) 57–68. doi:10.1109/TIP.2006.884956.

[34] D. Maturana, D. Mery, . Soto, Face recognition with decision tree-based local binary patterns, in: R. Kimmel, R. Klette, A. Sugimoto (Eds.), Computer Vision ACCV 2010, Vol. 6495 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2011, pp. 618–629. doi:10.1007/978-3-642-19282-1_49.
URL http://dx.doi.org/10.1007/978-3-642-19282-1_49

[35] B. Zhang, Y. Gao, S. Zhao, J. Liu, Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor, Image Processing, IEEE Transactions on 19 (2) (2010) 533–544.

[36] Z. Lei, S. Liao, M. Pietikainen, S. Li, Face recognition by exploring information jointly in space, scale and orientation, Image Processing, IEEE Transactions on 20 (1) (2011) 247–256. doi:10.1109/TIP.2010.2060207.

[37] D. Maturana, D. Mery, A. Soto, Learning discriminative local binary patterns for face recognition, in: Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, IEEE, 2011, pp. 470–475.

[38] N.-S. Vu, A. Caplier, Enhanced patterns of oriented edge magnitudes for face recognition and image matching, Image Processing, IEEE Transactions on 21 (3) (2012) 1352–1365.

[39] W. Zhang, S. Shan, W. Gao, X. Chen, H. Zhang, Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition, in: Computer Vision, 2005. ICCV 2005.

Tenth IEEE International Conference on, Vol. 1, 2005, pp. 786–791 Vol. 1. doi:10.1109/ICCV.2005.147.

[40] J. Ylioinas, A. Hadid, J. Kannala, M. Pietikainen, An in-depth examination of local binary descriptors in unconstrained face recognition, in: Pattern Recognition (ICPR), 2014 22nd International Conference on, IEEE, 2014, pp. 4471–4476.

[41] P. J. Phillips, J. R. Beveridge, B. Draper, G. Givens, A. J. Toole, D. S. Bolme, J. Dunlop, Y. M. Lui, H. Sahibzada, S. Weimer, et al., An introduction to the good, the bad, & the ugly face recognition challenge problem, in: Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on, IEEE, 2011, pp. 346–353.

[42] Y. M. Lui, D. Bolme, P. J. Phillips, J. R. Beveridge, B. Draper, et al., Preliminary studies on the good, the bad, and the ugly face recognition challenge problem, in: Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on, IEEE, 2012, pp. 9–16.

[43] E. Rahtu, J. Heikkilä, V. Ojansivu, T. Ahonen, Local phase quantization for blur-insensitive image analysis, Image and Vision Computing 30 (8) (2012) 501–512.

[44] L. Wolf, T. Hassner, Y. Taigman, Effective unconstrained face recognition by combining multiple descriptors and learned background statistics, Pattern Analysis and Machine Intelligence, IEEE Transactions on 33 (10) (2011) 1978–1990.

[45] R. Verschae, J. Ruiz-Del-Solar, M. Correa, Face recognition in unconstrained environments: A comparative study, in: Workshop on Faces in'Real-Life'Images: Detection, Alignment, and Recognition, 2008.

[46] H. J. Seo, P. Milanfar, Face verification using the lark representation,

Information Forensics and Security, IEEE Transactions on 6 (4) (2011) 1275–1286.

[47] G. Sharma, S. ul Hussain, F. Jurie, Local higher-order statistics (lhs) for texture categorization and facial analysis, in: Computer Vision–ECCV 2012, Springer, 2012, pp. 1–12.

[48] S. R. Arashloo, J. Kittler, Efficient processing of mrfs for unconstrained-pose face recognition, in: Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on, IEEE, 2013, pp. 1–8.

[49] N. McLaughlin, J. Ming, D. Crookes, Largest matching areas for illumination and occlusion robust face recognition, IEEE Transactions on Cybernetics PP (99) (2016) 1–13. doi:10.1109/TCYB.2016.2529300.