2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Template Regularized Sparse Coding for Face Verification

Hongyu Xu*, Jingjing Zheng†, Azadeh Alavi* and Rama Chellappa*
*Department of Electrical and Computer Engineering and the Center for Automation Research, UMIACS
University of Maryland, College Park, MD, USA
Email: {hyxu, azadeh, rama}@umiacs.umd.edu
†General Electric Global Research, Niskayuna, NY, USA
Email: jingjing.zheng@ge.com

***Abstract*—In this paper, we propose a novel regularized sparse coding approach for template-based unconstrained face verification. Unlike traditional verification tasks, which require the evaluation on image-to-image or video-to-video pairs, template-based face verification/recognition methods can exploit training and/or gallery data containing a mixture of both images or videos from the person of interest. The proposed regularized sparse coding approach addresses the adaptation to training and gallery data using three steps. First, we construct a reference dictionary, which represents the training set. Then we learn the discriminative sparse codes of the templates for verification through the proposed template regularized sparse coding approach. Finally, we measure the similarity between templates. An efficient algorithm is employed to learn the template regularized sparse codes. Extensive experiments on the template-based verification benchmark dataset show that the proposed approach outperforms several state-of-the-art methods.**

I. Introduction

Face identification and verification are two main tasks in face-based biometrics. Face identification aims to recognize a person from a set of gallery (images or videos) and match the closest one to the probe, while verification determines whether a given pair of images or videos is from the same subject or not. In this paper, we address the unconstrained face verification/recognition problem where the face images have been acquired under significant variations in pose, expressions, lighting conditions and background.

Numerous methods have been proposed for improving the performance of face verification systems. Most existing approaches can be categorized into feature-based and metric learning-based methods. The first category, which includes LBP [1], SIFT [2], Fisher vector faces [3] and most recently the deep features [4], aims to derive robust and discriminative descriptors to represent face images. The common objective of the second category is to learn a good metric from the training data [5], [6], [7]. Some representative methods include cosine similarity metric learning [8], pairwise constrained component analysis [9] and logistic discriminant metric learning [10]. While dictionary learning techniques have shown impressive performance for face recognition [11], [12], [13], [14], [15], there are only a few reported works for the face verification problem [16], [17], [18].

Recently, template-based face verification problem has gained more popularity in computer vision community. The problem of traditional face verification is to verify whether two images or videos in a pair belong to the same subject over image-to-image pairs as in Labeled Face in the Wild dataset [19], or over video-to-video pairs as in the Youtube Faces database [20], whereas template-based face verification performs verification over **templates** as introduced in [21]. In this context, a template is a mixture of different media data such as images or frames sampled from multiple image sets or video clips from the person of interest. Template representation is important in real world as it provides more flexibility and longitudinal access control of data from subjects.

In this paper, we tackle the problem of template-based face verification by taking advantage of dictionary learning techniques. This is due to the fact that image or video samples could be well represented by a learned dictionary and corresponding sparse codes. Yet dictionary learning methods have not been exploited for template-based face verification. Two issues arise when existing dictionary-based methods such as [17] are used for template-based face verification. First, the dictionary learned by random sampling of the training data is not able to adequately represent the training set of face templates when several hundreds subjects are involved. Second, the sparse codes of all the samples from the same template are independently calculated, even though these samples are from the same subject. This may degrade the performance, especially when each template has significantly varying number of samples acquired from unconstrained environments. It is better to exploit this intra-class relationship among samples from the same template.

To overcome the limitations discussed above, we propose a novel template regularized sparse coding framework for template-based unconstrained face verification. The proposed approach consists of three steps. First, we construct a reference dictionary to adequately represent the training set. Then we exploit the intra-class relationship of the template by regularizing the sparse codes of the samples in one template to be similar, which results in more discriminative sparse codes. Finally, we measure the similarity between templates. To summarize, we make the following contributions:

- We are the first to propose a dictionary learning framework for template-based face verification problem.
- Our method learns a reference dictionary, which ad-

978-1-5090-4847-2/16/$31.00 ©2016 IEEE

equately represents the training set. Furthermore, we construct two template adaptive dictionaries to adapt the pair of templates.
- We propose a novel template regularized sparse coding method, which is able to capture the information in the samples in one template. An efficient algorithm is employed to learn the discriminative sparse codes.
- We demonstrate that the proposed framework outperforms several state-of-the-art methods on benchmark dateset for template-based face verification.

II. Related Work

Template-based Face Verification: Several state-of-the-arts methods for template-based face verification are briefly reviewed [22], [23], [24]. [22] addressed the template-based face verification problem through Joint Bayesian Metric Learning [25], [26] of deep CNN features. The triplet similarity embedding method [23] learned an embedding matrix, which projects the original feature to a low-dimensional space. Template adaptation [24] learned two linear SVM classifiers, where each of them is designed using the positive features from one template in the pair to the large negative features from the training set. Then the final similarity score is calculated by fusing the two SVM margins evaluated on the other mated template.

Dictionary Learning: Dictionary learning has shown impressive performance in face recognition [11], [12], [13], [14], [15]. However, only a few works are reported for the face verification problem [17], [16], [18]. One of the first methods [17] which adopted dictionary learning for face verification measured the similarity between the pair of images over the sparse codes using a reference dictionary. Subsequently, this work was extended by learning the local sparse codes from the patches of the face images. Although effective, learning patch-based sparse codes is sensitive to local perturbations. [16] generalized the dictionary learning framework to verification problems by adding a pairwise constraint. However, it suffers from high computational complexity. Furthermore, all these methods addressed the verification problem in image-to-image settings and are not directly applicable to template-based face verification [21].

III. Proposed Method

In this section, we provide a detailed description of our template regularized sparse coding approach for the template-based face verification problem.

A. Task and Overall Approach

The definition of template-based face verification can be simplified as follows: given a training set and a pair of templates from the test set, the objective is to verify whether the pair of templates are from the same subject or not.

Our approach for template-based face verification that (1) learns a reference dictionary $\mathbf{D}^R$ (with the help of hierarchical clustering), and (2) learns more discriminative sparse codes for verification purposes through the proposed *template regularized sparse coding* method, and (3) defines two distance-measures between template pairs through *reference score* and *template adaptive score* for computing the final similarity score.

The proposed approach consists of three steps. First, we learn two types of dictionaries: a reference dictionary and template adaptive dictionaries. The reference dictionary is learned only from the training set, which is disjoint from any test templates. The reference dictionary is used for learning the sparse representations of the test templates. Two template adaptive dictionaries are constructed by augmenting the reference dictionary with each template in the test pair respectively. Adding only one template to construct the template adaptive dictionary would result in adapting the reference dictionary to better represent the other templates from the same subject.

Second, we perform sparse coding both on the reference dictionary and template adaptive dictionaries to obtain two types of sparse representations. In particular, we regularize the sparse codes of the samples in one template of the test pair to be similar to each other.

Third, by using the two sparse codes obtained as discussed above, we compute two different similarity scores: reference score and template adaptive score. The reference score is defined as the similarity between the sparse codes of two templates with respect to the reference dictionary. Template adaptive score measures the difference between two types of sparse codes of each template in the pair with respect to two types of dictionaries.

The motivation behind the template adaptive score is that, if two templates in a pair are from the same subject, then the sparse coding coefficients of samples from one template corresponding to the augmented part (the added dictionary atoms from the other template) will have a significantly high value, while other coefficients corresponding to the reference dictionary will be smaller. On the other side, if the two templates are not from the same subject, the regularized sparse codes of two templates will not change significantly. Therefore, a higher template adaptive score indicates that the template pair, very likely comes from the same subject.

We first present the notations used in the rest of paper. Let $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_P] \in \mathbb{R}^{d \times P}$ be the general template data matrix, where P is the total number of samples in the template (P varies from template to template). Each $\mathbf{x}_i \in \mathbb{R}^d, 1 \leq i \leq P$ is the feature encoded from image or video frames in the template with unit l_2-norm. We denote the training set as $\mathcal{T}$, and a pair of templates $\mathbf{X}^A = [\mathbf{x}_1^A, ..., \mathbf{x}_{P_A}^A] \in \mathbb{R}^{d \times P_A}$ and $\mathbf{X}^B = [\mathbf{x}_1^B, ..., \mathbf{x}_{P_B}^B] \in \mathbb{R}^{d \times P_B}$ from the test set.

B. Reference Dictionary and Template Adaptive Dictionaries Learning

The first step in the method is to learn a reference dictionary. Let n be the number of subjects in the training set and n_i be the number of templates from subject $i (\in [1, n])$. We define the data matrix $\mathbf{T}^i = [\mathbf{X}_1^i, ..., \mathbf{X}_{n_i}^i]$ to represent subject i, where $\mathbf{X}_j^i (j \in [1, n_i])$ is the j-th template from person i.

Algorithm 1 Adaptive selection of k_i representative samples

Input: Training data $\mathbf{T}^i = [\mathbf{X}^i_1, ..., \mathbf{X}^i_{n_i}]$ from subject i, stopping threshold τ.
Initialize: $k = 1$
while not converged **do**
Increase k to $k + 1$
Find k mediods $\{\mathbf{c}^i_1, ..., \mathbf{c}^i_k\}$ and corresponding clusters $\{\mathcal{C}^i_1, ..., \mathcal{C}^i_k\}$ by using "k-mediods" clustering algorithm [27].
Compute r as follows:

$$r = \max_{1 \le m \le k} \max_{\mathbf{x}^i_j \in \mathbf{C}^i_m} \|\mathbf{x}^i_j - \mathbf{c}^i_m\|_2 \tag{1}$$

Check the convergence condition: $r \le \tau$
end while
Output: k_i and representative samples $\{\mathbf{c}^i_1, ..., \mathbf{c}^i_{k_i}\}$

Consequently, we represent the entire training set by $\mathcal{T} = [\mathbf{T}^1, ..., \mathbf{T}^n]$.

A good reference dictionary should be able to represent the training set with a compact set of dictionary atoms. To make the reference dictionary adequately represent the training set, we perform hierarchical adaptive clustering on the training set. More specifically, for each subject data matrix $\mathbf{T}^i, i \in [1, n]$, we adaptively determine the value of k_i and select k_i most representative samples by alternating the following two steps: (a) Increasing k to $k + 1$ (b) Applying the "k-medoids" algorithm [27] on $\mathbf{T}^i$ until the stopping criteria in (1) is satisfied. The alternating procedure is illustrated in Algorithm 1.

After we select k_i representative samples from $\mathbf{T}^i, i \in [1, n]$ subject by subject, the reference dictionary $\mathbf{D}^R$ is constructed by concatenating all the representative samples learned in Algorithm 1, *i.e.* $\mathbf{D}^R = [\mathbf{c}^1_1, ..., \mathbf{c}^1_{k_1} | \ ... \ | \mathbf{c}^n_1, ..., \mathbf{c}^n_{k_n}]$. We can rewrite the reference dictionary as $\mathbf{D}^R = [\mathbf{d}^R_1, ..., \mathbf{d}^R_N] \in \mathbb{R}^{d \times N}$, where $N = k_1 + ... + k_n$ is the total number of atoms (columns) in the dictionary.

Furthermore, given a test pair of templates $\mathbf{X}^A = [\mathbf{x}^A_1, ..., \mathbf{x}^A_{P_A}] \in \mathbb{R}^{d \times P_A}$ and $\mathbf{X}^B = [\mathbf{x}^B_1, ..., \mathbf{x}^B_{P_B}] \in \mathbb{R}^{d \times P_B}$, we construct two template adaptive dictionaries $\mathbf{D}^A$, $\mathbf{D}^B$ by augmenting the reference dictionary with samples from each template as follows: $\mathbf{D}^A = [\mathbf{D}^R | \mathbf{X}^B] \in \mathbb{R}^{d \times (N + P_B)}$ and $\mathbf{D}^B = [\mathbf{D}^R | \mathbf{X}^A] \in \mathbb{R}^{d \times (N + P_A)}$.

C. Template Regularized Sparse Coding

In this section, we present our template regularized sparse coding algorithm for the reference dictionary $\mathbf{D}^R$ and template adaptive dictionaries $\mathbf{D}^A$ and $\mathbf{D}^B$. We learn the sparse codes of the samples in one template by regularizing them to be similar as they are all from the same subject. For simplicity of notation, we drop the superscript in $\mathbf{D}^R$, $\mathbf{D}^A$ and $\mathbf{D}^B$ and denote the given dictionary as $\mathbf{D}$. Let the template data matrix be $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_P] \in \mathbb{R}^{d \times P}$. The template regularized sparse codes are obtained as follows:

$$\begin{aligned} \mathbf{Z}^* = \arg\min_{\mathbf{Z}} \sum_{i=1}^{P} (\|\mathbf{x}_i - \mathbf{D}\mathbf{z}_i\|_2^2 + \lambda_1 \|\mathbf{z}_i\|_1 + \lambda_2 \|\mathbf{z}_i\|_2^2) \\ + \frac{\beta}{2} \sum_{i,j=1}^{P} (\|\mathbf{z}_i - \mathbf{z}_j\|_2^2 w_{i,j}) \end{aligned} \tag{2}$$

where $\mathbf{Z} = [\mathbf{z}_1, ..., \mathbf{z}_P]$ are the corresponding sparse codes of $\mathbf{X}$ and λ_1, λ_2, β are the regularization parameters. The term $\|\mathbf{z}_i\|_1$ is the sparsity regularization term and the term $\|\mathbf{z}_i\|_2^2$ ensures the stability of the solution as in [28]. The last term is called the *template regularization term*, which sums the weighted difference of sparse codes of any two samples in the template. Let $\mathbf{W}$ be the matrix with entry $w_{i,j}$ in the i-th row and j-th column.

Constructing Matrix W: Given the sparse codes $\mathbf{z}_i$ and $\mathbf{z}_j$ of any two samples $\mathbf{x}_i$ and $\mathbf{x}_j$, $w_{i,j}$ is defined as follows:

$$w_{i,j} = \begin{cases} e^{-\frac{1}{2}\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}, \text{ if } i \neq j \\ 0, \quad \text{otherwise} \end{cases} \tag{3}$$

It is inversely proportional to the Euclidean distance between their original feature (*i.e.* $\|\mathbf{x}_i - \mathbf{x}_j\|_2$). It means that when two samples are very close or similar in the original feature space, the penalty associated with the difference of their sparse codes will be large. As the pair of templates could have different template size, in order to reduce the effect of the template size, we further normalize each column in $\mathbf{W}$ by its l_2-norm.

Optimization: We now discuss the optimization of (2). Equation (2) is rewritten as

$$\begin{aligned} \mathbf{Z}^* = \arg\min_{\mathbf{Z}} \sum_{i=1}^{P} (\|\mathbf{x}_i - \mathbf{D}\mathbf{z}_i\|_2^2 + \lambda_1 \|\mathbf{z}_i\|_1 + \lambda_2 \|\mathbf{z}_i\|_2^2) \\ + \beta \operatorname{Tr}(\mathbf{Z}^T \mathbf{Z} \mathbf{L}) \end{aligned} \tag{4}$$

where $\mathbf{L}$ is the Laplacian matrix $\mathbf{L} = \mathbf{A} - \mathbf{W}$ and $\mathbf{A}$ is a diagonal matrix whose diagonal elements are the sum of row elements of $\mathbf{W}$, i.e. $a_{i,i} = \sum_{j=1}^{P} w_{i,j}$.

Motivated by [16], [29], we optimize $\mathbf{z}_i$ in a column by column fashion. Given dictionary $\mathbf{D}$, when updating $\mathbf{z}_i$ by fixing other $\mathbf{z}_j (j \neq i)$, the objective function of (2) with respect to $\mathbf{z}_i$ is reduced to:

$$\begin{aligned} \mathbf{z}^*_i = \arg\min_{\mathbf{z}_i} \|\mathbf{x}_i - \mathbf{D}\mathbf{z}_i\|_2^2 + \lambda_1 \|\mathbf{z}_i\|_1 + \lambda_2 \mathbf{z}_i^T \mathbf{z}_i \\ + 2\beta \mathbf{z}_i^T (\mathbf{Z}\mathbf{L}_i) - \beta \mathbf{z}_i^T \mathbf{z}_i \mathbf{L}_{i,i} \end{aligned} \tag{5}$$

The minimization of (5) is a $\mathbf{L}_1$-regularized least squares problem and we compute $\mathbf{z}_i$ by feature-sign search algorithm proposed in [29].

The analytical solution of $\mathbf{z}_i$ could be derived by setting the first derivative of (5) with respect to $\mathbf{z}_i$ to be zero:

$$\mathbf{z}^*_i = [\mathbf{D}^T\mathbf{D} + (\lambda_2 + \beta \mathbf{L}_{i,i})\mathbf{I}]^{-1} (\mathbf{D}^T \mathbf{x}_i - \beta \sum_{\substack{k=1 \\ k \neq i}}^{P} \mathbf{z}_k \mathbf{L}_{k,i} - \frac{1}{2}\lambda\boldsymbol{\theta}) \tag{6}$$

where $\boldsymbol{\theta}$ is the coefficient sign vector of $\mathbf{z}_i$. We choose a small value of β to ensure that the Hessian matrix $[\mathbf{D}^T\mathbf{D} + (\lambda_2 + \beta\mathbf{L}_{i,i})\mathbf{I}]$ is positive semidefinite, which guarantees the convexity of (4).

Thus, we learn four sets of regularized sparse codes of the templates ($\mathbf{X}^A$ or $\mathbf{X}^B$) with respect to the reference dictionary $\mathbf{D}^R$ and the template adaptive dictionary ($\mathbf{D}^A$ or $\mathbf{D}^B$) which are denoted as follows:

$$\begin{aligned}
&\mathbf{Z}^A = [\mathbf{z}_1^A, ..., \mathbf{z}_{P_A}^A] \in \mathbb{R}^{N\times P_A} \text{ coded with } \mathbf{D}^R \\
&\tilde{\mathbf{Z}}^A = [\tilde{\mathbf{z}}_1^A, ..., \tilde{\mathbf{z}}_{P_A}^A] \in \mathbb{R}^{(N+P_B)\times P_A} \text{ coded with } \mathbf{D}^A \\
&\mathbf{Z}^B = [\mathbf{z}_1^B, ..., \mathbf{z}_{P_B}^B] \in \mathbb{R}^{N\times P_B} \text{ coded with } \mathbf{D}^R \\
&\tilde{\mathbf{Z}}^B = [\tilde{\mathbf{z}}_1^B, ..., \tilde{\mathbf{z}}_{P_B}^B] \in \mathbb{R}^{(N+P_A)\times P_B} \text{ coded with } \mathbf{D}^B
\end{aligned} \tag{7}$$

D. Reference Score and Template Adaptive Score

After we learn the template regularized sparse representations using (2), we evaluate how similar the test templates are, by computing the reference score and the template adaptive score. The reference score is defined as the average of the cosine similarity between all the sample pairs from the two templates as follows:

$$\text{REF}(\mathbf{X}^A, \mathbf{X}^B) = \frac{1}{P_A \times P_B} \sum_{i=1}^{P_A}\sum_{j=1}^{P_B} \cos(\mathbf{z}_i^a, \mathbf{z}_j^b) \tag{8}$$

where $\cos(\mathbf{z}_i^A, \mathbf{z}_j^B)(i \in [1, P_A], j \in [1, P_B])$ is computed as the cosine similarity between two sparse codes as in [8]

$$\cos(\mathbf{z}_i^A, \mathbf{z}_j^B) = \frac{(\mathbf{z}_i^A)^T\mathbf{z}_j^B}{\|\mathbf{z}_i^A\|_2\|\mathbf{z}_j^B\|_2} \tag{9}$$

In addition, in order to exploit the full power of the template regularized sparse codes $\tilde{\mathbf{Z}}^A$ and $\tilde{\mathbf{Z}}^B$, we also compute the template adaptive score of the template pair [17]. Following the notation in (7), let us first define the sample adaptive score of one sample $\mathbf{x}_i^A$ in the template $\mathbf{X}^A$ as

$$\text{adapt}(\mathbf{x}_i^A) = 1 - \cos(\mathbf{z}_i^A, \tilde{\mathbf{z}}_i^A(1{:}\text{N})) \tag{10}$$

where cos metric is defined in (9). Similar to sample $\mathbf{z}_i^B$ in the template $\mathbf{X}^B$, we have $\text{adapt}(\mathbf{x}_i^B) = 1 - \cos(\mathbf{z}_i^B, \tilde{\mathbf{z}}_i^B(1{:}\text{N}))$. Note that the higher sample adaptive score indicates more significant change from the sparse code.

Therefore, the template adaptive score of the template pair is computed as:

$$\text{ADP}(\mathbf{X}^A, \mathbf{X}^B) = \frac{1}{P_A \times P_B} \sum_{i=1}^{P_a}\sum_{j=1}^{P_b} \frac{1}{2}[\text{adapt}(\mathbf{x}_i^A) + \text{adapt}(\mathbf{x}_j^B)] \tag{11}$$

Finally, the similarity score of the tested template pair is computed as the average of the reference score and the template adaptive score.

IV. Experiments

In this section, we present the results of the proposed dictionary approach on the challenging IARPA Janus Benchmark A(IJB-A) [21] dataset. We will first introduce the dataset and experimental settings. This is then followed by a discussion of the experimental results.

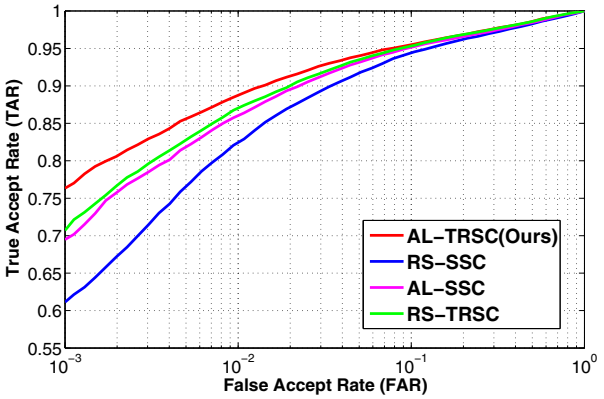


Fig. 1. The average ROC curves of different dictionary learning and sparse coding strategies for the IJB-A [21] verification protocol over 10 splits

Methods	FAR = 0.001	FAR = 0.01	FAR = 0.1
RS-SSC	0.613±0.059	0.824±0.026	0.944±0.007
AL-SSC	0.696±0.057	0.860±0.016	0.950±0.005
RS-TRSC	0.713±0.041	0.869±0.014	0.952±0.006
AL-TRSC(Ours)	**0.769±0.038**	**0.885±0.011**	**0.955±0.003**

TABLE I
VERIFICATION ACCURACY COMPARISON OF DIFFERENT DICTIONARY LEARNING AND SPARSE CODING STRATEGIES FOR THE IJB-A DATASET [21]. THE TRUE ACCEPT RATES(TAR) AT FALSE ACCEPT RATE (FAR) OF 0.001, 0.01 AND 0.1 ARE REPORTED.

A. Dataset and Settings

The IARPA Janus Benchmark A(IJB-A) [21] dataset contains $5,397$ images and $2,042$ videos, which sampled to $20,412$ frames from total 500 subjects. Each subject has 11.4 images and 4.2 video clips on average. The smallest representation unit of each subject constitutes the **template**, which comprises a mixture of still images and sampled video frames.

The evaluation of verification protocol from IJB-A is over 10 splits. Each split consists of training and testing sets without any overlapping subjects between them. The test set in one split contains around $11,748$ pairs of templates ($1,756$ genuine and $9,992$ poster pairs). True Accept Rates(TAP) at different False Accept Rates(FAR) are reported in the evaluation metric.

In our experiment, the faces are represented with deep features extracted using the network discussed in [22]. More specifically, the deep CNN network is trained on the CASIA-WebFace dataset [30] with non-overlapped $490,356$ face images of $10,548$ subjects to IJB-A dataset. We use the network presented in [22] to extract the 320-dimensional feature vector for each template in training and testing sets. Furthermore, following the setting in [24], in order to reduce the effect caused by the unbalanced size of different media (images or videos) in one template, we compute the mean feature to represent one video by averaging the features extracted from the same video clips. Finally, all the features in one template are normalized to have unit l_2-norm, which we call it the template media average features. The template media average features are used in all the experiments.

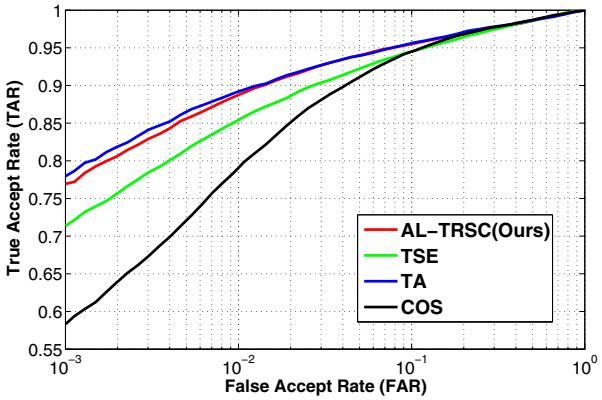


Fig. 2. The average ROC curves of state-of-the-art and baseline methods for the IJB-A [21] verification protocol over 10 splits

Methods	FAR = 0.001	FAR = 0.01	FAR = 0.1
GOTS	0.198±0.008	0.406±0.014	0.627±0.012
COS	0.586±0.059	0.791±0.052	0.942±0.008
[22]	-	0.818±0.037	**0.961±0.010**
TSE [23]	0.718±0.039	0.855±0.019	0.945±0.005
TA [24]	**0.779±0.023**	**0.889± 0.012**	***0.955±0.007***
AL-TRSC(Ours)	***0.769±0.038***	***0.885±0.011***	***0.955±0.003***

TABLE II
VERIFICATION ACCURACY COMPARISON WITH STATE-OF-THE-ART APPROACHES FOR THE IJB-A DATASET [21]. THE TRUE ACCEPT RATES(TAR) AT FALSE ACCEPT RATE (FAR) OF 0.001, 0.01 AND 0.1 ARE REPORTED.

B. Results and Analysis

We perform two series of experiments to evaluate our approach for template-based face verification on IARPA Janus Benchmark A(IJB-A) [21] dataset.

Comparison of Different Dictionary Learning and Sparse Coding Strategies. To demonstrate the improvement of our approach (**AL-TRSC**) over [17] in both dictionary learning and template regularized sparse coding, we compare it with three methods:

- Random Sample + Single Sparse Coding (RS-SSC) [17]. We randomly select samples from the training set to generate the reference dictionary and independently compute the sparse codes of all the samples without the regularization term in (2).
- Adaptive Leaning + Single Sparse Coding (AL-SSC). We learn the reference dictionary as described in Section III-B, followed by the same sparse coding strategy above.
- Random Sample + Template Regularized Sparse Coding (RS-TRSC). We construct the reference dictionary by random sampling of the training set. However, we learn the template regularized sparse codes described in Section III-C

We plot the average ROC curves in Figure 1 of the four methods for the IJB-A dataset over 10 splits. In addition, we report the average TAR at FAR= 0.001, 0.01 and 0.1 in Table I. First, our method (**AL-TRSC**) consistently outperforms AL-SSC, RS-TRSC and RS-SSC by a large margin. Compared with RS-TRSC, the reference dictionary, which is learned adaptively, is able to better represent the training set than random sampling. The AL-SSC algorithm only learns the sparse codes of all samples without template regularization. However, our method regularizes the sparse codes from one template to be close, which yields more discriminative sparse codes across template pairs. It is also noted that both AL-SSC and RS-TRSC achieve improvements over RS-SSC [17]. This demonstrates that both adaptive reference dictionary learning and template regularized sparse coding are indispensable for template-based face verification.

Comparison with State-of-the-art Approaches In order to evaluate the effectiveness of our approach (**AL-TRSC**) for template-based face verification, we further compare it with several state-of-the-art listed next: Joint Bayesian Metric Learning [22], Triplet Similarity Embedding (TSE) [23], Template Adaptation (TA) [24]. All the methods are implemented following the algorithm except [22]. The parameters are tuned based on the settings reported in their papers. We evaluate all the methods on the template media average features as a fair comparison, which is the same as the setting in [24][1].

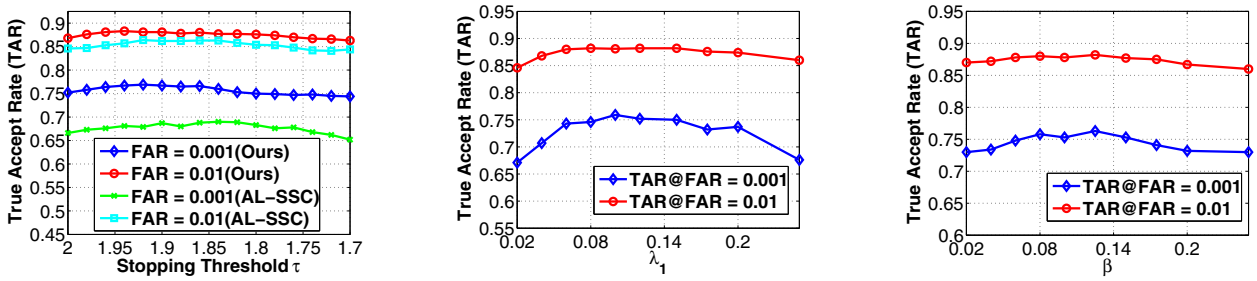
In addition, we also compare it with two baseline methods, the first one, COS computes the cosine similarity [8] from all the pair samples of two templates and average them to get the final similarity score between the two templates. The second baseline GOTS is from the commercial off-the-shelf matchers mentioned in the NIST FRVT study [31].

We plot the IJB-A average ROC curves over 10 splits of TSE [23], TA [24] and COS [8] in Figure 2. Furthermore, we also report the average TAR at FAR= 0.001, 0.01 and 0.1 in Table II. All the methods [22], [23], [24] and ours improve the performance over COS and GOTS by a wide margin. Moreover, it can been seen that our method outperforms metric-based methods [22], [23] and achieves results comparable to [24], which demonstrates the effectiveness of the proposed approach.

Parameter Sensitivity: In order to evaluate the effects of the stopping threshold τ in (1) and the hyper-parameters λ_1, λ_2, β in (2) of our method, we run different choice of parameters and plot the TAR with respect to the parameters at FAR = 0.001 and 0.01 in Figure 3.

Firstly, in Figure 3(a), it can been seen that both AL-SSC and AL-TRSC exhibit the same tendency with respect to τ. We observe that as τ decreases from 2.0 to 1.9, the verification performance improves. It is also interesting to note that when $\tau = 1.7$, the performance degenerates. With a large-sized reference dictionary, some atoms selected from the samples may not be useful for verification, thus affecting the regularized sparse coding. The final dictionary size is inverse proportional to the stopping threshold τ, and in order to balance the time and accuracy, we choose $\tau \in [1.85, 1.95]$

[1]Note that result DCNN$_{ft+m+c}$ reported in [22] didn't use template media average features, all the other methods TA [24], TSE [23] and COS are evaluated on the same template media average features



(a) stopping threshold τ (b) parameter λ_1 (c) parameter β

Fig. 3. The effects of stopping threshold τ, hyper-parameters λ_1 and β on IRAPA IJB-A dataset [21].

in all the experiments, which yields the reference dictionary size to be between 400 and 500.

We also evaluate our method by varying parameters λ_1 and β as shown in Figures 3(b) and 3(c). It is observed that the performance is more sensitive to the choice of λ_1, which is associated with sparse penalty. Our results are reported by setting $\lambda_1 \in [0.08, 0.12]$ and $\beta = \{0.15, 0.1\}$. In addition, our approach is insensitive to the regularization parameter λ_2, which is set to 0.05 throughout all the experiments.

V. Conclusion

In this paper, we presented a novel template regularized sparse coding approach for template-based face verification. First, we adaptively learned a reference dictionary to adequately represent the training set. Then template adaptive dictionaries are generated by adapting the reference dictionary with the test template pair. Second, we performed template regularized sparse coding on all the dictionaries to derive the discriminative template sparse codes for verification purpose. Finally, both the reference score and template adaptive score are used to measure the similarity of the pair templates. We extensively evaluated our approach on the benchmark IARPA IJB-A dataset for template-based face verification. The experimental results clearly demonstrate competitive performance over the state-of-the-art.

VI. Acknowledgement

This research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 2014-14071600012. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

References

[1] 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, 2006.

[2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[3] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Fisher vector faces in the wild," in *Proceedings of the British Machine Vision Conference (BMVC)*, 2013.

[4] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1701–1708.

[5] X. Cai, C. Wang, B. Xiao, X. Chen, and J. Zhou, "Deep nonlinear metric learning with independent subspace analysis for face verification," in *Proceedings of the 20th ACM Multimedia Conference (MM)*, 2012, pp. 749–752.

[6] Z. Cui, W. Li, D. Xu, S. Shan, and X. Chen, "Fusing robust face region descriptors via multiple metric learning for face recognition in the wild," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 3554–3561.

[7] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2007, pp. 209–216.

[8] H. V. Nguyen and L. Bai, "Cosine similarity metric learning for face verification." in *Asian Conference on Computer Vision (ACCV)*, vol. 6493, 2010, pp. 709–720.

[9] A. Mignon and F. Jurie, "PCCA: A new approach for distance learning from sparse pairwise constraints," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 2666–2672.

[10] M. Guillaumin, J. J. Verbeek, and C. Schmid, "Is that you? metric learning approaches for face identification," in *IEEE International Conference on Computer Vision (ICCV)*, 2009, pp. 498–505.

[11] L. Ma, C. Wang, B. Xiao, and W. Zhou, "Sparse representation for face recognition based on discriminative low-rank dictionary learning," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 2586–2593.

[12] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 2691–2698.

[13] G. Zhang, R. He, and L. S. Davis, "Jointly learning dictionaries and subspace structure for video-based face recognition," in *Asian Conference on Computer Vision (ACCV)*, 2014, pp. 97–111.

[14] H. Xu, J. Zheng, A. Alavi, and R. Chellappa, "Learning a structured dictionary for video-based face recognition," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016, pp. 1–9.

[15] H. Xu, J. Zheng, and R. Chellappa, "Bridging the domain shift by domain adaptive dictionary learning," in *Proceedings of the British Machine Vision Conference (BMVC)*, 2015, pp. 96.1–96.12.

[16] H. Guo, Z. Jiang, and L. S. Davis, "Discriminative dictionary learning with pairwise constraints," in *Asian Conference on Computer Vision (ACCV)*, 2012, pp. 328–342.

[17] H. Guo, R. Wang, J. Choi, and L. S. Davis, "Face verification using sparse representations," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2012, pp. 37–44.

[18] C. Duan, C. Chiang, and S. Lai, "Face verification with local sparse representation," *IEEE Signal Processing Letters*, vol. 20, no. 2, pp. 177–180, 2013.

[19] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, October 2007.
[20] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 529–534.
[21] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, and A. K. Jain, "Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1931–1939.
[22] J. Chen, V. M. Patel, and R. Chellappa, "Unconstrained face verification using deep CNN features." in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016.
[23] S. Sankaranarayanan, A. Alavi, and R. Chellappa, "Triplet similarity embedding for face verification," *CoRR*, vol. abs/1602.03418, 2016. [Online]. Available: http://arxiv.org/abs/1602.03418
[24] N. Crosswhite, J. Byrne, O. M. Parkhi, C. Stauffer, Q. Cao, and A. Zisserman, "Template Adaptation for Face Verification and Identification," *CoRR*, vol. abs/1603.03958, 2016. [Online]. Available: http://arxiv.org/abs/1603.03958
[25] D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun, "Bayesian face revisited: A joint formulation," in *European Conference on Computer Vision (ECCV)*, 2012, pp. 566–579.
[26] X. Cao, D. P. Wipf, F. Wen, G. Duan, and J. Sun, "A practical transfer learning algorithm for face verification," in *IEEE International Conference on Computer Vision (ICCV)*, 2013, pp. 3208–3215.
[27] H. Park and C. Jun, "A simple and fast algorithm for k-medoids clustering," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3336–3341, 2009.
[28] J. Mairal, F. R. Bach, and J. Ponce, "Task-driven dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 791–804, 2012.
[29] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*, 2007, pp. 801–808.
[30] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *CoRR*, vol. abs/1411.7923, 2014. [Online]. Available: http://arxiv.org/abs/1411.7923
[31] P. Grother and M. Ngan, "Face recognition vendor test (FRVT): Performance of face identification algorithms," *NIST Interagency Report 8009*, May 2014.