Image Set based Collaborative Representation for Face Recognition

Pengfei Zhu, *Student Member, IEEE,* Wangmeng Zuo, *Member, IEEE,* Lei Zhang, *Member, IEEE,* Simon C.K. Shiu, *Member, IEEE,* David Zhang, *Fellow, IEEE*

***Abstract*—With the rapid development of digital imaging and communication technologies, image set based face recognition (ISFR) is becoming increasingly important. One key issue of ISFR is how to effectively and efficiently represent the query face image set by using the gallery face image sets. The set-to-set distance based methods ignore the relationship between gallery sets, while representing the query set images individually over the gallery sets ignores the correlation between query set images. In this paper, we propose a novel image set based collaborative representation and classification method for ISFR. By modeling the query set as a convex or regularized hull, we represent this hull collaboratively over all the gallery sets. With the resolved representation coefficients, the distance between the query set and each gallery set can then be calculated for classification. The proposed model naturally and effectively extends the image based collaborative representation to an image set based one, and our extensive experiments on benchmark ISFR databases show the superiority of the proposed method to state-of-the-art ISFR methods under different set sizes in terms of both recognition rate and efficiency.**

***Index Terms*—image set, collaborative representation, set to sets distance, face recognition.**

I. Introduction

Image set based classification has been increasingly employed in face recognition [1], [2], [3], [4], [5], [6], [7], [8], [9], [10] and object categorization [11], [12] in recent years. Due to the rapid development of digital imaging and communication techniques, now image sets can be easily collected from multi-view images using multiple cameras [11], long term observations [7], personal albums and news pictures [13], etc. Meanwhile, image set based face recognition (ISFR) has shown superior performance to single image based face recognition since the many sample images in the gallery set can convey more within-class variations of the subject [8]. One special case of ISFR is video based face recognition, which collects face image sets from consecutive video sequences [1], [14], [15]. Similar to the work in [5], [8], in this paper we focus on the general case of ISFR without considering the temporal relationship of samples in each set.

The key issues in image set based classification include how to model a set and consequently how to compute the distance/similarity between query and gallery sets. Researchers have proposed parametric and non-parametric approaches for image set modeling. Parametric modeling methods model each set as a parametric distribution, and use Kullback-Leibler divergence to measure the similarity between the distributions [2], [7]. The disadvantage of parametric set modeling lies in the difficulty of parameter estimation, and it may fail when the estimated parametric model does not fit well the real gallery and query sets [11], [4], [8].

Many non-parametric set modeling methods have also been proposed, including subspace [11], [1], manifold [16], [17], [4], [12], [18], affine hull [5], [8], convex hull [5], and covariance matrix based ones [18], [19], [20]. The method in [11] employs canonical correlation to measure the similarity between two sets. A projection matrix is learned by maximizing the canonical correlations of within-class sets while minimizing the canonical correlations of between-class sets. The methods in [21] use manifold to model an image set and define a manifold-to-manifold distance (MMD) for set matching. MMD models each image set as a set of local subspaces and the distance between two image sets is defined as a weighted average of pairwise subspace to subspace distance. As MMD is a non-discriminative measure, Manifold Discriminant Analysis (MDA) is proposed to learn an embedding space by maximizing manifold margin [12]. The performance of subspace and manifold based methods may degrade much when the set has a small sample size but big data variations [8], [18]. In affine hull and convex hull based methods [5], [8], the between-set distance is defined as the distance between the two closest points of the two sets. When convex hull is used, the set to set distance is equivalent to the nearest point problem in SVM [22]. In [23], a method called sparse approximated nearest points (SANP) is proposed to measure the dissimilarity between two image sets. To reduce the model complexity of SANP, a reduced model, which is called regularized nearest points (RNP), is proposed by modeling each image set as a regularized hull [24]. However, the closest points based methods [5], [8], [25], [24] rely highly on the location of each individual sample in the set, and the model fitting can be heavily deteriorated by outliers [18]. A collaborative regularized nearest points (CRNP) method is proposed in [26] to extend RNP.

To improve the classification performance, the kernel trick can be introduced to map the image sets to high-dimensional subspaces, e.g., kernel mutual subspace method [27] and kernel discriminant transformation [28]. In [18], an image set is represented by a covariance matrix and a Riemannian kernel function is defined to measure the similarity between

Pengfei Zhu, Lei Zhang, Simon C.K. Shiu and David Zhang are with Department of Computing, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China. (e-mail: cspzhu, cslzhang, csckshiu, csdzhang@comp.polyu.edu.hk).
Wangmeng Zuo is with School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China. (email: cswmzuo@gmail.com).

two image sets by a mapping from the Riemannian manifold to a Euclidean space. With the kernel function between two image sets, traditional discriminant learning methods, e.g., linear discriminative analysis [29], partial least squares [30], kernel machines, can be used for image set classification [19], [20]. The disadvantages of covariance matrix based methods include the computational complexity of eigen-decomposition of symmetric positive-definite (SPD) matrices and the curse of dimensionality with limited number of training sets.

No matter how the set is modeled, in almost all the previous works [11], [1], [16], [17], [4], [12], [18], [5], [8], [24], the query set is compared to each of the gallery sets separately, and then classified to the class closest to it. Such a classification scheme does not consider the correlation between gallery sets, like the nearest neighbor or nearest subspace classifier in single image based face recognition. In recent years, the sparse representation based classification (SRC) [31] has shown interesting results in image based face recognition. SRC represents a query face as a sparse linear combination of samples from all classes, and classifies it to the class which has the minimal representation residual to it. Though SRC emphasizes much on the role of l_1-norm sparsity of representation coefficients, it has been shown in [32] that the collaborative representation mechanism (i.e., using samples from all classes to collaboratively represent the query image) is more important to the success of SRC. The so-called collaborative representation based classification (CRC) with l_2-regularization leads to similar results to SRC but with much lower computational cost [32]. In [33], feature weights are introduced to the representation model to penalize pixels with large error so that the model is robust to outliers. Moreover, a kernel sparse representation model is proposed for face recognition by mapping features to a high dimensional Reproducing Kernel Hilbert Space (RKHS), which further improves the recognition accuracy [34]. Similarly, a robust kernel representation model is proposed with iteratively reweighted algorithms [35].

One may apply SRC/CRC to ISFR by representing each image of the query set over all the gallery sets, and then using the average or minimal representation residual of the query set images for classification. However, such a scheme does not exploit the correlation and distinctiveness of sample images in the query set. If the average representation residual is used for classification, the discrimination of representation residuals by different classes will be reduced; if the minimal representation residual is used, the classification can suffer from the outlier images in the query set. In addition, there are redundancies in an image set. The redundancies will lead to great storage burden and computational complexity, and deteriorate the recognition performance.

In this paper, we propose a novel image set based collaborative representation and classification (ISCRC) approach for ISFR, as illustrated in Fig. 1. The query set, denoted by $\boldsymbol{Y}$ (each column of $\boldsymbol{Y}$ is an image in the set) is modeled as a hull $\boldsymbol{Ya}$ with the sum of coefficients in $\boldsymbol{a}$ being 1. Let $\boldsymbol{X}_k, k = 1, 2, ..., K$, be a gallery set. We then propose a collaborative representation based set (i.e., $\boldsymbol{Y}$) to sets (i.e., $\boldsymbol{X} = [\boldsymbol{X}_1, ..., \boldsymbol{X}_k, ..., \boldsymbol{X}_K]$) distance (CRSSD for short); that

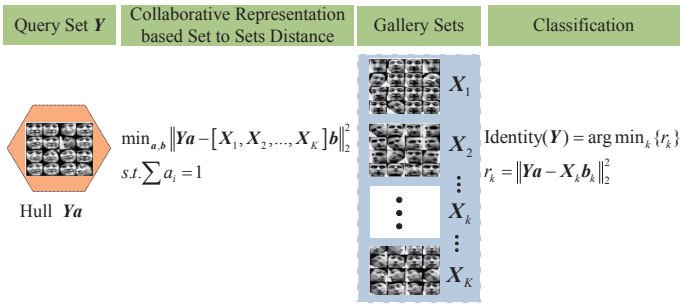

Fig. 1. Image set based collaborative representation and classification (ISCRC).

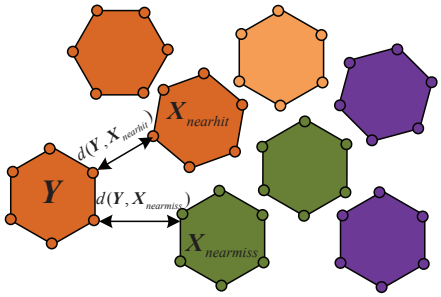

Fig. 2. Illustration of image set margin.

is, we represent the hull $\boldsymbol{Ya}$ over the gallery sets $\boldsymbol{X}$ as $\boldsymbol{Xb}$, where $\boldsymbol{b}$ is a coefficient vector. Consequently, we can classify the query set $\boldsymbol{Y}$ by checking which gallery set has the minimal representation residual to the hull $\boldsymbol{Ya}$. To get a stable solution to CRSSD, regularizations can be imposed on $\boldsymbol{a}$ and $\boldsymbol{b}$. In the proposed ISCRC, the gallery sets $\boldsymbol{X}_k$ can be compressed to a smaller size to remove the redundancy so that the time complexity of ISCRC can be much reduced without sacrificing the recognition rate. Our experiments on three benchmark ISFR databases show that the proposed ISCRC is superior to state-of-the-art methods in terms of both recognition rate and efficiency.

To better illustrate the motivation of ISCRC, we use an example to explain the superiority of ISCRC over set to set distance based classifiers (e.g., CHISD [5], SANP [8], RNP [24]) from a large margin perspective. Large margin principle has been widely used in classifier design (e.g., SVM [22], LVQ [36]), ensemble learning (e.g., AdaBoost [37]) and metric learning (e.g., MDA [12], LMNN [38]). In classification, large margin can lead to better generalization ability [39]. In [40], SRC is interpreted as a margin classifier and a margin is derived for SRC. Actually, in image set based classification, MDA [12], DCC [11] and CDL [18] all try to learn a discriminative set to set distance in a large margin manner, i.e., pull the similar image sets together while push the dissimilar image sets away. Similar to sample margin in nearest neighbor classifier, image set margin can be defined. Given a query set $\boldsymbol{Y}$ but multiple gallery sets $\boldsymbol{X}_k, k = 1, 2, ..., K$, as illustrated in Fig. 2, the image set margin is defined as:

$$margin_{\boldsymbol{Y}} = d(\boldsymbol{Y}, \boldsymbol{X}_{nearmiss}) - d(\boldsymbol{Y}, \boldsymbol{X}_{nearhit}) \quad (1)$$

where $\boldsymbol{X}_{nearhit}$ is the nearest gallery set of $\boldsymbol{Y}$ with the same class label, $\boldsymbol{X}_{nearmiss}$ is the nearest gallery set of $\boldsymbol{Y}$ with a different class label, $d(\boldsymbol{Y}, \boldsymbol{X}_{nearmiss})$ is the distance between $\boldsymbol{Y}$ and $\boldsymbol{X}_{nearmiss}$, and $d(\boldsymbol{Y}, \boldsymbol{X}_{nearhit})$ is the distance between $\boldsymbol{Y}$ and $\boldsymbol{X}_{nearhit}$. If $margin_{\boldsymbol{Y}}$ is positive, $\boldsymbol{Y}$ can be correctly classified; otherwise, $\boldsymbol{Y}$ would be misclassified. Hence, a large margin is desired in image set classification.

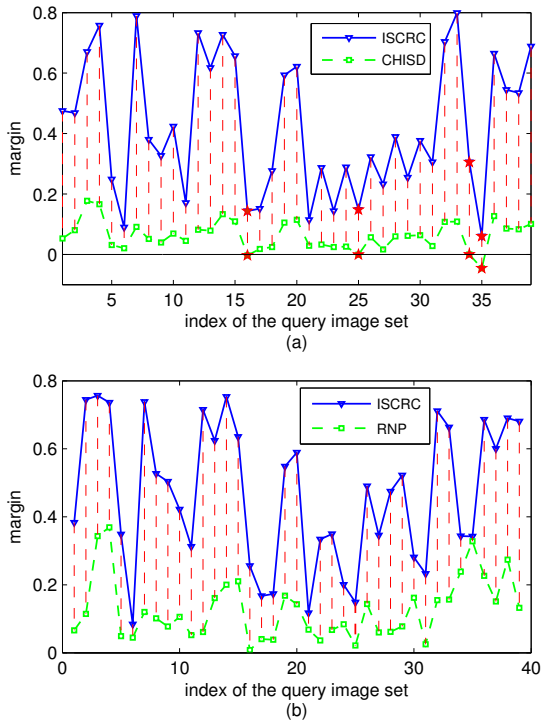

Fig. 3. Margin comparison between ISCRC and CHISD (a) and RNP (b).

Fig. 3 shows the margin comparison between the proposed ISCRC and hull based set to set distances (i.e., CHISD [5] and RNP [24]), where the Honda/USCD[1] database [14] is used. Fig. 3(a) is the comparison between ISCRC and convex hull based image set distance, i.e., CHISD. The image sets marked by pentagram are misclassified by CHISD with negative margin while correctly classified by ISCRC with positive margin. Besides, the margin of the other image sets are all enlarged, which represents better generalization ability in classification. Fig. 3(b) illustrates the comparison between ISCRC and regularized hull based image set distance, i.e., RNP. Although RNP classifies all the image sets correctly with positive margin, ISCRC results in much larger margin than RNP. Both comparisons show that the proposed ISCRC can lead to larger image set margin compared with set to set distance, indicating that ISCRC would get better generalization performance.

The rest of this paper is organized as follows. Section II discusses in detail the proposed CRSSD and ISCRC methods. Section III presents the regularized hull based ISCRC, followed by the convex hull based ISCRC in Section IV. Section V conducts experiments and Section VI gives our conclusions.

[1]http://vision.ucsd.edu/ leekc/HondaUCSDVideoDatabase/HondaUCSD.html

The main abbreviations used in the development of our method are summarized in Table I.

TABLE I
THE MAIN ABBREVIATIONS USED IN THIS PAPER

ISFR	image set based face recognition
SRC	sparse representation based classification
CRC	collaborative representation based classification
CRSSD	collaborative representation based set to sets distance
ISCRC	image set based collaborative representation and classification
RH-ISCRC	regularized hull based ISCRC
KCH-ISCRC	kernelized convex hull based ISCRC

II. COLLABORATIVE REPRESENTATION BASED SET TO SETS DISTANCE

We first introduce the hull based set to set distance in II-A, and then propose the collaborative representation based set to sets distance (CRSSD) in II-B. With CRSSD, the image set based collaborative representation and classification (ISCRC) scheme can be naturally proposed. In II-C and II-D, the convex hull and regularized hull based CRSSD are respectively presented.

A. Hull based set to set distance

In image set based classification, compared to the parametric modeling of image set, non-parametric modeling does not impose assumptions on the data distribution and inherits many favorable properties [11], [8], [18]. One simple non-parametric set modeling approach is the hull based modeling [5], [8], which models a set as the linear combination of its samples. Given a sample set $\boldsymbol{Y} = \{\boldsymbol{y}_1, ..., \boldsymbol{y}_i, ..., \boldsymbol{y}_{n_a}\}$, $\boldsymbol{y}_i \in \Re^d$, the hull of set $\boldsymbol{Y}$ is defined as: $H(\boldsymbol{Y}) = \{\sum a_i \boldsymbol{y}_i\}$. Usually, $\sum a_i = 1$ is required and the coefficients a_i are required to be bounded:

$$H(\boldsymbol{Y}) = \{\sum a_i \boldsymbol{y}_i \,|\, \sum a_i = 1, 0 \le a_i \le \tau\} \tag{2}$$

If $\tau = 1$, $H(\boldsymbol{Y})$ is a convex hull [41]. If $\tau < 1$, $H(\boldsymbol{Y})$ is a reduced convex hull [22]. For the convenience of expression, in the following development we call both the cases convex hull.

By modeling a set as a convex hull, the distance between set $\boldsymbol{Y} = \{\boldsymbol{y}_1, ..., \boldsymbol{y}_i, ..., \boldsymbol{y}_{n_a}\}$ and set $\boldsymbol{Z} = \{\boldsymbol{z}_1, ..., \boldsymbol{z}_j, ..., \boldsymbol{z}_{n_z}\}$ can be defined as follows:

$$\begin{array}{l} \min_{\boldsymbol{a},\boldsymbol{b}} \|\sum a_i \boldsymbol{y}_i - \sum b_j \boldsymbol{z}_j\|_2^2 \\ s.t. \sum a_i = 1, 0 \le a_i \le \tau \\ \quad \sum b_j = 1, 0 \le b_j \le \tau \end{array} \tag{3}$$

When the two sets have no intersection, the set to set distance in Eq. (3) becomes the distance between the nearest points in the two convex hulls (CHISD [5]), as illustrated in Fig. 4. It is not difficult to see that such a distance is equivalent to the distance computed by SVM [22]. If the discriminative function of SVM is $f = \boldsymbol{w}\boldsymbol{x} + b$, then $\boldsymbol{w} = \sum a_i \boldsymbol{y}_i - \sum b_j \boldsymbol{z}_j$ and the margin is $2/\|\boldsymbol{w}\|$. If we consider each image set as one class,

then maximizing margin between the two classes is equivalent to finding the set to set distance [42]. In image set based face recognition, there is usually no intersection between image sets of different persons. If there are intersections between two image sets, then τ can be set as below 1 and the resulting problem can be related with soft-margin SVM and v-SVM [43], [5]. Unfortunately, such a distance relies highly on the location of each individual sample and can be sensitive to outliers [18]. More detailed discussions about convex/affine hull based classifiers can be found in [22], [43], [5], [44].

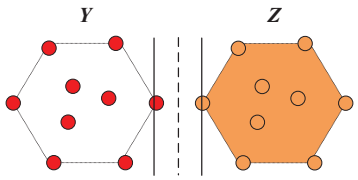

Fig. 4. Convex hull based set to set distance.

B. Collaborative representation based set to sets distance and classification

In image set based face recognition (ISFR), we have a query set $\boldsymbol{Y}$ but multiple gallery sets $\boldsymbol{X}_k, k = 1, 2, ..., K$. One fact in face recognition is that the face images from different people still have much similarity. If we compute the distance between $\boldsymbol{Y}$ and each $\boldsymbol{X}_k$ by using methods such as hull based set to set distance (refer to II-A), the correlation between different gallery sets will not be utilized. As we discussed in the Introduction section, inspired by the SRC [31] and CRC [32] methods in image based face recognition, here we propose a novel ISFR method, namely image set based collaborative representation and classification (ISCRC).

The key component of ISCRC is the collaborative representation based set to sets distance (CRSSD) defined as follows. Let $\boldsymbol{X} = [\boldsymbol{X}_1, ..., \boldsymbol{X}_k, ..., \boldsymbol{X}_K]$ be the concatenation of all gallery sets. We model each of $\boldsymbol{Y}$ and $\boldsymbol{X}$ as a hull, i.e., $\boldsymbol{Y}\boldsymbol{a}$ and $\boldsymbol{X}\boldsymbol{b}$, where $\boldsymbol{a}$ and $\boldsymbol{b}$ are coefficient vectors, and then we define the CRSSD between set $\boldsymbol{Y}$ and sets $\boldsymbol{X}$ as:

$$\min_{\boldsymbol{a},\boldsymbol{b}} \|\boldsymbol{Y}\boldsymbol{a} - \boldsymbol{X}\boldsymbol{b}\|^2 \ \ s.t. \sum a_i = 1 \tag{4}$$

where a_i is the i^{th} coefficeint in $\boldsymbol{a}$ and we let $\sum a_i = 1$ to avoid the trivial solution $\boldsymbol{a} = \boldsymbol{b} = \boldsymbol{0}$. In Eq. (4), the hull $\boldsymbol{Y}\boldsymbol{a}$ of the query set $\boldsymbol{Y}$ is collaboratively represented over the gallery sets; however, the coefficients in $\boldsymbol{a}$ will make the samples in $\boldsymbol{Y}$ be treated differently in the representation and the subsequent classification process. By minimizing the distance between $\boldsymbol{Y}\boldsymbol{a}$ and $\boldsymbol{X}\boldsymbol{b}$, the outliers (e.g., one frame with large corruptions/occlusions) in both the query image set $\boldsymbol{Y}$ and the gallery image sets $\boldsymbol{X}$ will be assigned with very small representation coefficients. Therefore, the impact of outliers can be much alleviated. Our experimental results in Section V showed that ISCRC is robust to face variations in different conditions.

Suppose that the coefficient vectors $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$ are obtained by solving Eq. (4), then we can write $\hat{\boldsymbol{b}}$ as $\hat{\boldsymbol{b}} = [\hat{\boldsymbol{b}}_1; ...; \hat{\boldsymbol{b}}_k; ...; \hat{\boldsymbol{b}}_K]$, where $\hat{\boldsymbol{b}}_k$ is is the sub-vector of coefficients associated with gallery set $\boldsymbol{X}_k$. Similar to the classification in SRC and CRC, we use the representation residual of hull $\boldsymbol{Y}\hat{\boldsymbol{a}}$ by each set $\boldsymbol{X}_k$ to determine the class label of $\boldsymbol{Y}$. The classifier in the proposed ISCRC is:

$$Identity(\boldsymbol{Y}) = argmin_k \{r_k\} \tag{5}$$

where $r_k = \left\|\boldsymbol{Y}\hat{\boldsymbol{a}} - \boldsymbol{X}_k\hat{\boldsymbol{b}}_k\right\|_2^2$.

Clearly, the solutions to $\boldsymbol{a}$ and $\boldsymbol{b}$ in Eq. (4) determine the CRSSD and hence the result of ISCRC. In order to get stable solutions, we could impose reasonable regularizations on $\boldsymbol{a}$ and $\boldsymbol{b}$. In the following sections II-C and II-D, we discuss the convex hull based CRSSD and regularized hull based CRSSD, respectively.

C. Convex hull based CRSSD

One important instantiation of CRSSD is the convex hull based CRSSD. In this case, both the hulls $\boldsymbol{Y}\boldsymbol{a}$ and $\boldsymbol{X}\boldsymbol{b}$ are required to be convex hulls, and then the distance in Eq. (4) becomes

$$\begin{aligned} &\min_{\boldsymbol{a},\boldsymbol{b}} \|\boldsymbol{Y}\boldsymbol{a} - \boldsymbol{X}\boldsymbol{b}\|^2 \\ &s.t. \sum a_i = 1, \sum b_j = 1, \\ &\quad 0 \le a_i \le \tau, i = 1, ..., n_a, \\ &\quad 0 \le b_j \le \tau, j = 1, ..., n_b \end{aligned} \tag{6}$$

where a_i and b_j are the i^{th} and j^{th} coefficients in $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively, n_a and n_b are the number of samples in set $\boldsymbol{Y}$ and sets $\boldsymbol{X}$, respectively, and $\tau \le 1$.

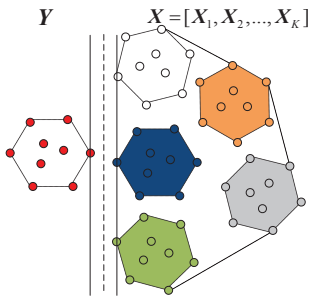

Fig. 5. Convex hull based CRSSD.

A geometric illustration of convex hull based CRSSD is shown in Fig. 5. Different from the CHISD method in [5], which models each gallery set as a convex hull, here we model all the gallery sets as one big convex hull. Similar to the closest points searching in SVM, convex hull based CRSSD aims to find the closest points in the query set $\boldsymbol{Y}$ and the whole gallery set $\boldsymbol{X}$ in a large margin manner. With convex hull based CRSSD, the corresponding ISCRC method can be viewed as a large margin based classifier in some sense. Nonetheless, the classification rules in SVM and ISCRC are very different.

D. l_p-norm regularized hull based CRSSD

The convex hull modeling of a set can be affected much by outlier samples in the set [18]. To make CRSSD more stable, the l_p-norm regularized hull can be used to model $\boldsymbol{Y}$ and $\boldsymbol{X}$.

For the query set $\boldsymbol{Y}$, we should keep the constraint $\sum a_i = 1$ to avoid the trivial solution, and the l_p-norm regularized hull of $\boldsymbol{Y}$ is defined as

$$H(\boldsymbol{Y}) = \{\sum a_i \boldsymbol{y}_i \,|\, \|\boldsymbol{a}\|_{l_p} < \delta\} \; s.t. \sum a_i = 1 \tag{7}$$

For the gallery set $\boldsymbol{X}$, its regularized hull is defined as:

$$H(\boldsymbol{X}) = \{\sum b_i \boldsymbol{x}_i \,|\, \|\boldsymbol{b}\|_{l_p} < \delta\} \tag{8}$$

Finally, the regularized hull based CRSSD between $\boldsymbol{Y}$ and $\boldsymbol{X}$ is defined as:

$$\begin{array}{l} \min_{\boldsymbol{a},\boldsymbol{b}} \|\boldsymbol{Y}\boldsymbol{a} - \boldsymbol{X}\boldsymbol{b}\|_2^2 \\ s.t. \|\boldsymbol{a}\|_{l_p} < \delta_1, \|\boldsymbol{b}\|_{l_p} < \delta_2, \sum a_i = 1 \end{array} \tag{9}$$

III. Regularized hull based ISCRC

In Section II, we introduced CRSSD, and presented two important instantiations of it, i.e., convex hull based CRSSD and regularized hull based CRSSD. With either one of them, the ISCRC (refer to Eq. (5)) can be implemented to perform ISFR. In this section, we discuss the minimization of regularized hull based CRSSD model, and the corresponding classification scheme is called regularized hull based ISCRC, denoted by RH-ISCRC. The minimization of convex hull based CRSSD and the corresponding classification scheme will be discussed in Section IV.

A. Main model

We can re-write the regularized hull based CRSSD model in Eq. (9) as its Lagrangian formulation:

$$\begin{array}{l} \min_{\boldsymbol{a},\boldsymbol{b}} \|\boldsymbol{Y}\boldsymbol{a} - \boldsymbol{X}\boldsymbol{b}\|_2^2 + \lambda_1 \|\boldsymbol{a}\|_{l_p} + \lambda_2 \|\boldsymbol{b}\|_{l_p} \\ s.t. \sum a_i = 1 \end{array} \tag{10}$$

where λ_1 and λ_2 are positive constants to balance the representation residual and the regularizer.

In ISFR, each gallery set $\boldsymbol{X}_k$ often has tens to hundreds of sample images so that the whole set $\boldsymbol{X}$ can be very big, making the computational cost to solve Eq. (10) very high. Considering the fact that the images in each set $\boldsymbol{X}_k$ have high redundancy, we can compress $\boldsymbol{X}_k$ into a much more compact set, denoted by $\boldsymbol{D}_k$, via dictionary learning methods [45], such as KSVD [46] and metaface learning [47]. Let $\boldsymbol{D} = [\boldsymbol{D}_1, ..., \boldsymbol{D}_k, ..., \boldsymbol{D}_K]$. We can then replace $\boldsymbol{X}$ by $\boldsymbol{D}$ in Eq. (10) to compute the regularized hull based CRSSD:

$$\begin{array}{l} (\hat{\boldsymbol{a}}, \hat{\boldsymbol{\beta}}) = \arg\min_{\boldsymbol{a},\boldsymbol{\beta}} \left\{ \begin{array}{l} \|\boldsymbol{Y}\boldsymbol{a} - \boldsymbol{D}\boldsymbol{\beta}\|_2^2 + \\ \lambda_1 \|\boldsymbol{a}\|_{l_p} + \lambda_2 \|\boldsymbol{\beta}\|_{l_p} \end{array} \right\} \\ s.t. \sum a_i = 1 \end{array} \tag{11}$$

where $\boldsymbol{\beta} = [\boldsymbol{\beta}_1; ...; \boldsymbol{\beta}_k; ...; \boldsymbol{\beta}_K]$ and $\boldsymbol{\beta}_k$ is the sub-vector of coefficients associated with $\boldsymbol{D}_k$. Based on our experimental results, compressing $\boldsymbol{X}_k$ into $\boldsymbol{D}_k$ significantly improve the speed with almost the same ISFR rate.

Either l_1-norm or l_2-norm can be used to regularize $\boldsymbol{a}$ and $\boldsymbol{\beta}$, while l_1-regularization will lead to sparser solutions but with more computational cost. Like in l_1-SVM [48] and SRC [31], sparsity can enhance the classification rate if the features are not informative enough. Note that if the query set $\boldsymbol{Y}$ has only one sample, then $\boldsymbol{a} = 1$ and the proposed model in Eq. (11) will be reduced to the SRC (for l_1-regularization) or CRC (for l_2-regularization) scheme. Next, we present the optimization of l_2-norm and l_1-norm regularized hull based ISCRC in Section III-B and Section III-C, respectively.

B. l_2-norm regularized hull based ISCRC

When l_2-norm is used to regularize $\boldsymbol{a}$ and $\boldsymbol{\beta}$, the problem in Eq. (11) has a closed-form solution. The Lagrangian function of Eq. (11) becomes

$$\begin{array}{l} L(\boldsymbol{a}, \boldsymbol{\beta}, \lambda_3) = \|\boldsymbol{Y}\boldsymbol{a} - \boldsymbol{D}\boldsymbol{\beta}\|_2^2 + \lambda_1 \|\boldsymbol{a}\|_2^2 + \lambda_2 \|\boldsymbol{\beta}\|_2^2 \\ +\lambda_3(\boldsymbol{e}\boldsymbol{a} - 1) \\ = \left\| [\boldsymbol{Y} \;\; -\boldsymbol{D}] \left[\begin{array}{c} \boldsymbol{a} \\ \boldsymbol{\beta} \end{array} \right] \right\|_2^2 + \left[\boldsymbol{a}^T \;\; \boldsymbol{\beta}^T \right] \left[\begin{array}{cc} \lambda_1 \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \lambda_2 \boldsymbol{I} \end{array} \right] \left[\begin{array}{c} \boldsymbol{a} \\ \boldsymbol{\beta} \end{array} \right] \\ +\lambda_3([\boldsymbol{e} \;\; \boldsymbol{0}] \left[\begin{array}{c} \boldsymbol{a} \\ \boldsymbol{\beta} \end{array} \right] - 1) \end{array} \tag{12}$$

where $\boldsymbol{e}$ is a row vector whose elements are 1.

Let $\boldsymbol{z} = \left[\begin{array}{c} \boldsymbol{a} \\ \boldsymbol{\beta} \end{array} \right]$, $\boldsymbol{A} = [\boldsymbol{Y} \;\; -\boldsymbol{D}]$, $\boldsymbol{B} = \left[\begin{array}{cc} \lambda_1 \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \lambda_2 \boldsymbol{I} \end{array} \right]$ and $\boldsymbol{d} = [\boldsymbol{e} \;\; \boldsymbol{0}]^T$. Then Eq. (12) becomes:

$$L(\boldsymbol{z}, \lambda_3) = \boldsymbol{z}^T \boldsymbol{A}^T \boldsymbol{A} \boldsymbol{z} + \boldsymbol{z}^T \boldsymbol{B} \boldsymbol{z} + \lambda_3(\boldsymbol{d}^T \boldsymbol{z} - 1) \tag{13}$$

There are

$$\frac{\partial L}{\partial \lambda_3} = \boldsymbol{d}^T \boldsymbol{z} - 1 = 0 \tag{14}$$

$$\frac{\partial L}{\partial z} = \boldsymbol{A}^T \boldsymbol{A} \boldsymbol{z} + \boldsymbol{B} \boldsymbol{z} + \lambda_3 \boldsymbol{d} = 0 \tag{15}$$

According to Eq. (14) and Eq. (15), we get the closed form solution to Eq. (12):

$$\hat{\boldsymbol{z}} = \left[\begin{array}{c} \hat{\boldsymbol{a}} \\ \hat{\boldsymbol{\beta}} \end{array} \right] = \boldsymbol{z}_0 / \boldsymbol{d}^T \boldsymbol{z}_0 \tag{16}$$

where $\boldsymbol{z}_0 = (\boldsymbol{A}^T \boldsymbol{A} + \boldsymbol{B})^{-1} \boldsymbol{d}$.

After $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{\beta}}$ are got, the distance between query set $\boldsymbol{Y}$ and a gallery set $\boldsymbol{X}_k$ is calculated as $r_k = \left\| \boldsymbol{Y}\hat{\boldsymbol{a}} - \boldsymbol{D}_k \hat{\boldsymbol{\beta}}_k \right\|_2^2$, and then the class label of $\boldsymbol{Y}$ is determined by Eq. (5). For RH-ISCRC-l_2, the main time consumption is to solve the inverse of matrix $(\boldsymbol{A}^T \boldsymbol{A} + \boldsymbol{B})$. Hence, the time complexity of RH-ISCRC-l_2 is $O((n_a + n_\beta)^3)$, where n_a is the number of sample images in $\boldsymbol{Y}$ and n_β is the number of atoms in $\boldsymbol{D}$.

The CRNP method [26] also collaboratively represents the query set over the gallery sets. The differences between the proposed RH-ISCRC-l_2 and CRNP lie in the optimization procedure and the classification rule. RH-ISCRC-l_2 has a closed-form solution while CRNP adopts the same optimization method as RNP [24], which iteratively converges to the global optimal solution. Besides, CRNP uses the same classification rule as RNP, which utilizes both the reconstruction error and rank of image set matrix. RH-ISCRC-l_2 only uses the reconstruction error for classification.

C. l_1-norm regularized hull based ISCRC

When l_1-norm regularization is used, we use the alternating minimization method, which is very efficient to solve multiple variable optimization problems [49]. For Eq. (11), we have the following augmented Lagrangian function:

$$L(\boldsymbol{a},\boldsymbol{\beta},\lambda) = \|\boldsymbol{Y}\boldsymbol{a}-\boldsymbol{D}\boldsymbol{\beta}\|_2^2 + \lambda_1\|\boldsymbol{a}\|_1 + \lambda_2\|\boldsymbol{\beta}\|_1 + <\lambda, \boldsymbol{e}\boldsymbol{a}-1> + \frac{\gamma}{2}\|\boldsymbol{e}\boldsymbol{a}-1\|_2^2 \quad (17)$$

where λ is the Lagrange multiplier, $\langle\cdot,\cdot\rangle$ is the inner product, and $\gamma > 0$ is the penalty parameter.

Then $\boldsymbol{a}$ and $\boldsymbol{\beta}$ are optimized alternatively with the other one fixed. More specifically, the iterations of minimizing $\boldsymbol{a}$ go as follows:

$$\begin{aligned}\boldsymbol{a}^{(t+1)} &= \arg\min_{\boldsymbol{a}} L(\boldsymbol{a},\boldsymbol{\beta}^{(t)},\lambda^{(t)}) \\ &= \arg\min_{\boldsymbol{a}} f(\boldsymbol{a}) + \frac{\gamma}{2}\left\|\boldsymbol{e}\boldsymbol{a}-1+\lambda^{(t)}/\gamma\right\|_2^2 \\ &= \arg\min_{\boldsymbol{a}} \left\|\tilde{\boldsymbol{Y}}\boldsymbol{a}-\boldsymbol{x}\right\|_2^2 + \lambda_1\|\boldsymbol{a}\|_1\end{aligned} \quad (18)$$

where $f(\boldsymbol{a}) = \left\|\boldsymbol{Y}\boldsymbol{a}-\boldsymbol{D}\boldsymbol{\beta}^{(t)}\right\|_2^2 + \lambda_1\|\boldsymbol{a}\|_{l_p}$, $\tilde{\boldsymbol{Y}} = \left[\boldsymbol{Y}; (\gamma/2)^{1/2}\boldsymbol{e}\right]$, $\boldsymbol{x} = [\boldsymbol{D}\boldsymbol{\beta}^{(t)}; (\gamma/2)^{1/2}(1-\lambda^{(t)}/\gamma)]$.

The problem in Eq. (18) can be easily solved by some representative l_1-minimization approaches such as LARS [50].

After $\boldsymbol{a}^{(t+1)}$ is updated, $\boldsymbol{\beta}^{(t+1)}$ can be obtained by solving another l_1-regularized optimization problem:

$$\begin{aligned}\boldsymbol{\beta}^{(t+1)} &= \arg\min_{\boldsymbol{\beta}} L(\boldsymbol{a}^{(t+1)},\boldsymbol{\beta},\lambda^{t}) \\ &= \arg\min_{\boldsymbol{\beta}} \left\|\boldsymbol{Y}\boldsymbol{a}^{(t+1)}-\boldsymbol{D}\boldsymbol{\beta}\right\|_2^2 + \lambda_2\|\boldsymbol{\beta}\|_1\end{aligned} \quad (19)$$

Once $\boldsymbol{a}^{(t+1)}$ and $\boldsymbol{\beta}^{(t+1)}$ are got, λ is updated as follows:

$$\lambda^{(t+1)} = \lambda^{(t)} + \gamma\left(\boldsymbol{e}\boldsymbol{a}^{(t+1)}-1\right) \quad (20)$$

The algorithm of RH-ISCRC-l_1 for ISFR is summarized in Table II and it converges. The problem in Eq. (17) is convex, and the subproblems in Eq. (18) and Eq. (19) are convex and can be solved using the LARS algorithm. It had been shown in [51], for the general convex problem, the alternating minimization approach would converge to the correct solution. One curve of the objective function value of RH-ISCRC-l_1 versus the iteration number is shown in Fig. 6. Honda/USCD database [14] is also used. The query set $\boldsymbol{Y}$ and each gallery set $\boldsymbol{X}_k$ has 200 frames. Note that one image set is acquired from one video clip and there is no intersection between the query set and each gallery set. We compress each set $\boldsymbol{X}_k$ into a dictionary $\boldsymbol{D}_k$ with 20 atoms by using the metaface learning method [47]. Since there are 20 gallery sets, the set $\boldsymbol{D} = [\boldsymbol{D}_1, ..., \boldsymbol{D}_k, ..., \boldsymbol{D}_{20}]$ has 20 × 20=400 atoms. From the figure we can see that RH-ISCRC-l_1 converges after about five iterations.

Since the complexity of sparse coding is $O(m^2 n^{\varepsilon})$, where m is the feature dimension, n is the atom number and $\varepsilon \geq 1.2$ [52], we can get that the time complexity of RH-ISCRC-l_1 is $O(lm^2({n_a}^{\varepsilon} + {n_\beta}^{\varepsilon}))$, where n_a is the number of samples in $\boldsymbol{Y}$, n_β is the number of atoms in $\boldsymbol{D}$ and l is the iteration number.

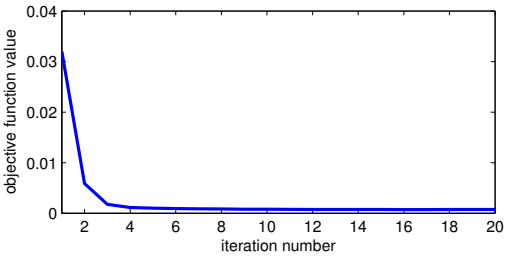

Fig. 6. Convergence of RH-ISCRC-l_1.

TABLE II
ALGORITHM OF RH-ISCRC FOR ISFR

Input: query set $\boldsymbol{Y}$; gallery sets $\boldsymbol{X} = [\boldsymbol{X}_1, ..., \boldsymbol{X}_k, ..., \boldsymbol{X}_K]$, λ_1 and λ_2.
Output: the label of query set $\boldsymbol{Y}$.
Initialize $\beta^{(0)}$, $\lambda^{(0)}$ and $0 \leftarrow t$.
Compress $\boldsymbol{X}_k$ to $\boldsymbol{D}_k$, $k = 1, 2, ..., K$ using metaface learning [47].
While $t < max\_num$ do
  Step 1: Update $\boldsymbol{a}$ by Eq. (18);
  Step 2: Update $\boldsymbol{\beta}$ by Eq. (19);
  Step 3: Update λ by Eq. (20);
  Step 4: $t \leftarrow t + 1$.
End while
Compute $r_k = \left\|\boldsymbol{Y}\hat{\boldsymbol{a}} - \boldsymbol{D}_k\hat{\boldsymbol{\beta}}_k\right\|_2^2$, $k = 1, 2, ...K$.
Identity($\boldsymbol{Y}$)=$\arg\min_k\{r_k\}$.

D. Examples and discussions

Let's use an example to better illustrate the classification process of RH-ISCRC. We use the Honda/USCD database [14]. The experiment setting is the same as Fig. 6. By Eq. (11), the computed coefficients in $\boldsymbol{a}$ and $\boldsymbol{\beta}$ are plotted in Fig. 7 (by l_1-regularization) and Fig. 8 (by l_2-regularization), respectively. The highlighted coefficients in the figures are associated with set $\boldsymbol{X}_{10}$, which has the same class label as $\boldsymbol{Y}$. Clearly, these coefficients are much more significant than the coefficients associated with the other classes. Meanwhile, from Fig. 7 and Fig. 8 we can see that l_1-regularized hull based CRSSD leads to sparser $\boldsymbol{a}$ and $\boldsymbol{\beta}$, implying that only few samples are dominantly involved in representation and classification.

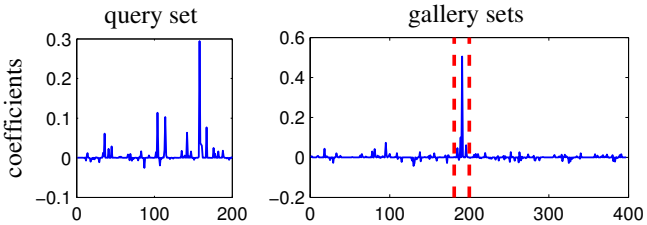

Fig. 7. The coefficient vectors $\hat{\boldsymbol{a}}$ (of $\boldsymbol{Y}$) and $\hat{\boldsymbol{\beta}}$ (of $\boldsymbol{D}$) by l_1-regularized hull based CRSSD.

In Fig. 9, we show the reconstructed faces by $\boldsymbol{Y}\hat{\boldsymbol{a}}$ with l_1-regularized hull based CRSSD. The distances between $\boldsymbol{Y}\hat{\boldsymbol{a}}$ and each $\boldsymbol{D}_k\hat{\boldsymbol{\beta}}_k$, i.e., r_k, are also given. We see that r_{10} is 0.03, which is the minimal one among all the gallery sets, meaning that ISCRC will make the correct recognition. Here

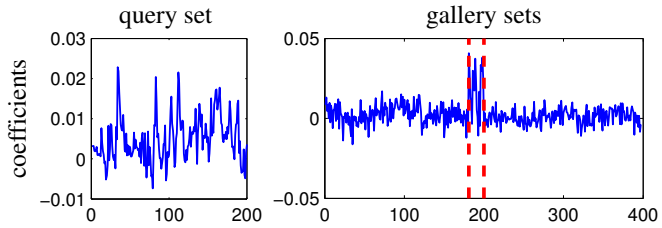

Fig. 8. The coefficient vectors $\hat{\boldsymbol{a}}$ (of $\boldsymbol{Y}$) and $\hat{\boldsymbol{\beta}}$ (of $\boldsymbol{D}$) by l_2-regularized hull based CRSSD.

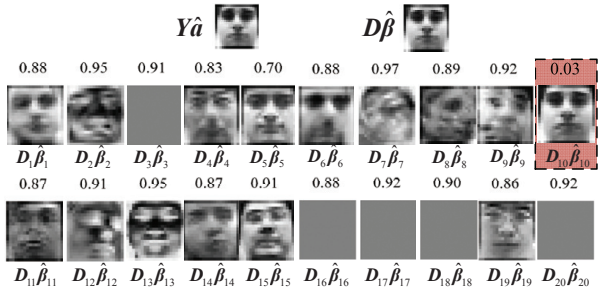

Fig. 9. Reconstructed faces $\boldsymbol{Y}\hat{\boldsymbol{a}}$, $\boldsymbol{D}\hat{\boldsymbol{\beta}}$, $\boldsymbol{D}_k\hat{\boldsymbol{\beta}}_k$ (we normalized each $\boldsymbol{D}_k\hat{\boldsymbol{\beta}}_k$ for better visualization). The number over each $\boldsymbol{D}_k\hat{\boldsymbol{\beta}}_k$ is the residual $r_k = \left\|\boldsymbol{Y}\hat{\boldsymbol{a}} - \boldsymbol{D}_k\hat{\boldsymbol{\beta}}_k\right\|_2^2$.

the relationships between ISCRC and manifold based methods can be revealed. MMD assumes that an image set can be modeled as a set of local subspaces so that the image set distance is defined as the weighted average distance between any two local subspaces [4]. The distance between two local subspaces is related to the cluster exemplar and principle angel. Correspondingly, ISCRC seeks for a local subspace ($\boldsymbol{Y}\hat{\boldsymbol{a}}$) in the query image set and a local subspace ($\boldsymbol{D}\hat{\boldsymbol{\beta}}$) in all the gallery sets, as shown in Fig. 7 . In classification, the distance between the query set and the template set of the k^{th} class is the distance between the local subspace ($\boldsymbol{Y}\hat{\boldsymbol{a}}$) and the local subspace $\boldsymbol{D}_k\hat{\boldsymbol{\beta}}_k$.

IV. Kernelized Convex Hull Based ISCRC

We then focus on how to compute the convex hull based CRSSD in Eq. (6) and use it for ISCRC. Since there can be many sample images in gallery sets, $\boldsymbol{X}$ can be a fat matrix (note that usually we use a low dimensional feature vector to represent each face image). Even we compress $\boldsymbol{X}$ into a more compact set $\boldsymbol{D}$, the system can still be under-determined. In Section 3 we imposed the l_p-norm regularization on $\boldsymbol{a}$ and $\boldsymbol{b}$ to make the solution stable. When the convex hull is used, however, the constraint may not be strong enough to get a stable solution of Eq. (6). In addition, if the underlying relationship between the query set and gallery sets is highly nonlinear, it is difficult to approximate the hull of query set as a linear combination of gallery sets.

One simple solution to solving both the above two problems is the kernel trick; that is, we can map the data into a higher dimensional space where the subjects can be approximately linearly separable. The mapped gallery data matrix in the high-dimensional space will be generally over-determined. In such a case, the convex hull constraint will be strong enough for a stable solution. The kernelized convex hull based CRSSD model is:

$$\begin{aligned} &\min_{\boldsymbol{a},\boldsymbol{\beta}} \left\|\phi(\boldsymbol{Y})\boldsymbol{a} - [\phi(\boldsymbol{D}_1), \phi(\boldsymbol{D}_2), ..., \phi(\boldsymbol{D}_K)]\,\boldsymbol{\beta}\right\|^2 \\ &s.t. \sum a_i = 1, \sum \beta_j = 1, \\ &\quad 0 \le a_i \le \tau, i = 1, ..., n_a, \\ &\quad 0 \le \beta_j \le \tau, j = 1, ..., n_\beta. \end{aligned} \tag{21}$$

The above minimization can be easily solved by the standard quadratic optimization (QP [53]) method. The solution exhibits global and quadratic convergence, as proved in [53]. Different kernel functions can be used, e.g., linear kernel and Gaussian kernel. We call the corresponding method kernelized convex hull based ISCRC, denoted by KCH-ISCRC. The classification rule is the same as RH-ISCRC with $r_k = \left\|\phi(\boldsymbol{Y})\hat{\boldsymbol{a}} - \phi(\boldsymbol{D}_k)\hat{\boldsymbol{\beta}}_k\right\|_2^2$. As convex hull based CRSSD is to solve a convex QP problem, the time complexity of KCH-ISCRC is $O((n_\beta + n_a)^3)$, which is similar to SVM. The algorithm of KCH-ISCRC is given in Table III. To reduce the computational cost, the kernel matrix $k(\boldsymbol{D}, \boldsymbol{D})$ can be computed and stored. When a query set $\boldsymbol{Y}$ comes, we only need to calculate $k(\boldsymbol{Y}, \boldsymbol{Y})$ and $k(\boldsymbol{Y}, \boldsymbol{D})$.

TABLE III
Algorithm of KCH-ISCRC for ISFR

Input: query set $\boldsymbol{Y}$; gallery sets $\boldsymbol{X} = [\boldsymbol{X}_1, ..., \boldsymbol{X}_k, ..., \boldsymbol{X}_K]$, τ.
Output: the label of query set $\boldsymbol{Y}$.
Compress $\boldsymbol{X}_k$ to $\boldsymbol{D}_k$, $k = 1, 2, ..., K$ by metaface learning [24];
Solve the QP problem in Eq. (21);
Compute $r_k = \left\|\phi(\boldsymbol{Y})\hat{\boldsymbol{a}} - \phi(\boldsymbol{D}_k)\hat{\boldsymbol{\beta}}_k\right\|_2^2$, $k = 1, 2, ...K$;
Identity($\boldsymbol{Y}$)=$\arg\min_k\{r_k\}$.

Like in Fig. 7 and Fig. 8, in Fig. 10 we show the coefficient vectors $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{\beta}}$ solved by Eq. (21). The Gaussian kernel is used and the experimental setting is the same as that in Figs. 7 and 8 (the only difference is that each compressed gallery set $\boldsymbol{D}_k$ has 50 atoms). We can see that the coefficients associated with gallery set $\boldsymbol{D}_{10}$ are larger than the other gallery sets, resulting in a smaller representation residual and hence the correct recognition.

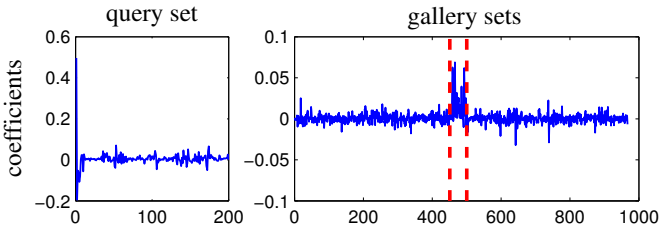

Fig. 10. The coefficient vectors $\hat{\boldsymbol{a}}$ (of $\boldsymbol{Y}$) and $\hat{\boldsymbol{\beta}}$ (of $\boldsymbol{D}$) by kernelized convex hull based CRSSD.

V. Experimental Analysis

We used the Honda/UCSD [14], CMU Mobo [54], and Youtube Celebrities [55] datasets to test the performance of the proposed method. The comparison methods fall into four categories:

C1. Subspace and manifold based methods: Mutual Subspace Method (MSM) [1], Discriminant Canonical Correlations (DCC[2]) [11], Manifold-Manifold Distance (MMD[3]) [4], and Manifold Discriminant Analysis (MDA[4]) [12].

C2. Affine/convex hull based methods: Affine Hull based Image Set Distance (AHISD[5]) [5], Convex Hull based Image Set Distance (CHISD[6]) [5], Sparse Approximated Nearest Points (SANP[7]) [8], and Regularized Nearest Points (RNP) [24].

C3. Representation based methods: Sparse Representation based Classifier (SRC) [31], Collaborative Representation based Classifier (CRC) [32]. We tested to use the average and minimal representation residual of query set for classification and found that average residual works better. Hence in this paper, the average residual is used in SRC/CRC for classification.

C4. Kernel methods: KSRC (Kernel SRC) [56], KCRC (Kernel CRC) [35], AHISD [5], and CHISD [5]. For KSRC and KCRC, the average residual is used for classification.

For the proposed methods, RH-ISCRC is compared with those non-kernel methods and KCH-ISCRC is compared with those kernel methods.

A. Parameter setting

For competing methods, the important parameters were empirically tuned according to the recommendations in the original literature for fair comparison. For DCC [11], if there is only one set per class, then the training set is divided into two sets since at least two sets per class are needed in DCC. For MMD, the number of local models is set following the work in [4]. For MDA, there are three parameters, i.e., the number of local models, the number of between-class NN local models and the subspace dimension. The three parameters are configured according to the work in [12]. For SANP, we adopted the same parameters as [8]. For SRC, CRC, KSRC and KCRC, λ that balances the residual and regularization is tuned from $[0.01, 0.001, 0.0001]$. For AHISD and CHISD, C is set as 100. For all kernel methods, Gaussian kernel ($k(x, y) = \exp(-\|x - y\|_2^2/2\delta^2)$) is used, and δ is set as 5. The experiments of 50 frames, 100 frames and 200 frames per set are conducted on the three databases. If the number of samples in the set is less than the given number, then all the samples in the set are used.

For the proposed RH-ISCRC, we set $\lambda_1 = 0.001$, $\lambda_2 = 0.001$, $\lambda = 2.5/n_a$ (n_a is the number of samples in the query set), $\gamma = \lambda/2$. The number of atoms in the compressed set $\boldsymbol{D}_k$ is set as 20 on Honda/UCSD and 10 on CMU MoBo and YouTube. For KCH-ISCRC, $\tau = 1$ and the number of atoms in each $\boldsymbol{D}_k$ is set as 50 for all datasets. The sensitivity of the proposed methods to parameters will be discussed in Section V-F.

[2]http://www.iis.ee.ic.ac.uk/ tkkim/code.htm
[3]http://www.jdl.ac.cn/user/rpwang/research.htm
[4]http://www.jdl.ac.cn/user/rpwang/research.htm
[5]http://www2.ogu.edu.tr/ mlcv/softwareimageset.html
[6]http://www2.ogu.edu.tr/ mlcv/softwareimageset.html
[7]https://sites.google.com/site/yiqunhu/cresearch/sanp

B. Honda/UCSD

The Honda/UCSD dataset consists of 59 video sequences involving 20 different subjects [14]. The Viola-Jones face detector [57] is used to detect the faces in each frame and resize the detected faces to 20×20 images. Some examples of Honda/UCSD dataset are shown in Figure 11. Histogram equalization is utilized to reduce the illumination variations. Our experiment setting is the same as [14][8]: 20 sequences are set aside for training and the remaining 39 sequences for testing. The intensity is used as the feature.

Fig. 11. Some examples of Honda/UCSD dataset

TABLE IV
RECOGNITION RATES ON HONDA/UCSD (%)

Non-kernel	50	100	200	Year
MSM [1]	74.36	79.49	89.74	1998
DCC [11]	76.92	84.62	94.87	2007
MMD [4]	69.23	87.18	94.87	2008
MDA [12]	82.05	94.87	97.44	2009
SRC [31]	84.62	92.31	92.31	2009
AHISD [5]	82.05	84.62	89.74	2010
CHISD [5]	82.05	84.62	92.31	2010
SANP [8]	84.62	92.31	94.87	2011
CRC [32]	84.62	94.87	94.87	2011
RNP [24]	87.18	94.87	**100.0**	2011
RH-ISCRC-l_1	**89.74**	**97.44**	**100.0**	
RH-ISCRC-l_2	**89.74**	**97.44**	**100.0**	
Kernel	50	100	200	Year
AHISD [5]	84.62	84.62	82.05	2010
CHISD [5]	84.62	87.18	89.74	2010
KSRC [56]	87.18	**97.44**	97.44	2009
KCRC [35]	82.05	94.87	94.87	2012
KCH-ISCRC	**89.74**	94.87	**100.0**	

The experimental results are listed in Table IV. We can see that for those non-kernel methods, the proposed RH-ISCRC outperforms much all the other methods. Note that in [5], kernel CHISD achieves 100% recognition accuracy when all the frames in one video clip are used. In this paper, following the experiment setting of SANP [8], we reported the accuracy using different number of frames per set. When 200 frames per set are used, both RH-ISCRC and KCH-ISCRC achieve 100% accuracy, which shows the superiority to CHISD and AHISD. For the kernel based method, the proposed KCH-ISCRC performs the best except for the case when 100 frames per set are used. We can also see that on this dataset, RH-ISCRC-l_1 and RH-ISCRC-l_2 achieve the same recognition rate, which implies that on this dataset the l_2-norm regularization is strong enough to yield a good solution to the regularized hull based CRSSD in Eq. (11).

C. CMU MoBo

The CMU Mobo[8] (Motion of Body) dataset [54] was originally established for human pose identification and it contains 96 sequences from 24 subjects. Four video sequences are collected per subject, each of which corresponds to a walking pattern. Again, the Viola-Jones face detector [57] is used to detect the faces and the detected face images are resized to 40 $\times$ 40. The LBP feature is used, which is the same as the work in [5] and [8].

One video sequence per subject is selected for training while the rest are used for testing. Ten-fold cross validation experiments are conducted and the average recognition results are shown in Table V. We can clearly see that the proposed methods outperform the other methods under different frames per set. On this dataset and the Honda/UCSD dataset, the proposed non-kernel RH-ISCRC and the kernel based KCH-ISCRC have similar ISFR rates.

TABLE V
RECOGNITION RATES ON CMU MOBO(%)

Non-kernel	50	100	200	Year
MSM [1]	84.3 ± 2.6	86.6±2.2	89.9±2.4	1998
DCC [11]	82.1± 2.7	85.5±2.8	91.6±2.5	2007
MMD [4]	86.2 ±2.9	94.6±1.9	96.4±0.7	2008
MDA [12]	86.2 ±2.9	93.2±2.8	95.8±2.3	2009
SRC [31]	91.0 ±2.1	91.8±2.7	96.5±2.5	2009
AHISD [5]	91.6 ±2.8	94.1±2.0	91.9±2.6	2010
CHISD [5]	91.2 ±3.1	93.8±2.5	96.0±1.3	2010
SANP [8]	91.9 ±2.7	94.2±2.1	97.3±1.3	2011
CRC [32]	89.6 ±1.8	92.4±3.7	96.4±2.8	2011
RNP [24]	91.9 ±2.5	94.7±1.2	97.4±1.5	2013
RH-ISCRC-l_1	**93.5±2.8**	**96.5±1.9**	**98.7±1.7**	
RH-ISCRC-l_2	**93.5±2.8**	**96.4±1.9**	**98.4±1.7**	
Kernel	50	100	200	Year
AHISD [5]	88.9±1.7	92.4±2.8	93.5±4.2	2010
CHISD [5]	91.5±2.0	93.4±4.0	97.4±1.9	2010
KSRC [56]	91.6 ±2.8	94.1±2.0	96.8±2.0	2010
KCRC [35]	91.2 ±3.1	93.4±2.9	96.6±2.6	2012
KCH-ISCRC	**94.2 ±2.1**	**96.4±2.3**	**98.4±1.9**	

D. YouTube Celebrities

The YouTube Celebrities[9] is a large scale video dataset collected for face tracking and recognition, consisting of 1,910 video sequences of 47 celebrities from YouTube [55]. As the videos were captured in unconstrained environments, the recognition task becomes much more challenging due to the larger variations in pose, illumination and expressions. Some examples of YouTube Celebrities dataset are shown in Figure 12. The face in each frame is also detected by the Viola-Jones face detector and resized to a 30 $\times$ 30 gray-scale image. The intensity value is used as feature. The experiment setting is the same as [8], [12], [18]. Three video sequences per subject are selected for training and six for testing. Five-fold cross validation experiments are conducted.

The experimental results are shown in Table VI. It can be seen that among the non-kernel methods, the proposed RH-ISCRC-l_1 achieves the highest recognition rate, while among the kernel based methods, the proposed KCH-ISCRC performs the best. Since this Youtube Celebrities dataset was established under uncontrolled environment, there are significant variations among the query and gallery sets, and therefore the l_1-regularization is very helpful to improve the stability and discrimination of the solution to Eq. (11). As a consequence, RH-ISCRC-l_1 leads to much better results than RH-ISCRC-l_2 on this dataset. On the other hand, the kernel based KCH-ISCRC leads to better results than RH-ISCRC in this experiment. Besides, the number of frames per set also affect the performance of ISCRC. When number of frames is small, the improvement by ISCRC is more significant.

Fig. 12. Some examples of YouTube Celebrities dataset

TABLE VI
RECOGNITION RATES ON YOUTUBE (V1 %)

Non-kernel	50	100	200	Year
MSM [1]	54.8±8.7	57.4±7.7	56.7±6.9	1998
DCC [11]	57.6±8.0	62.7±6.8	65.7±7.0	2007
MMD [4]	57.8±6.6	62.8±6.2	64.7±6.3	2008
SRC [31]	61.5±6.9	64.4±6.8	66.0±6.7	2009
MDA [12]	58.5±6.2	63.3±6.1	65.4±6.6	2009
AHISD [5]	57.5±7.9	59.7±7.2	57.0±5.5	2010
CHISD [5]	58.0±8.2	62.8±8.1	64.8±7.1	2010
SANP [8]	57.8±7.2	63.1±8.0	65.6±7.9	2011
CRC [32]	56.5±7.4	59.5±6.6	61.4±6.4	2011
RNP [24]	59.9 ±7.3	63.3±8.1	64.4±7.8	2013
RH-ISCRC-l_1	**62.3±6.2**	**65.6±6.7**	**66.7±6.4**	
RH-ISCRC-l_2	57.4±7.2	60.7±6.5	61.4±6.4	
Kernel	50	100	200	Year
AHISD [5]	57.2±7.5	59.6±7.4	61.8±7.3	2010
CHISD [5]	57.9±8.3	62.6±8.1	64.9±7.2	2010
KSRC [56]	61.4±7.0	65.9±6.9	67.8±6.4	2010
KCRC [35]	57.5±7.9	60.6±6.8	62.7±7.7	2012
KCH-ISCRC	**64.5±7.6**	**67.4±8.0**	**69.7±7.4**	

E. Time comparison

Then let's compare the efficiency of competing methods. The Matlab codes of all competing methods are obtained from the original authors, and we run them on an Intel(R) Core(TM) i7-2600K (3.4GHz) PC. The average running time per set on CMU MoBo (200 frames per set) is listed in Table VII. We can see that the proposed RH-ISCRC-l_2 is the fastest among all competing methods except for RNP, while RH-ISCRC-l_1 also has a fast speed. Among all the kernel based methods, the proposed KCH-ISCRC is much faster than others. Overall, the proposed RH-ISCRC and KCH-ISCRC methods have not only high ISFR accuracy but also high efficiency than the competing methods.

[8]http://www.ri.cmu.edu/publication_view.html?pub_id=3904

[9]http://seqam.rutgers.edu/site/index.php?option=com_content&view=article&id=64&Itemid=80

TABLE VII
AVERAGE RUNNING TIME PER SET ON CMU MOBO (s)

Non-kernel	Time	Kernel	Time
MSM [1]	0.338	AHISD [5]	18.546
DCC [11]	0.349	CHISD [5]	18.166
MMD [4]	3.216	KSRC [56]	35.508
SRC [31]	5.301	KCRC [35]	6.543
MDA [12]	2.035	KCH-ISCRC	**2.03**
AHISD [5]	31.365		
CHISD [5]	18.029		
SANP [8]	11.124		
CRC [32]	0.684		
RNP [24]	**0.113**		
RH-ISCRC-l_1	0.788		
RH-ISCRC-l_2	**0.280**		

F. Parameter sensitivity analysis

To verify if the proposed methods are sensitive to parameters, in this section we present the recognition accuracies with different parameter values. For RH-ISCRC, there are two parameters, λ_1 and λ_2 in Eq. (17), which need to be set. For KCH-ISCRC, there is only one parameter τ in Eq. (5). We show the recognition accuracies versus the parameters on the CMU MoBo dataset in Fig. 13, Fig. 14 and Fig. 15, respectively, for RH-ISCRC-l_1, RH-ISCRC-l_2 and KCH-ISCRC. The different colors correspond to different accuracies, as shown in the color bar. λ_1 and λ_2 are selected from $\{0.0005, 0.001, 0.01, 0.05\}$. In Fig. 13 and Fig. 14, the top subfigure is for 50 frames per set, the middle is for 100 frames per set and the bottom corresponds to 200 frames per set. From Fig. 13, we can see that the accuracy of RH-ISCRC-l_1 is very stable when λ_1 varies from 0.0005 to 0.05 and λ_2 varies from 0.0005 to 0.01. When λ_2 is increased to 0.05, the recognition performance would degrade. Fig. 14 shows that RH-ISCRC-l_2 is insensitive to the values of λ_1 and λ_2. For example, in the experiments of 100 and 200 frames per set, the accuracy variation is within 0.5% for different λ_1 and λ_2. Considering the performance of both RH-ISCRC-l_1 and RH-ISCRC-l_2, λ_1 and λ_2 can both be set as 0.001. With this parameter setting, the accuracy is very stale in different experiments. For KCH-ISCRC, its recognition accuracies with different values of τ are shown in Fig. 15. τ is set as $\{1, 2, 5, 10, 50, 100\}$. One can see that KCH-ISCRC is insensitive to τ. Hence, we simply set τ as 1.

The dictionary learning technique is used in our method to compress each image set to reduce the time complexity when representing a query image set. The number of atoms in the dictionary needs to be defined before dictionary learning. If the number of atoms is too small, the representation power of the dictionary will be reduced; if the number of atoms is large, the system tends to be under-determined and thus the solution may be less stable. We tested our algorithm by varying the number of atoms (for each sub-dictionary $\boldsymbol{D}_k$) from 5 to 50. The recognition accuracies versus the number of atoms on the CMU MoBo dataset are shown in Figs. 16-18. From Fig. 16 and Fig. 17, we can see that the recognition accuracies of both RH-ISCRC-l_1 and RH-ISCRC-l_2 vary little if the number of atoms is set within [10, 20]. From Fig. 18, we can see that for KCH-ISCRC the variation of recognition accuracies is within 0.5% under different number of atoms. This is because the feature dimension is relatively high in the kernel space and thus the solution is affected little by the dictionary size. Based on the above observation, in all our experiments we set the number of atoms as 10 or 20 for RH-ISCRC-l_1 and RH-ISCRC-l_2, and 50 for KCH-ISCRC.

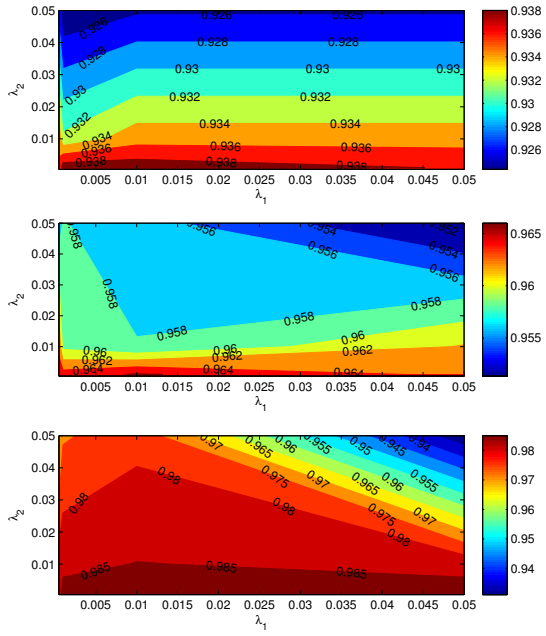

Fig. 13. Recognition accuracy of RH-ISCRC-l_1 on CMU MoBo with different λ_1 and λ_2. Different colors represent different accuracy. Top: 50 frames per set; middle: 100 frames per set; bottom: 200 frames per set.

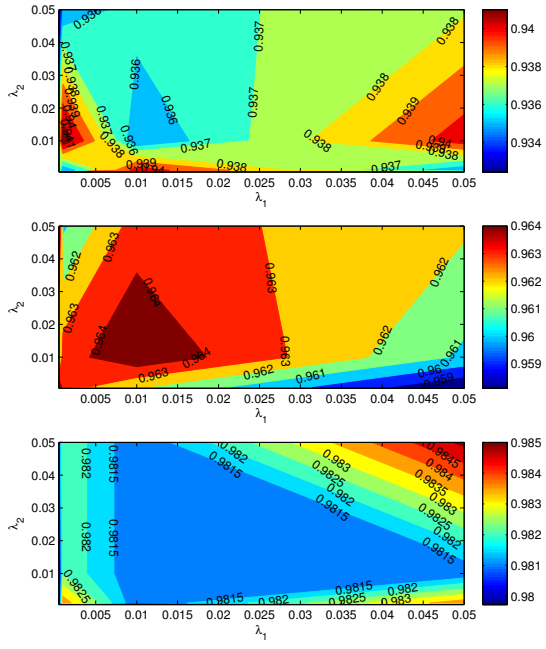

Fig. 14. Recognition accuracy of RH-ISCRC-l_2 on CMU MoBo with different λ_1 and λ_2. Different colors represent different accuracy. Top: 50 frames per set; middle: 100 frames per set; bottom: 200 frames per set.

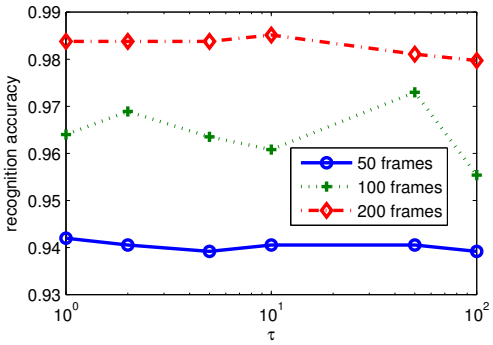

Fig. 15. Recognition accuracy of KCH-ISCRC on CMU MoBo with different τ.

VI. Conclusion

We proposed a novel image set based collaborative representation and classification (ISCRC) scheme for image set based face recognition (ISFR). The query set was modeled as a convex or regularized hull, and a collaborative representation based set to sets distance (CRSSD) was defined by representing the hull of query set over all the gallery sets. The CRSSD considers the correlation and distinction of sample images within the query set and the relationship between the gallery sets. With CRSSD, the representation residual of the hull of query set by each gallery set can be computed and used for classification. Experiments on the three benchmark ISFR databases showed that the proposed ISCRC is superior to state-of-the-art ISFR methods in terms of both recognition rates and efficiency.

Acknowledgment

The authors thank T. Kim for sharing the source code of DCC, S. Gao for the source code of KSRC and Y. Hu for the source code of SANP. We thank R. Wang for sharing the source code of MMD and MDA, and the cropped faces of the Honda/UCSD dataset and YouTube Celebrities dataset. We also thank H. Cevikalp for sharing the source code of AHISD/CHISD and providing the LBP features for the Mobo dataset. This work is partly supported by Hong Kong RGC PPR fund (K-QZ09).

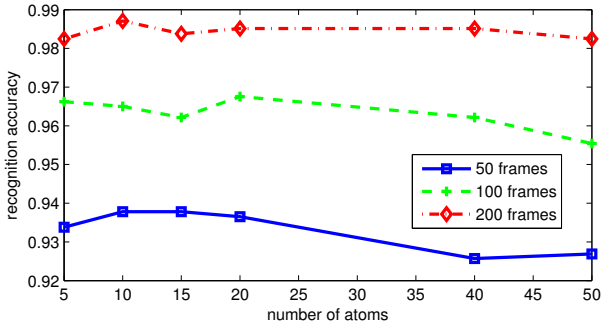

Fig. 16. Recognition accuracy of RH-ISCRC-l_1 on CMU MoBo with different number of atoms per set.

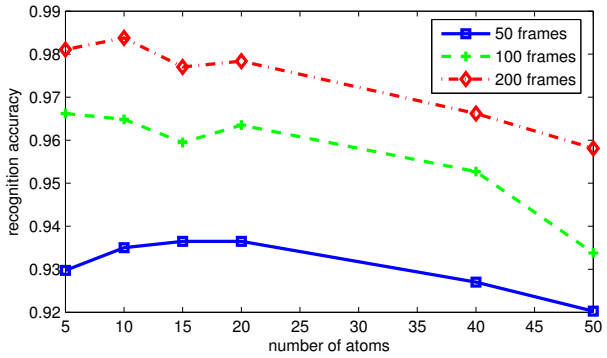

Fig. 17. Recognition accuracy of RH-ISCRC-l_2 on CMU MoBo with different number of atoms per set.

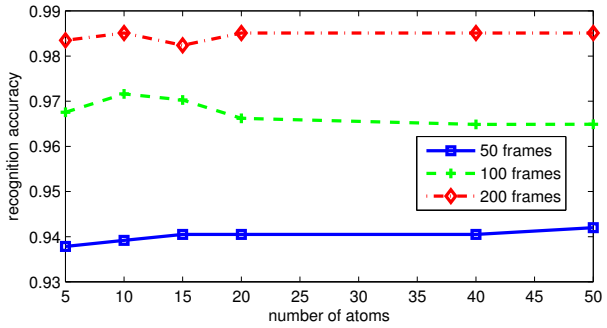

Fig. 18. Recognition accuracy of KCH-ISCRC on CMU MoBo with different number of atoms per set.

References

[1] O. Yamaguchi, K. Fukui, and K.-i. Maeda, "Face recognition using temporal image sequence," in *Automatic Face and Gesture Recognition, IEEE Conference on*. IEEE, 1998, pp. 318–323.

[2] O. Arandjelovic, G. Shakhnarovich, J. Fisher, R. Cipolla, and T. Darrell, "Face recognition with image sets using manifold density divergence," in *Computer Vision and Pattern Recognition, IEEE Conference on*. IEEE, 2005, pp. 581–588.

[3] M. Nishiyama, M. Yuasa, T. Shibata, T. Wakasugi, T. Kawahara, and O. Yamaguchi, "Recognizing faces of moving people by hierarchical image-set matching," in *Computer Vision and Pattern Recognition, IEEE Conference on*. IEEE, 2007, pp. 1–8.

[4] R. Wang, S. Shan, X. Chen, and W. Gao, "Manifold-manifold distance with application to face recognition based on image set," in *Computer Vision and Pattern Recognition, IEEE Conference on*. IEEE, 2008, pp. 1–8.

[5] H. Cevikalp and B. Triggs, "Face recognition based on image sets," in *Computer Vision and Pattern Recognition, IEEE Conference on*. IEEE, 2010, pp. 2567–2573.

[6] N. Poh, C. H. Chan, J. Kittler, S. Marcel, C. McCool, E. A. Rúa, J. L. A. Castro, M. Villegas, R. Paredes, V. Struc *et al.*, "An evaluation of video-to-video face verification," *Information Forensics and Security, IEEE Transactions on*, vol. 5, no. 4, pp. 781–801, 2010.

[7] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in *Computer Vision and Pattern Recognition, IEEE Conference on*. IEEE, 2011, pp. 529–534.

[8] Y. Hu, A. S. Mian, and R. Owens, "Sparse approximated nearest points for image set classification," in *Computer Vision and Pattern Recognition, IEEE Conference on*. IEEE, 2011, pp. 121–128.

[9] Z. Cui, S. Shan, H. Zhang, S. Lao, and X. Chen, "Image sets alignment for video-based face recognition," in *Computer Vision and Pattern Recognition, IEEE Conference on*. IEEE, 2012, pp. 2626–2633.

[10] Y.-C. Chen, V. M. Patel, P. J. Phillips, and R. Chellappa, "Dictionary-based face recognition from video," in *Computer Vision–ECCV 2012*. Springer, 2012, pp. 766–779.

[11] T. Kim, J. Kittler, and R. Cipolla, "Discriminative learning and recognition of image set classes using canonical correlations," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 6, pp. 1005–1018, 2007.

[12] R. Wang and X. Chen, "Manifold discriminant analysis," in *Computer Vision and Pattern Recognition, IEEE Conference on*. IEEE, 2009, pp. 429–436.

[13] H. J. Seo and P. Milanfar, "Face verification using the lark representation," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 4, pp. 1275–1286, 2011.

[14] K.-C. Lee, J. Ho, M.-H. Yang, and D. Kriegman, "Video-based face recognition using probabilistic appearance manifolds," in *Computer Vision and Pattern Recognition, IEEE Conference on*. IEEE, 2003, pp. 313–320.

[15] J. Stallkamp, H. K. Ekenel, and R. Stiefelhagen, "Video-based face recognition on real-world data," in *Computer Vision, IEEE International Conference on*. IEEE, 2007, pp. 1–8.

[16] A. Hadid and M. Pietikainen, "From still image to video-based face recognition: an experimental analysis," in *Automatic Face and Gesture Recognition, IEEE Conference on*. IEEE, 2004, pp. 813–818.

[17] W. Fan and D.-Y. Yeung, "Locally linear models on face appearance manifolds with application to dual-subspace based classification," in *Computer Vision and Pattern Recognition, IEEE Conference on*. IEEE, 2006, pp. 1384–1390.

[18] R. Wang, H. Guo, L. S. Davis, and Q. Dai, "Covariance discriminative learning: A natural and efficient approach to image set classification," in *Computer Vision and Pattern Recognition, IEEE Conference on*. IEEE, 2012, pp. 2496–2503.

[19] R. Caseiro, P. Martins, J. F. Henriques, F. S. Leite, and J. Batista, "Rolling riemannian manifolds to solve the multi-class classification problem," in *Computer Vision and Pattern Recognition, IEEE Conference on*. IEEE, 2013, pp. 41–48.

[20] S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi, "Kernel methods on the riemannian manifold of symmetric positive definite matrices," in *Computer Vision and Pattern Recognition, IEEE Conference on*. IEEE, 2013, pp. 73–80.

[21] R. Wang, S. Shan, X. Chen, Q. Dai, and W. Gao, "Manifold-manifold distance and its application to face recognition with image sets," *Image Processing, IEEE Transactions on*, vol. 21, no. 10, pp. 4466–4479, 2012.

[22] D. Burges and C. Crisp, "A geometric interpretation of v−svm classifiers," in *Advances in Neural Information Processing Systems*, 2000, pp. 244–250.

[23] Y. Hu, A. S. Mian, and R. Owens, "Face recognition using sparse approximated nearest points between image sets," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 10, pp. 1992–2004, 2012.

[24] M. Yang, P. Zhu, L. Van Gool, and L. Zhang, "Face recognition based on regularized nearest points between image sets," in *Automatic Face and Gesture Recognition, IEEE Conference on*. IEEE, 2013, pp. 1–7.

[25] Y. Wu, M. Minoh, M. Mukunoki, and S. Lao, "Set based discriminative ranking for recognition," in *Computer Vision–ECCV 2012*. Springer, 2012, pp. 497–510.

[26] Y. Wu, M. Minoh, and M. Mukunoki, "Collaboratively regularized nearest points for set based recognition," in *Proceedings of the British Machine Vision Conference*. BMVA Press, 2013, pp. 1–10.

[27] K. Fukui, B. Stenger, and O. Yamaguchi, "A framework for 3d object recognition using the kernel constrained mutual subspace method," in *Computer Vision–ACCV 2006*. Springer, 2006, pp. 315–324.

[28] W.-S. Chu, J.-C. Chen, and J.-J. J. Lien, "Kernel discriminant transformation for image set-based face recognition," *Pattern Recognition*, vol. 44, no. 8, pp. 1567–1580, 2011.

[29] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural computation*, vol. 12, no. 10, pp. 2385–2404, 2000.

[30] R. Rosipal and N. Krämer, "Overview and recent advances in partial least squares," in *Subspace, Latent Structure and Feature Selection*. Springer, 2006, pp. 34–51.

[31] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 2, pp. 210–227, 2009.

[32] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Computer Vision, IEEE International Conference on*. IEEE, 2011, pp. 471–478.

[33] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Regularized robust coding for face recognition," *Image Processing, IEEE Transactions on*, vol. 22, no. 5, pp. 1753–1766, 2013.

[34] S. Gao, I. W.-H. Tsang, and L.-T. Chia, "Sparse representation with kernels," *Image Processing, IEEE Transactions on*, vol. 22, no. 2, pp. 423–434, 2013.

[35] M. Yang, L. Zhang, S.-K. Shiu, and D. Zhang, "Robust kernel representation with statistical local features for face recognition," *Neural Networks and Learning Systems, IEEE Transactions on*, vol. 24, no. 6, pp. 900–912, 2013.

[36] K. Crammer, R. Gilad-Bachrach, A. Navot, and N. Tishby, "Margin analysis of the lvq algorithm," in *Advances in neural information processing systems*, 2002, pp. 462–469.

[37] S. Rosset, J. Zhu, and T. Hastie, "Boosting as a regularized path to a maximum margin classifier," *The Journal of Machine Learning Research*, vol. 5, pp. 941–973, 2004.

[38] J. Blitzer, K. Q. Weinberger, and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," in *Advances in neural information processing systems*, 2005, pp. 1473–1480.

[39] R. Gilad-Bachrach, A. Navot, and N. Tishby, "Margin based feature selection-theory and algorithms," in *Machine Learning, International Conference on*. ACM, 2004, pp. 43–50.

[40] Z. Wang, J. Yang, N. Nasrabadi, and T. Huang, "A max-margin perspective on sparse representation-based classification," in *Computer Vision, IEEE International Conference on*. IEEE, 2013.

[41] R. Rockafellar, *Convex analysis*. Princeton Univ Pr, 1997, vol. 28.

[42] H. Cevikalp, B. Triggs, H. S. Yavuz, Y. Küçük, M. Küçük, and A. Barkana, "Large margin classifiers based on affine hulls," *Neurocomputing*, vol. 73, no. 16, pp. 3160–3168, 2010.

[43] K. P. Bennett and E. J. Bredensteiner, "Duality and geometry in svm classifiers," in *Machine Learning, International Conference on*. ACM, 2000, pp. 57–64.

[44] X. Peng and Y. Wang, "Geometric algorithms to large margin classifier based on affine hulls," *Neural Networks and Learning Systems, IEEE Transactions on*, vol. 23, no. 2, pp. 236–246, 2012.

[45] V. M. Patel, T. Wu, S. Biswas, P. J. Phillips, and R. Chellappa, "Dictionary-based face recognition under variable lighting and pose," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 3, pp. 954–965, 2012.

[46] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit," *CS Technion*, 2008.

[47] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Metaface learning for sparse representation based face recognition," in *Image Processing, IEEE International Conference on*. IEEE, 2010, pp. 1601–1604.

[48] J. Zhu, S. Rosset, T. Hastie, and R. Tibshirani, "1-norm support vector machines." in *Advances in neural information processing systems*, 2003, pp. 49–56.

[49] A. Gunawardana and W. Byrne, "Convergence theorems for generalized alternating minimization procedures," *The Journal of Machine Learning Research*, vol. 6, pp. 2049–2073, 2005.

[50] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[51] U. Niesen, D. Shah, and G. W. Wornell, "Adaptive alternating minimization algorithms," *Information Theory, IEEE Transactions on*, vol. 55, no. 3, pp. 1423–1429, 2009.

[52] S. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale l1-regularized least squares," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606–617, 2007.

[53] T. F. Coleman and Y. Li, "A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables," *SIAM Journal on Optimization*, vol. 6, no. 4, pp. 1040–1058, 1996.

[54] R. Gross and J. Shi, "The cmu motion of body (mobo) database," *Technical Report*, vol. 27, no. 1, pp. 1–13, 2001.

[55] M. Kim, S. Kumar, V. Pavlovic, and H. Rowley, "Face tracking and recognition with visual constraints in real-world videos," in *Computer Vision and Pattern Recognition, IEEE Conference on*. IEEE, 2008, pp. 1–8.

[56] S. Gao, I. W.-H. Tsang, and L.-T. Chia, "Kernel sparse representation for image classification and face recognition," in *Computer Vision–ECCV 2010*. Springer, 2010, pp. 1–14.

[57] P. Viola and M. Jones, "Robust real-time face detection," *Interational Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

Pengfei Zhu received his B. S. and M. S. from Harbin Institute of Technology, Harbin, China in 2009 and 2011, respectively. He is now a PhD candidate with The Hong Kong Polytechnic University. His research interests are focused on machine learning and computer vision.

Wangmeng Zuo (M'09) received the Ph.D. degree in computer application technology from the Harbin Institute of Technology, Harbin, China, in 2007. From July 2004 to December 2004, from November 2005 to August 2006, and from July 2007 to February 2008, he was a Research Assistant at the Department of Computing, Hong Kong Polytechnic University, Hong Kong. From August 2009 to February 2010, he was a Visiting Professor in Microsoft Research Asia. He is currently an Associate Professor in the School of Computer Science and Technology, Harbin Institute of Technology. His current research interests include image modeling and low level vision, discriminative learning, biometrics, and computer vision. Dr. Zuo has published about 50 papers in those areas. Dr. Zuo is an Associate Editor of the IET Biometrics and the Scientific World.

Lei Zhang (M'04) received the B.Sc. degree in 1995 from Shenyang Institute of Aeronautical Engineering, Shenyang, P.R. China, the M.Sc. and Ph.D degrees in Control Theory and Engineering from Northwestern Polytechnical University, Xi'an, P.R. China, respectively in 1998 and 2001. From 2001 to 2002, he was a research associate in the Dept. of Computing, The Hong Kong Polytechnic University. From Jan. 2003 to Jan. 2006 he worked as a Postdoctoral Fellow in the Dept. of Electrical and Computer Engineering, McMaster University, Canada. In 2006, he joined the Dept. of Computing, The Hong Kong Polytechnic University, as an Assistant Professor. Since Sept. 2010, he has been an Associate Professor in the same department. His research interests include Image and Video Processing, Computer Vision, Pattern Recognition and Biometrics, etc. Dr. Zhang has published about 200 papers in those areas. Dr. Zhang is currently an Associate Editor of IEEE Trans. on CSVT and Image and Vision Computing. He was awarded the 2012-13 Faculty Award in Research and Scholarly Activities. More information can be found in his homepage http://www4.comp.polyu.edu.hk/ cslzhang/.

Simon C.K. Shiu (M'90) received the M.Sc. degree in computing science from the University of Newcastle Upon Tyne, Newcastle Upon Tyne, U.K., the M.Sc. degree in business systems analysis and design from City University London, London, U.K., and the Ph.D. degree in computing from The Hong Kong Polytechnic University, Hong Kong, in 1985, 1986, 1997, respectively. He is currently an Assistant Professor with the Department of Computing, The Hong Kong Polytechnic University. From 1985 to 1990, he was a System Analyst and the Project Manager with several business organizations at Hong Kong. His current research interests include case-based reasoning, machine learning, and soft computing. He co-authored the book Foundations of Soft Case-Based Reasoning (Hoboken, NJ, USA: Wiley, 2004). Dr. Shiu was a Guest Co-Editor of a special issue on soft case-based reasoning of the Applied Intelligence. He is a member of the British Computer Society.

David Zhang (M'89-SM'95-F'08) received the Graduation degree in computer science from Peking University, Beijing, China, the M.Sc. degree in computer science in 1982, and the Ph.D. degree in 1985 from the Harbin Institute of Technology (HIT), Harbin, China. In 1994, he received the second Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, Ontario, Canada.

From 1986 to 1988, he was a Postdoctoral Fellow with Tsinghua University and then an Associate Professor at the Academia Sinica, Beijing. He is currently a Chair Professor at the Hong Kong Polytechnic University, Kowloon, Hong Kong, where he is the Founding Director of the Biometrics Technology Centre (UGC/CRC) supported by the Hong Kong SAR Government in 1998. He also serves as a Visiting Chair Professor at Tsinghua University, Beijing, and an Adjunct Professor at Peking University, Shanghai Jiao Tong University, Shanghai, China, HIT, and the University of Waterloo. He is the author of more than 10 books and 250 journal papers.

Dr. Zhang is the Founder and Editor-in-Chief, International Journal of Image and Graphics (IJIG); Book Editor, Springer International Series on Biometrics (KISB); Organizer, the International Conference on Biometrics Authentication (ICBA); and Associate Editor of more than ten international journals. He is a Croucher Senior Research Fellow, Distinguished Speaker of the IEEE Computer Society, and a Fellow of International Association of Pattern Recognition.