International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 2, Issue 5, May 2015

# Face Recognition Using Collaborative Image Set

Sandeep G.S[1], Kotreshi S.N[2], Maruthi S.T[3], Arun Kumar B.T[4]
Asst. Professor, Dept of Computer Science & Engineering,
G M Institute of Technology, Davangere - 577006, Karnataka State, India

*Abstract*- **Digital imaging and communication technologies are developing in recent year; the image set based face recognition (ISFR) is becoming increasingly important. The difficulty in ISFR is how effectively and efficiently represents the query image set by using the gallery 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.**

## INTRODUCTION

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 [7], long term observations [5], personal albums and news pictures [9], 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 ca

n convey more within-class variations of the subject [6]. 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.

There are two proposed methods namely 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], [5]. 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 [7], [4], [6].

Many non-parametric set modeling methods have also been proposed, including subspace [7], [1], manifold [11], [12], [4], [8], [13], affine hull [4], [6], convex hull [4], and covariance matrix based ones [13], [14], [15].

To improve the classification performance, the kernel trick can be introduced to map the image sets to high-dimensional subspaces. In [13], an image set is represented by a covariance matrix and a Riemannian kernel function is defined to measure the similarity between 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 [16].

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 Y (each column of Y is an image in the set) is modeled as a hull Ya with the sum of coefficients in a being 1. Let Xk, K = 1, 2,…K be a gallery set. We then propose a collaborative representation based set (i.e.,Y) to sets

(i.e.,X=[X1…., Xk….XK) distance(CRSSD) that is

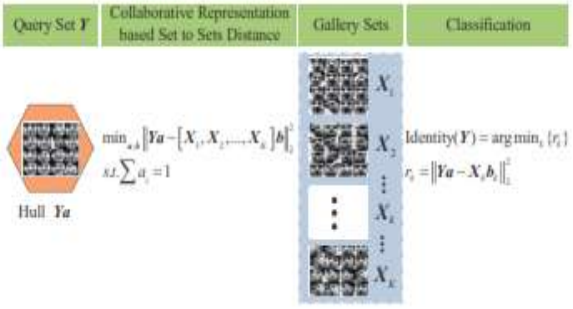


Fig1.Image set based collaborative representation and classification (ISCRC).

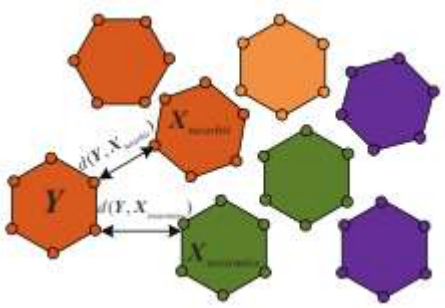


Fig. 2. Illustration of image set margin.

we represent the hull Ya over the gallery sets X as Xb, where b is a coefficient vector. Consequently, we can classify the query set Y by checking which gallery set has the minimal representation residual to the hull Ya. To get a stable solution to CRSSD, regularizations can be imposed on a and b. In the proposed ISCRC, the gallery sets $X_k$ 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.

Actually, in image set based classification, MDA [8], DCC [7] and CDL [13] 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 Y but multiple gallery sets $X_k$, where k = 1, 2, …….. K, as illustrated in Fig. 2, the image set margin is defined as:

Margin Y=d(Y, Xnearmiss) - d(Y, Xnearhit) (1)

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

## 2. COLLABORATIVE REPRESENTATION BASED SET TO SETS DISTANCE

We first introduce the hull based set to set distance and then propose the collaborative representation based set to sets distance. With CRSSD, the image set based collaborative representation and classification (ISCRC) scheme can be naturally proposed. The convex hull and regularized hull based CRSSD are respectively presented.

### A. Hull based set to set distance

One simple non-parametric set modeling approach is the hull based modeling which models a set as the linear combination of its samples.

Given a sample set $Y = \{y_1,\ldots..y_i,\ldots\ldots..y_{na}\}$, yi Є $R^d$, the hull of set Y is defined as:

H(Y) = $\{ \sum aiyi \}$. Usually, $\sum ai$ =1 is required and the coefficients ai are required to be bounded.

$$H(Y) = \{ \sum aiyi \mid \sum ai = 1, 0 <= a_i <= \tau \} \quad (2)$$

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 $Y=\{y_1,\ldots..y_i,\ldots..y_{na}\}$and set $Z=\{z_1,\ldots..z_i,\ldots..z_{nz}\}$, can be defined as follows:

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

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. 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 [17]. In image set based face recognition, there is usually no intersection between image sets of different persons.

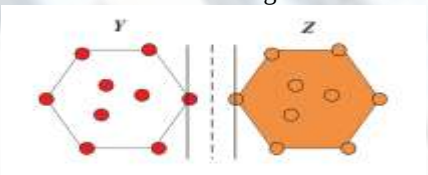


Fig 3. Convex hull based set to set distance.

### B. Collaborative representation based set to sets distance and classification

The key component of ISCRC is the collaborative representation based set to sets distance (CRSSD) defined as follows. Let $X=[X_1,\ldots\ldots X_k,\ldots\ldots X_K]$, be the concatenation of all gallery sets. We model each of Y and X as a hull, i.e., Ya and $X_b$, where a and b are coefficient vectors, and then we define the CRSSD between set Y and sets X as:

$$\min_{a,b} \|\boldsymbol{Ya} - \boldsymbol{Xb}\|^2 \ s.t. \sum a_i = 1 \quad (4)$$

By minimizing the distance between Ya and $X_b$, the outliers (e.g., one frame with large corruptions/occlusions) in both the query image set Y and the gallery image sets X will be assigned with very small representation coefficients. Therefore, the impact of outliers can be much alleviated.

### C. Convex hull based CRSSD

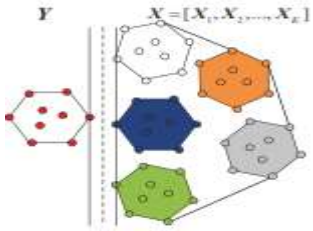


Fig. 4. Convex hull based CRSSD.

Different from the CHISD method in [4], 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 Y and the whole gallery set X in a large margin manner.

**D. $l_p$-norm regularized hull based CRSSD**

To make CRSSD more stable, the $l_p$-norm regularized hull can be used to model Y and X. For the query set Y, we should keep the constraint $\sum ai$ =1 to avoid the trivial solution and the lp-norm regularized hull of Y is defined as

$$H(\boldsymbol{Y}) = \{\sum a_i \boldsymbol{y}_i \,|\, \|\boldsymbol{a}\|_{l_p} < \delta\} \; s.t. \sum a_i = 1 \quad \textbf{(5)}$$

For the gallery set X, its regularized hull is defined as:.

$$H(\boldsymbol{X}) = \{\sum b_i \boldsymbol{x}_i \,|\, \|\boldsymbol{b}\|_{l_p} < \delta\} \quad \textbf{(6)}$$

Finally, the regularized hull based CRSSD between Y and X is defined as:

$$\min_{\boldsymbol{a},\boldsymbol{b}} \|\boldsymbol{Y}\boldsymbol{a} - \boldsymbol{X}\boldsymbol{b}\|_2^2 \quad s.t. \|\boldsymbol{a}\|_{l_p} < \delta_1, \|\boldsymbol{b}\|_{l_p} < \delta_2, \sum a_i = 1 \quad \textbf{(7)}$$

## 3. REGULARIZED HULL BASED ISCRC

In Section 2, 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 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.

**A. Main model**

In ISFR, each gallery set X often has tens to hundreds of sample images so that the whole set X can be very big, making the computational cost to solve very high. Considering the fact that the images in each set $X_k$ have high redundancy, we can compress X into a much more compact set, denoted by $D_k$, via dictionary learning methods, such as KSVD and metaface learning. Let D = [$D_1$……$D_k$…….$D_K$]. We can then replace X by D. To compute the regularized hull based CRSSD.

$$(\hat{\boldsymbol{a}}, \hat{\boldsymbol{\beta}}) = \arg\min_{\boldsymbol{a},\boldsymbol{\beta}} \left\{ \begin{array}{l} \|\boldsymbol{Y}\boldsymbol{a} - \boldsymbol{D}\boldsymbol{\beta}\|_2^2 + \\ \lambda_1 \|\boldsymbol{a}\|_{l_p} + \lambda_2 \|\boldsymbol{\beta}\|_{l_p} \end{array} \right\} \quad s.t. \sum a_i = 1 \quad \textbf{(8)}$$

Either $l_1$-norm or $l_2$-norm can be used to regularize α and β, while $l_1$-regularization will lead to sparser solutions but with more computational cost. Next, we present the optimization of $l_2$-norm and $l_1$-norm regularized hull based ISCRC.

**B. $l_2$-norm regularized hull based ISCRC**

When $l_2$-norm is used to regularize α and β has a closed-form solution. The Lagrangian function becomes

$$L(\boldsymbol{a},\boldsymbol{\beta},\lambda_3) = \|\boldsymbol{Y}\boldsymbol{a} - \boldsymbol{D}\boldsymbol{\beta}\|_2^2 + \lambda_1 \|\boldsymbol{a}\|_2^2 + \lambda_2 \|\boldsymbol{\beta}\|_2^2 + \lambda_3(e\boldsymbol{a} - 1)$$

$$= \left\| [\boldsymbol{Y} \;\; -\boldsymbol{D}] \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{\beta} \end{bmatrix} \right\|_2^2 + [\boldsymbol{a}^T \;\; \boldsymbol{\beta}^T] \begin{bmatrix} \lambda_1 \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \lambda_2 \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{\beta} \end{bmatrix} + \lambda_3([\boldsymbol{e} \;\; \boldsymbol{0}] \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{\beta} \end{bmatrix} - 1)$$

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

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

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

There are

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

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

The time complexity of RH-ISCRC-l2 is $O((n_\alpha + n_\beta)^3)$, where n is the number of sample images in Y and $n_\beta$ is the number of atoms in D.

**C. $l_1$-norm regularized hull based ISCRC**

When $l_1$ -norm regularization is used, which is very efficient to solve multiple variable optimization problems. The algorithm of RH-ISCRC-$l_1$ for ISFR is summarized in below table and it converges. One curve of the objective function value of RH-ISCRC-$l_1$ versus the iteration number. Honda/USCD database is also used. The query set Y and each gallery set $X_k$ 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 Xk into a dictionary D with 20 atoms by using the metaface learning method.

**Input**: query set Y ;
gallery sets X = [X1,…Xk,…..XK] λ1 and λ2.
**Output**: the label of query set Y.
Initialize β(0), $\lambda^{(0)}$, 0 ← t.
Compress Xk to Dk, k=1,2,…….k using metaface learning.
While t < max num do
Step 1: Update α ;
Step 2: Update β ;
Step 3: Update λ ;
Step 4: t ← t + 1.
End while

Compute $r_k = \left\| \boldsymbol{Y}\hat{\boldsymbol{a}} - \boldsymbol{D}_k \hat{\boldsymbol{\beta}}_k \right\|_2^2, k = 1, 2, \ldots K.$

Identity(Y)=arg mink { rk }

## 4. KERNELIZED CONVEX HULL BASED ISCRC

Since there can be many sample images in gallery sets, X can be a fat matrix. Even we compress X into a more compact set D, the system can still be under-determined. The lp-norm regularization on a and 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. The 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. Mapping 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 kernalized convex hull based CRSSD model is,

$$\begin{aligned} &\min_{\alpha,\beta} \|\phi(\boldsymbol{Y})\boldsymbol{\alpha} - [\phi(\boldsymbol{D}_1), \phi(\boldsymbol{D}_2), ..., \phi(\boldsymbol{D}_K)]\,\boldsymbol{\beta}\|^2 \\ &s.t. \sum \alpha_i = 1, \sum \beta_j = 1, \\ &\quad 0 \le \alpha_i \le \tau, i = 1, ..., n_\alpha, \\ &\quad 0 \le \beta_j \le \tau, j = 1, ..., n_\beta. \end{aligned} \qquad \textbf{(9)}$$

The solution exhibits global and quadratic convergence. 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

$$r_k = \left\| \Phi(Y)\ \hat{a} - \Phi(D_k)\ \beta_k \right\|_2^2 .$$

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(D, D) can be computed and stored. When a query set Y comes, we only need to calculate k(Y,Y) and k(Y,D).

**TABLE III**

**Algorithm of KCH-ISCRC for ISFR**

**Input:** query set Y ;

gallery sets X= [ X1, ….. Xk,……..XK], τ,

**Output:** the label of query set Y.

Compress $X_k$ to $D_k$, k=1,2,…..k by metaface learning;

Solve the QP problem.

Compute $r_k = \left\| \Phi(Y)\ \hat{a} - \Phi(D_k)\ \beta_k \right\|_2^2$ . k=1,2,…….k;

Identity (Y)=arg min{$r_k$}.

## 5. EXPERIMENTAL ANALYSIS

The comparison methods fall into four categories:

C1. Subspace and manifold based methods: Mutual Subspace Method (MSM) [1], Discriminant Canonical Correlations (DCC2) [7], Manifold-Manifold Distance (MMD) [3], and Manifold Discriminant Analysis (MDA4) [8].

C2. Affine/convex hull based methods: Affine Hull basedImage Set Distance (AHISD5) [4], Convex Hull based Image Set Distance (CHISD6) [4], Sparse Approximated Nearest Points (SANP7) [6], and Regularized Nearest Points (RNP) [20].

C3. Representation based methods: Sparse Representation based Classifier (SRC) [21], 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) [22], KCRC (Kernel CRC) [35], AHISD [4], and CHISD [4]. 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 [7], 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 [3]. 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.

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(-$\left\| x - y \right\|_2^2 / 2\delta^2$ )) 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 λ1 = 0:001, λ2= 0:001,λ = 2.5/na (n is the number of samples in the query set), $\gamma$ = λ/2. The number of atoms in the compressed set Dk 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 D is set as 50 for all datasets.

### B. Honda/UCSD

The Honda/UCSD dataset consists of 59 video sequences involving 20 different subjects [10]. The Viola-Jones face detector [19] is used to detect the faces in each frame and resize the detected faces to 20x20 images. Histogram equalization is utilized to reduce the illumination variations. **2**0 sequences are set aside for training and the remaining 39 sequences for testing. The intensity is used as the feature.

Fig 5. 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 proposed RH-ISCRC outperforms much all the other methods. Kernel CHISD achieves 100% recognition accuracy when all the frames in one video clip are used.

When 200 frames per set are used, both RH-ISCRC and KCHISCRC 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 $l_1$-norm regularization is strong enough to yield a good solution to the regularized hull based CRSSD.

**C. CMU MoBo**

The CMU Mobo (Motion of Body) dataset 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 [19] is used to detect the faces and the detected face images are resized to

40 X 40. The LBP feature is used.

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 KCHISCRC 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 [18]. As 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 6. The face in each frame is also detected by the Viola-Jones ace detector and resized to a 30 X 30 gray-scale image. The intensity value is used as feature. The experiment setting is same as [6], [8], [13]. Three video sequences per subject selected for training and six for testing. Five-fold cross validation experiments are conducted. The experimental results are shown in Table VI. It can seen that among the non-kernel methods, the proposed

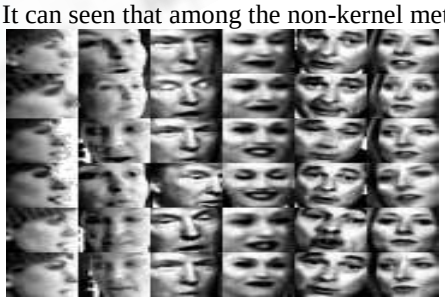

**Fig 6. Some examples of YouTube Celebrities** dataset RH-ISCRC-l achieves the highest recognition rate, while among the kernel based methods, the proposed KCH-ISCRC performs the best. Since this Youtube Celebrities dataset was established under uncontrolled environment, there are significant variations among the query and gallery sets, and therefore the $l_1$-regularization is very helpful to improve the stability and discrimination of the solution. 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 affects the performance of ISCRC. When number of frames is small, the improvement by ISCRC is more significant.

TABLE VI
RECOGNITION RATES ON YOUTUBE (VI %)

| 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-L2 is the fastest among all competing methods except for RNP, while RH-SCRC-L2 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.

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, λ1 and λ2. For KCH-ISCRC, there is only one parameter $\tau$ .When λ2 is increased to 0.05, the recognition performance would degrade.

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 D) from 5 to 50.

## 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.

## 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] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

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

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

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

[17] H. Cevikalp, B. Triggs, H. S. Yavuz, Y. K ¨uc¸¨uk, M. K¨uc¸¨uk, and A. Barkana, "Large margin classifiers based on affine hulls," Neurocomputing, vol. 73, no. 16, pp. 3160–3168, 2010.

[18] 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.

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

[20] 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.

[21] 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.

[22] 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.