

# Euler Sparse Representation for Image Classification

Yang Liu,<sup>1</sup> Quanxue Gao,<sup>1\*</sup> Jungong Han,<sup>2</sup> Shujian Wang<sup>1</sup>

<sup>1</sup>State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an, China

<sup>2</sup>School of Computing and Communications, Lancaster University, United Kingdom  
xidianliuyang@gmail.com, qxgao@xidian.edu.cn, jungonghan77@gmail.com, wangsj079@163.com

## Abstract

Sparse representation based classification (SRC) has gained great success in image recognition. Motivated by the fact that kernel trick can capture the nonlinear similarity of features, which may help improve the separability and margin between nearby data points, we propose Euler SRC for image classification, which is essentially the SRC with Euler sparse representation. To be specific, it first maps the images into the complex space by Euler representation, which has a negligible effect for outliers and illumination, and then performs complex SRC with Euler representation. The major advantage of our method is that Euler representation is explicit with no increase of the image space dimensionality, thereby enabling this technique to be easily deployed in real applications. To solve Euler SRC, we present an efficient algorithm, which is fast and has good convergence. Extensive experimental results illustrate that Euler SRC outperforms traditional SRC and achieves better performance for image classification.

## Introduction

Image classification is one of the active topics in pattern recognition and machine learning. In many applications, each image is usually represented as a point in the high-dimensional image space. However, many works have demonstrated that the naturally generated images may reside on a lower dimensional manifold. Thus, it is beneficial to first perform a dimensionality reduction (DR) prior to utilizing any technique for image classification and retrieval. Principal component analysis (PCA) (Turk and Pentland 1991), linear discriminant analysis (LDA) (Belhumeur, Hespanha, and Kriegman 1997), locality preserving projection (LPP) (Niyogi 2004) and neighborhood preserving embedding (NPE) (He et al. 2005) are four of the most prevalent techniques for DR. PCA is an unsupervised technique and used to extract the most expressive features. LDA is considered to be capable of extracting the most discriminating features. Different from PCA and LDA, which preserve the global geometric structure, LPP and NPE characterize the local geometric structure of images.

Motivated by these methods, many related approaches have been developed for dimensionality reduction, although

the motivations of these methods are different, they employ squared  $\ell_2$ -norm as the distance metric in the objective function to measure the similarity between images. It is commonly known that squared  $\ell_2$ -norm is not robust in the sense that outlying measurements can arbitrarily skew the solution from the desired solution (Ke and Kanade 2005; Candès et al. 2011; Oh et al. 2016). To improve the robustness of subspace learning methods, some  $\ell_1$ -norm based subspace learning methods have been developed, which can be broadly divided into two categories:  $\ell_1$ -norm based PCA technique and  $\ell_1$ -norm based LDA technique. As can be easily derived from the name, this sort of DR algorithms employs  $\ell_1$  normal as the distance metric in the objective function to measure the similarity between data, where the representatives include L1-PCA (Ke and Kanade 2005), PCA-L1 (Kwak 2008), and LDA-L1 (Liu et al. 2016).

Another important area requiring further investigation is the design of classifier for image classification. Nearest neighbor (NN) classifier has been widely used in classification stage by using Euclidean distance measure. It is easy to see that Euclidean distance is very sensitive to shape variation. To handle this problem, Li and Lu (Li and Lu 1999) proposed the nearest feature line (NFL) classifier. Motivated by the impressive results, Chien and Wu (Chien and Wu 2002) proposed two robust classifiers: nearest feature plane (NFP) and nearest feature space (NFS) to improve face recognition system. Gao and Wang (Gao and Wang 2007) proposed center-based nearest neighbor classifier to deal with pattern classification. Wang and Zhang (Wang and Zhang 2004) proposed linear generalization subspace (LGS) to achieve higher recognition rate using only limited training images. Zheng *et al.* (Zheng, Zhao, and Zou 2004) developed two classifiers (nearest neighbor line and nearest neighbor plane) to improve the robustness of classification.

Recently, sparse representation has shown its superior performance in image classification and computer vision (Zhang et al. 2016; Wright et al. 2009; Gao, Tsang, and Chia 2010). Wright *et al.* (Wright et al. 2009) proposed a sparse representation based classification (SRC) algorithm for face recognition, and obtained impressive results in experiments. Inspired by SRC, sparse representation has been widely used in image recognition and denoising. To improve performance of SRC, kernel sparse representation has become an active topic in image recognition (Harandi et al.

\*Corresponding author: Q. Gao.

Copyright © 2018, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

2012; Gao, Tsang, and Chia 2010). Despite its relatively better performance, an essential mapping from the image to a high-dimensional hidden space required by those methods is computationally expensive, thus impeding their usage in real-world applications.

Inspired by the fact that kernel trick can capture the nonlinear similarity of features, which can suppress outliers, we propose Euler SRC for image classification. Euler SRC is essentially the SRC in Euler space. It comes from the same DR family tree as the kernel sparse representation, but they are clearly different in the way of organizing mapping. More specifically, Euler SRC first maps the images into the complex space by Euler representation, which has a negligible effect for outliers and illumination, and then performs complex SRC with Euler representation. The main advantage of our method is that Euler representation not only is explicit but also does not increase the dimensionality of image space, enabling the fast and simple computation. Extensive experiments illustrate the effectiveness of proposed approach.

### Sparse representation based classification

Classification is one of the particularly simple applications for sparse representation. It assumes that the training samples from a single class do lie on a subspace. Thus, for any test sample  $\mathbf{y} \in \mathbf{R}^m$ , it will approximately lie in the linear span of the samples, which has the same class label with  $\mathbf{y}$ . Assume that we have a new matrix  $\mathbf{X} \in \mathbf{R}^{m \times N}$  for the entire training set sampled from the  $C$  classes, then  $\mathbf{y}$  can be represented as a linear representation of the entire training samples. The corresponding coefficient vector  $\mathbf{s} \in \mathbf{R}^N$  can be obtained by solving the following objective function (Candès and Tao 2006).

$$\min_{\mathbf{s}} \|\mathbf{s}\|_1 \quad \text{s.t.} \quad \mathbf{X}\mathbf{s} = \mathbf{y} \quad (1)$$

where  $\|\cdot\|_1$  denotes the  $\ell_1$ -norm, which counts the sum of modulus of each element in a vector.

The objective function (1) illustrates that  $\mathbf{y}$  can be exactly represented. In real applications, data are noisy, implying that it is impossible to exactly express  $\mathbf{y}$  as a sparse superposition of the training samples. In this case, the objective function (1) is modified as follows (Wright et al. 2010).

$$\min_{\mathbf{s}} \frac{1}{2} \|\mathbf{X}\mathbf{s} - \mathbf{y}\|_2^2 + \alpha \|\mathbf{s}\|_1, \quad \alpha > 0 \quad (2)$$

After obtaining the coefficient vector  $\mathbf{s}$  by the objective function (2), SRC classifies the test sample  $\mathbf{y}$  according to residual error. To be specific, for each class  $i$ , let  $\delta_i : \mathbf{R}^N \rightarrow \mathbf{R}^N$  be the characteristic function that selects the coefficients associated with the  $i$ th class.  $\delta_i(\mathbf{s}) \in \mathbf{R}^N$  is a new vector whose only nonzero entries are the entries in  $\mathbf{s}$  that are associated with class  $i$ . Then, we assign  $\mathbf{y}$  to be the  $j$ th class if

$$r_j(\mathbf{y}) = \min_i r_i(\mathbf{y}) = \|\mathbf{y} - \mathbf{X}\delta_i(\mathbf{s})\|_2 \quad (3)$$

It is easy to see that SRC has a good classification performance in image recognition, especially for images with occlusion. As can be seen in the objective function (2), SRC employs squared  $\ell_2$ -norm as the distance metric in the first

term in the objective function (2). It is commonly known that the variations between images, which have the same class label under different illumination and time, is larger than the change of the image identity under the squared  $\ell_2$ -norm distance metric. Thus, this affects the robustness of SRC and reduces the flexibility of SRC. To solve this problem, we introduce Euler SRC in the following section.

## Euler SRC

### Motivation

To achieve robustness, the contribution of distance metric to the criterion function (2) should reduce the effect of large distance. Motivated by the fact that kernel trick can capture the nonlinear similarity of features, which can suppress outliers, we propose Euler SRC for image classification. Prior to formulating the proposed Euler SRC, we first introduce cosine kernel function, i.e., cosine distance metric and then analyze its advantages.

**Definition 1** (Fitch et al. 2005; Liwicki et al. 2013) Given two arbitrary vectors  $\mathbf{x}_j$  and  $\mathbf{x}_q \in \mathbf{R}^m$ , the cosine distance metric between them is

$$d(\mathbf{x}_j, \mathbf{x}_q) = \sum_{c=1}^m \{1 - \cos(\alpha\pi(\mathbf{x}_j(c) - \mathbf{x}_q(c)))\} \quad (4)$$

where  $\mathbf{x}_j(c)$  is the  $c$ th element of  $\mathbf{x}_j$ .

Let us consider two motivating examples in which different dissimilarity measures are applied to the images shown in Figure 1. Figure 1 shows one image, which is marked by A, and the other eight images that are its 4 nearest neighbor images having the same class label with it and its 4 nearest neighbor images having different class labels with it. Table 1 and Table 2 list the distance between A and the other 8 images under different distance metrics. As can be seen in Figure 1, Table 1 and Table 2, the  $\ell_2$ -norm associates a smaller margin ( $8.6263-7.1618 = 1.1645$ ). In contrast, the use of the cosine-based measure results in a large margin ( $488.2649-451.0045=37.2604$ ). It shows that cosine distance metric can help obtain a large margin. Thus, cosine distance metric, which is defined in Eq. (4), is an effective kernel function. It is not only robust to outliers but also suitable for classification compared with the squared  $\ell_2$ -norm.

Table 1: Comparison of normalized dissimilarity measures between images having the same class label in Figure 1.

Dissimilarity	A-B	A-C	A-D	A-E
Euclidean	5.1609	5.7977	6.173	7.1618
Cosine-based	192.9056	296.278	320.8623	451.0045

Table 2: Comparison of normalized dissimilarity measures between images having different class labels in Figure 1.

Dissimilarity	A-F	A-G	A-H	A-K
Euclidean	8.6263	8.7619	9.0132	9.0332
Cosine-based	488.2649	682.7639	570.5157	683.3511

Combining the aforementioned analysis, we can get the following interesting conclusion:

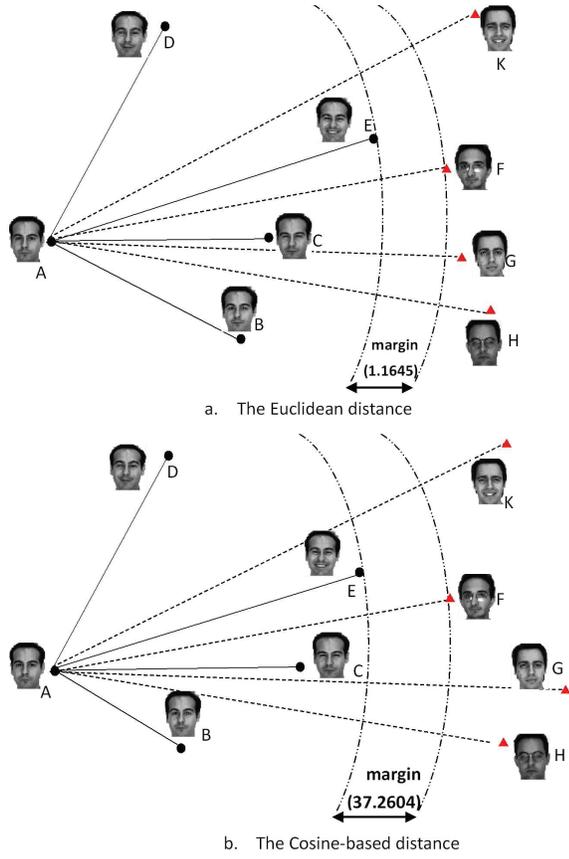


Figure 1: Original image and its 4 nearest images having the same class label and its 4 nearest images having different class labels with it. Images with the same shape belong to the same person.

- First, the cosine distance metric enlarges the distance between all image so it can help obtain a large margin and make it easy to realize classification.
- Second, the distance between the same class images was enlarged a smaller multiple than the distance between the different class images by the cosine distance metric. Therefore, compared with Euclidean distance, cosine distance metric not only can help obtain a large margin but also improve the separability between different class image.

Combining the aforementioned analysis for cosine distance metric, we can easy build a robust formulation for SRC. To be specific, we use cosine distance instead of squared  $\ell_2$ -norm as the distance metric in the image pixel space.

By simple algebra, Eq. (4) becomes

$$\begin{aligned}
 d(\mathbf{x}_j, \mathbf{x}_q) &= \sum_{c=1}^m \{1 - \cos(\gamma\pi(\mathbf{x}_j(c) - \mathbf{x}_q(c)))\} \\
 &= \left\| \frac{1}{\sqrt{2}}(e^{i\gamma\pi\mathbf{x}_j} - e^{i\gamma\pi\mathbf{x}_q}) \right\|_2^2 \\
 &= \|\mathbf{z}_j - \mathbf{z}_q\|_2^2
 \end{aligned} \tag{5}$$

where

$$\mathbf{z}_j = \frac{1}{\sqrt{2}} \begin{bmatrix} e^{i\gamma\pi\mathbf{x}_j(1)} \\ \vdots \\ e^{i\gamma\pi\mathbf{x}_j(c)} \end{bmatrix} = \frac{1}{\sqrt{2}} e^{i\gamma\pi\mathbf{x}_j} \tag{6}$$

is called Euler representation of  $\mathbf{x}_j$ .

Eq. (5) illustrates that cosine distance metric between two images in pixel space is equivalent to squared  $\ell_2$ -norm between the corresponding two vectors with Euler representation. Moreover, compared with cosine distance metric, squared  $\ell_2$ -norm is easy to be performed in solving the optimization problem. Thus, our proposed robust method first maps the intensity values  $\mathbf{x}_j$  normalized in  $[0, 1]$  onto the complex representation  $\mathbf{z}_j \in \mathbb{C}^m$  and then performs complex SRC in Euler space.

### Euler SRC

Denote by  $\mathbf{A} = \frac{1}{\sqrt{2}} e^{i\alpha\pi\mathbf{X}}$  and  $\mathbf{b} = \frac{1}{\sqrt{2}} e^{i\gamma\pi\mathbf{y}}$ , according to the aforementioned analysis, Euler SRC aims to seek the coefficient vector  $\mathbf{z}$  by solving the following objective function.

$$\min_{\mathbf{z}} \frac{1}{2} \|\mathbf{A}\mathbf{z} - \mathbf{b}\|_2^2 + \alpha \|\mathbf{z}\|_1, \quad \alpha > 0 \tag{7}$$

where  $\|\mathbf{z}\|_2^2$  is defined as  $\|\mathbf{z}\|_2^2 = \mathbf{z}^* \mathbf{z}$ , where  $\mathbf{z}^*$  is the complex conjugate transpose of  $\mathbf{z}$  and  $\|\mathbf{z}\|_1$  is defined as  $\|\mathbf{z}\|_1 = \sum_{i=1}^m |z_i|$  with  $z_i$  the  $i$ th component of  $\mathbf{z}$ .

Before solving the objective function (7), we first introduce the following definition.

**Definition 2** (Fletcher 2013). Subgradient of  $\|\mathbf{z}\|_1$ , denoted by  $\partial\|\mathbf{z}\|_1$ , as follows:

$$\partial\|\mathbf{z}\|_1 = \left\{ \mathbf{u} \mid u_i = \frac{z_i}{|z_i|} \text{ if } z_i \neq 0, \text{ and } |u_i| \leq 1 \text{ otherwise} \right\} \tag{8}$$

**Theorem 1** (Complex SRC for sparse coding): For the following function

$$\arg \min_{\mathbf{z}} (\|\mathbf{b} - \mathbf{A}\mathbf{z}\|_2^2 + \alpha \|\mathbf{z}\|_1) \tag{9}$$

where matrix  $\mathbf{A}$  and  $\mathbf{b}$  are complex valued matrix and vector respectively. The sparse representation  $\mathbf{z}$  corresponding to the input signal  $\mathbf{b}$  can be iteratively solved by

$$\mathbf{z} = |\mathbf{Z}| \mathbf{A}^* (\mathbf{A} |\mathbf{Z}| \mathbf{A}^* + \alpha \mathbf{I})^{-1} \mathbf{b} \tag{10}$$

where  $\mathbf{Z}$  is a square diagonal matrix built upon the components of vector  $\mathbf{z}$ , i.e.,  $\mathbf{Z} = \text{diag}(\mathbf{z})$ .  $|\mathbf{Z}|$  is also a square diagonal matrix whose diagonal elements are the abstract value of the corresponding diagonal elements of  $\mathbf{Z}$ .

**Proof:** The Lagrangian function of the problem (9) is

$$L(\mathbf{z}) = \frac{1}{2} \|\mathbf{b} - \mathbf{A}\mathbf{z}\|_2^2 + \alpha \|\mathbf{z}\|_1 \tag{11}$$

The KKT condition for optimal solution of the problem (11) specifies that the gradient of  $L$  w.r.t.  $\mathbf{z}_i$  must be zero, i.e.,

$$\frac{\partial L}{\partial \mathbf{z}} = \mathbf{A}^* (\mathbf{A}\mathbf{z} - \mathbf{b}) + \alpha \mathbf{u} = 0 \tag{12}$$

Note that,  $\mathbf{u}$  is related to  $\mathbf{z}$ .

To get the root of the model (12), we adopt a fixed point iteration algorithm to solve it. The core of fixed point algorithm is to convert the equation to the form  $\mathbf{z} = g(\mathbf{z})$ .

According to the definition of subgradient of  $\|\mathbf{z}\|_1$ , we have

$$\mathbf{u} = |\mathbf{Z}|^{-1} \mathbf{z} \quad (13)$$

In Eq.(13), it may have  $\frac{0}{0}$  when the elements of  $\mathbf{Z}$  have zero. We only use it for the sake of subsequent derivation and do not calculate it. Moreover, it will vanish in the final representation (See Eq. (14)).

Substituting (13) into (12), and by simple algebra, we get

$$\begin{aligned} \mathbf{z} &= (\mathbf{A}^* \mathbf{A} + \alpha |\mathbf{Z}|^{-1})^{-1} \mathbf{A}^* \mathbf{b} \\ &= \{(\mathbf{A}^* \mathbf{A} |\mathbf{Z}| + \alpha \mathbf{I}) |\mathbf{Z}|^{-1}\}^{-1} \mathbf{A}^* \mathbf{b} \\ &= |\mathbf{Z}| (\mathbf{A}^* \mathbf{A} |\mathbf{Z}| + \alpha \mathbf{I})^{-1} \mathbf{A}^* \mathbf{b} \end{aligned} \quad (14)$$

According to  $(\mathbf{A}^* \mathbf{A} |\mathbf{Z}| + \alpha \mathbf{I}) \mathbf{A}^* = \mathbf{A}^* (\mathbf{A} |\mathbf{Z}| \mathbf{A}^* + \alpha \mathbf{I})$ , we have

$$(\mathbf{A}^* \mathbf{A} |\mathbf{Z}| + \alpha \mathbf{I})^{-1} \mathbf{A}^* = \mathbf{A}^* (\mathbf{A} |\mathbf{Z}| \mathbf{A}^* + \alpha \mathbf{I})^{-1} \quad (15)$$

Combining (14) and (15) yields

$$\mathbf{z} = |\mathbf{Z}| \mathbf{A}^* (\mathbf{A} |\mathbf{Z}| \mathbf{A}^* + \alpha \mathbf{I})^{-1} \mathbf{b} \quad (16)$$

As can be seen in Eq. (16), we can iteratively solve the complex sparse vector of the model (9) by Eq. (16). ■

According to the fixed point iteration algorithm and Eq. (16), we summarize the pseudo code for solving the objective function (7) in Algorithm 1.

---

#### Algorithm 1: Euler SRC algorithm

---

**Input:** Training Sample Matrix  $\mathbf{X} \in \mathbf{R}^{m \times N}$ , Parameter  $\gamma$  ( $0 < \gamma < 2$ ),  $\alpha$  ( $\alpha > 0$ ),  $t = 1$ ,  $\tau = 10^{-6}$ .

**Initialize**  $\mathbf{z}^1 = \mathbf{1}$ .

**Procedure**

1. Map each column vector in  $\mathbf{X}$  in  $[0,1]$ .
  2. Calculate  $\mathbf{A} = \frac{1}{\sqrt{2}} \exp(\gamma * \pi i * \mathbf{X})$ .
  3. If  $m > N$ , calculate  $\mathbf{z}^{t+1}$  by Eq. (14), otherwise, calculate  $\mathbf{z}^{t+1}$  by Eq. (16).
  4.  $t = t + 1$ , go to step 3 until  $\|\mathbf{z}^{t+1} - \mathbf{z}^t\|_1 < \tau$ .
  5. Return  $\mathbf{z}^t$ .
- 

## Experiments

In this section, we validate our proposed method on four databases (COIL20, AR, PIE and LFWcrop), and compare it with the recently proposed algorithms SRC (Wright et al. 2009) and Riemannian Sparse Representation (RSR) (Harandi et al. 2012). In the experiments,  $\alpha = 0.5$ , PCA and down-sampling are selected as preprocessing for each approach respectively, where PCA reduces dimensionality to 200. In Euler SRC, we set  $\gamma = 1.9$  in the following experiments.

### Experiments on Image Classification

The AR database (Martinez 1998) contains over 4000 color face images of 126 people, including frontal views of faces with different facial expressions, lighting conditions and occlusions. The pictures of 120 individuals (65 men and 55 women) are taken in two sessions (separated by two weeks).

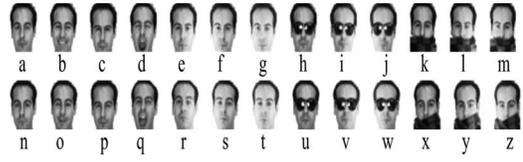


Figure 2: Some samples in the AR dataset.

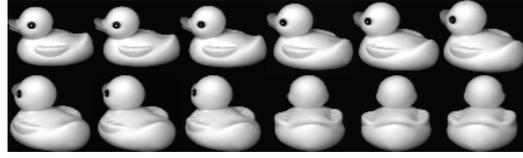


Figure 3: Some samples in the COIL20 dataset.



Figure 4: Some samples in the PIE dataset.



Figure 5: Some samples in the LFWCrop dataset.

Each session contains 13 color images, in which 6 of them are with occlusion and the other 7 full facial images are with different facial expressions and lighting conditions. In the experiments, we manually cropped the face portion of the image and then normalized it to  $50 \times 40$  pixels. Figure 2 shows some normalized images of one person. We do three group experiments in this database. Experiment 1: We consider the corrupted training images with sunglasses. We select facial images (a)-(g) and image (h) with sunglasses for training, and the remaining full facial images and the images with sunglasses, which include 2 images from session one and 3 images from session two, for testing. In summary, we have 960 training images and 1440 testing images. Experiment 2: We consider the impact of occlusion images corrupted by scarfs, which occlude about 40% of the face image, for image classification. In the experiments, we select images from (a) to (g) and image (k) per person for training, and images from (n) to (t) and image (x) per person for testing. Thus, we again end up with 960 training images and 1440 testing images. Experiment 3: We consider the impact of occlusion images corrupted by sunglasses and scarfs. In this experiment, we select 7 full facial images from (a) to (g) and two corrupted images (h) and (k) per person for training, and the remaining images for testing. Thus, we have 1080

Table 3: The average classification rate (%) and standard deviation on the AR dataset using PCA as a preprocessing.

Methods	SRC	RSR	Euler SRC
Exp.1	80.00±0.49	81.26±0.39	<b>82.98±0.42</b>
Exp.2	77.74±1.12	81.29±0.33	<b>83.63±0.83</b>
Exp.3	78.72±0.19	79.90±0.20	<b>81.96±1.15</b>

Table 4: The average classification rate (%) and standard deviation on the AR dataset using down-sampling as a preprocessing.

Methods	SRC	RSR	Euler SRC
Exp.1	77.80±0.40	79.01±0.39	<b>80.32±0.89</b>
Exp.2	75.90±0.19	77.00±0.29	<b>77.25±0.29</b>
Exp.3	76.21±0.11	80.23±0.17	<b>81.37±1.14</b>

Table 5: The average classification rate (%) and standard deviation on the COIL20, PIE and LFWCrop datasets using PCA as a preprocessing.

Methods	SRC	RSR	Euler SRC
COIL20	98.71±0.77	98.93±0.27	<b>99.28±0.50</b>
PIE	96.76±1.55	97.01±1.23	<b>97.74±1.53</b>
LFWCrop	50.40±3.04	53.62±1.01	<b>55.13±2.86</b>

Table 6: The average classification rate (%) and standard deviation on the COIL20, PIE and LFWCrop datasets using down-sampling as a preprocessing.

Methods	SRC	RSR	Euler SRC
COIL20	99.05±0.60	99.51±0.31	<b>99.63±0.29</b>
PIE	96.72±1.64	97.03±0.91	<b>97.99±1.82</b>
LFWCrop	51.03±2.10	51.06±1.01	<b>52.20±1.86</b>

training images and 2040 testing images. All of the aforementioned experiments are repeated 3 times.

The COIL20 database (Nene et al. 1996) includes 1440 color images of 20 objects (72 images per object). The object has a wide variety of complex geometric and reflectance characteristics. This database, called Columbia Object Image Library (COIL-20), was used in a real-time 20 object recognition system. Each object was placed in a stable configuration at approximately the center of the turntable. The turntable was then rotated through 360 degrees and 72 images were taken per object; one at every 5 degrees of rotation. Some sample images of one object are shown in Figure 3. In this database, we randomly select 10 images per subject for training, and the remaining images for testing. All of experiments are repeated 10 times.

The CMU PIE database (Sim, Baker, and Bsat 2002) contains 68 subjects with 41368 face images as a whole. Face images of each person are captured by 13 synchronized cameras and 21 flashes, under 43 different illumination conditions, and with 4 different expressions. In the experiment, we select pose C05 (a nearly frontal pose) as gallery which

includes 68 persons and 49 images for each person. All images were manually aligned, cropped, and resized to  $64 \times 64$  pixels. Some sample images of one person are shown in Figure 4. In this database, we randomly select 7 images per subject for training, and the remaining images for testing. All of experiments are repeated 10 times.

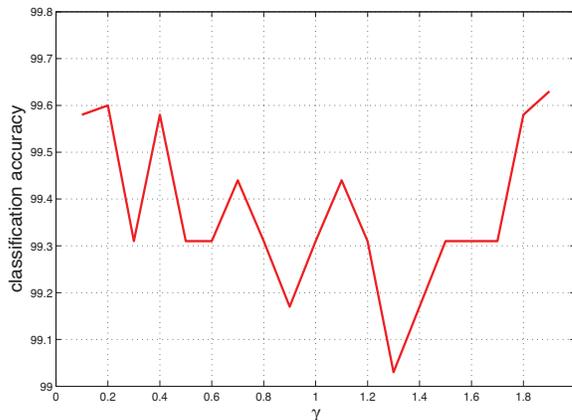


Figure 6: The average accuracy of Euler SRC vs.  $\gamma$  on the COIL20 database

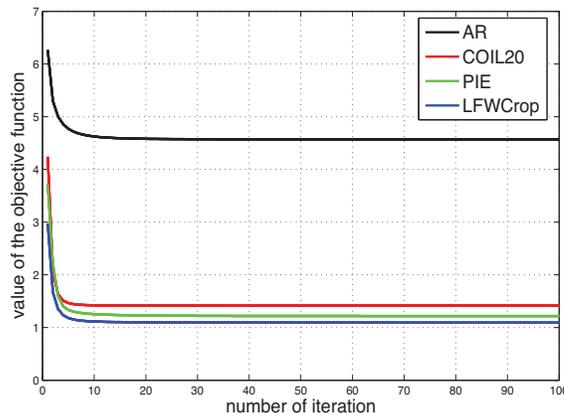


Figure 7: Convergence curve of Euler SRC on four databases.

The LFWCrop database (Sanderson and Lovell 2009) is a cropped version of the Labeled Faces in the Wild (LFW) dataset (Huang et al. 2007). In the vast majority of images, almost all of the background is omitted. The extracted area was then scaled to a size of  $64 \times 64$  pixels. The cropped faces in LFWCrop exhibit real-life conditions, including misalignment, scale variations, in-plane as well as out-of-plane rotations. Some sample images of one person are shown in Figure 5. In the experiment, we choose person who has more than 20 photos but less than 100 photos as the sub-dataset, which contains 57 classes and 1883 images. For each person, we randomly choose ninety percent of images for training, and the remaining images for testing. We repeat this process 10 times.

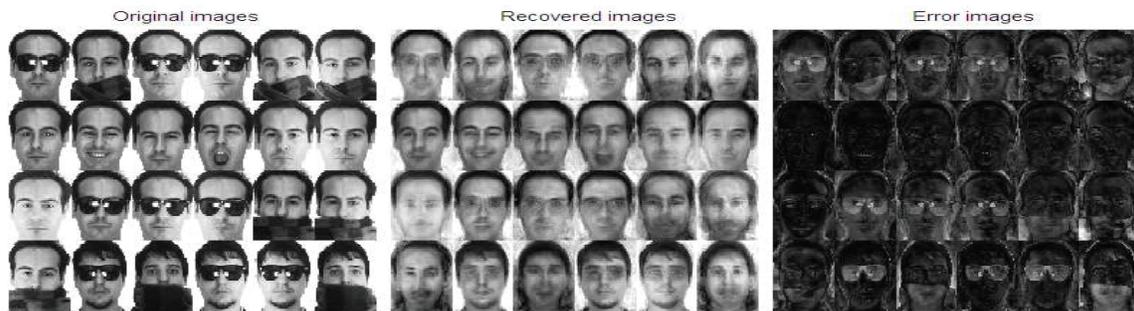


Figure 8: Face recovery results of Euler SRC on the AR dataset.



Figure 9: Face recovery results of SRC on the AR dataset.

Table 3 and Table 4 list the average classification accuracy and standard deviation on the AR dataset. Table 5 and Table 6 list the average classification accuracy and standard deviation on the COIL20, PIE and LFWCrop datasets, respectively. Figure 6 plots the recognition accuracy of Euler SRC vs. parameter  $\gamma$  on the COIL20 database. Figure 7 shows the convergence curve of Euler SRC on the AR, COIL20, PIE and LFWCrop databases respectively.

Comparing the aforementioned experiments, we have the several interesting observations as follows.

- Table 3 to Table 6 show that Euler SRC is overall superior to SRC and RSR. This is probably because that Euler space can help improve the separability of data and obtain a large margin for different classes. Moreover, SRC and RSR are performed in image pixels space, which is very sensitive to outliers and illumination.
- Figure 6 shows that when  $\gamma = 1.9$ , we get a best classification accuracy on the COIL20 database. This is probably because that Euler representation can well suppress outliers. It is consistent with (Liwicki et al. 2013). Figure 6 also illustrates that parameter *gamma* has a small effect for the performance of Euler SRC.
- Figure 7 illustrates that our proposed iterative algorithm monotonically decreases the objective function in each iteration and can get converged within a few iterations, although we cannot prove it in theoretical analysis. We will study it in our future work.

### Experiments on Face Representation

To illustrate the robustness of Euler SRC, We evaluate Euler SRC and SRC for face representation on the AR database. The training images is the same as that in the aforementioned experiments in the AR dataset. The original face images, recovered faces and error faces of Euler SRC and SRC are described in Figure 8 and Figure 9, respectively. As can be seen in Figure 8 and Figure 9, Euler SRC well remove the occlusions from the noise images. It means that Euler SRC is superior to SRC for image reconstruction. The reason may be that Euler representation is more robust to outliers than image pixels representation.

### Conclusions

We present a robust Euler sparse representation, which is called Euler SRC, for image classification. Different from kernel SRC, Euler SRC maps the images into a Euler space by explicit Euler representation and does not increase the dimensionality of image space. Thus, it is easy to realize Euler SRC in real applications. We provide an algorithm to find an exact sparse representation, which can best optimize the objective function. Experiments on the AR, COIL20, PIE and LFWCrop databases illustrate the advantages of the proposed approach.

### Acknowledgements

This work is supported by National Natural Science Foundation of China under Grant 61271296 and 61773302, and the 111 Project of China (B08038).

## References

- Belhumeur, P. N.; Hespanha, J. P.; and Kriegman, D. J. 1997. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7):711–720.
- Candès, E. J., and Tao, T. 2006. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE transactions on information theory* 52(12):5406–5425.
- Candès, E. J.; Li, X.; Ma, Y.; and Wright, J. 2011. Robust principal component analysis? *Journal of the ACM (JACM)* 58(3):11.
- Chien, J. T., and Wu, C. C. 2002. Discriminant waveletfaces and nearest feature classifiers for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(12):1644–1649.
- Fitch, A. J.; Kadyrov, A.; Christmas, W. J.; and Kittler, J. 2005. Fast robust correlation. *IEEE Transactions on Image Processing* 14(8):1063–1073.
- Fletcher, R. 2013. *Practical methods of optimization*. John Wiley & Sons.
- Gao, Q. B., and Wang, Z. Z. 2007. Center-based nearest neighbor classifier. *Pattern Recognition* 40(1):346–349.
- Gao, S.; Tsang, I. W.-H.; and Chia, L.-T. 2010. Kernel sparse representation for image classification and face recognition. In *European Conference on Computer Vision*, 1–14.
- Harandi, M. T.; Sanderson, C.; Hartley, R.; and Lovell, B. C. 2012. Sparse coding and dictionary learning for symmetric positive definite matrices: A kernel approach. In *Computer Vision—ECCV 2012*. Springer. 216–229.
- He, X.; Cai, D.; Yan, S.; and Zhang, H.-J. 2005. Neighborhood preserving embedding. In *IEEE International Conference on Computer Vision*, volume 2, 1208–1213. Beijing, China: IEEE.
- Huang, G. B.; Ramesh, M.; Berg, T.; and Learned-Miller, E. 2007. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, Technical Report 07-49, University of Massachusetts, Amherst.
- Ke, Q., and Kanade, T. 2005. Robust  $L_1$  norm factorization in the presence of outliers and missing data by alternative convex programming. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, 739–746. San Diego, CA: IEEE.
- Kwak, N. 2008. Principal component analysis based on  $l_1$ -norm maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(9):1672–1680.
- Li, S. Z., and Lu, J. 1999. Face recognition using the nearest feature line method. *IEEE Transactions on Neural Networks* 10(2):439–443.
- Liu, Y.; Gao, Q.; Gao, X.; Nie, F.; and Li, Y. 2016. A non-greedy algorithm for  $l_1$ -norm lda. *IEEE Transactions on Image Processing* 69(1):DOI: 10.1109/TIP.2016.2621667.
- Liwicki, S.; Tzimiropoulos, G.; Zafeiriou, S.; and Pantic, M. 2013. Euler principal component analysis. *International Journal of Computer Vision* 101(3):498–518.
- Martinez, A. M. 1998. The ar face database. *CVC Technical Report 24*.
- Nene, S. A.; Nayar, S. K.; Murase, H.; et al. 1996. Columbia object image library (coil-20). Technical report, Technical report CUCS-005-96.
- Niyogi, X. 2004. Locality preserving projections. In *Neural information processing systems*, volume 16, 153.
- Oh, T. H.; Tai, Y. W.; Bazin, J. C.; Kim, H.; and Kweon, I. S. 2016. Partial sum minimization of singular values in robust pca: Algorithm and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38(4):744–758.
- Sanderson, C., and Lovell, B. C. 2009. *Multi-Region Probabilistic Histograms for Robust and Scalable Identity Inference*. Springer Berlin Heidelberg.
- Sim, T.; Baker, S.; and Bsat, M. 2002. The cmu pose, illumination, and expression (pie) database. In *IEEE International Conference on Automatic Face and Gesture Recognition*, 46–51.
- Turk, M., and Pentland, A. 1991. Eigenfaces for recognition. *Journal of cognitive neuroscience* 3(1):71–86.
- Wang, H., and Zhang, L. 2004. Linear generalization probe samples for face recognition. *Pattern Recognition Letters* 25(8):829–840.
- Wright, J.; Yang, A. Y.; Ganesh, A.; Sastry, S. S.; and Ma, Y. 2009. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(2):210–227.
- Wright, J.; Ma, Y.; Mairal, J.; Sapiro, G.; Huang, T. S.; and Yan, S. 2010. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE* 98(6):1031–1044.
- Zhang, Z.; Li, F.; Zhao, M.; Zhang, L.; and Yan, S. 2016. Joint low-rank and sparse principal feature coding for enhanced robust representation and visual classification. *IEEE Transactions on Image Processing* 25(6):2429–2443.
- Zheng, W.; Zhao, L.; and Zou, C. 2004. Locally nearest neighbor classifiers for pattern classification. *Pattern Recognition* 37(6):1307–1309.