# Component SPD matrices: A low-dimensional discriminative data descriptor for image set classification

**Kai-Xuan Chen**[1,2], **Xiao-Jun Wu**[1,2] (✉)

**Abstract**  In pattern recognition, the task of image set classification has often been performed by representing data using symmetric positive definite (SPD) matrices, in conjunction with the metric of the resulting Riemannian manifold. In this paper, we propose a new data representation framework for image sets which we call *component symmetric positive definite* representation (CSPD). Firstly, we obtain sub-image sets by dividing the images in the set into square blocks of the same size, and use a traditional SPD model to describe them. Then, we use the Riemannian kernel to determine similarities of corresponding sub-image sets. Finally, the CSPD matrix appears in the form of the kernel matrix for all the sub-image sets; its $i,j$-th entry measures the similarity between the $i$-th and $j$-th sub-image sets. The Riemannian kernel is shown to satisfy Mercer's theorem, so the CSPD matrix is symmetric and positive definite, and also lies on a Riemannian manifold. Test on three benchmark datasets shows that CSPD is both lower-dimensional and more discriminative data descriptor than standard SPD for the task of image set classification.

**Keywords**  symmetric positive definite (SPD) matrices; Riemannian kernel; image classification, Riemannian manifold

## 1  Introduction

Image set classification has received wide attention in the domains of artificial intelligence and pattern

1  School of IoT Engineering, Jiangnan University, Wuxi 214122, China. E-mail: K.-X. Chen, kaixuan_chen_jnu@163.com; X.-J. Wu, xiaojun_wu_jnu@163.com (✉).

2  Jiangsu Provincial Engineering Laboratory of Pattern Recognition and Computational Intelligence, Jiangnan University, Wuxi 214122, China.

recognition [1–8]. An image set contains a large number of images taken under different conditions, allowing more robust discrimination than use of single-shot images [9–13]. Representations of image sets used for the task of image classification commonly include Gaussian mixture models [14], linear subspaces [1], and covariance descriptors (CovDs) [7, 8, 15]. The latter have been widely applied, e.g., in virus recognition [13], object detection [4], and gesture classification [5]. The traditional SPD model is based on CovDs, and SPD matrices lie on a non-linear manifold called the SPD manifold.
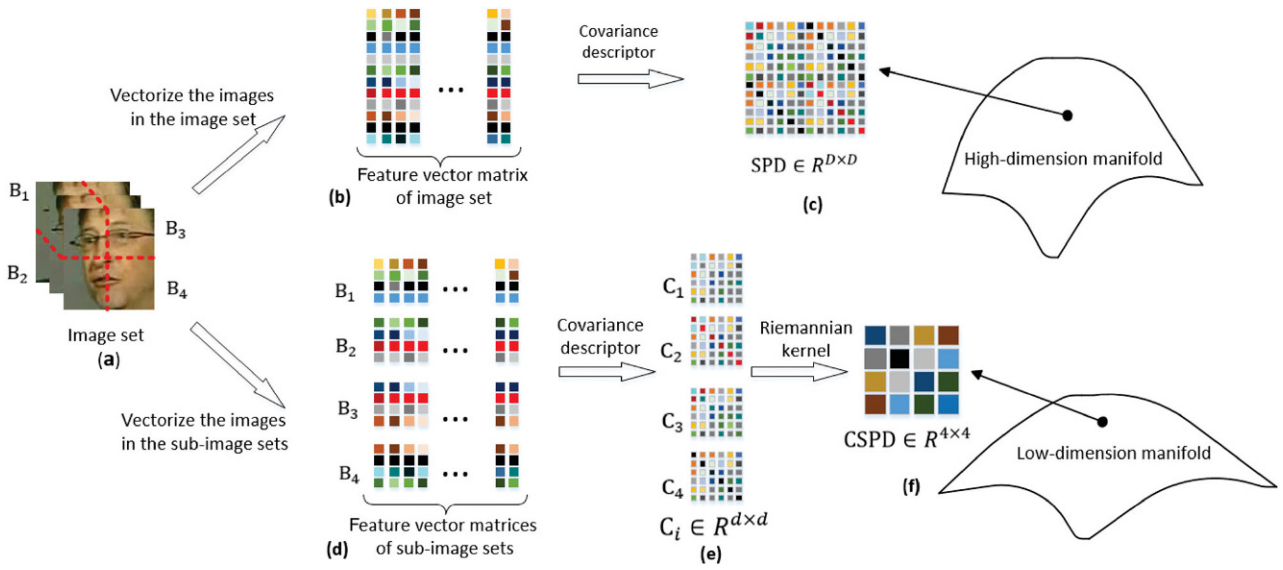
The dimensionality of traditional SPD matrices [2, 4–8] used for representing image sets is relatively high, which results in excessive computation and low efficiency of algorithms. Dimensionality reduction is always important in computer vision and machine learning. Classical methods, such as principal component analysis (PCA) [16] and linear discriminant analysis (LDA) [17] are widely used in various applications. However, as SPD matrices lie on a non-linear Riemannian manifold, these methods are unsuitable for analyzing SPD matrices. Recently, the work extending dimensionality reduction to Riemannian manifolds has received wide attention. Bidirectional covariance matrices (BCM) [8] and SPDML [2] are dimensionality reduction methods which can be used with the SPD manifold. BCM [8] applies two-directional two-dimensional PCA [18] directly to the SPD matrices to obtain low-dimensional descriptors. SPDML [2] embeds the high-dimensional SPD matrices into a lower-dimensional, more discriminative SPD manifold through projection.

In this paper, we propose a new framework to obtain low-dimensional, discriminative descriptors for representing image sets. Figure 1 shows pipelines

**Fig. 1** Processing an image set using the traditional SPD approach and our CSPD approach. Above: traditional SPD. The SPD matrix is computed using a covariance descriptor; it lies on a non-linear SPD manifold. Below: our CSPD approach. It firstly divides images in the set into square blocks of the same size. The $i$-th sub-image set $B_i$ is represented by a traditional SPD model. The Riemannian kernel is then used to describe the similarity between the sub-image sets. The final CSPD appears in the form of the Riemannian kernel matrix of the representations of sub-image sets.

for producing descriptors of image sets, for both our framework and the traditional SPD model. Given an image set with $n$ images, in traditional SPD, the images are vectorized to obtain an image set matrix $S = [s_1, \ldots, s_n]$ (see Fig. 1(b)), where $s_i \in R^D$ represents the $i$-th image in the image set. Using CovDs results in a $D \times D$ SPD matrix (see Fig. 1(c)) for the representation of the entire image set. Its $i, j$-th element is the covariance between the $i$-th and $j$-th rows of the image set matrix $S$ (see Fig. 1(b)). Unlike the traditional SPD model, we describe the image set by measuring similarities between sub-image sets. We firstly divide the images in the set into $d \times d$ blocks of the same size, and obtain $d^2$ sub-image sets. We then vectorize the sub-images (see Fig. 1(d)) and use covariance descriptors to represent them (Fig. 1(e)). Finally, the CSPD (Fig. 1(f)) takes the form of the Riemannian kernel matrix; its $i, j$-th element denotes the similarity between the $i$-th and $j$-th sub-image sets. The dimensionality of the CSPD matrix is $d^2 \times d^2$, which depends on the number of sub-image sets.

The rest of this paper is organized as follows. In Section 2, briefly overview the geometry of the SPD manifold and some related classical Riemannian metrics. In Section 3, we present the traditional SPD model and our CSPD model, and introduce some SPD manifold-based classification algorithms

used in our experiments. In Section 4, we present experimental results, which show that CSPD is a lower-dimensional and more discriminative data descriptor than standard SPD for the task of image set classification. In Section 5, we consider the effects of block size. In Section 6, we present our conclusions and discuss future directions.

## 2  Related work

In this section, we overview the geometry of the SPD manifold and some related classical Riemannian metrics. We adopt the following notation: $S_n^+$ is the space spanned by real $n \times n$ SPD matrices, $S_n$ is the tangent space spanned by real $n \times n$ symmetric matrices at the point of the identity matrix $I_n \in R^{n \times n}$, and $T_P S_n^+$ is the tangent space spanned by real $n \times n$ symmetric matrices at the point $P \in S_n^+$.

### 2.1  SPD manifold

As described in Ref. [2], the SPD manifold spanned by SPD matrices is a non-linear Riemannian manifold, and does not satisfy the scalar multiplication axiom of a vector space. For example, the matrix resulting by multiplying an SPD matrix by a negative scalar does not lie on $S_n^+$ [11]. Thus, similarity of two SPD matrices cannot be meaningfully computed using the Euclidean metric, and instead, Riemannian metrics are a better tool for analyzing SPD matrices.

A variety of Riemannian metrics have been proposed for SPD manifolds. In particular, the affine invariant Riemannian metric (AIRM) [2, 8, 19] is the most widely studied Riemannian metric, and has the property of affine invariance. The Stein divergence and Jeffrey divergence [2, 10, 20], which are efficient metrics to measure geodesic distance on the SPD manifold, are equivalent to Bregman divergence for some special seed function. The log-Euclidean metric (LEM) [7, 8] gives the similarity between two SPD matrices by computing their distance in tangent space. We now consider AIRM and LEM in detail, as these two metrics are used in our experiments.

## 2.2 Affine invariant Riemannian metric

$S_n^+$ is the Riemannian manifold spanned by SPD matrices; it can be viewed as a convex cone in $n(n + 1)/2$ dimensional Euclidean space [2]. The similarity between two SPD matrices on the manifold can be described by the length of the geodesic curve linking them, which is analogous to the straight line between two points in vector space. For point $P$ on the SPD manifold, the affine invariant Riemannian metric can be defined through two tangent vectors $u, v \in T_P S_n^+$ as

$$\langle u, v \rangle_P \triangleq \left\langle P^{-\frac{1}{2}} u P^{-\frac{1}{2}}, P^{-\frac{1}{2}} v P^{-\frac{1}{2}} \right\rangle = \mathrm{tr}\left(P^{-1} u P^{-1} v\right) \tag{1}$$

The AIRM geodesic distance $d_{\mathrm{AIRM}}$ between two points $Sp_i$ and $Sp_j$ on the SPD manifold can be written as

$$d_{\mathrm{AIRM}}(Sp_i, Sp_j) = \| \log\left(Sp_i^{-\frac{1}{2}} Sp_j Sp_i^{-\frac{1}{2}}\right) \|_{\mathrm{F}} \tag{2}$$

where $\| \cdot \|_{\mathrm{F}}$ denotes the Frobenius norm, $\log(\cdot)$ is the matrix logarithm operator.

## 2.3 Log-Euclidean metric

The Log-Euclidean metric [4, 7, 8, 11] is a bivariant Riemannian metric coming from Lie group multiplication on SPD matrices [11]. The distance $d_{\mathrm{LogED}}$ between two points $Sp_i$ and $Sp_j$ on the SPD manifold computed by this metric can be written as

$$d_{\mathrm{LogED}} = \| \log\left(Sp_i\right) - \log\left(Sp_j\right) \|_{\mathrm{F}} \tag{3}$$

The LEM can be viewed as the distance between points in the domain of the matrix logarithm, which is the tangent space of the SPD manifold obtained by logarithmic mapping [7, 8]:

$$\varphi_{\log} : S_n^+ \to S_n, \quad Sp \to \log(Sp), \quad Sp \in S_n^+ \tag{4}$$

where $S_n$ is a vector space. Figure 2 illustrates the concept of logarithm mapping. Furthermore,
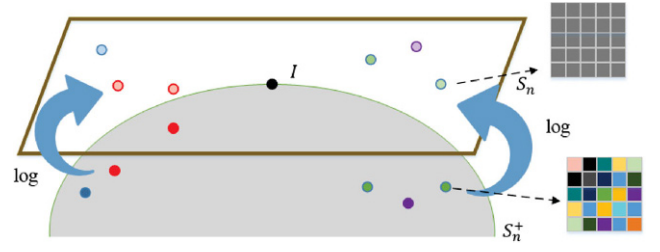


**Fig. 2**  Logarithmic mapping.

the associated Riemannian kernel function can be computed by the inner product of points in the tangent space [7, 11]:

$$k_{\mathrm{LogE}}(Sp_i, Sp_j) = \langle Sp_i, Sp_j \rangle_{\mathrm{LogE}}$$
$$= \mathrm{tr}(\log(Sp_i) \log(Sp_j)) \tag{5}$$

For the all points $Sp_1, \ldots, Sp_N \in S_n^+$, $k_{\mathrm{LogE}}$ is a symmetric function because $k_{\mathrm{LogE}}(Sp_i, Sp_j) = k_{\mathrm{LogE}}(Sp_j, Sp_i)$. From Ref. [7], we have

$$\sum_{i,j} a_i a_j k_{\mathrm{LogE}}(Sp_i, Sp_j)$$
$$= \sum_{i,j} a_i a_j \mathrm{tr}(\log(Sp_i), \log(Sp_j))$$
$$= \mathrm{tr}\left[\left(\sum_i a_i \log(Sp_i)\right)^2\right]$$
$$= \left\|\sum_i a_i \log(Sp_i)\right\|_{\mathrm{F}}^2 \geqslant 0, \quad a_i \in R, \quad \forall i \in N \tag{6}$$

Equation (6) shows that the Log-Euclidean kernel guarantees the positive definiteness of the Riemannian kernel which thus satisfies Mercer's theorem.

## 3 SPD and CSPD models

In this section, we briefly describe the traditional SPD model [7, 8] and then introduce our proposed CSPD model in detail. We also discuss classification algorithms based on these approaches.

### 3.1 Traditional SPD model

For an image set matrix $S = [s_1, \ldots, s_n]$ with $n$ images, let $s_i \in R^D$ be a feature vector obtained by vectorizing the $i$-th image in the set. The final descriptor of the image set is a $D \times D$ SPD matrix [2, 7, 8] computed from covariance descriptors:

$$C = \frac{1}{n} \sum_{i=1}^{n} (s_i - \bar{s})(s_i - \bar{s})^{\mathrm{T}} = \frac{1}{n} S J_n S^{\mathrm{T}} \tag{7}$$

where $\bar{s} = \frac{1}{n} \sum_{i=1}^{n} s_i$ is the mean vector of $S$. $J_n = I_n - \frac{1}{n} 1_n 1_n^{\mathrm{T}}$ is the centering matrix, and $1_n$ is a column vector of $n$ ones [2]. $J_n$ is a symmetric matrix with rank $n - 1$, and $J_n^2 = J_n$. In general, the number of the

images is often smaller than the dimensionality of the feature vector, so we need to add a small perturbation to avoid covariance matrix singularity [7]:

$$C^* = C + \lambda I \qquad (8)$$

where $\lambda$ is set to $10^{-3}\text{tr}(C)$ and $I$ is the identity matrix [7].

In this way, the image sets are represented as SPD matrices, which lie on the SPD manifold.

## 3.2 Component SPD model

Our proposed framework, which describes the similarities between sub-image sets, offers lower-dimensional and more discriminative descriptors for image sets than SPD. We firstly divide the images in the set into $d \times d$ square blocks of the same size, and each sub-image set is described by covariance descriptors (see Eq. (7)). This results in $d^2$ SPD matrices representing sub-image sets. The final CSPD matrix is the Riemannian kernel matrix of these SPD matrices, and its dimensionality is $d^2 \times d^2$.

See Fig. 1(below) again. In this figure, the images were divided into $2 \times 2$ square blocks to form 4 sub-image sets: $B_1, \ldots, B_4$. There are 4 corresponding covariance descriptors: $C_1, \ldots, C_4$. The CSPD lies in $\in R^{4 \times 4}$, and is a matrix describing the similarity between these 4 sub-image sets. In order to measure similarity between the sub-image sets, we use the Log-Euclidean kernel (Eq. (5)) to compute the similarity of the covariance descriptors (Fig. 1(f)):

$$\text{CSPD}_{i,j} = k_{\text{LogE}}(C_i, C_j) = \text{tr}(\log(C_i)\log(C_j)) \quad (9)$$

where $\text{CSPD}_{i,j}$ measures the similarity between the $i$-th and $j$-th sub-image sets. Note that as Eq. (6) shows that the log-Euclidean kernel guarantees positive definiteness of the Riemannian kernel, the CSPD matrix also lies on the SPD manifold.

## 3.3 Classification algorithms based on the SPD manifold

The nearest neighbor (NN) algorithm is one of the simplest methods for classification and regression used in computer vision and pattern recognition. Nearest neighbour classification algorithms based on AIRM and LEM have been utilized with the SPD manifold [8], and these simple classification algorithms clearly show the advantages of our CSPD model (see later).

Covariance discriminative learning (CDL) has been proposed for image set classification [7], and classical classification algorithms are directly based on the SPD manifold. They derive a kernel function that maps the SPD matrices from the Riemannian manifold to Euclidean space through the LEM metric. This allows classical classification algorithms operating on a linear space to be exploited in the kernel formulation. Linear discriminant analysis (LDA) and partial least squares (PLS) in this linear space have been used for the task of classification [7].

We also consider the Riemannian sparse coding algorithm LogEKSR [11], which takes the Riemannian geometry of the SPD manifold into account and applies sparse representation and dictionary learning to SPD matrices by mapping the SPD matrices into reproducing kernel Hilbert space (RKHS). Note that the log-Euclidean kernels in this algorithm, which are the derivatives of those in Eq. (7) and meet Mercer's theorem, are respectively the polynomial kernel, exponential kernel, and Gaussian kernel.

# 4 Results and discussion

## 4.1 Preliminaries

In order to verify the effectiveness of our model, we have carried out experiments on three tasks: object categorization, hand gesture recognition, and virus cell classification using the three datasets ETH-80 [4], the Cambridge hand gesture dataset (CG) [5], and the virus dataset [13] respectively. In our experiments, we compared the accuracies resulting from our proposed CSPD model with those from the traditional SPD model using the same classification algorithms. To do so, we used the most common nearest neighbor classifiers based on AIRM [2, 8] and LEM [4, 8, 11], as described earlier. As well as these two NN classifiers, we made use of classical Riemannian classification algorithms LogEKSR [11] and CDL [7], which are the efficient methods on SPD manifolds. In the following, we use the following notation:

- NN-AIRM$_{\text{SPD}}$: AIRM-based nearest neighbor classifier on the SPD manifold spanned by traditional SPD matrices.
- NN-AIRM$_{\text{CSPD}}$: AIRM-based nearest neighbor classifier on the SPD manifold spanned by our proposed CSPD matrices.
- NN-LogED$_{\text{SPD}}$: LEM-based nearest neighbor classifier on the SPD manifold spanned by traditional SPD matrices.

- NN-LogED$_{\text{CSPD}}$: LEM-based nearest neighbor classifier on the SPD manifold spanned by our proposed CSPD matrices.
- CDL$_{\text{SPD}}$: CD on the SPD manifold spanned by traditional SPD matrices.
- CDL$_{\text{CSPD}}$: CDL on the SPD manifold spanned by our proposed CSPD matrices.
- LogEKSR$_{\text{SPD}}$: LogEKSR on the SPD manifold spanned by traditional SPD matrices.
- LogEKSR$_{\text{CSPD}}$: LogEKSR on the SPD manifold spanned by our proposed CSPD matrices.

In our experiments, we re-sized all the images to $24 \times 24$, allowing them to be divided into $2 \times 2$, $3 \times 3$, $4 \times 4$, $6 \times 6$, $8 \times 8$, and $12 \times 12$ blocks. With this image size, the dimensionality of the traditional SPD is $576 \times 576$. Instead, the dimensionality of the CSPD will be $4 \times 4$, $9 \times 9$, $16 \times 16$, $36 \times 36$, $64 \times 64$, and $144 \times 144$. It is thus clear that our approach provides a lower-dimensional data representation. Next, we use the results of the experiments to verify the discriminative power of our model.

### 4.2 Object categorization

The ETH-80 dataset has eight categories of images: apples, pears, tomatoes, cows, dogs, horses, cups, and cars. Each class has 10 image sets, and each image set comprises 41 images from different angles. Figure 3(top) shows some images in the ETH-80 dataset. For each class, we randomly chose 2 image sets as training data, and the remaining image sets were used as test data. We give average accuracies and standard deviations of the 10 cross validation experiments. Table 1 shows the performance of both our proposed CSPD model and the traditional SPD model using four classification algorithms. The results for our CSPD model here use $6 \times 6$ blocks.
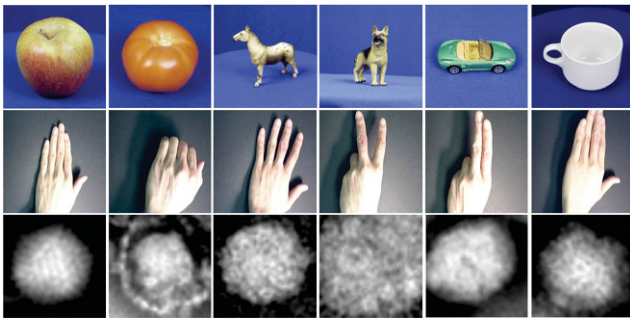


**Fig. 3** Images from the three datasets. Top: ETH-80 [4]. Middle: CG [5]. Bottom: virus [13].

**Table 1** Recognition rates with standard deviations, ETH-80 dataset

| Method | Accuracy |
| --- | --- |
| NN-AIRM$_{\text{SPD}}$ | $58.22 \pm 6.35$ |
| NN-AIRM$_{\text{CSPD}}$ | $84.92 \pm 4.54$ |
| NN-LogED$_{\text{SPD}}$ | $64.48 \pm 6.25$ |
| NN-LogED$_{\text{CSPD}}$ | $87.52 \pm 3.85$ |
| CDL$_{\text{SPD}}$ | $78.66 \pm 7.07$ |
| CDL$_{\text{CSPD}}$ | $88.92 \pm 3.95$ |
| LogEKSR$_{\text{SPD}}$ | $86.94 \pm 4.58$ |
| LogEKSR$_{\text{CSPD}}$ | $\mathbf{89.92 \pm 3.84}$ |

### 4.3 Hand gesture recognition

The Cambridge hand gesture dataset consists of a set of high resolution color sequences acquired by a Senz3D sensor showing an image sequence of hand gestures defined by 3 primitive hand shapes and 3 primitive motions. This dataset has 900 image sets in 9 classes with 100 image sets in each class (see Fig. 3(middle)). For the task of hand gesture recognition, 20 image sets of each class were randomly selected as training data, and the remaining image sets were used as test data. Ten-fold cross validation experiments were carried out on this dataset. We give the average accuracies with standard deviations for the ten experiments in Table 2. The results for our CSPD model are again based on $6 \times 6$ blocks.

### 4.4 Virus cell classification

The virus dataset contains 15 categories, each category having 5 image sets, each with 20 pictures taken from different angles (see Fig. 3(bottom)). We arbitrarily chose 3 for training and the rest for testing. Table 3 shows the results using our proposed CSPD model and the traditional SPD model using four classification algorithms. The results for our CSPD model are this time based on $4 \times 4$ blocks.

**Table 2** Recognition rates with standard deviations, hand gesture dataset

| Method | Accuracy |
| --- | --- |
| NN-AIRM$_{\text{SPD}}$ | $51.77 \pm 2.47$ |
| NN-AIRM$_{\text{CSPD}}$ | $76.39 \pm 1.81$ |
| NN-LogED$_{\text{SPD}}$ | $67.49 \pm 1.60$ |
| NN-LogED$_{\text{CSPD}}$ | $80.60 \pm 1.48$ |
| CDL$_{\text{SPD}}$ | $89.23 \pm 2.00$ |
| CDL$_{\text{CSPD}}$ | $90.84 \pm 1.20$ |
| LogEKSR$_{\text{SPD}}$ | $89.69 \pm 1.19$ |
| LogEKSR$_{\text{CSPD}}$ | $\mathbf{91.02 \pm 1.54}$ |

**Table 3**  Recognition rates with standard deviations, virus cell dataset

| Method | Accuracy |
|---|---|
| NN-AIRM$_\text{SPD}$ | 27.57 ± 4.34 |
| NN-AIRM$_\text{CSPD}$ | 33.67 ± 6.33 |
| NN-LogED$_\text{SPD}$ | 25.97 ± 4.62 |
| NN-LogED$_\text{CSPD}$ | 41.07 ± 6.10 |
| CDL$_\text{SPD}$ | 45.30 ± 5.65 |
| CDL$_\text{CSPD}$ | **54.50 ± 7.38** |
| LogEKSR$_\text{SPD}$ | 47.13 ± 4.58 |
| LogEKSR$_\text{CSPD}$ | 53.77 ± 6.44 |

### 4.5  Analysis

For all three datasets, the results for all four classifiers show that our proposed CSPD can provide more discriminative and robust features than traditional SPD. Especially for the ETH-80 dataset, the recognition rates of the two NN classifiers based on our proposed CSPD, NN-AIRM$_\text{CSPD}$ and NN-LogED$_\text{CSPD}$, are higher than for the two state-of-the-art classification algorithms CDL and LogEKSR, based on traditional SPD. Also, the accuracies of the two classifiers CDL and LogEKSR show that our proposed CSPD is more discriminative, and LogEKSR$_\text{CSPD}$ achieves the best result with an accuracy of 89.92% and standard deviation of 3.84. For the hand gesture and virus datasets, the advantages are not as obvious as for ETH-80. Nevertheless, LogEKSR$_\text{CSPD}$ still achieves the best result with an accuracy of 91.02% and standard deviation of 1.54 for the hand gesture dataset, while CDL$_\text{CSPD}$ achieves the best result with an accuracy of 54.50% and standard deviation of 7.38 on the virus dataset.

## 5  Effect of block size

Next, we will present the effects of varying block size on accuracy and running for a fixed classification algorithm. We show results using the ETH-80 dataset as an example. The notation used is as follows:
- SPD$^\text{TR}$: traditional SPD data representation.
- CSPD$^{n \times n}$: CSPD descriptor obtained by dividing the image into $n \times n$ blocks.

### 5.1  Effects of block size on accuracy

In order to display the effects of block size, we give the average accuracies achieved for traditional SPD as well as for 6 different CSPD descriptors arising from different block sizes. Table 4 shows the average

**Table 4**  Effect of block size on average accuracy

| Method | NN-AIRM | NN-LogED | CDL | LogEKSR |
|---|---|---|---|---|
| SPD$^\text{TR}$ | 58.22 | 64.48 | 78.66 | 86.94 |
| CSPD$^{2\times2}$ | 56.72 | 63.48 | 62.98 | 59.58 |
| CSPD$^{3\times3}$ | 73.73 | 79.89 | 81.02 | 81.63 |
| CSPD$^{4\times4}$ | 82.31 | 85.52 | 86.45 | 86.44 |
| CSPD$^{6\times6}$ | 84.92 | 87.52 | 88.92 | 89.92 |
| CSPD$^{8\times8}$ | 85.83 | 86.66 | 88.03 | 89.34 |
| CSPD$^{12\times12}$ | 86.86 | 84.19 | 87.52 | 88.69 |

accuracies achieved with different data descriptors and NN-AIRM, NN-LogED, CDL, and LogEKSR classification algorithms. The data in each row are the average recognition rates using the same data descriptor and different classification algorithms; the columns give the average recognition rates for the same classification algorithm with different data descriptors. The recognition rates using the CSPD model are lower than for the traditional SPD model when using $2 \times 2$ blocks. However, the algorithms have better recognition rates than for SPD when using our proposed CSPD model and larger blocks.

In order to show the robustness of our proposed CSPD model, we give average standard deviations of ten experiments in Table 5. The data in the row are the standard deviations of accuracy using the same data descriptor with different classification algorithms, while columns give standard deviations of accuracy for the same classification algorithm with different data descriptors. The results show that standard deviations from our CSPD model are generally lower than for the traditional SPD model, for the same classification algorithm, especially for larger block sizes.

The results in these two tables show that, typically, our CSPD model works best when the images were divided into $6 \times 6$ blocks, justifying the use of this size in the results in Tables 1 and 2. The results of classification algorithms based on the CSPD model in Table 3 were obtained by dividing the images in

**Table 5**  Effect of block size on standard deviation of accuracy

| Method | NN-AIRM | NN-LogED | CDL | LogEKSR |
|---|---|---|---|---|
| SPD$^\text{TR}$ | 6.35 | 6.25 | 7.07 | 4.58 |
| CSPD$^{2\times2}$ | 5.66 | 5.67 | 6.66 | 5.99 |
| CSPD$^{3\times3}$ | 5.47 | 4.18 | 4.40 | 4.37 |
| CSPD$^{4\times4}$ | 4.91 | 4.78 | 4.27 | 3.74 |
| CSPD$^{6\times6}$ | 4.54 | 3.85 | 3.95 | 3.84 |
| CSPD$^{8\times8}$ | 4.00 | 4.31 | 3.82 | 4.12 |
| CSPD$^{12\times12}$ | 4.24 | 4.23 | 4.18 | 3.72 |

the virus cell dataset into $4 \times 4$ blocks, which also works well.

## 5.2 Effects of block size on run time

CSPD matrices have lower dimensionality than traditional SPD matrices. This property saves run time. We consider the efficiency of our CSPD model in two ways: (i) the run time using different data representation models while using the same classifier, and (ii) the time needed to compute data descriptors (SPD or CSPD).

Table 6 shows the time needed to compute data descriptors (SPD or CSPD). Clearly, the time needed for CSPD is less than for traditional SPD when the image set is divided into $2 \times 2$, $3 \times 3$, $4 \times 4$, and $6 \times 6$ blocks. In general, all descriptors take comparable time.

Secondly, we compare classification time using different data descriptors in Table 7. Each row gives run time for the same data descriptor while using different classification algorithms, while columns give run time using the same classification algorithm with different data descriptors.

These two tables show that our CSPD model takes least time when the images were divided into $2 \times 2$ blocks. Even with $12 \times 12$ blocks, our CSPD model takes far less time than traditional SPD while using the same classification algorithm. We can thus conclude that our CSPD model improves the efficiency of algorithms significantly.

## 6 Conclusions and future work

In this paper, we have proposed the component symmetric positive definite (CSPD) model as a novel descriptor for image sets. Its superior time performance is due to its lower dimensionality, and it also shows better discriminative ability, providing higher recognition rates than those from traditional SPD when using the same classification algorithm. The latter is clearly demonstrated by results from two nearest neighbor classification algorithms. In future, we hope to devise further data descriptors for image set classification.

## References

[1] Huang, Z.; Wang, R.; Shan, S.; Chen, X. Projection metric learning on Grassmann manifold with application to video based face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 140–149, 2015.

[2] Harandi, M.; Salzmann, M.; Hartley, R. Dimensionality reduction on SPD manifolds: The emergence of geometry-aware methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 40, No. 1, 48–62, 2017.

[3] Chang, F.-J.; Nevatia, R. Image set classification via template triplets and context-aware similarity embedding. In: *Computer Vision–ACCV 2016. Lecture Notes in Computer Science, Vol. 10115.* Lai, S. H.; Lepetit, V.; Nishino, K.; Sato, Y. Eds. Springer Cham, 231–247, 2016.

[4] Huang, Z.; Wang, R.; Shan, S.; Li, X.; Chen, X. Log-Euclidean metric learning on symmetric positive definite manifold with application to image set classification. In: Proceedings of the 32nd International Conference on Machine Learning, Vol. 37, 720–729, 2015.

[5] Faraki, M.; Harandi, M. T.; Porikli, F. Image set classification by symmetric positive semi-definite matrices. In: Proceedings of the IEEE Winter Conference on Applications of Computer Vision, 1–8, 2016.

[6] Chen, Z.; Jiang, B.; Tang, J.; Luo, B. Image set representation and classification with attributed covariate-relation graph model and graph sparse representation classification. *Neurocomputing* Vol. 226, 262–268, 2017.

**Table 6** Time needed to compute data descriptors

| Method | Time |
|---|---|
| $\mathrm{SPD^{TR}}$ | 3.03 s |
| $\mathrm{CSPD^{2\times2}}$ | 2.80 s |
| $\mathrm{CSPD^{3\times3}}$ | 2.73 s |
| $\mathrm{CSPD^{4\times4}}$ | 2.47 s |
| $\mathrm{CSPD^{6\times6}}$ | 2.98 s |
| $\mathrm{CSPD^{8\times8}}$ | 3.17 s |
| $\mathrm{CSPD^{12\times12}}$ | 3.64 s |

**Table 7** Time taken by different data descriptors with the same classification algorithm

| Method | NN-AIRM | NN-LogED | CDL | LogEKSR |
|---|---|---|---|---|
| $\mathrm{SPD^{TR}}$ | 31063.0 ms | 4168.9 ms | 348.6 ms | 877.5 ms |
| $\mathrm{CSPD^{2\times2}}$ | 13.6 ms | 3.4 ms | 8.1 ms | 2.6 ms |
| $\mathrm{CSPD^{3\times3}}$ | 20.6 ms | 3.9 ms | 9.2 ms | 4.6 ms |
| $\mathrm{CSPD^{4\times4}}$ | 37.9 ms | 5.5 ms | 10.0 ms | 5.9 ms |
| $\mathrm{CSPD^{6\times6}}$ | 137.8 ms | 15.6 ms | 10.9 ms | 6.0 ms |
| $\mathrm{CSPD^{8\times8}}$ | 422.0 ms | 42.2 ms | 13.0 ms | 12.9 ms |
| $\mathrm{CSPD^{12\times12}}$ | 1634.1 ms | 211.6 ms | 33.4 ms | 58.0 ms |

[7]  Wang, R.; Guo, H.; Davis, L. S.; Dai, Q. Covariance discriminative learning: A natural and efficient approach to image set classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2496–2503, 2012.

[8]  Ren, J.; Wu, X. Bidirectional covariance matrices: A compact and efficient data descriptor for image set classification. In: *Intelligence Science and Big Data Engineering. Image and Video Data Engineering. Lecture Notes in Computer Science, Vol. 9242.* He, X. et al. Eds. Springer Cham, 186–195, 2015.

[9]  Cherian, A.; Sra, S. Riemannian dictionary learning and sparse coding for positive definite matrices. *IEEE Transactions on Neural Networks and Learning Systems* Vol. 28, No. 12, 2859–2871, 2017.

[10]  Harandi, M. T.; Hartley, R.; Lovell, B.; Sanderson, C. Sparse coding on symmetric positive definite manifolds using Bregman divergences. *IEEE Transactions on Neural Networks and Learning Systems* Vol. 27, No. 6, 1294–1306, 2016.

[11]  Li, P.; Wang, Q.; Zuo, W.; Zhang, L. Log-Euclidean kernels for sparse representation and dictionary learning. In: Proceedings of the IEEE International Conference on Computer Vision, 1601–1608, 2013.

[12]  Wang, Q.; Li, P.; Zuo, W.; Zhang, L. RAID-G: Robust estimation of approximate infinite dimensional Gaussian with application to material recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4433–4441, 2016.

[13]  Faraki, M.; Harandi, M. T.; Porikli, F. Approximate infinite-dimensional region covariance descriptors for image classification. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 1364–1368, 2015.

[14]  Arandjelovic, O.; Shakhnarovich, G.; Fisher, J.; Cipolla, R.; Darrell, T. Face recognition with image sets using manifold density divergence. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, 581–588, 2005.

[15]  Tuzel, O.; Porikli, F.; Meer, P. Region covariance: A fast descriptor for detection and classification. In: *Computer Vision–ECCV 2006. Lecture Notes in Computer Science, Vol. 3952.* Leonardis, A.; Bischof, H.; Pinz, A. Eds. Springer Berlin Heidelberg, 589–600, 2006.

[16]  Moore, B. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Transactions on Automatic Control* Vol. 26, No. 1, 17–32, 1981.

[17]  Izenman, A. J. Linear discriminant analysis. In: *Modern Multivariate Statistical Techniques. Springer Texts in Statistics.* Springer New York, 237–280, 2013.

[18]  Zhang, D.; Zhou, Z.-H. $(2D)^2$PCA: Two-directional two-dimensional PCA for efficient face representation and recognition. *Neurocomputing* Vol. 69, Nos. 1–3, 224–231, 2005.

[19]  Pennec, X.; Fillard, P.; Ayache, N. A Riemannian framework for tensor computing. *International Journal of Computer Vision* Vol. 66, No. 1, 41–66, 2006.

[20]  Harandi, M.; Salzmann, M. Riemannian coding and dictionary learning: Kernels to the rescue. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3926–3935, 2015.

**Kai-Xuan Chen** is a student in the School of IoT Engineering at Jiangnan University. His research interests are in machine learning and pattern recognition.

**Xiao-Jun Wu** is a professor in the School of IoT Engineering at Jiangnan University. He has his Ph.D. degree in pattern recognition and intelligent systems. He has published more than 150 papers on pattern recognition, computer vision, fuzzy systems, neural networks, and intelligent systems.