Beyond Mahalanobis Metric: Cayley-Klein Metric Learning

Yanhong Bi, Bin Fan, Fuchao Wu
Institute of Automation, Chinese Academy of Sciences
Beijing, 100190, China
{yanhong.bi, bfan, fcwu}@nlpr.ia.ac.cn

Abstract

Cayley-Klein metric is a kind of non-Euclidean metric suitable for projective space. In this paper, we introduce it into the computer vision community as a powerful metric and an alternative to the widely studied Mahalanobis metric. We show that besides its good characteristic in non-Euclidean space, it is a generalization of Mahalanobis metric in some specific cases. Furthermore, as many Mahalanobis metric learning, we give two kinds of Cayley-Klein metric learning methods: MMC Cayley-Klein metric learning and LMNN Cayley-Klein metric learning. Experiments have shown the superiority of Cayley-Klein metric over Mahalanobis ones and the effectiveness of our Cayley-Klein metric learning methods.

1. Introduction

Distance metric plays an important role in many computer vision and pattern recognition tasks, such as classification [4, 30, 17], retrieval [2, 7] and clustering [19]. The most widely used distance metric is the Euclidean metric, which considers the input space as an isotropic one. However, such an isotropic assumption may not hold in many practical applications. For this reason, Euclidean metric can not fairly reflect the underlying relationships between input instances, which further limits its performance in many applications [6, 8, 9, 24, 30].

A simple and popular solution is to replace the Euclidean metric by Mahalanobis metric. While the Euclidean metric treats all the data dimensions equally, Mahalanobis could take the correlation among different data dimensions into consideration. Basically, Mahalanobis metric can be viewed as Euclidean metric on a global linear transformed input space. How to estimate such linear transformations on the input space is at core of Mahalanobis metric learning, which aims to obtain a distance metric better modeling the underlying relationship among input data [31, 25, 30, 5, 16, 21].

Except for learning a positive semidefinite matrix for Mahalanobis metric, few attempts have been made for a proper non-Euclidean metric. In this paper, we present a non-Euclidean metric beyond Mahalanobis framework. The core idea lies in a novel distance metric defined based on the non-Euclidean geometry discovered by A. Cayley and F. Klein in the 19th century. The so called Cayley-Klein metric, induced by an invertible symmetric matrix, is a metric in projective space defined using a cross-ratio. We show in this paper that in special case the Cayley-Klein metric can be considered as a generalized Mahalanobis metric. Furthermore, by integrating it into two typical supervised metric learning paradigms (MMC [31] and LMNN [30]), we obtain two Cayley-Klein metric learning methods. Experi-

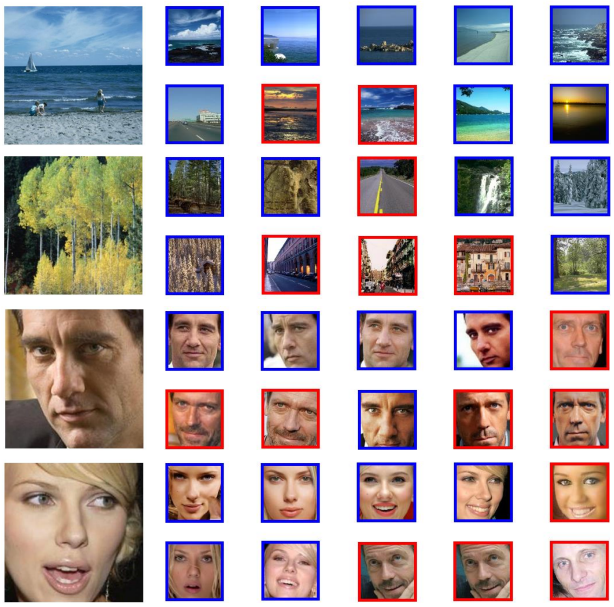

Figure 1. Results of similarity search on OSR (first two rows) and PubFig (last two rows). We present for each query the 5 nearest neighbors returned by our method (first row) and by LMNN (second row). Results in blue correspond to images in the same class as the query while results in red are images from different classes. Our method can return more semantically relevant images.

mental results show that the Cayley-Klein metric is a powerful alternative to currently widely used distance metric, and can be used in many applications. Figure 1 gives an example of image retrieval by using Cayley-Klein metric and Mahalanobis metric with LMNN learning strategy respectively. For each query image, it returns the most 5 similar images according to their used metrics. As can be seen, our method consistently outperforms the state of the art by returning more semantically relevant images.

The remainder of this paper is organized as follows. Firstly, Section 2 reviews previous approaches on metric learning. Then, Section 3 elaborates the Cayley-Klein metric and its properties, followed by Section 4 that describes how to learn a specific Cayley-Klein metric by leveraging on labeled samples. Encouraging results of Cayley-Klein metric as well as comparisons to Mahalanobis metric are reported in Section 5. Finally, Section 6 concludes this paper.

2. Related work

Metric learning aims to learn a specific distance function for a particular task, and is proven to be very useful when dealing with problems that rely on distances. In the Mahalanobis metric learning framework, the central task is to learn a positive semidefinite matrix $\mathbf{M}$ to fit the squared Mahalanobis distance $d^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T \mathbf{M} (\mathbf{x} - \mathbf{y})$.

Perhaps the simplest case of metric learning arises in the context of k-NN classification using Mahalanobis distances. The Mahalanobis metric can be equivalently viewed as a global linear transformation of the input space and then precedes k-NN classification using Euclidean distances. Therefore, many classical dimensionality reduction methods can be viewed as a kind of metric learning. For example, the well known PCA [12] finds a linear transformation to map the input data to a lower dimensional space such that the transformed data space captures the original information as much as possible. Other supervised methods [28, 20] utilize label information to discover such linear transformation in order to maximally separate each class.

Generally, the class label information is presented in a set of constraints incorporated in objective functions. Some methods use pairwise constraints which have input training data as similar and dissimilar pairs. The optimal distance metric is supposed to keep instances in similar constraints close, and simultaneously instances in dissimilar constraints well separated [5, 21, 31]. Besides pairwise constraints, there are methods proposed to learn the optimal distance metric among triplet-wise training data [2, 7, 25, 30], even with quadruplet-wise constraints [16].

Two of the most typical metric learning methods are MMC [31] and LMNN [30]. In [31], Xing *et al*. proposed to cast the metric learning problem for clustering as a convex optimization problem, whose global optimized solution can be efficiently solved. This is the first attempt of using convex optimization for solving this problem in the literature. Specifically, MMC tries to maximize the distances between pairs of instances with different labels and constrain the sum over distances of pairs of identical labeled instances. LMNN [30] is a technique for Mahalanobis metric learning in the k-NN classification setting by semidefinite programming. Its learning target is to make the k-nearest neighbors always in the same class while instances from different classes are separated by a large margin. The model of LMNN is based on two simple intuitions for robust k-NN classification: first, each instance should share the same label as its k nearest neighbors; second, instances with different labels should be widely separated. As a result, LMNN attempts to learn a linear transformation of the input space so as to make the training inputs satisfy these properties. Since there is no parametric assumption about the structure or distribution of the input data, it performs rather well.

Recently, Riemannian metric and manifold learning are becoming popular. Cheng [3] aims at learning a rectangular similarity matrix and tackles the metric learning problem in a Riemannian optimization framework. Huang *et al*. [11] proposed to learn a Euclidean-to-Riemannian metric for point-to-set classification. Besides, there are other popular metric learning methods, such as non-linear metric learning methods [32, 29, 13, 1], information theory based methods [27, 5], and so on [10, 26, 18].

In this paper, we investigate a kind of non-Euclidean metric learning problem. It is defined on the Cayley-Klein metric, a kind of metric in projective space. Compared to existing metrics, Cayley-Klein metric has several advantages. First, the Cayley-Klein space is a special case of Riemannian space with fixed curvature. Second, the Cayley-Klein metric has an explicit definition while a general Riemannian metric does not have. Finally, it is a generalization of Mahalanobis metric by extending the metric definition based on a linear transformation to a fractional transformation. As a result, Cayley-Klein metric is more general compared to Euclidean and Mahalanobis metrics. Our main contributions are: 1) We propose a non-Euclidean distance to instead Mahalanobis based on a family of metrics named Cayley-Klein; 2) We find the relationship between Cayley-Klein metric and Mahalanobis metric, and prove that in some specific cases the Cayley-Klein metric is a generalized version of Mahalanobis metric; 3) We elaborate how to learn Cayley-Klein metric by optimizing over two kinds of popular objectives, one based on the pairwise training examples and the other based on the triplet-wise training data; 4) We show in experiments that Cayley-Klein metric outperforms other widely used metrics on challenging Computer Vision tasks.

3. Cayley-Klein metric

In mathematics, non-Euclidean geometry arose when the parallel postulate of Euclidean geometry was set aside. There are two traditional non-Euclidean geometries: elliptic geometry and hyperbolic geometry. The essential difference among these two non-Euclidean geometries and Euclidean geometry is the property of parallel lines. According to the Euclid's fifth postulate, *i.e*. the parallel postulate, for any given line l and a point $\mathbf{x}$ (not on l) on a 2-dimensional plane, there is exactly one line through $\mathbf{x}$ that does not intersect l. By comparison, there are many lines through $\mathbf{x}$ that intersect l in non-Euclidean geometry. More specifically, any line through $\mathbf{x}$ would intersect l in elliptic geometry. In hyperbolic geometry, although there are infinite lines through $\mathbf{x}$ intersect l, there also have infinite lines that do not intersect l.

A. Cayley noted that distance between points inside a conic could be defined in terms of logarithm and projective cross-ratio function. His method was then exploited by F. Klein to describe the non-Euclidean geometries [14] in 1871. As a result, it is called Cayley-Klein metric, which can provide working models for elliptic and hyperbolic geometries, as well as Euclidean geometry.

In this section, we focus on the Cayley-Klein metric and present some of its essential properties.

3.1. Definition

Given an invertible symmetric matrix $\Psi \in \mathbb{R}^{(n+1)\times(n+1)}$, its bilinear representation of $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ can be denoted by $\psi(\mathbf{x}, \mathbf{y})$:

$$\psi(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T, 1)\Psi\begin{pmatrix}\mathbf{y}\\1\end{pmatrix},\ \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \tag{1}$$

Instead of $\psi(\mathbf{x}, \mathbf{y})$, we take $\psi_{\mathbf{xy}}$ for short hereinafter.

If matrix Ψ is positive definite, then $\psi_{\mathbf{xx}} > 0$, we can define $\rho_E(\mathbf{x}, \mathbf{y}) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^+$ as:

$$\rho_E(\mathbf{x}, \mathbf{y}) = \frac{k}{2i}\log\left(\frac{\psi_{\mathbf{xy}} + \sqrt{\psi_{\mathbf{xy}}^2 - \psi_{\mathbf{xx}}\psi_{\mathbf{yy}}}}{\psi_{\mathbf{xy}} - \sqrt{\psi_{\mathbf{xy}}^2 - \psi_{\mathbf{xx}}\psi_{\mathbf{yy}}}}\right)(k > 0) \tag{2}$$

If matrix Ψ is indefinite, set $\mathbb{B}^n = \{\mathbf{x} \in \mathbb{R}^n | \psi_{\mathbf{xx}} < 0\}$, we can define $\rho_H(\mathbf{x}, \mathbf{y}) : \mathbb{B}^n \times \mathbb{B}^n \to \mathbb{R}^+$ as:

$$\rho_H(\mathbf{x}, \mathbf{y}) = -\frac{k}{2}\log\left(\frac{\psi_{\mathbf{xy}} + \sqrt{\psi_{\mathbf{xy}}^2 - \psi_{\mathbf{xx}}\psi_{\mathbf{yy}}}}{\psi_{\mathbf{xy}} - \sqrt{\psi_{\mathbf{xy}}^2 - \psi_{\mathbf{xx}}\psi_{\mathbf{yy}}}}\right)(k > 0) \tag{3}$$

It can be shown that $\rho_E(\mathbf{x}, \mathbf{y})$ and $\rho_H(\mathbf{x}, \mathbf{y})$ are two metrics on $\mathbb{R}^n$ and $\mathbb{B}^n$ respectively as they satisfy the following metric axioms:

- $\rho(\mathbf{x}, \mathbf{y}) \geq 0$ (Non-negativity)
- $\rho(\mathbf{x}, \mathbf{y}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{y}$ (Identity of indiscernibles)
- $\rho(\mathbf{x}, \mathbf{y}) = \rho(\mathbf{y}, \mathbf{x})$ (Symmetry)
- $\rho(\mathbf{x}, \mathbf{z}) \leq \rho(\mathbf{x}, \mathbf{y}) + \rho(\mathbf{y}, \mathbf{z})$ (Triangle inequality)

$(\mathbb{R}^n, \rho_E)$ is called elliptic geometry space and $(\mathbb{B}^n, \rho_H)$ is called hyperbolic geometry space. ρ_E and ρ_H together constitute the Cayley-Klein metric. For convenience, they can be written in an unified form as following:

$$\rho(\mathbf{x}, \mathbf{y}) = \frac{k}{2}\left|\log\left(\frac{\psi_{\mathbf{xy}} + \sqrt{\psi_{\mathbf{xy}}^2 - \psi_{\mathbf{xx}}\psi_{\mathbf{yy}}}}{\psi_{\mathbf{xy}} - \sqrt{\psi_{\mathbf{xy}}^2 - \psi_{\mathbf{xx}}\psi_{\mathbf{yy}}}}\right)\right|(k > 0) \tag{4}$$

where $1/k$ $(-1/k)$ is related to the curvature of elliptic (hyperbolic) space.

According to the above definition, the Cayley-Klein metric only relies on a symmetric matrix Ψ. In other word, given a symmetric matrix, one can have a specific Cayley-Klein metric. Therefore, Ψ is called the Cayley-Klein metric matrix.

3.2. Invariance properties

According to Klein, the characteristic of any geometry is determined by the type of correspondence under which its relations are invariant, *e.g*. Euclidean geometry is invariant under "similarity transformations". This concept of similarity, which plays a vital role in Euclidean geometry, however, has no analogue in either of the non-Euclidean geometries, *i.e*. elliptic geometry and hyperbolic geometry. In this subsection, we show some invariant properties of Cayley-Klein metric.

Proposition 1: Given two points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n(\mathbb{B}^n)$, let $\mathbf{z}_+$ and $\mathbf{z}_-$ be the points at which the straight line determined by $\mathbf{x}$ and $\mathbf{y}$ intersects the quadric surface $\Omega = \{\mathbf{z} | \psi(\mathbf{z}, \mathbf{z}) = 0\}$, then:

$$\rho(\mathbf{x}, \mathbf{y}) = \frac{k}{2}\left|\log r(\mathbf{xy}, \mathbf{z}_+\mathbf{z}_-)\right| \tag{5}$$

where $r(\mathbf{xy}, \mathbf{z}_+\mathbf{z}_-)$ is the cross-ratio of this quadruple of points $\{\mathbf{x}, \mathbf{y}, \mathbf{z}_+, \mathbf{z}_-\}$:

$$r(\mathbf{xy}, \mathbf{z}_+\mathbf{z}_-) = \frac{(\mathbf{x} - \mathbf{z}_+)(\mathbf{y} - \mathbf{z}_-)}{(\mathbf{x} - \mathbf{z}_-)(\mathbf{y} - \mathbf{z}_+)} \tag{6}$$

Given a Cayley-Klein metric matrix Ψ, let us consider the following matrix group:

$$\mathbb{G}(\Psi) = \{\mathbf{G} \in \mathbb{R}^{(n+1)\times(n+1)} | \mathbf{G}^{-T}\Psi\mathbf{G}^{-1} = \Psi\} \tag{7}$$

For any matrix $\mathbf{G} = (g_{ij}) \in \mathbb{G}(\Psi)$, it defines a linear fractional transformation $\mathbf{x}' = g(\mathbf{x})$ as following:

$$g(\mathbf{x}) = \frac{\begin{pmatrix}\sum_{i=1}^n g_{1i}x_i + g_{1(n+1)}\\ \sum_{i=1}^n g_{2i}x_i + g_{2(n+1)}\\ \vdots \\ \sum_{i=1}^n g_{ni}x_i + g_{n(n+1)}\end{pmatrix}}{\sum_{i=1}^n g_{(n+1)i}x_i + g_{(n+1)(n+1)}} \tag{8}$$

It can be shown that the quadric surface $\Omega = \{\mathbf{x}|\psi(\mathbf{x},\mathbf{x}) = 0\}$ is invariant under this linear fractional transformation, *i.e.* $\psi(\mathbf{x},\mathbf{x}) = 0 \Leftrightarrow \psi(g(\mathbf{x}), g(\mathbf{x})) = 0 : \forall \mathbf{G} \in \mathbb{G}(\Psi)$.

Since the linear fractional transform is capable of preserving cross-ratio, combined with **Proposition 1**, it holds that: $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n(\mathbb{B}^n)$, $\rho(g(\mathbf{x}), g(\mathbf{y})) = \rho(\mathbf{x}, \mathbf{y}) : \forall \mathbf{G} \in \mathbb{G}(\Psi)$. Therefore, Cayley-Klein metric is invariant to the transformation group $\mathbb{G}(\Psi)$.

Proposition 2: For any $\mathbf{G} \in \mathbb{G}(\Psi)$, there exists a $(n+1)$-dimensional antisymmetric matrix $\mathbf{W}$ satisfying:

$$\mathbf{G} = (\Psi + \mathbf{W})^{-1}(\Psi - \mathbf{W}) \tag{9}$$

Therefore, transformation group $\mathbb{G}(\Psi)$ actually has $n(n+1)/2$ essential parameters.

These invariance properties can simplify Cayley-Klein metric computation by using the normal form obtained with a fractional transformation. What is more, it gives theoretical foundation to the potential speedup of learning algorithm, because it is not necessary to iterate over corresponding invariant group as all the results in this group are identical.

3.3. Generalized Mahalanobis metric

Based on the Cayley-Klein metric described before, here we propose a special form of Cayley-Klein metric which we call generalized Mahalanobis metric since it approaches Mahalanobis metric in an extreme case.

Given a set of N data points $\{\mathbf{x}_i\}_{i=1}^N \subset \mathbb{R}^n$, we denote $\mathbf{m}$ as its mean and Σ as its inverse covariance. By definition, Mahalanobis metric is defined as:

$$d_\Sigma(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T \Sigma (\mathbf{x}_i - \mathbf{x}_j)} \tag{10}$$

To obtain a Cayley-Klein metric, we use $\mathbf{m}$ and Σ to define two reversible symmetric matrices $\mathbf{G}^\pm$ as:

$$\mathbf{G}^\pm = \begin{pmatrix} \Sigma & -\Sigma\mathbf{m} \\ -\mathbf{m}^T\Sigma & \mathbf{m}^T\Sigma\mathbf{m} \pm k^2 \end{pmatrix} \ (k > 0) \tag{11}$$

A typical value of k is around 3 in our experiments. $\mathbf{G}^+$ is positive definite while $\mathbf{G}^-$ is indefinite. According to Eq. (1), the bilinear form of $\mathbf{G}^\pm$ is given by:

$$\begin{aligned} \sigma^\pm(\mathbf{x}_i, \mathbf{x}_j) &= (\mathbf{x}_i^T, 1)\mathbf{G}^\pm \begin{pmatrix} \mathbf{x}_j \\ 1 \end{pmatrix} \\ &= (\mathbf{x}_i - \mathbf{m})^T \Sigma (\mathbf{x}_j - \mathbf{m}) \pm k^2 \ (k > 0) \end{aligned} \tag{12}$$

Based on the definition of Cayley-Klein metric, we have:

$$d_E(\mathbf{x}_i, \mathbf{x}_j) = \frac{k}{2i} \log \left(\frac{\sigma^+_{\mathbf{x}_i\mathbf{x}_j} + \sqrt{\sigma^{+2}_{\mathbf{x}_i\mathbf{x}_j} - \sigma^+_{\mathbf{x}_i\mathbf{x}_i} \cdot \sigma^+_{\mathbf{x}_j\mathbf{x}_j}}}{\sigma^+_{\mathbf{x}_i\mathbf{x}_j} - \sqrt{\sigma^{+2}_{\mathbf{x}_i\mathbf{x}_j} - \sigma^+_{\mathbf{x}_i\mathbf{x}_i} \cdot \sigma^+_{\mathbf{x}_j\mathbf{x}_j}}} \right) \tag{13}$$

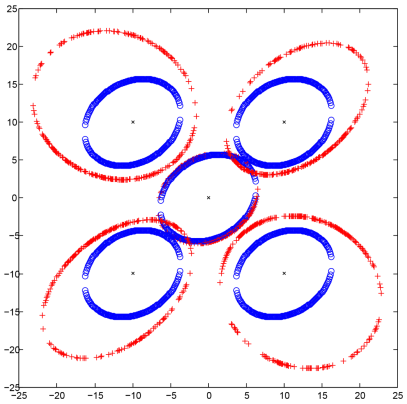

Figure 2. Equidistant distribution of five fixed points for Cayley-Klein metric and Mahalanobis metric. Under Mahalanobis metric (marked as "o" in blue), all points with unit distance to a fixed point (marked as "x" in black) form an ellipse, whose center is the fixed point. This ellipse is identical for any fixed point, wherever its location. On the contrary, under Cayley-Klein metric (marked as "+" in red), all points with unit distance to the origin form a shape similar to ellipse. However, this shape differs when the fixed point moves.

$$d_H(\mathbf{x}_i, \mathbf{x}_j) = -\frac{k}{2} \log \left(\frac{\sigma^-_{\mathbf{x}_i\mathbf{x}_j} + \sqrt{\sigma^{-2}_{\mathbf{x}_i\mathbf{x}_j} - \sigma^-_{\mathbf{x}_i\mathbf{x}_i} \cdot \sigma^-_{\mathbf{x}_j\mathbf{x}_j}}}{\sigma^-_{\mathbf{x}_i\mathbf{x}_j} - \sqrt{\sigma^{-2}_{\mathbf{x}_i\mathbf{x}_j} - \sigma^-_{\mathbf{x}_i\mathbf{x}_i} \cdot \sigma^-_{\mathbf{x}_j\mathbf{x}_j}}} \right) \tag{14}$$

Note that in $d_H(\mathbf{x}_i, \mathbf{x}_j)$, data points $\mathbf{x}$ should satisfy: $\{\mathbf{x} : \sigma^-(\mathbf{x}, \mathbf{x}) < 0\}$.

Proposition 3: The Cayley-Klein metrics $d_E(\mathbf{x}_i, \mathbf{x}_j)$, $d_H(\mathbf{x}_i, \mathbf{x}_j)$ and Mahalanobis metric $d_\Sigma(\mathbf{x}_i, \mathbf{x}_j)$ have the following relationship:

$$\lim_{k \to +\infty} d_E(\mathbf{x}_i, \mathbf{x}_j) = d_\Sigma(\mathbf{x}_i, \mathbf{x}_j) = \lim_{k \to +\infty} d_H(\mathbf{x}_i, \mathbf{x}_j) \tag{15}$$

We call $d_E(\mathbf{x}_i, \mathbf{x}_j)$ and $d_H(\mathbf{x}_i, \mathbf{x}_j)$ as elliptic and hyperbolic Mahalanobis metrics, because they are deduced from the elliptic and hyperbolic metrics respectively. On the other side, since $1/k$ $(-1/k)$ is related to the curvature of elliptic (hyperbolic) space, k approaches infinite is corresponded to the Mahalanobis space. As a result, d_E and d_H are considered together as *generalized Mahalanobis* metric, which we would use for initialization in Cayley-Klein metric learning as described in next section.

Figure 2 illustrates the difference between Cayley-Klein metric and Mahalanobis metric in 2-dimensional space. Under Mahalanobis metric, the equidistant distribution of a fixed point is an ellipse and stays unchanged when the fixed point changes its location. On the contrary, in case of Cayley-Klein metric, the shape and scale of this equidistant distribution would change, depending on the location of the fixed point.

4. Cayley-Klein metric learning

In this section, we show how to learn an appropriate Cayley-Klein metric matrix by leveraging on labeled data. Specifically, two kinds of popular objectives are considered respectively. One is based on MMC [31], while the other is on the basis of LMNN [30]. Their corresponding methods are called MMC Cayley-Klein metric learning and LMNN Cayley-Klein metric learning, which we would describe in this section. For simplicity, we consider the elliptic Cayley-Klein metric described in Section 3 since it is defined on a symmetric positive definite matrix.

Given a symmetric positive definite matrix $\mathbf{G}$, its bilinear form used in Cayley-Klein metric can be represented as:

$$\sigma(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T, 1)\mathbf{G}\begin{pmatrix}\mathbf{x}_j \\ 1\end{pmatrix} \triangleq \sigma_{ij} \tag{16}$$

Accordingly, the elliptic Cayley-Klein metric is:

$$d_{CK}(\mathbf{x}_i, \mathbf{x}_j) = \frac{k}{2i}\log\left(\frac{\sigma_{ij} + \sqrt{\sigma_{ij}^2 - \sigma_{ii}\sigma_{jj}}}{\sigma_{ij} - \sqrt{\sigma_{ij}^2 - \sigma_{ii}\sigma_{jj}}}\right)(k > 0) \tag{17}$$

In the following, we will use $d_{CK}(\mathbf{x}_i, \mathbf{x}_j)$ as a metric to measure the distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ and show how to learn the metric matrix $\mathbf{G}$ by using pairwise and triplet-wise constraints respectively.

4.1. MMC Cayley-Klein metric learning

As one classical method for metric learning, MMC takes pairs of similar and dissimilar training data as input. It aims to maximize distances between dissimilar pairs while constraining distances between similar pairs to be small under a certain metric. Instead of Mahalanobis metric, we use its pairwise constraint based objective function to learn a Cayley-Klein metric.

4.1.1 Objective function

Similar to MMC, in order to minimize distances between similar pairs and simultaneously maximize distances between dissimilar pairs, we have the following optimization problem with the Cayley-Klein metric:

$$\begin{aligned} &\textbf{maximize} && \textstyle\sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{D}} d_{CK}(\mathbf{x}_i, \mathbf{x}_j) \\ &\textbf{subject to} && (a) \quad \textstyle\sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{S}} d_{CK}(\mathbf{x}_i, \mathbf{x}_j) \le 1 \\ & && (b) \quad \mathbf{G} > 0 \end{aligned} \tag{18}$$

where $\mathcal{D}$ is the set of dissimilar training pairs and $\mathcal{S}$ is the set of similar ones.

Note that the first constraint is to make the problem feasible and bounded, and the second constraint enforces that $\mathbf{G}$ is a positive definite matrix. We call this method CK-MMC.

4.1.2 Solver

The optimization problem in Eq. (18) can be solved by the gradient ascent algorithm. At each iteration, we take a gradient ascent step on the objective function

$$\varepsilon(\mathbf{G}) = \textstyle\sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{D}} d_{CK}(\mathbf{x}_i, \mathbf{x}_j) \tag{19}$$

w.r.t. $\mathbf{G}$ and then project $\mathbf{G}$ into the sets $C_1 = \{\mathbf{G} : \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{S}} d_{CK}(\mathbf{x}_i, \mathbf{x}_j) \le 1\}$ and $C_2 = \{\mathbf{G} : \mathbf{G} > 0\}$ iteratively.

For simplicity, let $\mathbf{C}_{ij} = (\mathbf{x}_i^T, 1)^T(\mathbf{x}_j^T, 1)$. Furthermore, we can denote $\sigma(\mathbf{x}_i, \mathbf{x}_j)$ as:

$$\sigma(\mathbf{x}_i, \mathbf{x}_j) = tr(\mathbf{C}_{ij}\mathbf{G}) \tag{20}$$

Denoting the matrix $\mathbf{G}$ at the t-th iteration as $\mathbf{G}^t$, we can derive the gradient of the objective function at the t-th iteration as:

$$\begin{aligned}\mathcal{G}^t = \frac{\varepsilon(\mathbf{G})}{\partial\mathbf{G}}|_{\mathbf{G}^t} = \frac{k}{2i}\sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{D}}\Bigg(&\frac{2\mathbf{C}_{ij}}{\sqrt{\sigma_{ij}^2 - \sigma_{ii}\sigma_{jj}}} \\ &- \frac{\sigma_{ij}\mathbf{C}_{ii}}{\sigma_{ii}\sqrt{\sigma_{ij}^2 - \sigma_{ii}\sigma_{jj}}} - \frac{\sigma_{ij}\mathbf{C}_{jj}}{\sigma_{jj}\sqrt{\sigma_{ij}^2 - \sigma_{ii}\sigma_{jj}}}\Bigg)\end{aligned} \tag{21}$$

The whole gradient ascending process is summarized in Algorithm 1.

Algorithm 1 MMC Cayley-Klein Metric Learning

Input: training data and their labels, step size η.
Output: Cayley-Klein metric matrix $\mathbf{G}$

1: Initialization: $\mathbf{G}_0 = \mathbf{G}^+$ according to Eq. (11)
2: **repeat**
3: **repeat**
4: $\mathbf{G} = \arg\min_{\mathbf{G}'}\{\|\mathbf{G}' - \mathbf{G}\|_F : \mathbf{G}' \in C_1\}$
5: $\mathbf{G} = \arg\min_{\mathbf{G}'}\{\|\mathbf{G}' - \mathbf{G}\|_F : \mathbf{G}' \in C_2\}$
6: **until** $\mathbf{G}$ converges
7: compute gradient $\mathcal{G}^t$ (Eq. (21))
8: $\mathbf{G} = \mathbf{G} + \eta \cdot \mathcal{G}^t$
9: **until** stopping criterion (*e.g.* convergence)
10: **return** $\mathbf{G}$

In the inner iteration (line 3 to 7), projection onto C_1 involves minimizing a quadratic objective with non-linear constraints which is solved by interior point method. The second projection onto C_2 is implemented by eigenvalue decomposition and truncating small eigenvalues by a small positive threshold, which we set as 0.1 in our experiments.

4.2. LMNN Cayley-Klein metric learning

LMNN gives a metric learning technique by using triplet-wise constraints, which is more general than the pairwise ones. In the following we describe how to learn a

Cayley-Klein metric according to the learning paradigm of LMNN.

4.2.1 Objective function

The basic idea of LMNN is to make the k-nearest neighbors of a data point lie in the same class as the data point, and meanwhile make data points from different classes are separated by a large margin. By using Cayley-Klein metric in this framework, we have the following optimization problem:

$$\begin{aligned} \textbf{minimize} \quad & \sum_{i,j\to i} (d_{CK}(\mathbf{x}_i,\mathbf{x}_j) + \mu \textstyle\sum_l (1-y_{il})\xi_{ijl}) \\ \textbf{subject to} \quad & (a) \quad d_{CK}(\mathbf{x}_i,\mathbf{x}_l) - d_{CK}(\mathbf{x}_i,\mathbf{x}_j) \geq 1-\xi_{ijl} \\ & (b) \quad \xi_{ijl} \geq 0 \\ & (c) \quad \mathbf{G} > 0 \end{aligned} \tag{22}$$

We call this method CK-LMNN. Here the notation $j \to i$ is to indicate that $\mathbf{x}_j$ is a target neighbor of $\mathbf{x}_i$. $y_{ij} \in \{0,1\}$ indicates whether $\mathbf{x}_i$ and $\mathbf{x}_j$ have the same class label. $\xi_{ijl} \geq 0$ denotes the amount by which a differently labeled data $\mathbf{x}_l$ invades the "perimeter" around $\mathbf{x}_i$ defined by its target neighbor $\mathbf{x}_j$. The constant μ controls the trade-off between the two terms in the objective function.

4.2.2 Solver

To ensure the symmetry of $\mathbf{G}$, we consider working on its decomposition instead of $\mathbf{G}$, that is $\mathbf{G} = \mathbf{L}^T\mathbf{L}$ with $\mathbf{L} \in \mathbb{R}^{(n+1)\times(n+1)}$. For convenience we denote $\xi_{ijl}(\mathbf{L})$ as:

$$\xi_{ijl}(\mathbf{L}) = [1 + d_{CK}(\mathbf{x}_i,\mathbf{x}_j) - d_{CK}(\mathbf{x}_i,\mathbf{x}_l)]_+ \tag{23}$$

where $[z]_+ = z$ if $z \geq 0$ and $[z]_+ = 0$ if $z < 0$.

By considering the constraint (a) in Eq. (22) into the objective function, it becomes:

$$\varepsilon(\mathbf{L}) = \sum_{i,j\to i} d_{CK}(\mathbf{x}_i,\mathbf{x}_j) + \mu \sum_{i,j\to i,l} (1-y_{il})\xi_{ijl}(\mathbf{L}) \tag{24}$$

As $\varepsilon(\mathbf{L})$ is derivative with respect to $\mathbf{L}$, this optimization can be solved by the gradient descent algorithm.

With notation of $\mathbf{C}_{ij} = (\mathbf{x}_i^T, 1)^T(\mathbf{x}_j^T, 1)$, we have:

$$\sigma(\mathbf{x}_i,\mathbf{x}_j) = tr(\mathbf{C}_{ij}\mathbf{G}) = tr\left(\mathbf{C}_{ij}(\mathbf{L}^T\mathbf{L})\right) \tag{25}$$

Denoting the matrix $\mathbf{L}$ at the t-th iteration as $\mathbf{L}^t$, we can derive the gradient of Eq. (24) w.r.t. $\mathbf{L}$ at the t-th iteration as:

$$\begin{aligned} \mathcal{L}^t &= \frac{\partial \varepsilon(\mathbf{L})}{\partial \mathbf{L}}|_{\mathbf{L}^t} \\ &= \sum_{i,j\to i} \frac{\partial d_{CK}(\mathbf{x}_i,\mathbf{x}_j)}{\partial \mathbf{L}}|_{\mathbf{L}^t} + \mu \sum_{i,j\to i,l} (1-y_{il}) \frac{\partial \xi_{ijl}(\mathbf{L})}{\partial \mathbf{L}}|_{\mathbf{L}^t} \end{aligned} \tag{26}$$

where

$$\begin{aligned} \frac{\partial d_{CK}(\mathbf{x}_i,\mathbf{x}_j)}{\partial \mathbf{L}}|_{\mathbf{L}^t} = \frac{k}{2i} \cdot 2\mathbf{L} \Bigg(& \frac{2\mathbf{C}_{ij}}{\sqrt{\sigma_{ij}^2 - \sigma_{ii}\sigma_{jj}}} \\ & - \frac{\sigma_{ij}\mathbf{C}_{ii}}{\sigma_{ii}\sqrt{\sigma_{ij}^2 - \sigma_{ii}\sigma_{jj}}} - \frac{\sigma_{ij}\mathbf{C}_{jj}}{\sigma_{jj}\sqrt{\sigma_{ij}^2 - \sigma_{ii}\sigma_{jj}}} \Bigg) \end{aligned} \tag{27}$$

and

$$\frac{\partial \xi_{ijl}(\mathbf{L})}{\partial \mathbf{L}}|_{\mathbf{L}^t} = \begin{cases} 0, & \xi_{ijl}(\mathbf{L}) < 0 \\ \frac{\partial d_{CK}(\mathbf{x}_i,\mathbf{x}_j)}{\partial \mathbf{L}}|_{\mathbf{L}^t} - \frac{\partial d_{CK}(\mathbf{x}_i,\mathbf{x}_l)}{\partial \mathbf{L}}|_{\mathbf{L}^t}, & \xi_{ijl}(\mathbf{L}) > 0 \end{cases} \tag{28}$$

After convergence, $\mathbf{G}$ is simply obtained as $\mathbf{G} = \mathbf{L}^T\mathbf{L}$. The learning scheme is described in Algorithm 2.

Algorithm 2 LMNN Cayley-Klein Metric Learning

Input: training data and their labels, step size η.
Output: Cayley-Klein metric matrix $\mathbf{G}$

1: Initialization: $\mathbf{G}_0 = \mathbf{G}^+$ according to Eq. (11)
2: compute $\mathbf{L}$: $\mathbf{G} = \mathbf{L}^T\mathbf{L}$
3: **repeat**
4: compute gradient $\mathcal{L}^t$ (Eq. (26))
5: $\mathbf{L} = \mathbf{L} - \eta \cdot \mathcal{L}^t$
6: **until** stopping criterion (*e.g.* convergence)
7: **return** $\mathbf{G} = \mathbf{L}^T\mathbf{L}$

5. Experiments

In this section, we conduct various experiments to show the effectiveness of the Cayley-Klein metric. Firstly, we demonstrate the advantages of Cayley-Klein metric over traditional Euclidean and Mahalanobis metrics, showing how the classification accuracy can be improved by using Cayley-Klein metric. Then we show how Cayley-Klein metric learning can beat state of the art metric learning methods in image classification tasks.

5.1. Effectiveness of Cayley-Klein metric

In this subsection, we conduct a 3-nearest neighbors (3-NN) classification task to examine the effectiveness of Euclidean metric, Mahalanobis metric, generalized Mahalanobis metric (denoted as G-Mahalanobis), learned Mahalanobis metrics (MMC and LMNN) and the learned Cayley-Klein metrics (CK-MMC and CK-LMNN).

Datasets: In this experiment, we use eight different datasets from the UCI Machine Learning Repository at `http://archive.ics.uci.edu/ml/datasets.html`, containing the Iris, Wine, Sonar, Vowel, Balance,

Datasets	Data points	Attributes	Classes
Iris	150	4	3
Wine	178	13	3
Sonar	208	60	2
Vowel	528	10	11
Balance	625	4	3
Pima	768	8	2
Seg	2310	19	7
Letter	20000	16	26

Table 1. Characteristics of UCI Datasets.

Pima, Segmentation, and Letter datasets. All features are first normalized over the training data to have zero mean and unit variance and the test data features are normalized using the corresponding training mean and variance. The number of data points, feature dimensions, and the number of classes for each dataset are summarized in Table 1.

Set up: Due to the small numbers of data points in the Iris, Wine and Sonar datasets, we perform leave-one-out cross-validation to measure the performance of different metrics. For the Vowel, Balance and Pima, which are larger than the first three, we randomly divide the dataset into a training set of 250 data points and a test set of the remaining data points (278 for the Vowel, 375 for the Balance and 518 for the Pima). Then we repeat this procedure 10 times independently and record the average accuracies for these three datasets. For the largest two datasets, namely the Segmentation and Letter, we perform 10-fold cross-validation. That is to randomly divide the dataset into 10 sets of equal size and use one of them, in turn, as a test set and the remaining nine together as a training set. We repeat this procedure 10 times independently and record the average accuracies.

Results: Table 2 shows the classification accuracies for the seven evaluated metrics. For the first three datasets, since we perform leave-one-out cross-validation, results are presented by mean accuracy. For the rest five datasets, results are reported by mean accuracy and standard deviation.

From the 2nd to the 4th columns of the table, namely the classification accuracies obtained by Euclidean metric, Mahalanobis metric and generalized Mahalanobis metric, we can see that generalized Mahalanobis metric outperforms both Euclidean and Mahalanobis metrics, which validates the effectiveness of Cayley-Klein metric.

Considering all the evaluated metrics, it turns out that CK-LMNN achieves the best performance, closely followed by the other three learned metrics, namely CK-MMC, LMNN, and MMC. For all the datasets, CK-LMNN and CK-MMC improve the performance by 5%-20% compared to the Euclidean metric. Especially for the Wine dataset, they increase the accuracy obtained by the Euclidean metric by a margin over 20%. The good results of CK-MMC and CK-LMNN are not surprising as both of them use labeled information compared to others.

Table 3 compares the Cayley-Klein metric learning results with three kinds of initializations: identity matrix, random matrix, and generalized Mahalanobis calculated as in Section 3.3. It is clear that G-Mahalanobis gives a good initialization of the supervised Cayley-Klein metric learning. For the remaining experiments, we use G-Mahalanobis for initialization.

5.2. Image classification

Since our Cayley-Klein metric learning has a similar objective function to MMC or LMNN, we make a comparison to them on image classification tasks. More specifically, we focus on images represented by relative attributes [23].
Datasets: Two widely used datasets are adopted in our experiments: *Outdoor Scene Recognition Dataset (OSR)* [22] and a subset of *Public Figure Face Database (PubFig)* [15]. OSR contains 2688 images from 8 scene categories, while the subset of PubFig contains 772 images from 8 random identities. As in [23], 512-dimensional gist [22] descriptor is used to generate a 6-dimensional relative attributes for OSR. Similarly, a concatenation of the gist descriptor and a 45-dimensional Lab color histogram is used for generating an 11-dimensional relative attributes for PubFig. These relative attributes are used as input features to the metric learning methods evaluated in this paper. We use the publicly available codes of [23] to compute relative attributes.

Set up: We randomly select 30 images per class to learn a distance metric, and use the remaining images for testing. In the test stage, we use a 3-NN classifier based on the learned distance metric. We repeat this procedure 30 times and report the average classification accuracy across all categories. We use the publicly available codes [31, 30, 13] of MMC, LMNN and GB-LMNN as baselines. In LMNN, GB-LMNN and CK-LMNN, the k-nearest neighbors is set to be 3, which is the default value suggested by LMNN.

Results: The classification results on the OSR and PubFig are listed in Table 4. As can be seen, by using Cayley-Klein metric instead of Mahalanobis metric, both CK-MMC and CK-LMNN outperform their competitors (CK-MMC VS. MMC, and CK-LMNN VS. LMNN/GB-LMNN). CK-MMC increases the accuracies of MMC by 3.1% and 2.2% on the OSR and PubFig datasets, respectively. While comparing to the original LMNN algorithm, the accuracies achieved by CK-LMNN are increased by 2.8% and 2.5% on the OSR and PubFig datasets, respectively. Even for the recently proposed non-linear LMNN (GB-LMNN), our method has a better performance. Such a superior performance of Cayley-Klein metric based methods demonstrates the importance of a proper metric for image classification.

Table 5 shows the running times on OSR and PubFig for different methods, which are average results of 30 runs. Generally speaking, using Cayley-Klein metric requires

Datasets	Euclidean	Mahalanobis	G-Mahalanobis	MMC	CK-MMC	LMNN	CK-LMNN
Iris	90.0	93.3	96.7	95.9	96.8	96.6	**97.0**
Wine	74.2	88.9	94.9	91.1	95.2	95.4	**95.5**
Sonar	77.4	80.3	83.6	81.9	85.0	86.9	**87.1**
Vowel	85.0 ± 3.46	86.1 ± 3.91	87.3 ± 3.82	89.3 ± 1.35	92.1 ± 1.20	95.0 ± 1.72	**96.1 ± 1.76**
Balance	80.3 ± 1.90	82.1 ± 1.87	84.6 ± 1.96	86.1 ± 1.50	86.5 ± 1.51	87.7 ± 1.29	**87.9 ± 1.38**
Pima	70.4 ± 2.31	72.1 ± 1.34	73.9 ± 1.59	72.8 ± 1.97	74.3 ± 1.80	74.8 ± 1.33	**75.2 ± 1.26**
Seg	95.3 ± 2.56	96.9 ± 1.84	98.0 ± 1.70	96.9 ± 0.96	98.3 ± 0.98	97.4 ± 0.91	**99.7 ± 0.92**
Letter	93.6 ± 2.53	96.9 ± 1.87	98.1 ± 1.29	96.5 ± 1.29	98.5 ± 1.00	97.0 ± 0.83	**99.8 ± 0.90**

Table 2. Classification accuracies (mean in % for the first three datasets; mean and standard deviation in % for the rest five datasets) on UCI Datasets. Generalized Mahalanobis metric outperforms both Euclidean and Mahalanobis metrics, and CK-LMNN achieves the best performance.

Datasets	CK-MMC(**I**)	CK-MMC(**R**)	CK-MMC(**G**)	CK-LMNN(**I**)	CK-LMNN(**R**)	CK-LMNN(**G**)
Iris	96.4	95.6	**96.8**	96.8	96.5	**97.0**
Wine	92.3	91.0	**95.2**	94.8	93.9	**95.5**
Sonar	83.1	81.9	**85.0**	85.2	82.8	**87.1**
Vowel	91.5 ± 1.93	90.1 ± 3.00	**92.1 ± 1.20**	95.7 ± 2.03	91.3 ± 2.81	**96.1 ± 1.76**
Balance	85.3 ± 2.02	85.1 ± 2.97	**86.5 ± 1.51**	86.0 ± 2.31	85.7 ± 2.83	**87.9 ± 1.38**
Pima	73.4 ± 2.78	72.8 ± 3.01	**74.3 ± 1.80**	74.1 ± 1.90	73.8 ± 2.51	**75.2 ± 1.26**
Seg	98.1 ± 1.72	97.8 ± 1.89	**98.3 ± 0.98**	98.7 ± 1.41	98.2 ± 1.76	**99.7 ± 0.92**
Letter	98.0 ± 1.45	98.0 ± 1.81	**98.5 ± 1.00**	99.0 ± 1.32	98.8 ± 1.50	**99.8 ± 0.90**

Table 3. Comparison of different initializations. **I** represents initialization by an identity matrix, **R** means by a random matrix and **G** denotes by the generalized Mahalanobis matrix calculated as in Section 3.3.

Method	OSR	PubFig
MMC [31]	64.0 ± 1.6	80.3 ± 1.0
CK-MMC	**67.1 ± 1.6**	**82.5 ± 1.0**
LMNN [30]	67.3 ± 1.3	78.8 ± 1.4
GB-LMNN [13]	69.3 ± 1.3	79.9 ± 1.6
CK-LMNN	**70.1 ± 1.1**	**81.3 ± 1.7**

Table 4. Classification accuracies (mean and standard deviation in %) obtained on OSR and PubFig. CK-MMC and CK-LMNN have a clear improvement over MMC and LMNN respectively.

Method	Training time		Testing time	
	OSR	PubFig	OSR	PubFig
MMC [31]	3.46s	3.30s	0.22s	0.23s
CK-MMC	4.77s	4.87s	0.30s	0.31s
LMNN [30]	1.87s	6.20s	0.27s	0.29s
GB-LMNN [13]	1.12s	1.41s	0.19s	0.22s
CK-LMNN	3.06s	3.34s	0.47s	0.32s

Table 5. Running times on OSR and PubFig.

a litter more time in testing as more operations are involved in computing Cayley-Klein metric according to its definition. While for training, CK-MMC and CK-LMNN is comparable or better than MMC and LMNN, this is because although Cayley-Klein requires more computations but it actually needs less iterations to convergence. Among all the evaluated methods, GB-LMNN is the most efficient.

Figure 1 presents some recognition results of CK-LMNN and LMNN [30]. We show for each query the 5 most similar images using the metric learned by CK-LMNN (first row) and LMNN (second row) respectively. Please see the supplementary material for more results. It is clear that our method could return more semantically relevant images.

6. Conclusion

This paper introduces the Cayley-Klein metric as a powerful alternative to the widely used Mahalanobis metric in the community of metric learning. Cayley-Klein metric is a more general metric in non-Euclidean space. We prove in this paper that by carefully designing the Cayley-Klein metric matrix, it approaches Mahalanobis metric in extreme case. Moreover, two Cayley-Klein metric learning methods are proposed to use labeled training data to learn an appropriate metric for a given task. Experiments on image classification have shown that better performance can be obtained by using Cayley-Klein metric. One of the future work is to apply Cayley-Klein metric to other visual applications.

Acknowledgements

This work is supported by the National Nature Science Foundation of China (No. 61375043, 61272394, and 61203277).

References

[1] M. S. Baghshaha and S. B. Shouraki. Non-linear metric learning using pairwise similarity and dissimilarity constraints and the geometrical structure of data. *Pattern Recognition*, 43(8):2982–2992, 2010.

[2] G. Chechik, V. Sharma, U. Shalit, and S. Bengio. Large scale online learning of image similarity through ranking. *Journal of Machine Learning Research*, 11:1109–1135, 2010.

[3] L. Cheng. Riemannian similarity learning. In *ICML*, pages 540–548, 2013.

[4] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.

[5] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In *ICML*, pages 209–216, 2007.

[6] C. Domeniconi and D. Gunopulos. Adaptive nearest neighbor classification using support vector machines. In *NIPS*, 2002.

[7] A. Frome, Y. Singer, F. Sha, and J. Malik. Learning globally-consistent local distance functions for shape-based image retrieval and classification. In *ICCV*, pages 1–8, 2007.

[8] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov. Neighbourhood components analysis. In *NIPS*, 2005.

[9] T. Hastie and R. Tibshirani. Discriminant adaptive nearest neighbor classification. *IEEE Pattern Analysis and Machine Intelligence*, 18(6):607–616, 1996.

[10] J. Hu, J. Lu, and Y.-P. Tan. Discriminative deep metric learning for face verification in the wild. In *CVPR*, pages 1875–1882, 2014.

[11] Z. Huang, R. Wang, S. Shan, and X. Chen. Learning euclidian-to-riemannian metric for point-to-set classification. In *CVPR*, pages 1677–1684, 2014.

[12] I. T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics, 2 edition, 2002.

[13] D. Kedem, S. Tyree, K. Q. Weinberger, and F. Sha. Non-linear metric learning. In *NIPS*, 2012.

[14] F. Klein. über die sogenannte nicht-euklidische geometrie. *Mathematische Annalen*, 4:573–625, 1871.

[15] N. Kumar, A. Berg, P. Belhumeur, and S. Nayar. Attribute and simile classifiers for face verification. In *ICCV*, 2009.

[16] M. T. Law, N. Thome, and M. Cord. Fantope regularization in metric learning. In *CVPR*, pages 1051–1058, 2014.

[17] D. Lim, B. McFee, and G. Lanckriet. Robust structural metric learning. In *ICML*, volume 28, pages 615–623, 2013.

[18] J. Lu, X. Zhou, Y.-P. Tan, Y. Shang, and J. Zhou. Neighborhood repulsed metric learning for kinship verification. In *CVPR*, pages 2594–2601, 2012.

[19] J. B. MacQueen. On convergence of k-means and partitions with minimum average variance. *The Annals of Mathematical Statistics*, 1965.

[20] G. J. McLachlan. *Discriminant Analysis and Statistical Pattern Recognition*. Wiley Interscience, 2004.

[21] A. Mignon and F. Jurie. Pcca: A new approach for distance learning from sparse pairwise constraints. In *CVPR*, pages 2666–2672, 2012.

[22] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001.

[23] D. Parikh and K. Grauman. Relative attributes. In *ICCV*, pages 503–510, 2011.

[24] J. Peng, D. Heisterkamp, and H. Dai. Adaptive kernel metric nearest neighbor classification. In *International Conference on Pattern Recognition*, volume 3, pages 33–36, 2002.

[25] M. Schultz and T. Joachims. Learning a distance metric from relative comparisons. In *NIPS*, 2003.

[26] C. Shen, J. Kim, L. Wang, and A. van den Hengel. Positive semidefinite metric learning with boosting. In *NIPS*, pages 1007–1036, 2009.

[27] N. Shental, T. Hertz, D. Weinshall, and M. Pavel. Adjustment learning and relevant component analysis. In *ECCV*, pages 776–790, 2002.

[28] M. Sugiyama. Local fisher discriminant analysis for supervised dimensionality reduction. In *ICML*, pages 905–912, 2006.

[29] J. Wang, H. T. Do, A. Woznica, and A. Kalousis. Metric learning with multiple kernels. In *NIPS*, 2011.

[30] K. Weinberger and L. Saul. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, 10:207–244, 2009.

[31] E. Xing, A. Ng, M. Jordan, and S. Russell. Distance metric learning, with application to clustering with side-information. In *NIPS*, pages 505–512, 2002.

[32] D.-Y. Yeung and H. Chang. A kernel approach for semisupervised metric learning. *IEEE Transactions on Neural Networks*, 18(1):141–149, 2007.