K-means Principal Geodesic Analysis on Riemannian Manifolds

Youshan Zhang(✉)

Computer Science and Engineering, Lehigh University, Bethlehem, PA 18015, USA
yoz217@lehigh.edu

Abstract. Principal geodesic analysis (PGA) has been proposed for data dimensionality reduction on manifolds. However, a single PGA model is limited to data with a single modality. In this paper, we are the first to propose a generalized K-means-PGA model on manifolds. This model can analyze multi-mode nonlinear data, which applies to more manifolds (Sphere, Kendall's shape and Grassmannian). To show the applicability of our model, we apply our model to spherical, Kendall's shape and Grassmannian manifolds. Our K-means-PGA model offers correct geometry geodesic recovery than K-means-PCA model. We also show shape variations from different clusters of human corpus callosum and mandible data. To demonstrate the efficiency of our model, we perform face recognition using ORL dataset. Our model has a higher accuracy (99.5%) than K-means-PCA model (95%) in Euclidean space.

Keywords: K-means · Principal Geodesic Analysis · Manifolds

1 Introduction

Principal component analysis (PCA) has been widely used in analyzing high-dimensional data. It also has other applications. For example, object recognition has attracted more attention in the field of pattern recognition and artificial intelligence. Many literature addressed this issue that included steps of object prepossessing, feature extraction and recognition [1–3]. The first appearance-based method applied the PCA model to recognize images [1]. Some techniques added additionally incorporate facial shape, which included morphable model and active appearance model [2,3]. Other models considered all images as a Lambertian object, and recognized images from orthonormal basis [4,5]. A statistical model was proposed to show variations of normal surface direction, which used PCA to transfer surfaces norm to the tangent space [6,7]. However, PCA has difficulties in analyzing nonlinear data. Nowadays, the convolutional neural network has been widely used in object recognition. However, it always needs a sufficient number of training data to get a good classifier [8,9]. Principal geodesic analysis (PGA) has been proposed to analyze the nonlinear data on Riemannian manifolds [10], and it also can be used to analyze facial shape variations [11–13]. However, previous work did not provide a general model for all manifolds,

K. Arai et al. (Eds.): FTC 2019, AISC 1069, pp. 578–589, 2020.
https://doi.org/10.1007/978-3-030-32520-6_42

which is only applicable to spherical manifold. PGA model also cannot deal with multi-cluster data, but there are always different modes of data.

In this paper, we aim to explore whether K-means-PGA model is applicable to general manifolds for shape analysis and whether our model can achieve a higher accuracy than K-means-PCA model for face recognition. Our contributions are in two-fold:

(1). A general K-means-PGA model for more manifolds.
(2). The detailed calculations of K-means-PGA model follow the geometry of data. Also, it has a higher accuracy than the cases in Euclidean space.

We first define a more general statistical model, which is applicable to more manifolds. We then draw the concept of K-means-PCA in Euclidean space and define a general objective function of K-means-PGA model on manifolds. We also introduce the detailed computations of the exponential map and logarithmic map on three manifolds. To show the applicability of our model, we apply our model to the spherical and Kendall's shape manifolds. Our K-means-PGA model offers correct geometry geodesic recovery than K-means-PCA since K-means-PGA allows the model to capture the specific geometry of data. In addition, we show shape variations from different clusters of human corpus callosum and mandible data. To demonstrate the efficiency of our model, we perform the face recognition using ORL data set. We further explore whether the dimensionality reduction can affect the recognition rate of our model. We compare our method with K-means-PCA model, empirical results show that the recognition rate can achieve 99.5% with a reduced dimensionality as low as 67 (vs. original 360).

2 K-means Principal Component Analysis

In this section, we briefly review the K-means principal component analysis model in the Euclidean space.

2.1 K-means on Euclidean Space

The purpose of the K-means algorithm is to partition the given data points into a predefined amount of k clusters. K-means algorithm aims to optimize the objective function of Eq. 1. The algorithm is described in Algorithm 1, it starts with a random set of k center-points μ_k. In the update step, all data x are assigned to their nearest center-point in Eq. 2. If multiple centers have the same distance to the data, we will randomly choose one.

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} ||x_n - \mu_k||^2, r_{nk} = \begin{cases} 1 & x_n \in S_k \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

where x is the observed data, μ_k is the k centers. K is the number of clusters, and S_k is one cluster data.

$$S_k = \left\{ x_n : \left\| x_n - \mu_k \right\|^2 \leq \left\| x_n - \mu_j \right\|^2,\ j \neq k \right\}, \quad \mu_k^{'} = \frac{1}{|S_k|} \sum_{x_j' \in S_k} x_j', \tag{2}$$

We then recalculate the mean of assigned observations to original center-points $\mu_k^{'}$, where $|S_k|$ is the number of cluster S_k.

Algorithm 1. K-means on Euclidean space

Input: $x_1, \cdots, x_N \in X_{i=1}^N$, and the number of clusters k
Output: k clusters data X_k^K
1: Randomly choose k centroids from data set X
2: Compute distances from each data to each centroid, and assign each data to its closest centroid
3: Replace each centroid by the mean of partitioned data in step 2
4: Repeat step 2 and 3 until all k centroids are chosen

2.2 K-means-PCA Algorithm

After calculating k clusters data from Algorithm 1, we then obtain the K-means-PCA algorithm as following:

Algorithm 2. K-means-PCA

Input: k clusteres data X_k^K from Alg. 1
Output: Eigenvectors ve_k and eigenvalues λ_k of each cluster
1: **For** $k = 1$ to K
2: $\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_{ik}$
3: $S = \frac{1}{N_k} \sum_{i=1}^{N_k} (x_{ik} - \mu_k)(x_{ik} - \mu_k)^T$
4: $ve_k, \lambda_k =$ eigenvectors/eigenvalues of S // using singular value decomposition
5: **end**

3 K-means Principal Geodesic Analysis

3.1 Riemannian Geometry

In this section, we first review three basic concepts (geodesic, exponential and logarithmic map) of Riemannian geometry (more details are provided by others [14–16]).

Geodesic. Let (M, g) be a Riemannian manifold, where g is a Riemannian metric on the manifold M. A curve $\gamma(t) : [0, 1] \rightarrow M$ and let $\gamma'(t) = d\gamma/dt$ be its velocity. The operation $D \cdot /dt$ is called a *covariant derivative* (also called a connection on M), which is denoted as $\nabla_{\gamma'(t)}$ or $\nabla_{\gamma'}$. A vector field $V(t)$ along γ is parallel if $\frac{DV(t)}{dt} = \nabla_{\gamma'} V = 0$. We call γ a geodesic if $\gamma'(t)$ is parallel along γ, that is: $\gamma'' = \frac{D\gamma'}{dt} = \nabla_{\gamma'}\gamma' = 0$, which means the acceleration vector (directional derivative) γ'' is normal to $T_{\gamma(t)}M$ (the tangent space of M at $\gamma(t)$). A geodesic is also a curve $\gamma \in M$ that locally minimizes $E(\gamma) = \int_0^1 ||\gamma'(t)||^2 dt$. Here $|| \cdot ||$ is

called *Riemannian norm*, for any points $p \in M$, and $v \in T_pM$, $||v||$ is defined by: $||v|| = \sqrt{g_p(v,v)}$. $g_p(u,v)$ is called *Riemannian inner product* of two tangent vectors $u, v \in T_pM$, which can also be denoted by $\langle u,v\rangle_p$ or simply $\langle u,v\rangle$. The norm of velocity in a geodesic γ is constant, that is: $||\gamma'(t)|| = c$ [15]. Note that geodesics are straight lines in Euclidean space ($\mathbb{R}^n$).

Exponential Map. For any point $p \in M$ and its tangent vector v, let $\mathcal{D}(p)$ be the open subset of T_pM defined by: $\mathcal{D}(p) = \{v \in T_pM | \gamma(1)$ is defined, where γ is the unique geodesic with initial conditions $\gamma(0) = p$ and $\gamma'(0) = v$. The *exponential map* is the map $\mathrm{Exp}_p : \mathcal{D}(p) \to M$ defined by: $\mathrm{Exp}_p(vt_{=1}) = \gamma(1)$, which means the exponential map returns the points at $\gamma(1)$ when $t = 1$. If $\omega \in \mathcal{D}(p)$, then the line segment $\{t\omega | 0 \le t \le 1\}$ is constrained to be in $\mathcal{D}(p)$. We then define: $\mathrm{Exp}_p(vt) = \gamma(t)$, where $0 \le t \le 1$.

Logarithmic Map. Given two points p and $p' \in M$, the *logarithmic map* takes the point pair (p, p') and maps them into the tangent space T_pM, and it is an inverse of the exponential map: $\mathrm{Log}(p,p') \to T_pM$. $\mathrm{Log}(p,p')$ can also be denoted as: $\mathrm{Log}_p p'$. Because Log is an inverse of the exponential map, we can also write: $p' = \mathrm{Exp}(p, \mathrm{Log}(p,p'))$. The *Riemannian distance* is defined as $d(p,p') = \|\mathrm{Log}_p(p')\|$. In Euclidean space ($\mathbb{R}^n$), the logarithmic map is the subtraction operation: $\mathrm{Log}_p(p') = p' - p$ [17].

3.2 K-means on Manifold

We define a general objective function J for K-means algorithm on manifolds.

$$J = \sum_{n=1}^{N}\sum_{k=1}^{K} ||\,\mathrm{Log}_{x_n}(\mu_k)||^2 = \sum_{n=1}^{N}\sum_{k=1}^{K} d(x_n,\mu_k)^2 \tag{3}$$

Alternatively, we aim to minimize the following energy function:

$$E = \arg\min_J \sum_{n=1}^{N}\sum_{k=1}^{K} d(x_n,\mu_k)^2, \tag{4}$$

where μ_k is the intrinsic mean, which is calculated in Algorithm 4, $|S_k|$ is the number of cluster S_k. By using Algorithm 3, we obtain the k clusters data on manifolds.

3.3 Principal Geodesic Analysis (PGA)

Similar to the PCA model, PGA was proposed by Fletcher (2004), it aims to reduce the dimensionality of nonlinear data [10]. Differ from the Algorithm 2, we need to calculate the intrinsic mean of data. We then perform the PGA in Algorithm 4.

Algorithm 3. K-means on manifold

Input: $x_1, \cdots, x_N \in X_{i=1}^N$, number of clusters k
Output: k clusters data X_k^K

1: Randomly choose k centroids from data set X.
2: Compute the **Riemannian distances** (using Log map) from each data to each centroid, and assign each data to its closest centroid
3: Replace each centroid by the **intrinsic mean** of partitioned data points in step 2
4: Repeat step 2 and 3 until all k centroids are chosen

Algorithm 4. Principal Geodesic Analysis

Input: $x_1, \cdots, x_N \in M$
Output: Eigenvectors ve and eigenvalues λ of input data

1: $\mu_0 = x_1$
2: **Do**
$\Delta\mu = \frac{\tau}{N} \sum_{i=1}^N \mathrm{Log}_{\mu_j x_i}$
3: $\mu_{j+1} = \mathrm{Exp}_{\mu_j}(\Delta\mu)$
4: **While** $||\Delta\mu|| > \epsilon$ // calculate intrinsic mean of $\{x_i\}$
5: $x_i' = \mathrm{Log}_{\mu}(x_i)$
6: $S = \frac{1}{N} \sum_{i=1}^N x_i' x_i'^T$
7: $ve_k, \lambda_k =$ eigenvectors/eigenvalues of S

3.4 K-means-PGA Algorithm

After calculating the data of each cluster from Algorithm 3, and we can combine it with Algorithm 4 to obtain the K-means-PGA algorithm as following:

Algorithm 5. K-means-PGA

Input: k clusters data X_k^K from Alg. 4
Output: Eigenvectors ve_k and eigenvalues λ_k of each cluster

1: **For** $k = 1$ to K
2: $\mu_k =$ intrinsic mean of $\{x_{ik}\}$
3: $x_{ik}' = \mathrm{Log}_{\mu}(x_{ik})$
4: $S = \frac{1}{N} \sum_{i=1}^N x_{ik}' x_{ik}'^T$
5: $ve_k, \lambda_k =$ eigenvectors/eigenvalues of S
6: **end**

3.5 Computation of Exp and Log Map Manifolds

Sphere Manifold. One of the well-known spherical manifolds is the 3D sphere (2D surface embedding in 3D space). Let r be the radius of the sphere, u the azimuth angle and v the zenith angle. Then any points on 3D sphere can be expressed by: $X = (r\sin(u)\sin(v), r\cos(u)\sin(v), r\cos(v))$. The generalized $n-1$ dimensional hyper-sphere embedded in $\mathbb{R}^n$ Euclidean space $(x_1, x_2, \cdots, x_n)$ has the constraint of: $\sum_{i=1}^n x_i^2 = r^2$, here r is the radius of such a hyper-sphere,

we set $r = 1$. Let p and p' be such points on on a sphere embedded in $\mathbb{R}^n$, and let v be a tangent vector at p.

The Log map between two points p, p' on the sphere can be computed as follows:

$$\begin{aligned} v = \mathrm{Log}(p, p') = \frac{\theta \cdot L}{||L||}, \quad \theta = \arccos(\langle p, p' \rangle), \\ L = (p' - p \cdot \langle p, p' \rangle) \end{aligned} \tag{5}$$

where $p \cdot \langle p, p' \rangle$ denotes the projection of the vector p' onto p. $||L||$ is the *Riemannian norm* as defined in Sect. 3.1.

Given base point p, and its estimated tangent vector v from Eq. 5 and t, we can compute the Exp map as:

$$\mathrm{Exp}(p, vt) = \cos\theta \cdot p + \frac{\sin\theta}{\theta} \cdot vt, \ \theta = ||vt||. \tag{6}$$

This explanation is based on that of Wilson and Hancock (2010); see them for details of Log map and Exp map on the Sphere manifold.

Kendall's Shape Manifold. A more complex manifold was studied by David G. Kendall [18]. Kendall's shape can provide a geometric setting for analyzing arbitrary sets of landmarks. The landmark points in Kendall's space are a collection of points in Euclidean space. Before calculating the Log and Exp maps, we use some transformations (scale and rotation) to get the pre-shape of data.

Given two shapes $p, p' \in V$ with $d \times n$ matrix (d is the dimension of the shape, n is the number of points), to construct Kendall's shape space, first, we remove the translation and scale of the shape. To get the pre-shape of an object, we eliminate the centroid (subtract the row means from each row) and scale it into unit norm (divide by the Frobenius norm). Then, we remove the rotation of the shape using Orthogonal Procrustes Analysis (OPA) [19]. OPA solves the problem of finding the rotation R^* that minimizes distance between p and p'.

$$R^* = \arg\min_{R \in SO(d)} ||Rp - p'||,$$

where SO means a special orthogonal group. OPA performs singular value decomposition of $p \cdot p'^T$; let $[U, S, V] = \mathrm{SVD}(p \cdot p'^T)$, then the $R* = UV^T$.

The Log map between two shapes p, p' of Kendall's shape manifold is also given by finding the rotation between p and p' first. To find the rotation of p', we calculate the singular value decomposition of $p \cdot p^T$; then we find the rotation of p' according to $p'_{(Rot)} = \mathrm{rotation} \cdot p'$. Then the Log map is given by:

$$\begin{aligned} v = \mathrm{Log}(p, p') = \frac{\theta \cdot L}{||L||}, \quad \theta = \arccos(\langle p, p'_{(Rot)} \rangle), \\ L = (X_{T(Rot)} - p \cdot \langle p, p'_{(Rot)} \rangle), \end{aligned} \tag{7}$$

where $\langle p, p'_{(Rot)} \rangle$ denotes the projection of the vector p' onto p.

The Exp on Kendall's shape is the same as sphere manifold (Eq. 6). Refer to Kendall (1984) to see the details of Log map and Exp map Kendall's manifold.

Grassmannian Manifold: Before we introduce the Exp and Log map of Grassmannian manifold, we should get the subspace of the data, since Grassmannian manifold $\mathbb{G}_{N,d}$ is defined as $d-$dimension subspace. Therefore, it is necessary to use Principal Geodesic Analysis to get the subspace (submanifold) of Grassmannian manifold. Algorithm 4 describes how to calculate the intrinsic mean and PGA reduction of Grassmannian manifold. Please refer to [10] for more details. Suppose that we get the subspace from PGA of the data (that is p and p'), again to get the sampling from p to p', we need to calculate the Log map and Exp map. Differ from [20], and we use the following calculations:

Then the Log map is given by:

$$\begin{aligned} &X = (I - p \cdot p^T) \cdot p' \cdot (p^T \cdot p')^{-1}, \quad [U, s, V] = \text{SVD}(X), \\ &v = \text{Log}(p, p') = U \cdot \theta \cdot V^T, \ \theta = \arctan s, \end{aligned} \tag{8}$$

where I is the identify matrix, and θ is the principal angles between p and p'.

The exponential map is computed by base point p, and the estimated initial velocity v, that is,

$$[U, \theta t, V] = \text{SVD}(vt), \quad \text{Exp}(p, vt) \quad = p \cdot V \cdot \cos(\theta t) + U \cdot \sin(\theta t) \cdot V^T, \tag{9}$$

Refer to [21] and [22] to see the details of geometry on Grassmannian manifold.[1]

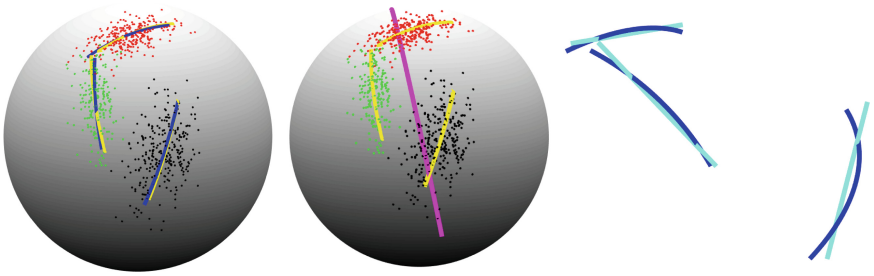
4 Experiments

In this section, we demonstrate the effectiveness of our K-means-PGA model by using both synthetic data and real data.

4.1 Sphere

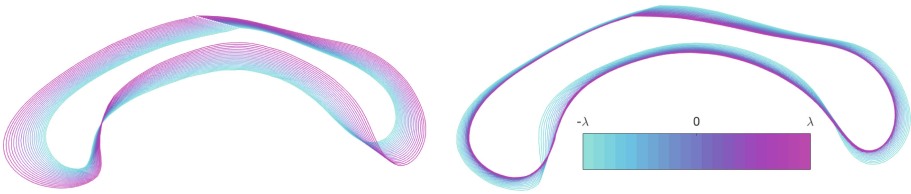
To validate our model in sphere manifold, we randomly simulate 764 data points on a unit sphere S^2 with three clusters. We can visualize the true geodesic from the Fig. 1. The blue line is the estimated geodesic of K-means-PGA model, while the yellow line is true geodesic. We can find that estimated geodesics are almost overlaying with true geodesic, which demonstrates the correctness of our model. We also compare the estimated results of our model with K-means-PCA in the Euclidean space. As shown in Fig. 1c, estimated geodesics of K-means-PGA are curves on the sphere, but the estimated geodesic of K-means-PCA is straight lines, which cannot recover true geodesics on the sphere. Figure 1b illustrates the limitation of PGA model since it cannot estimate geodesics of multimode data since only one geodesic can be estimated. This result also illustrates the advantage of K-means-PGA model than the traditional PGA model.

[1] Source code is available at: https://github.com/heaventian93/Kmeans-Principal-Geodesic-Analysis.



(a) K-means-PGA (b) PGA (c) K-means-PCA

Fig. 1. (a) The estimated geodesic using K-means-PGA. The blue line is the estimated geodesic, and the yellow line is the ground truth. (b) The estimated geodesic (magenta color) of PGA model, which is the limitation of PGA model since it only shows one curve. (c) The comparison of estimated geodesics of K-means-PGA and K-means-PCA. The blue line is the estimated geodesic in (a). Cyan lines are estimated geodesics using K-means-PCA, which are straight lines.



(a) Cluster one (b) Cluster two

Fig. 2. The shape variations of two clusters corpus callosum using K-means-PGA along the first geodesic with $[-\sqrt{\lambda}, \sqrt{\lambda}]$.

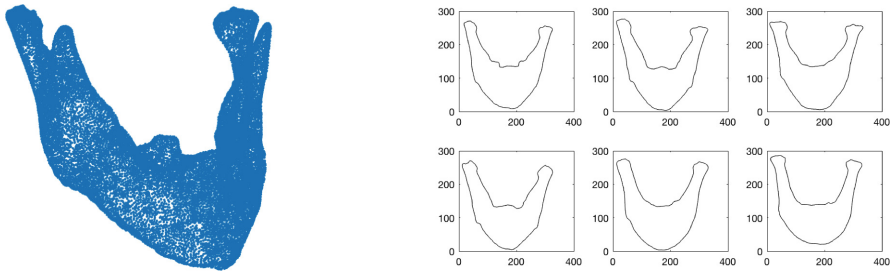
4.2 Kendall's Shape

Corpus Callosum Data. To show K-means-PGA model for demonstrating classification for the real shape changes, we apply it to corpus callosum data. The corpus callosum data are derived from MRI scans of human brains. The data is from the ABIDE database,[2] which is used in Hiess et al., (2015). The valid data contains 1100 patients; we extract the boundary of each corpus callosum shape, and each shape contained 2×142 points. As shown in Fig. 2, it demonstrates corpus callosum shape variations of two different clusters, which are generated from the points along the estimated principal geodesics: $\text{Exp}(\mu, \alpha_i v e_i)$, where $\alpha_i = [-\sqrt{\lambda}, \sqrt{\lambda}]$, for $i = 1$, which means the largest the eigenvalue. It shows that the anterior, mid-caudate and posterior of corpus callosum from the unhealthy group (Fig. 2a) is significantly larger than the healthy group (Fig. 2b).

Mandible Shape. The mandible data is extracted from a collection of CT scans of human mandibles. It contains 77 subjects and the age is from 0 to 19 years.

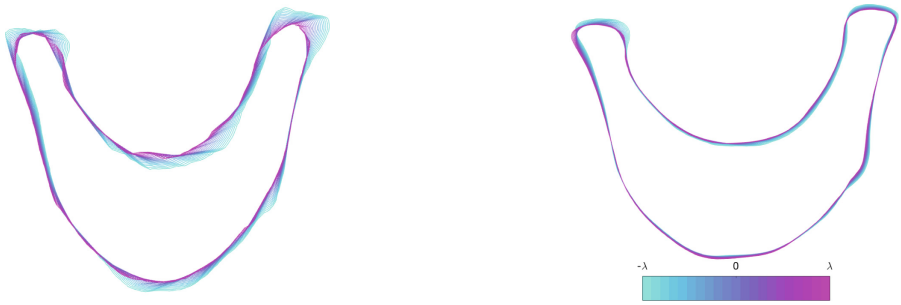
[2] https://sites.google.com/site/hpardoe/cc_abide.

We sample 2×400 points on the boundaries. Figure 3b shows examples of 3D raw data of mandible and 2D shape data. Shape variations of mandible data from male (Fig. 4a) and female group (Fig. 4b) are shown in Fig. 4. In general, male mandibles have larger shape variations in the temporal crest, middle part and the base than that in female mandibles, which is consistent with previous studies [23].



(a) 3D mandible shapes (b) 2D mandible shapes

Fig. 3. Example of 3D and 2D mandible shapes.



(a) Cluster one (b) Cluster two

Fig. 4. The shape variations of two clusters mandible data using K-means-PGA along the first geodesic with $[-\sqrt{\lambda}, \sqrt{\lambda}]$.

4.3 Grassmannian Manifold

We use the ORL face database for examining face recognition task. This dataset is composed of 400 images with the size of 112×92. There are 40 persons and ten images for each person [24]. For the experiments, we use the ten-folds-cross-validation, we chose 90% of images of each category as training data and the rest as test data. As shown in Fig. 5, we observe eight eigenface examples which are chosen from the first component of the K-means-PGA model. From Fig. 6,

the face recognition percentage of K-means-PGA model (99.5%) is significantly higher than K-means-PCA model (95%), which illustrates the superiority of our model. The K-means-PGA model reduces the dimensionality of input images and only 67 dimensionalities out of 360 can extract enough number of features for face recognition.

Fig. 5. The first eigenvector of each cluster, which shows the most important feature of each category.

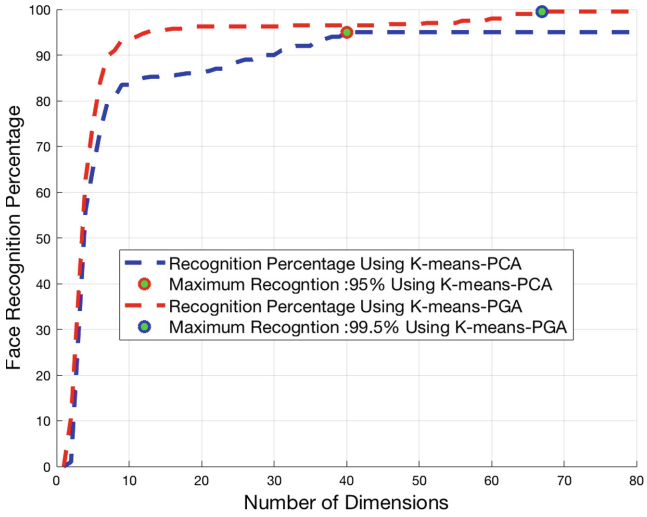


Fig. 6. The comparison of face recognition percentage of K-means-PGA and K-means-PCA with different number of dimensionalities.

5 Discussion

The experiments for sphere and Kendall's shape demonstrate that K-means-PGA model is suitable for shape analysis, especially for nonlinear data. Also, our model achieves a higher facial recognition accuracy than K-means-PCA model since the K-means-PGA model can capture the intrinsic geometry of data. To choose a proper manifold for different datasets, we suggest that sphere manifold is suitable for data, which lies on hypersphere; Kendall's manifold is the

best choice if data are 2D landmark shapes, and Grassmannian manifold will have a good performance if data should be prepossessed in the subspace.

However, there are two major limitations of our model. First, the computation time is $\mathcal{O}(i \times k \times n \times d)$, i is the number of iteration, k is the number of clusters, n is the number of samples, and d is the dimensionality of each sample feature. If $i \times k \times n \times d$ is bigger, our model will consume longer computation time. Secondly, the number of cluster k cannot be automatically determined. In our experiment, we determine the number of clusters k using the minimal error, which defined in Eq. 4.

6 Conclusion

In this paper, we first define a general model for K-means-PGA model on different Riemannian manifolds, and we give closed-form expressions of Log map and Exp map on three manifolds. We conduct four experiments on three manifolds, and experimental results show that K-means-PGA model is able to analyze shape variations in multimodal data, and the accuracy of K-means-PGA model is relatively higher than the model in Euclidean space. For future work, we will apply our model in more manifolds, and analyze detailed changes of data (e.g., gesture recognition in images).

References

1. Turk, M., Pentland, A.: Eigenfaces for recognition. J. Cogn. Neurosci. **3**(1), 71–86 (1991)
2. Timothy, F.C., Gareth, J.E., Christopher, J.T.: Active appearance models. IEEE Trans. Pattern Anal. Mach. Intell. **23**(6), 681–685 (2001)
3. Blanz, V., Vetter, T.: Face recognition based on fitting a 3D morphable model. IEEE Trans. Pattern Anal. Mach. Intell. **25**(9), 1063–1074 (2003)
4. Basri, R., Jacobs, D.: Lambertian reflectance and linear subspaces. In: Proceedings Eighth IEEE International Conference on Computer Vision 2001. ICCV 2001, vol. 2, pp. 383–390. IEEE (2001)
5. Peter, N.B., David, J.K.: What is the set of images of an object under all possible illumination conditions? Int. J. Comput. Vision **28**(3), 245–260 (1998)
6. William, A.P.S., Edwin, R.H.: Recovering facial shape using a statistical model of surface normal direction. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 1914–1930 (2006)
7. William, A.P.S., Edwin, R.H.: Face recognition using 2.5D shape information. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 1407–1414. IEEE (2006)
8. Yin, X., Liu, X.: Multi-task convolutional neural network for pose-invariant face recognition. IEEE Trans. Image Process. **27**(2), 964–975 (2018)
9. Ding, C., Tao, D.: Trunk-branch ensemble convolutional neural networks for video-based face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **40**(4), 1002–1014 (2018)

10. Fletcher, P.T., Lu, C., Pizer, S.M., Joshi, S.: Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Trans. Med. Imaging **23**(8), 995–1005 (2004)
11. Dickens, M.P., Smith, W.A.P., Wu, J., Hancock, E.R.: Face recognition using principal geodesic analysis and manifold learning. In: Iberian Conference on Pattern Recognition and Image Analysis, pp. 426–434. Springer (2007)
12. Pennec, X.: Probabilities and statistics on Riemannian manifolds: a geometric approach. PhD thesis, INRIA (2004)
13. Jing, W., William, A.P.S., Edwin, R.H.: Weighted principal geodesic analysis for facial gender classification. In: Iberoamerican Congress on Pattern Recognition, pp. 331–339. Springer (2007)
14. Perdigão do Carmo, M.: Riemannian Geometry. Birkhauser (1992)
15. Gallier, J.: Notes on differential geometry and lie groups (2012)
16. Pennec, X.: Statistical computing on manifolds: from Riemannian geometry to computational anatomy. In: Emerging Trends in Visual Computing, pp. 347–386. Springer (2009)
17. Zhang, M., Fletcher, P.T.: Probabilistic principal geodesic analysis. In: Advances in Neural Information Processing Systems, pp. 1178–1186 (2013)
18. David, G.K.: Shape manifolds, procrustean metrics, and complex projective spaces. Bull. Lond. Math. Soc. **16**(2), 81–121 (1984)
19. John, C.G.: Generalized procrustes analysis. Psychometrika **40**(1), 33–51 (1975)
20. Kyle, A.G., Anuj, S., Xiuwen, L., Paul Van, D.: Efficient algorithms for inferences on Grassmann manifolds. In: 2003 IEEE Workshop on Statistical Signal Processing, pp. 315–318. IEEE (2003)
21. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM J. Matrix Anal. Appl. **20**(2), 303–353 (1998)
22. Absil, P.-A., Mahony, R., Sepulchre, R.: Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. Acta Applicandae Mathematica **80**(2), 199–220 (2004)
23. Moo, K.C., Anqi, Q., Seongho, S., Houri, K.V.: Unified heat Kernel regression for diffusion, Kernel smoothing and wavelets on manifolds and its application to mandible growth modeling in CT images. Med. Image Anal. **22**(1), 63–76 (2015)
24. Guodong, G., Stan, Z.L., Kapluk, C.: Face recognition by support vector machines. In: Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000, pp. 196–201. IEEE (2000)