

# Dimensionality Reduction and Clustering on Statistical Manifolds

Sang-Mook Lee and A. Lynn Abbott  
Virginia Polytechnic Institute and State University  
Blacksburg, VA 24061, USA  
LSMOOK@vt.edu, abbott@vt.edu

Philip A. Araman  
U.S. Forest Service, Southern Research Station  
Blacksburg, VA 24060, USA  
paraman@vt.edu

## Abstract

*Dimensionality reduction and clustering on statistical manifolds is presented. Statistical manifold [16] is a 2D Riemannian manifold which is statistically defined by maps that transform a parameter domain onto a set of probability density functions. Principal component analysis (PCA) based dimensionality reduction is performed on the manifold, and therefore, estimation of a mean and a variance of the set of probability distributions are needed. First, the probability distributions are transformed by an isometric transform that maps the distributions onto a surface of hyper-sphere. The sphere constructs a Riemannian manifold with a simple geodesic distance measure. Then, a Fréchet mean is estimated on the Riemannian manifold to perform the PCA on a tangent plane to the mean. Experimental results show that clustering on the Riemannian space produce more accurate and stable classification than the one on Euclidean space.*

## 1. Introduction

Texture segmentation is of fundamental importance in image analysis and pattern recognition tasks and have been studied extensively [1,2,3,4]. Example approaches include transform methods [5,6], stochastic techniques [7,8], and combined techniques [9]. Also, curve evolution techniques are gaining in popularity [10,11,12,13]. Most of the reported methods deal with image models that have two or more regions and associated probability density functions. In [11,14], statistics of image regions are modeled with parametric methods, while Kim et al. [13] use Parzen's density estimates as region descriptors and then utilize an information theoretic approach to image segmentation. A mixture of parametric and nonparametric methods has been proposed in [4], where different techniques are applied to different feature spaces.

Meanwhile, Sochen *et al.* [15] introduced a geometric framework by which images and image feature spaces are considered as 2-dimensional manifolds. Later, a statistical

manifold framework has been studied for texture image segmentation [16], which is substantially different from the work of [15] that constructs non-statistical manifolds. However, there was a drawback in the statistical manifold framework, which introduces boundary offsets in some situations when creating a metric tensor map, consequently decreasing efficiency in boundary detection applications. The drawback has been overcome [17] by using a diffusion scheme on statistical manifolds and a substantial improvement has made over the initial work. The result is a more robust framework for localizing texture boundaries. As a related work, a vector probability diffusion scheme has been introduced by [18], where the probabilities are treated as a vector field and then a minimization problem is solved on the field.

In this paper, we introduce a principal component analysis (PCA) based dimensionality reduction and texture clustering scheme applied to the statistical manifold framework. Statistical manifold is defined as an embedding map that assigns each image coordinate with a set of probability density functions (PDFs) of features. A multinomial distribution representation is used for the statistical manifold to accommodate multimodal distributions in discrete spaces. Being PDFs whose discrete sum are unity, multinomial distributions are confined on a hyper-plane of  $n$ -simplex, which is nice because it is possible to employ linear methods for dimensionality reduction. Therefore, we first use an ordinary PCA method to reduce dimensionality of the multinomial distribution. It is not required for the reduced ones to have unity of summation, and thus their linearity is not guaranteed. On the other hand, an isometric transformation maps the multinomial distributions into points on a sphere, and a distance between two points on the sphere are measured by the arc length of a great circle connecting the points [19]. Applying a linear method to this manifold is not straightforward since a mean point and its tangent plane must be estimated on the spherical manifold, which leads to the use of special projections and to the estimation of Fréchet mean [20].

In the next section, we review the definition of statistical manifolds and the diffusion scheme on statistical

manifolds. Then, the dimensionality reduction on statistical manifolds and a clustering method are discussed in section 3. Section 4 presents some segmentation results, and then section 5 concludes the paper.

## 2. Statistical Manifolds

A Riemannian manifold  $\mathcal{M}^p$  is an abstract surface of arbitrary dimension  $p$  with a proper choice of metric. Then, an image  $I(\mathbf{x})$  parameterized in  $\mathbf{R}^2$ , that is,  $\mathbf{x}=(x,y)\in\mathbf{R}^2$ , is viewed as a 2-dimensional Riemannian manifold,  $\mathcal{M}^2$ , embedded in  $\mathbf{R}^n$  with an embedding map  $(\mathbf{x}, I(\mathbf{x}))$ , where  $n=3$  for intensity images and 5 for color images. Similarly,  $m$ -dimensional feature spaces of an image can be considered as  $\mathcal{M}^2$  embedded in  $\mathbf{R}^{m+2}$  [15].

### 2.1. Statistical embedding

Statistically defined manifold has been introduced in [16], where each feature at a local coordinate  $\mathbf{x}\in\mathbf{R}^2$  is represented by a set of PDFs rather than by deterministic values. Parametric estimation methods can be used for the feature statistics, but in most cases they are not suitable to model multimodal distributions. A Gaussian mixture model could be used for multimodal cases, but it bears high computational complexity. Thus, here only nonparametric methods, such as simple normalized histogram or Gaussian kernel based estimation are considered.

Accordingly, for an  $M$ -dimensional feature space, the embedding map becomes  $(\mathbf{x}, f(\theta^1;\mathbf{x}), \dots, f(\theta^M;\mathbf{x}))$  called a statistical embedding. Means and variances of each feature can be used directly as features, constructing ordinary non-statistical image or feature manifolds. This directly leads to the work of [15], namely non-statistical embedding and non-statistical manifold. Figure 1 illustrates the difference between non-statistical and statistical manifolds. Statistical manifolds associate each parameter location  $x_p$  with a set of PDFs of features while non-statistical manifolds map the parametric space onto a set of scalar values of features.

An example of the statistical embedding is depicted in Figure 2. The PDFs in the first column are estimated from a point inside the body of the cheetah and the set on the second column from outside the body. It is clear that the PDFs for gray value, for instance, for both points are different: bi-modal and mono-modal.

### 2.2. Manifold of multinomial distributions

In the statistical embedding described above, PDF for a feature  $\theta$  can be modeled with a multinomial distribution specified by a parameterization  $z$ :

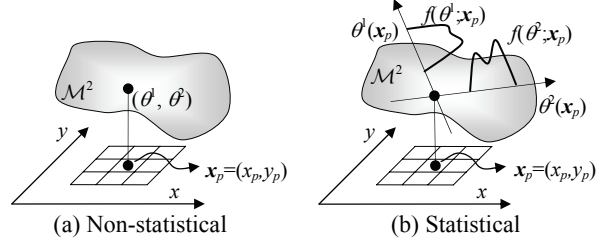


Figure 1. Unlike non-statistical manifolds which map the parametric space onto a set of scalar values of features, statistical manifolds associate each parameter location  $x_p$  with a set of PDFs of features.

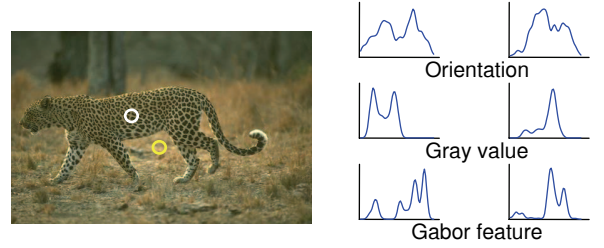


Figure 2. Statistical embedding. The PDFs are estimated from inside (first column) and outside (second) of the body.

$$p(\theta, z) = \sum_{i=1}^{n+1} \delta(\theta - i) z_i, \quad \theta \in \{1, 2, \dots, n+1\}, \quad \sum_{i=1}^{n+1} z_i = 1, \quad z_i > 0. \quad (1)$$

Then, a statistical manifold  $\mathcal{S} = \{p(\theta, z)\}$  can be identified as a  $n$ -simplex in  $\mathbf{R}^{n+1}$  whose coordinate system is  $(z_1, \dots, z_{n+1})$ . That is, the multinomial distributions are laid on the surface of the simplex (Figure 3a).

Meanwhile, Fisher's information matrix defined as

$$g_{ij}(z) = E_z[\partial_i \ell_z \partial_j \ell_z] = \int_{\Theta} \partial_i \log f(\theta, z) \partial_j \log f(\theta, z) d\theta, \quad (2)$$

provides a geometry under a statistical manifold, where  $\partial_i = \partial / \partial z_i$  and  $\Theta$  is a parameter space [21]. Alternatively, the matrix can be represented as

$$g_{ij}(z) = 4 \int_{\Theta} \partial_i \sqrt{f(\theta, z)} \partial_j \sqrt{f(\theta, z)} d\theta. \quad (3)$$

Then an isometric transformation of  $p(\theta, z)$ ,

$$\xi_i = 2\sqrt{z_i}, \quad i=1, \dots, n+1, \quad (4)$$

maps the  $n$ -simplex onto a surface of  $n+1$ -dimensional sphere,  $\mathbb{S}^n$ , of radius 2. This leads to the fact that the Fisher information corresponds to the inner product of tangent vectors to the sphere. Then, the information distance of two distributions  $p$  and  $p'$  is defined as

$$d(p, p') = 2 \cos^{-1} \left( \sum_{i=1}^{n+1} \sqrt{z_i z'_i} \right) \quad (5)$$

which is the arc length of a great circle connecting the distributions (Figure 3b). Here, the length is a geodesic distance. The geodesic distance of (5) is used for a diffusion process on statistical manifolds.

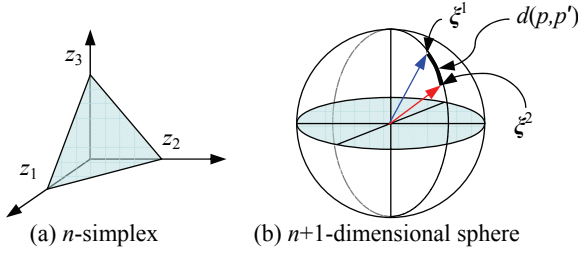


Figure 3. Multinomial distributions as points on an  $n$ -simplex (a) and a sphere (b).

### 2.3. Diffusion on statistical manifold

A classical anisotropic diffusion process proposed by Perona and Malik [22] can be used on the statistical manifolds. Explicitly, the diffusion equation on statistical manifold  $\mathcal{M}^2$  can be defined as

$$f_t = \text{div}(c(d(f, f')) \nabla f) = c(d(f, f')) \Delta f + \nabla c \cdot \nabla f, \quad (6)$$

where  $\text{div}$  is the divergence operator, and  $\nabla$  and  $\Delta$  respectively represent the gradient and Laplacian operators. We denote the diffused manifold as  $\tilde{\mathcal{M}}^2$ . Diffusion on non-statistical manifolds use an edge estimate  $\|\nabla \theta\|$  as an argument of a monotonically decreasing function  $c(\cdot)$  to achieve a conduction coefficient. However, on statistical manifold, the geodesic distance is used for the conduction coefficient since edges on a statistical manifold can be identified by the geodesic distance. Then, following the discretization scheme in [22] and with the choice of

$$c(d(p, p')) = \exp(-d(p, p')^2 / K^2), \quad (7)$$

the diffusion process on a statistical manifold becomes straightforward and produces promising results for further processing.

Figure 4 shows a significant difference between statistical and non-statistical manifolds using a synthetic image (2a) generated from two known PDFs (2b) of the same mean and variance of intensity. To identify the texture boundary, a metric tensor map, defined as a determinant of metric tensor, is calculated. The metric tensor for statistical manifold is based on PDF dissimilarity measures such as Kullback-Leibler ( $K-L$ ) divergence and will be defined later in section 2.4. Tensor maps based on non-statistical manifolds, which take only parametric information of a PDF into account, failed to locate the desired texture boundary (2c, 2d), unless prior knowledge of the PDFs are provided. In contrast, the one on statistical manifolds results in a successful localization of the texture boundary (2e). One drawback of using PDF dissimilarity measures is that they induce offsets from true boundaries (the yellow mark), or sometimes produce thick

lines when the shape of PDFs gradually change over a boundary. The offset and the thickness of detected boundary depends on window size used to estimate PDFs. Also, the use of  $K-L$  divergence normally results in two thin lines for texture boundaries, which is equally undesirable. This will be shown in a later section. Figure 2f shows the result of a diffusion process on a statistical manifold. Compare the locations of the mark and the detected boundary. The diffusion corrects offsets falsely induced in the previous boundary detection.

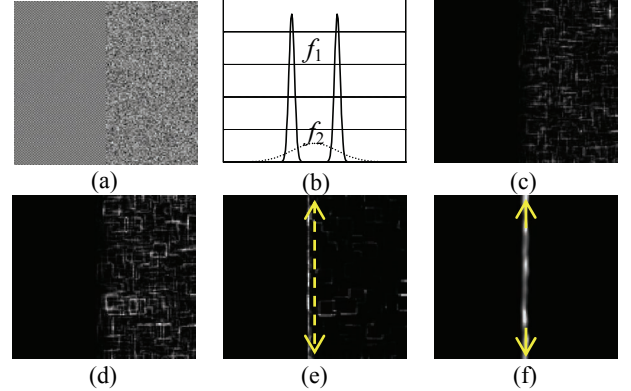


Figure 4. Statistical manifolds can be used to separate regions that have the same mean and variance but look different. The test image (a) was generated from two different PDFs,  $f_1$  and  $f_2$  in (b). A technique based on non-statistical manifolds failed to accurately locate the boundary of the two regions (c,d), but statistical manifold framework successfully identifies the texture boundary (e). In addition, diffusion on the statistical manifold produces a better result (f).

### 2.4. Metric tensor

A metric tensor matrix contains information related to the geometric structure of a manifold and is used to measure distances on manifolds. The determinant of the metric tensor matrix is a good indicator used for edge detection in various image processing applications.

The calculation of metric tensor matrix requires partial derivatives with respect to the parametric domain. Lee *et al.* [16] used  $K-L$  divergence as an approximation of the partial derivatives of PDF  $f(\mathbf{x})$  at location  $\mathbf{x}=(x, y)$ . That is,

$$\begin{aligned} \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} &\cong KL_{\mathbf{x}}(f) \\ &= KL(f(\mathbf{x}), f(\mathbf{x}+\delta \mathbf{x})) \\ &= \frac{1}{2} [\mathbf{kl}(f(\mathbf{x}), f(\mathbf{x}+\delta \mathbf{x})) + \mathbf{kl}(f(\mathbf{x}+\delta \mathbf{x}), f(\mathbf{x}))] \end{aligned} \quad (8)$$

where

$$\mathbf{kl}(f(\mathbf{x}), f(\mathbf{x}+\delta \mathbf{x})) = \sum_i f_i(\mathbf{x}) \log(f_i(\mathbf{x}) / f_i(\mathbf{x}+\delta \mathbf{x})) \quad (9)$$

Then, the metric tensor matrix for a statistical manifold is defined as

$$\tau(\mathbf{x}) = \begin{pmatrix} 1+KL_x KL_x & KL_x KL_y \\ KL_x KL_y & 1+KL_y KL_y \end{pmatrix}, \quad (10)$$

and its determinant measures statistical dissimilarity of nearby features on manifolds. The determinant of  $\tau(\mathbf{x})$  is much larger than unity when the manifold has a high statistical gradient, while the value is close to unity at locations where the manifold is statistically stationary.

### 3. Clustering on statistical manifolds

Clustering a set of multinomial distributions may be carried out straightforward by applying a simple *k-means* algorithm with an appropriate distance measure. Also, kernel based methods could be used to implicitly handle the distance measure. However, due to the curse of high dimension and its instability induced in clustering results, benefits are often acquired when a dimension of input space is reduced. Figure 5 shows an example of instability in clustering when used full rank feature space. The result on the left is produced by applying an ordinary *k-means* clustering algorithm to the PDFs in statistical manifold framework. Not to mention a computational complexity, clustering result is not trustworthy. In contrast, reduced features space with three principal components of PCA scheme produces more accurate clustering results (on the right in figure 5) compared to the full feature space.

In this section, we investigate a PCA based dimension reduction technique for a set of multinomial distributions. The distributions are first transformed by (4) onto a surface of  $n+1$  dimensional sphere. The surface is a Riemannian sub-manifold in  $\mathbf{R}^{n+1}$  and is denoted as  $\mathbb{S}^n$  here.

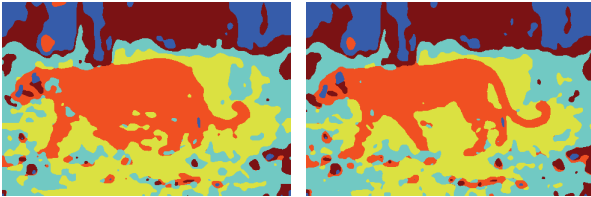


Figure 5. An example shows instability of clustering with a large input dimension.

#### 3.1. Fréchet means

Given a set of multinomial distributions,  $\mathbf{L} = \{p_1, p_2, \dots, p_N\}$ ,  $p_i \in \mathcal{S} = \{p(\theta, z), i=1, \dots, N\}$ , a mean distribution of the set can be obtained by independently computing the arithmetic mean of each coordinate. Then, the arithmetic mean minimizes the sum of squared distance between the

mean and the sample data set. However, our experiments show better clustering results when a Fréchet mean is used. Fréchet mean minimizes the sum of squared distance along geodesics on Riemannian manifolds and uniquely exists when the distributions are limited to a sufficiently small part of the manifold [23,24]. Multinomial distributions mapped onto  $\mathbb{S}^n$  satisfy this condition since all coordinates are positive.

A gradient descent algorithm established by [23] and expressed differently by others [24,25,26] is illustrated in Figure 6. First, the points  $\xi^i \in \mathbb{S}^n$  for  $i=1, \dots, N$  (Figure 6a) are projected onto a tangent plane  $T_y \mathbb{S}^n$  at  $y_t$  by an inverse projection (Figure 6b),

$$v^i = \exp_{y_t}^{-1}(\xi^i), \quad (11)$$

where  $\exp_y(v)$  projects a point  $v$  on a tangent plane at  $y$  onto a sphere. An expectation is calculated on the tangent plane and projected back onto the sphere by a projection  $\exp_y$  (Figure 6c),

$$y_{t+1} = \exp_{y_t}(\mathbf{E}[v^i]). \quad (12)$$

Explicit expressions for these projections are given as [27],

$$\begin{aligned} \exp_y^{-1}(\xi) &= [1 - (\mathbf{y} \cdot \xi)^2]^{-1/2} (\xi - (\mathbf{y} \cdot \xi)\mathbf{y}) \arccos(\mathbf{y} \cdot \xi) \\ \exp_y(v) &= \cos(\|v\|)\mathbf{y} + \sin(\|v\|)\mathbf{v} / \|v\| \quad (v \in T_y \mathbb{S}^n, v \neq 0) \end{aligned} \quad (13)$$

With a proper selection of starting point, the algorithm converges within a few iterations. Figure 7 shows the difference of the Fréchet and the arithmetic means for the image shown in Figure 2a.

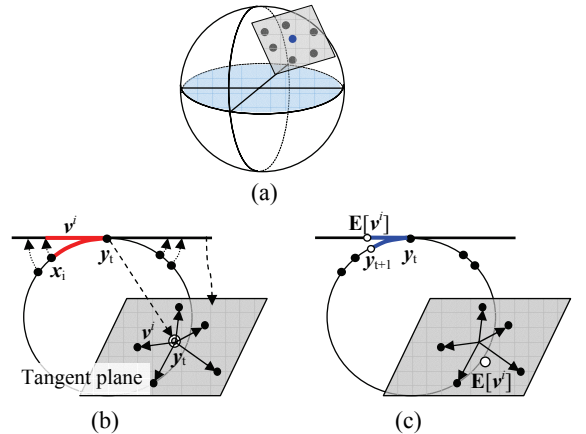


Figure 6. Iterative method to estimate Fréchet mean on a sphere. Points (a) on the sphere are projected on a tangent plane (b) at an initial point of mean estimate. An expectation is calculated and projected back onto the sphere (c). Iterate the procedure until the mean converges.

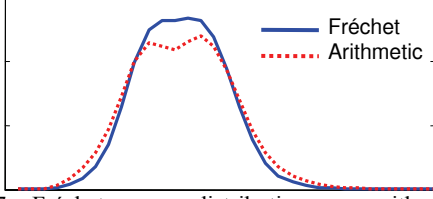


Figure 7. Fréchet mean distribution vs. arithmetic mean distribution. They are different in general.

### 3.2. Dimensionality reduction

Let  $\hat{y} \in \mathbb{S}^n$  be a Fréchet mean estimated by the iteration method described in the previous section, and let  $V = \{v^i\}$ ,  $i = 1, \dots, N$ , be the projections at the mean. Then, the set  $V$  spans the entire  $T_{\hat{y}}\mathbb{S}^n$  when  $N \gg n$ , and an ordinary singular value decomposition can be applied to extract eigenspace of the tangent plane. This simply leads to a PCA based dimensionality reduction of the tangent plane, and the clustering can be applied to this reduced space. The shape of the reduced space is arbitrary and nonlinear, and accordingly, a nonlinear technique such as LLE (locally linear embedding) [28] could be a possible choice for the clustering. However, it is unrealistic in terms of speed to use algorithms of complexity of  $O(N^2)$  when  $N$  is large. So, in this paper, a simple  $k$ -means method is used for clustering at the cost of misclassification. Figure 8 shows the case. The object is labeled as green in Figure 8a, but some of background is misclassified as objects. Among the misclassification, the ones indicated with ellipses should belong to background. These particular areas correspond to the portion indicated in Figure 8b, which can be correctly classified with LLE method.

## 4. Results

Several textured or mixed images are tested for clustering. PDFs are estimated with 32 bins and with different features for different images. Then, they are diffused with 20 iterations. Each iteration requires less than 2 seconds for images of size  $256 \times 256$  on a 2.8GHz Pentium 4 machine with Matlab™ implementation.

The top row of the figure 9 shows the tensor maps on manifolds  $\mathcal{M}^2$  (left) and  $\tilde{\mathcal{M}}^2$  (right) for the test image in figure 2. The diffusion process removes offsets and merges two thin lines across boundaries. Clustering results on both manifolds are depicted right below the tensor maps. As expected, more smooth and accurate clustering is achieved on  $\tilde{\mathcal{M}}^2$ .

Next experiment compares stability of clustering in between Euclidean  $\mathbf{R}^{n+1}$  (first column) and  $\mathbb{S}^n$  (second column) spaces. From top to bottom in Figure 10, 3, 5, and 7 principal components are used for clustering, respectively. It is evident that the  $\mathbb{S}^n$  space produces more accurate results than Euclidean space. Also, clustering stability in

Euclidean space decreases as the number of principal components increase. Finally, in Figure 11, the clustering method is tested with various natural images presented in [29]. For most cases, color information alone is used to construct feature PDFs.

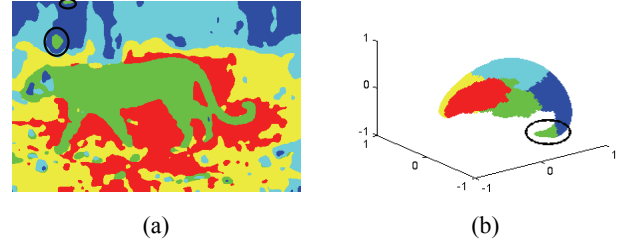


Figure 8. A case of misclassification.

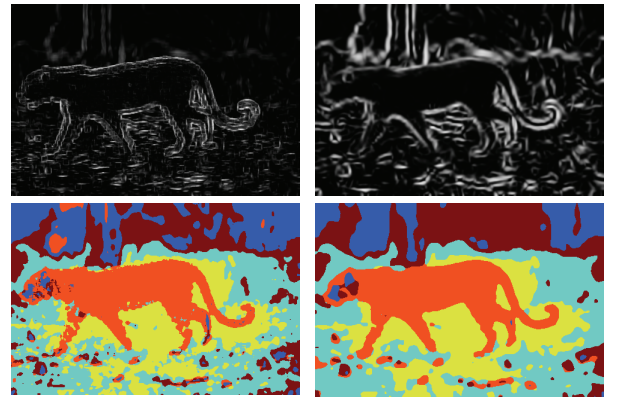


Figure 9. Intensity values are used as features. The diffusion process generates smooth and strong boundary on statistical manifolds, consequently producing accurate clustering results.

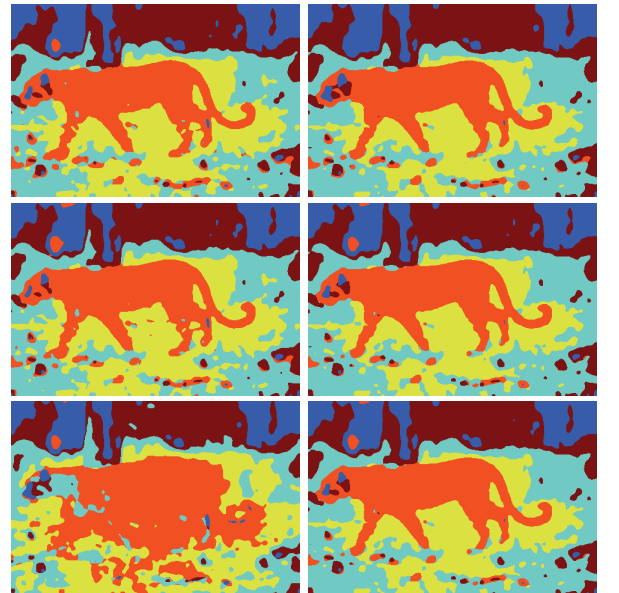


Figure 10. Clustering results and stability comparison in between Euclidean and  $\mathbb{S}^n$  spaces.



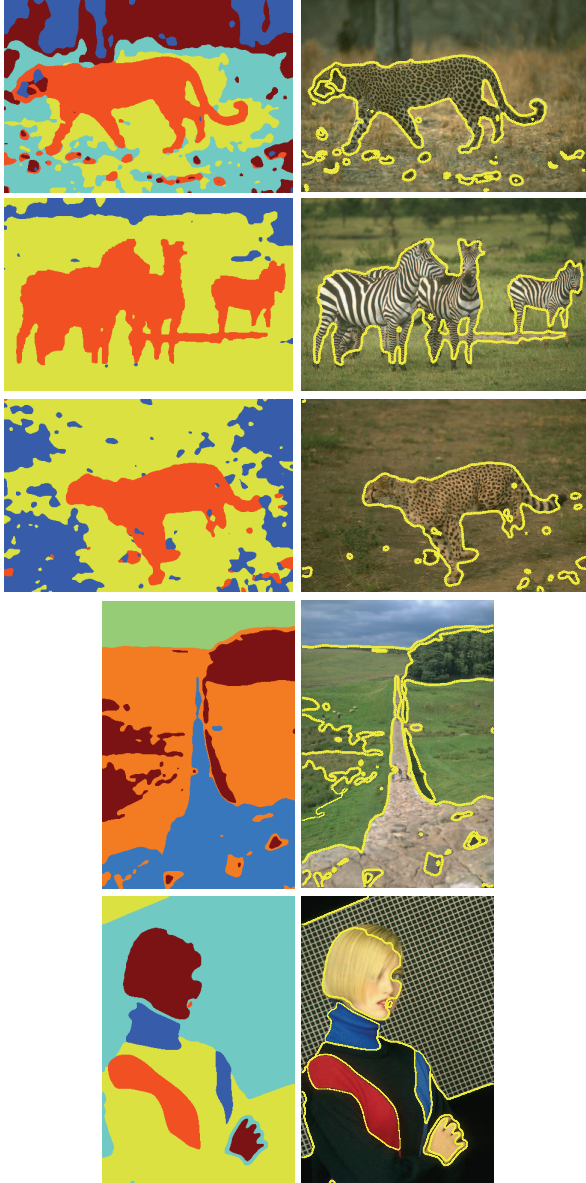


Figure 11. Natural image segmentation with the clustering on statistical manifolds. Except the first sample which uses intensity value as feature, color information in RGB alone is used as feature.

## 5. Conclusion

A clustering scheme on statistically defined manifolds is presented. The algorithm handles textured or mixed images with one framework. Our experiment on natural image segmentation shows promising results in classifying textured or mixed images, even if we used only one feature, that is, color information. The framework could be more powerful if multiple feature spaces are considered.

## References

- [1] R.M. Haralick, "Statistical and structural approaches to texture," *IEEE Proceedings*, vol. 67, no. 5, pp. 786-804, 1979.
- [2] A. Jain and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters," *Pattern Recognition*, 24, pp.1167-1186, 1991.
- [3] G. Haley and B.S. Manjunath, "Rotation invariant texture classification using a complete space-frequency model," *IEEE Trans. on Image Processing*, vol. 8, no. 2, pp. 255-269, 1999.
- [4] M. Rousson, T. Brox, and R. Deriche, "Active unsupervised texture segmentation on a diffusion based feature space," *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 2, pp. II-699-704, June 2003.
- [5] A. Laine and J. Fan, "Texture classification by wavelet packet signatures," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 1186-1191, 1993.
- [6] M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE Trans. on Image Processing*, vol. 4, pp. 1549-1560, November 1995.
- [7] C. Bouman and M. Shapiro, "A multiscale random field model for Bayesian image segmentation," *IEEE Trans. on Image Processing*, vol. 3, pp. 162-177, 1994.
- [8] B. Manjunath and R. Chellappa, "Unsupervised texture segmentation using Markov random fields models," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 478-482, May 1991.
- [9] L. Lepisto, I. Kunttu, J. Autio, and A. Visa, "Classification of non-homogenous texture images by combining classifiers," *Proc. of IEEE Intl. Conf. on Image Processing*, vol. 1, pp. I-981-984, 2003.
- [10] G. Sapiro, "Vector (self) snakes: a geometric framework for color, texture, and multiscale image segmentation," *Proc. of IEEE Intl. Conf. on Image Processing*, vol. 1, pp. 817-820, September 1996.
- [11] N. Paragios and R. Deriche, "Geodesic active regions for supervised texture segmentation," *Intl. Journal of Computer Vision*, vol. 46, no. 3, pp. 223-247, 2002.
- [12] T.F. Chan and L.A. Vese, "Active contours without edges," *IEEE Trans. on Image Processing*, vol. 10, no. 2, pp. 266-277, 2001.
- [13] J. Kim, J.W. Fisher III, A. Yezzi, Jr., M. Cetin, and A.S. Willsky, "Nonparametric methods for image segmentation using information theory and curve evolution," *Proc. of IEEE Intl. Conf. on Image Processing*, vol. 3, pp. 797 -800, 2002.
- [14] M Rousson and R. Deriche, "A variational framework for active and adaptive segmentation of vector valued images," *Proc. of Workshop on Motion and Video Computing*, pp. 56-61, 2002.
- [15] N. Sochen, R. Kimmel and R. Malladi, "A general framework for low level vision," *IEEE Trans. on Image Processing*, vol. 7, no. 3, pp. 310-318, 1998.
- [16] S. Lee, A.L. Abbott, N.A. Clark and P.A. Araman, "Active contours on statistical manifolds and texture segmentation," *Proc. of IEEE Intl. Conf. on Image Processing*, vol. 3, pp. 828-831, 2005.

- [17] S. Lee, A.L. Abbott, N.A. Clark and P.A. Araman, "Diffusion on statistical manifolds," *Proc. of IEEE Intl. Conf. on Image Processing*, vol. 1, pp. 233-236, 2006.
- [18] A. Pardo and G. Sapiro, "Vector probability diffusion," *IEEE Signal Processing Letters*, vol.8, no.4, pp.106-109, 2001.
- [19] R.E. Kass and P.W. Vos, *Geometrical Foundations of Asymptotic Inference*, Wiley-Interscience Publications, 1997.
- [20] H. Le, "Locating Frechet means with application to shape spaces", *Adv. Appl. Prob.*, vol.33, pp. 324-338, 2001.
- [21] S. Amari, *Differential-Geometrical Methods in Statistics*, Lecture Notes in Statistics, Springer-Verlag, 1985.
- [22] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no.7, pp. 629-639, 1990.
- [23] H. Karcker, "Riemann center of mass and mollifier smoothing," *Communications on Pure and Applied Mathematics*, vol. 30, pp. 509-541, 1977.
- [24] W. S. Kendall, "Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence", *Proceedings on London Mathematics Society*, vol. 61, pp. 371-406, 1990.
- [25] X. Pennec, "Probabilities and statistics on Riemannian manifolds: basic tools for geometric measurements",
- [26] W. Mio, D. Badlyans, and X. Liu, "A computational approach to Fisher information geometry with applications to image analysis", *Lecture Notes in Computer Science 3757*, pp. 18-33, 2005.
- [27] R. Bhattacharya and V. Patrangenaru, "Nonparametric estimation of location and dispersion on Riemannian manifolds", *Journal of Statistical Planning and Inference*, vol. 108, pp. 23-35, 2002.
- [28] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding", *Science*, vol.290 no.5500, pp. 2323-2326, Dec.22, 2000.
- [29] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics", *Proc. 8th Int'l Conf. Computer Vision*, vol.2, pp.416-423, 2001.