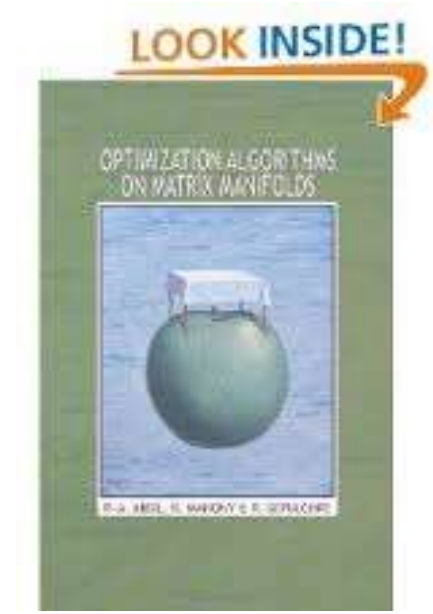
Kernel Methods on Manifolds

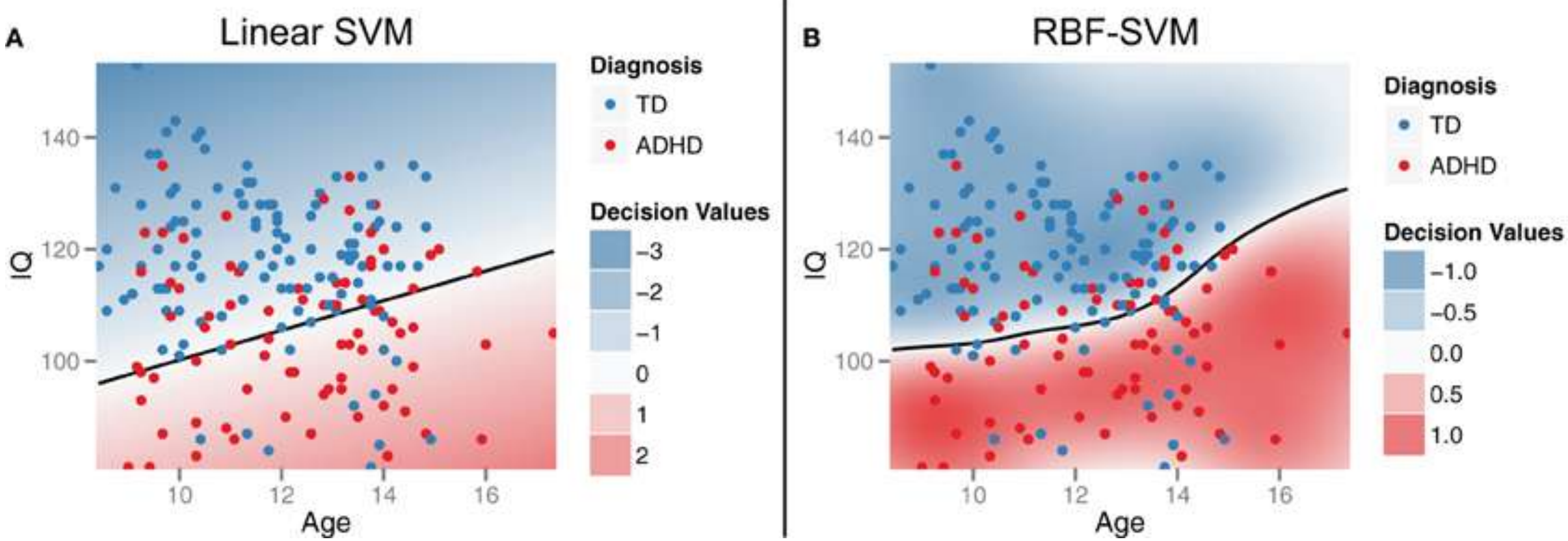
Richard Hartley, Sadeep Jayasumana,
Mehrtash Harandi, Mathieu Salzmann
Hongdong Li, Khurrum Aftab, Fatih Porikli
Conrad Sanderson

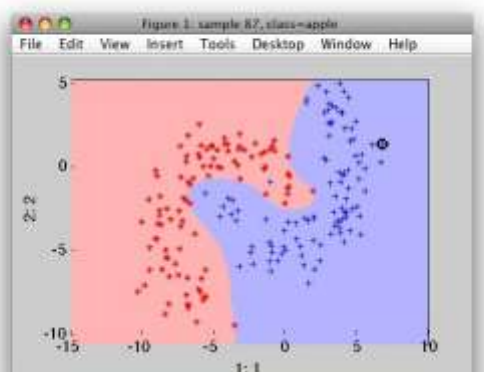
CVPR-2017

Optimization methods on Manifolds.

- Rotation averaging (SO3)
- Weiszfeld Algorithm on Riemannian manifolds
- General IRLS algorithms on manifolds







Linear and Kernel SVM

Kernels and kernel algorithms

- A kernel is like a "similarity measure" defined on points in some set.

$$K(x, y) \text{ for } x, y \in S$$

- If $K(x, y)$ is "large" then x and y are similar, if $K(x, y)$ is small, they are dissimilar.
- Analogous to inner product $< x, y >$.
- If a symmetric kernel is positive definite then it is essentially the same as an inner product.
- Applications
  - Kernel SVM
  - Kernel PCA
  - Kernel Fisher Discriminant Analysis
  - Dictionary learning (object recognition)

NICTA

Positive-definite Kernel

- A kernel $K : X \times X \to \mathbb{R}$ is called positive definite if for all real numbers c_i,

$$\sum_{i=1}^{n} c_i\, c_j\, K(X_i, X_j) \geq 0$$

for all choices of $X_1, X_2, \ldots X_n \in S$

- Theorem: If a symmetric kernel is positive definite, then it is just like an inner product: there exists a map $\Phi : X \to H$, a Hilbert space, such that

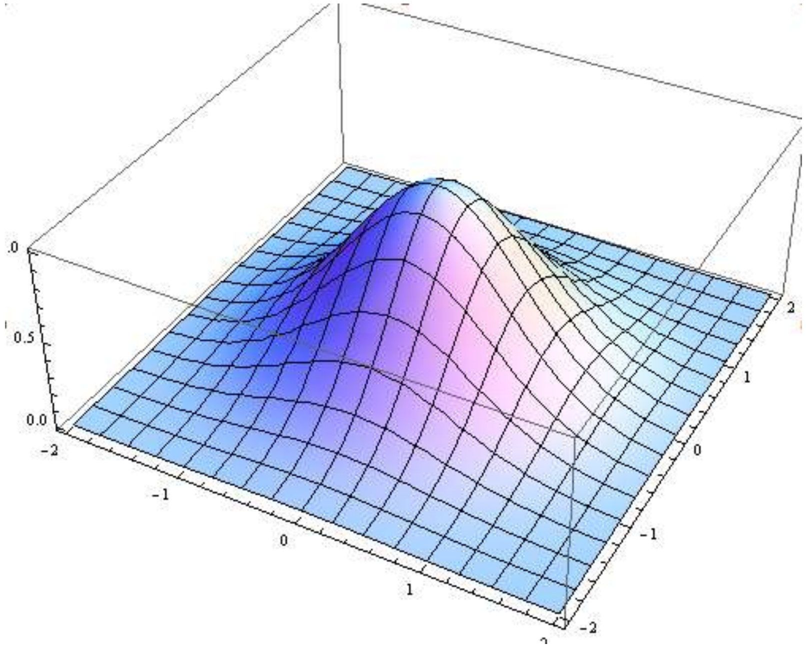
$$K(X,Y) = \langle \Phi(X), \Phi(Y) \rangle_H \ .$$

Radial Basis Function Kernel

- Commonly used kernel:

$$\begin{aligned} K(x,y) &= e^{-\|x-y\|^2/\sigma^2} \\ &= e^{-d(x,y)^2/\sigma_2} \end{aligned}$$

- This is always a positive definite kernel for all σ, if $\|\cdot\|$ is a norm in a Hilbert space (or Euclidean space)

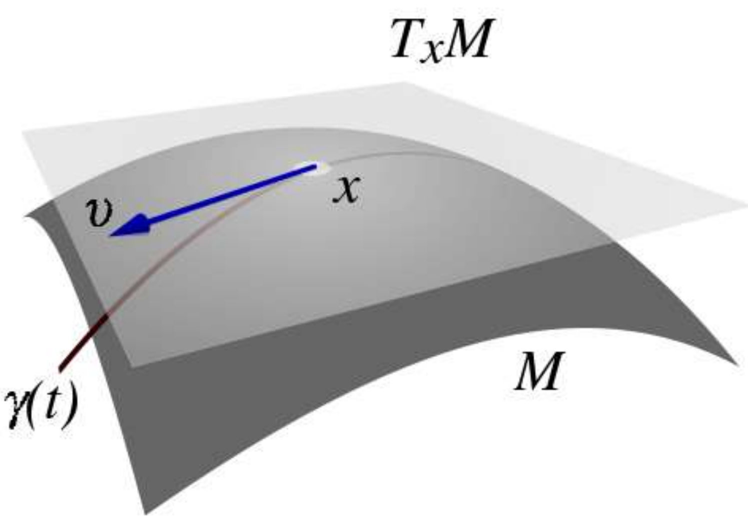


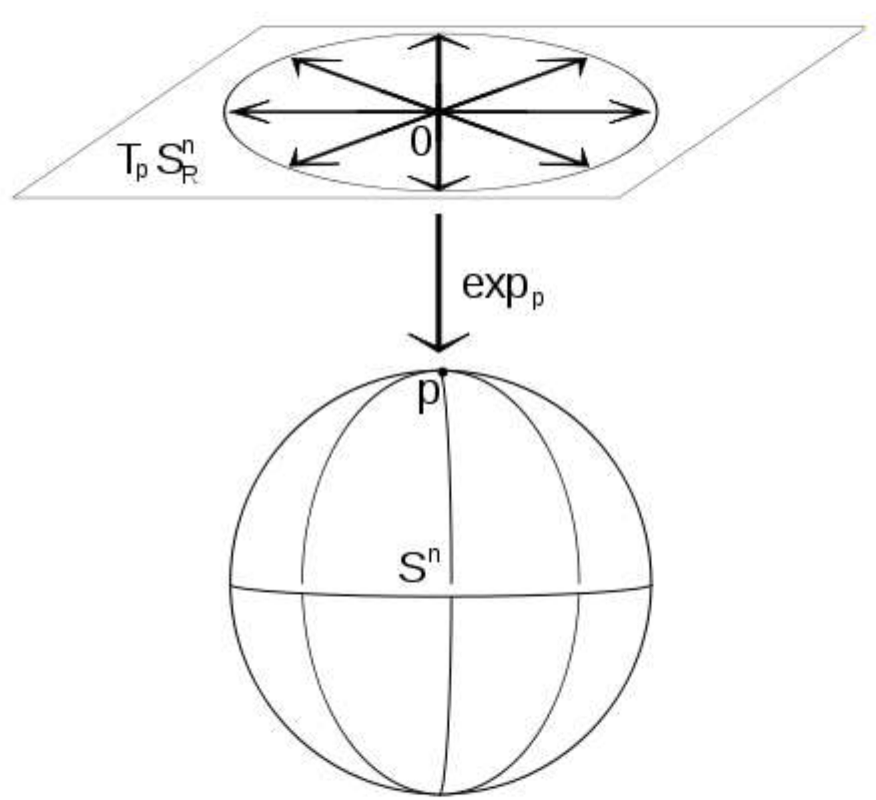
Examples of manifolds

- $\mathbb{R}^n$
- Sphere S^n
- Rotation space $SO(3)$ – used in rotation averaging
- Positive definite matrices – "covariance features"
- Grassman Manifolds – used to model sets of images
- Essential manifold – structure and motion
- Shape manifolds – capture the shape of an object

Kernels in the tangent space

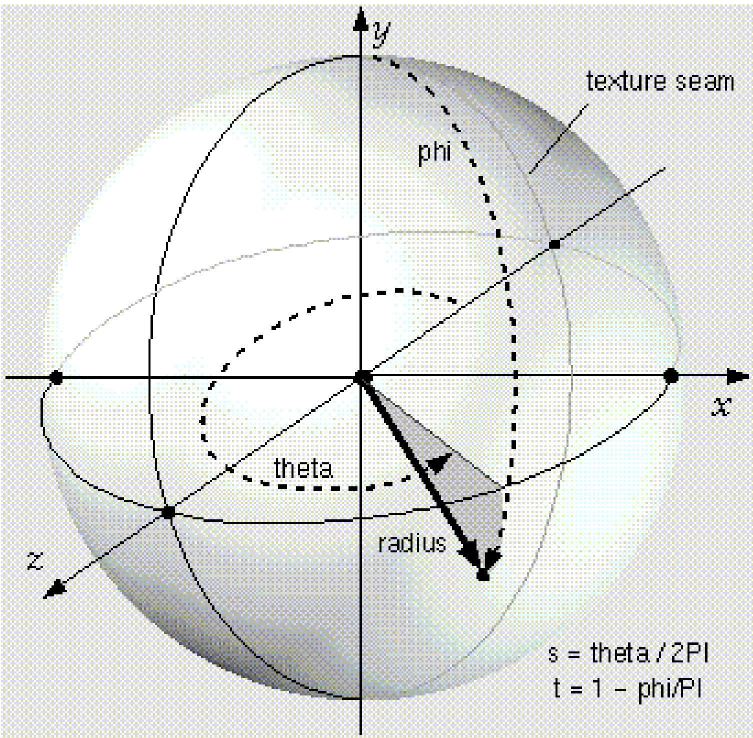
- Map from the manifold to the tangent space using the logarithm map.
- Carry out kernel learning methods in the tangent space.



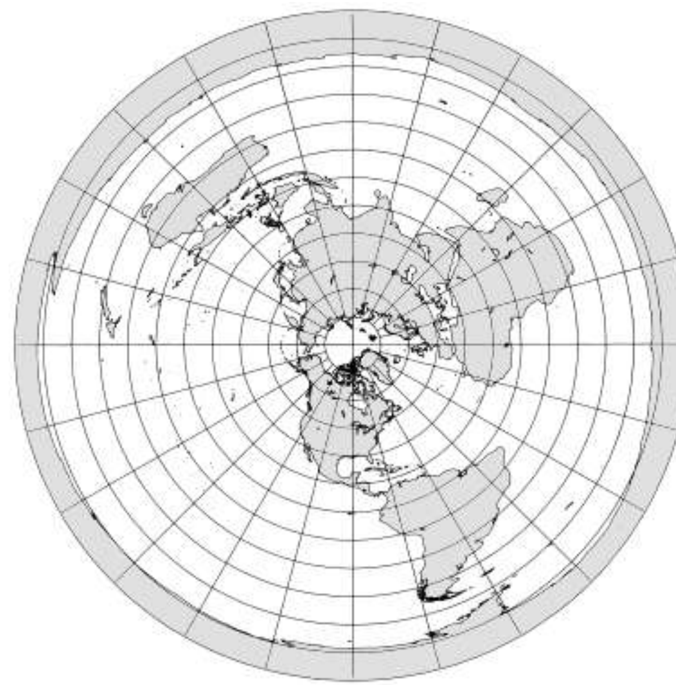


Why this is not a good idea at all

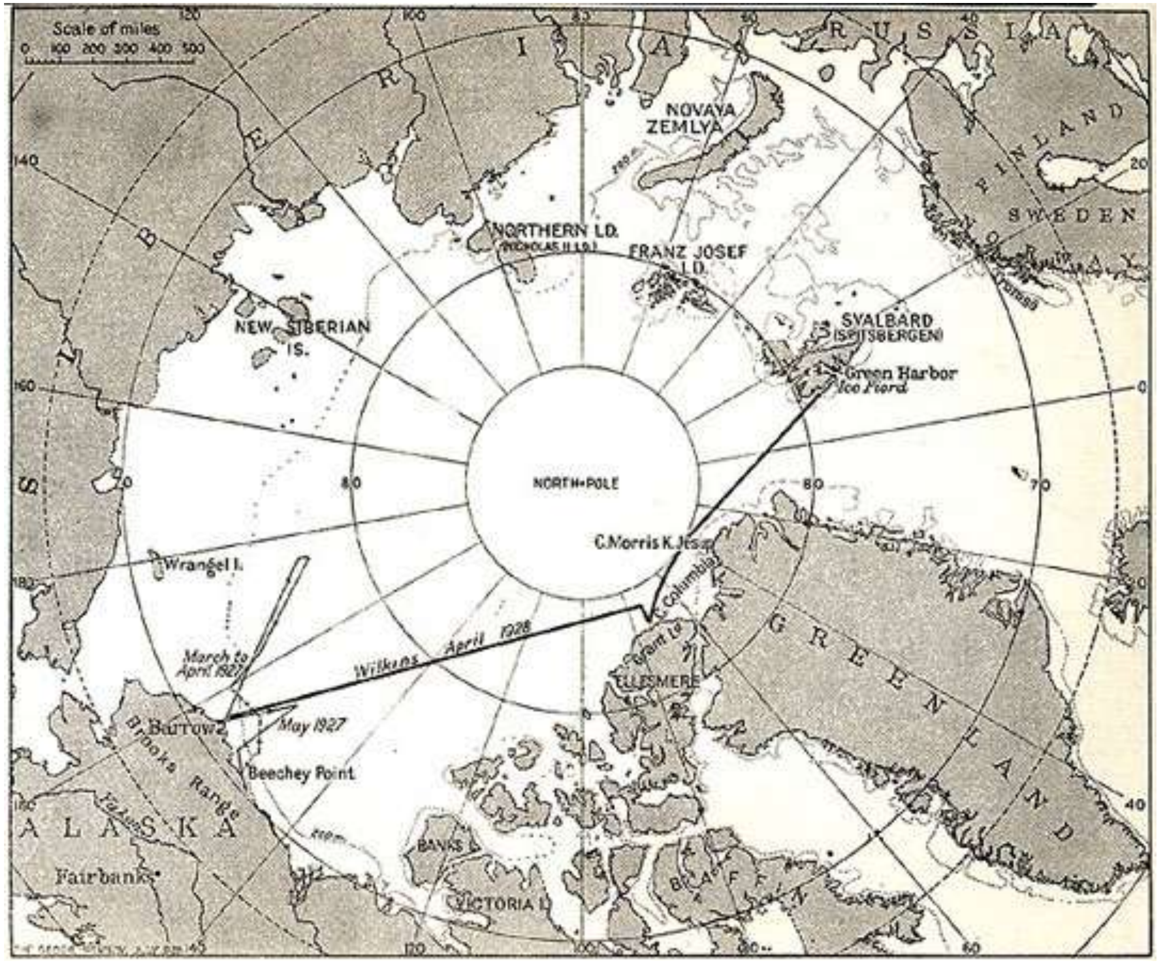
Angle-axis representation of Rotations



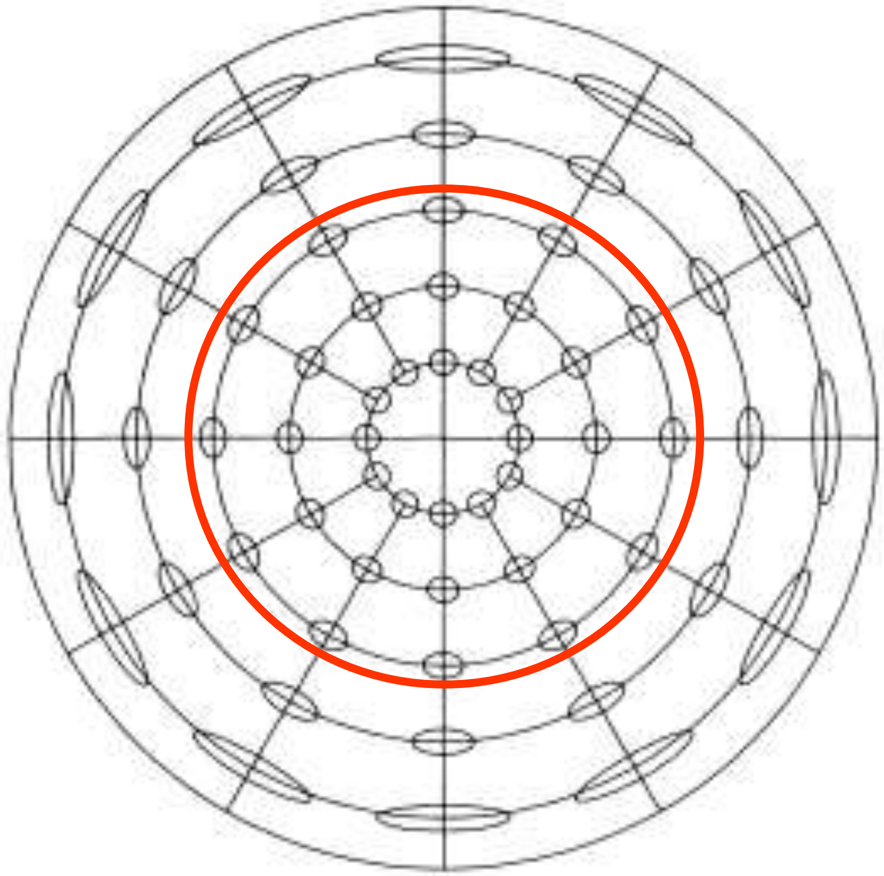
Flatten out the meridians
(longitude lines)



Azimuthal Equidistant Projection



Azimuthal Equidistant Projection



Tissot Indicatrix – shows distortion

Why is the RBF kernel positive definite?

1. Kernel is positive definite on R^n.
2. How do we generalize this?
3. Can we extend this to
   (a) Metric spaces: Distance function $d(x, y)$ defined.
   (b) Normed vector spaces?
   (c) Manifolds?

When is the RBF kernel positive definite

- Consider a "distance function" $d(X, Y)$ defined on a set S (metric space)
- Theorem: The radial basis function

$$K(X, Y) = e^{-d(X,Y)^2/\sigma^2}$$

is a positive definite kernel for all σ, if and only if S can be isometrically embedded in a Hilbert Space.

$$d(X, Y) = \|\phi(X), \phi(Y)\|_H$$

- (Technical point) It is not enough that H be a Banach space. The inner product is needed.

A negative result

Theorem. R^n is the only complete manifold M for which the RBF kernel

$$k(x,y) = e^{\frac{-d_g^2(x,y)}{\sigma^2}}$$

is a kernel for all σ.

Here, $d_g(x,y)$ is the geodesic distance on the manifold.

Solution: Find distance metrics on manifolds that do lead to RBF kernels. "Asymptotically geodesic distances".

1. Monotonic function of geodesic distance.
2. In the limit equal to a geodesic distance for small distances.

Positive Definite Matrices

- The Positive definite $n \times n$ matrices form a cone (not a linear subspace).
- Affine invariance:

$$d(X, Y) = d(A^\top X A, A^\top Y A)$$

- We can define an "affine invariant" Riemannian metric.
- Other metrics:
  - Logarithm:

$$d(X, Y) = \|\log(X) - \log(Y)\|_F$$

  - Stein Metric:

$$d(X, Y)^2 = -\log\det(XY) + 2\log\det((X+Y)/2)$$

Kernels on Positive Definite Matrices

Metric Name	Formula	Geodesic Distance	Positive Definite Gaussian Kernel $\forall\sigma > 0$
Log-Euclidean	$\lVert \log(\mathbf{S}_1) - \log(\mathbf{S}_2) \rVert_F$	**Yes**	**Yes**
Affine-Invariant	$\lVert \log(\mathbf{S}_1^{-1/2}\mathbf{S}_2\mathbf{S}_1^{-1/2}) \rVert_F$	Yes	No
Cholesky	$\lVert \mathrm{chol}(\mathbf{S}_1) - \mathrm{chol}(\mathbf{S}_2) \rVert_F$	No	Yes
Power-Euclidean	$\frac{1}{\alpha}\lVert \mathbf{S}_1^\alpha - \mathbf{S}_2^\alpha \rVert_F$	No	Yes
Root Stein Divergence	$\left[\log\det\left(\frac{1}{2}\mathbf{S}_1 + \frac{1}{2}\mathbf{S}_2\right) - \frac{1}{2}\log\det(\mathbf{S}_1\mathbf{S}_2)\right]^{1/2}$	No	No

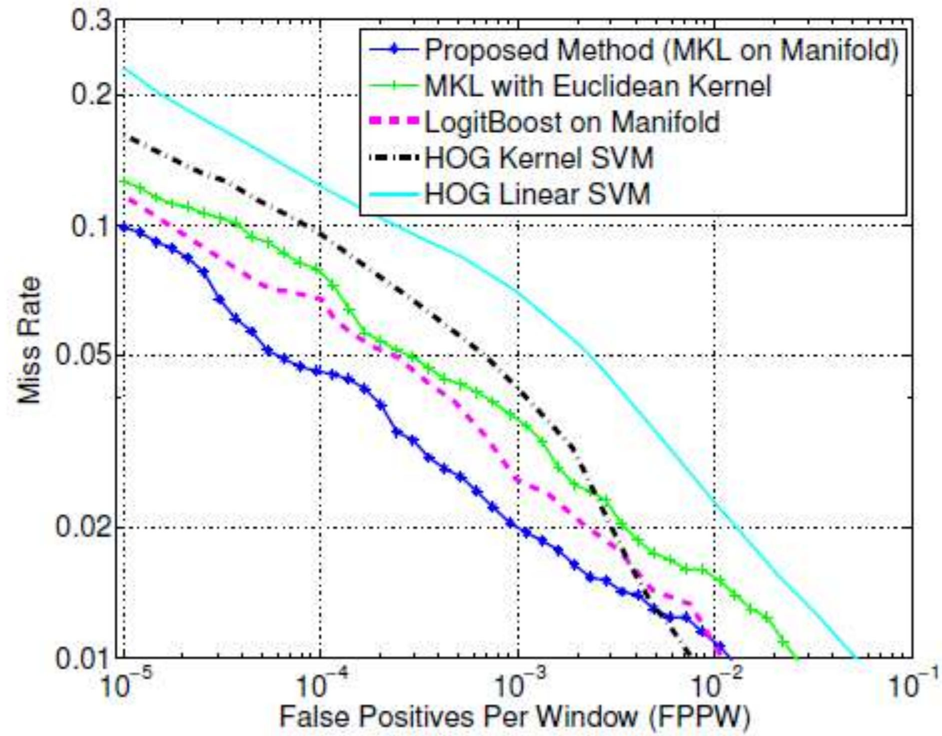
Nb. of classes	Euclidean		Cholesky		Power-Euclidean		Log-Euclidean	
	KM	KKM	KM	KKM	KM	KKM	KM	KKM
3	72.50	79.00	73.17	82.67	71.33	84.33	75.00	**94.83**
4	64.88	73.75	69.50	84.62	69.50	83.50	73.00	**87.50**
5	54.80	70.30	70.80	82.40	70.20	82.40	74.60	**85.90**
6	50.42	69.00	59.83	73.58	59.42	73.17	66.50	**74.50**
7	42.57	68.86	50.36	69.79	50.14	69.71	59.64	**73.14**
8	40.19	68.00	53.81	69.44	54.62	68.44	58.31	**71.44**

Pedestrian detection

Table: Sample images from INRIA dataset

Pedestrian detection

- Covariance descriptor is used as the region descriptor following Tuzel et al., 2008.
- Multiple covariance descriptors are calculated per detection window, an SVM + MKL framework is used to build the classifier.



Visual object categorization

Table: Sample images from ETH-80 dataset

Visual object categorization

- ETH-80 dataset, 5×5 covariance descriptors.
- Manifold k-means and manifold kernel k-means with different metrics.

Nb. of classes	Euclidean		Cholesky		Power-Euclidean		Log-Euclidean	
	KM	KKM	KM	KKM	KM	KKM	KM	KKM
3	72.50	79.00	73.17	82.67	71.33	84.33	75.00	**94.83**
4	64.88	73.75	69.50	84.62	69.50	83.50	73.00	**87.50**
5	54.80	70.30	70.80	82.40	70.20	82.40	74.60	**85.90**
6	50.42	69.00	59.83	73.58	59.42	73.17	66.50	**74.50**
7	42.57	68.86	50.36	69.79	50.14	69.71	59.64	**73.14**
8	40.19	68.00	53.81	69.44	54.62	68.44	58.31	**71.44**

Texture recognition

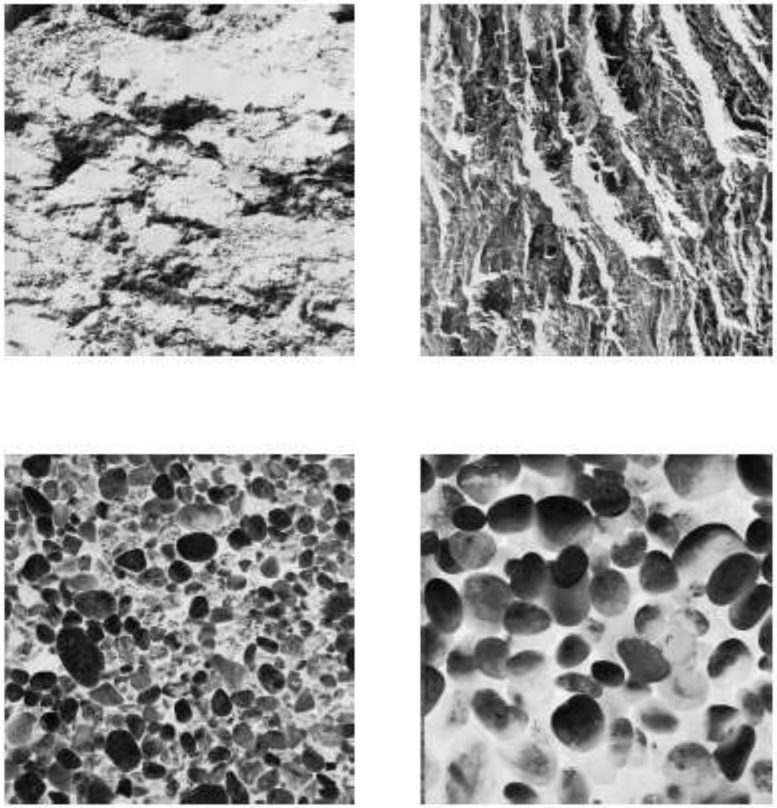
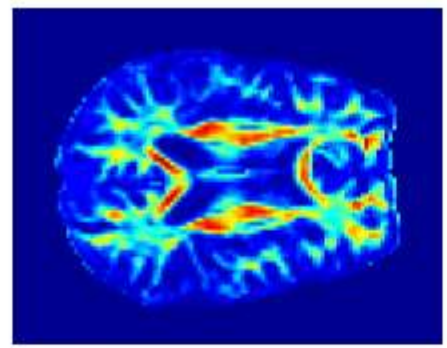


Table: Sample images from Brodatz

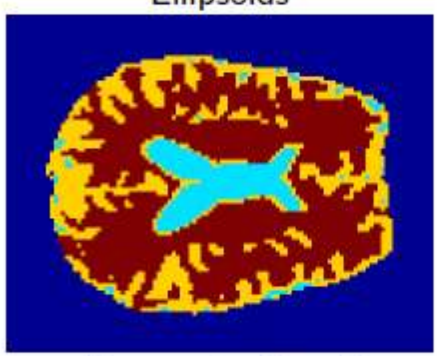
DTI segmentation

- Diffusion tensor at the voxel is directly used as the descriptor.
- Kernel k-means is utilized to cluster points on Sym_d^+, yielding a segmentation of the DTI image.

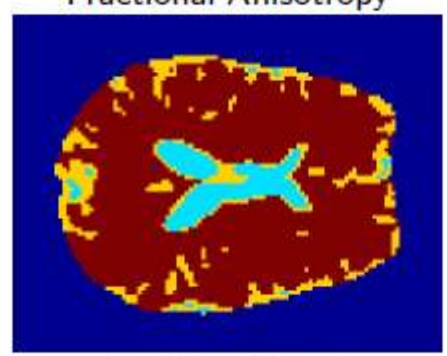
Ellipsoids



Fractional Anisotropy



Riemannian kernel



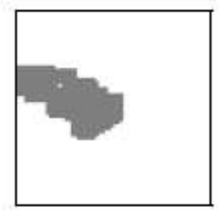
Euclidean kernel

Motion segmentation

- The structure tensor (3×3) was used as the descriptor.
- Kernel k-means clustering of the tensors yields the segmentation.
- Achieves better clustering accuracy than methods that work in a low dimensional space.

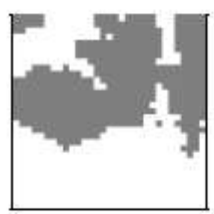
Frame 1

Frame 2

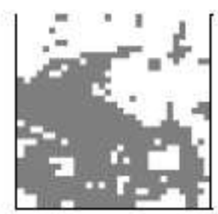


KKM on Sym_3^+

LLE on Sym_3^+



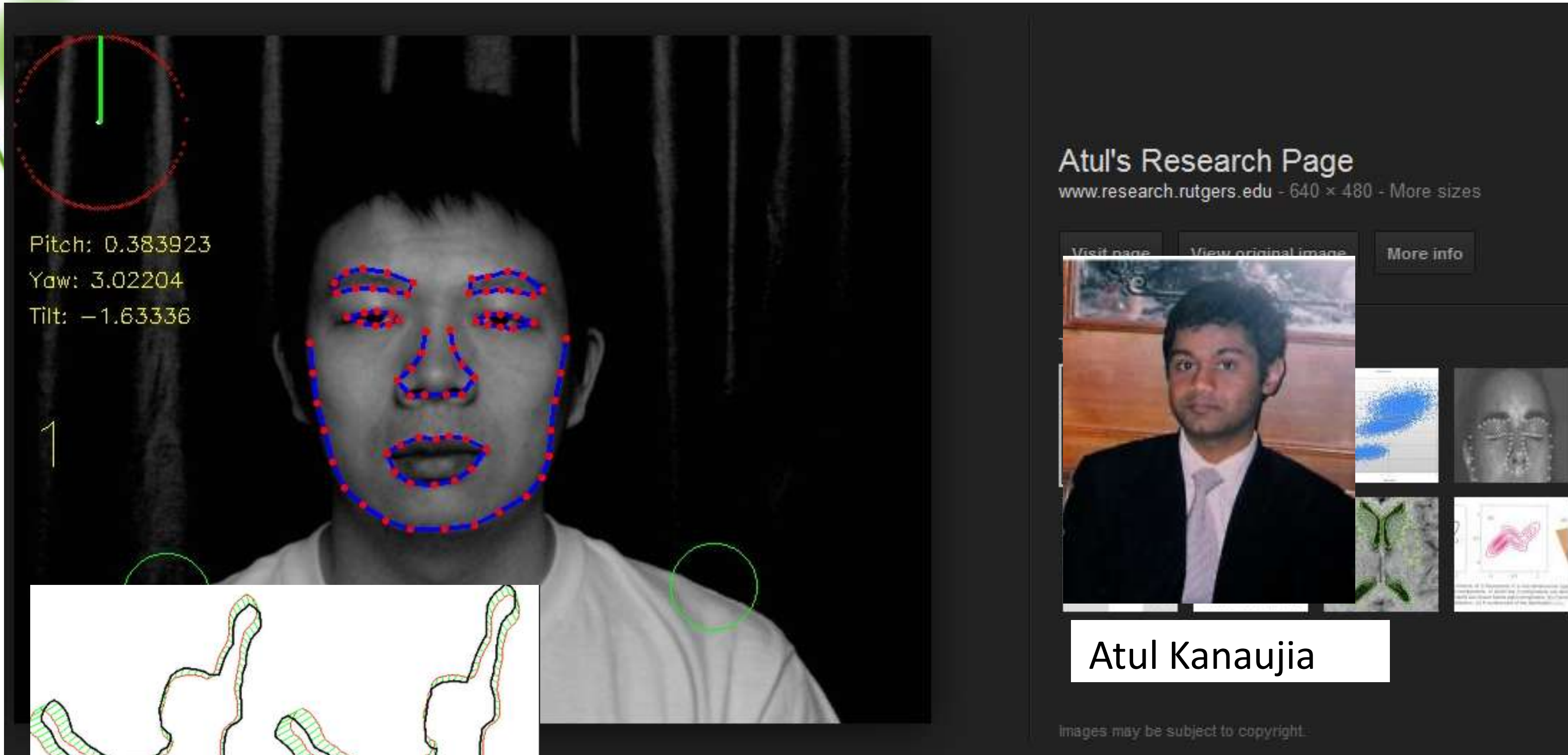
LE on Sym_3^+



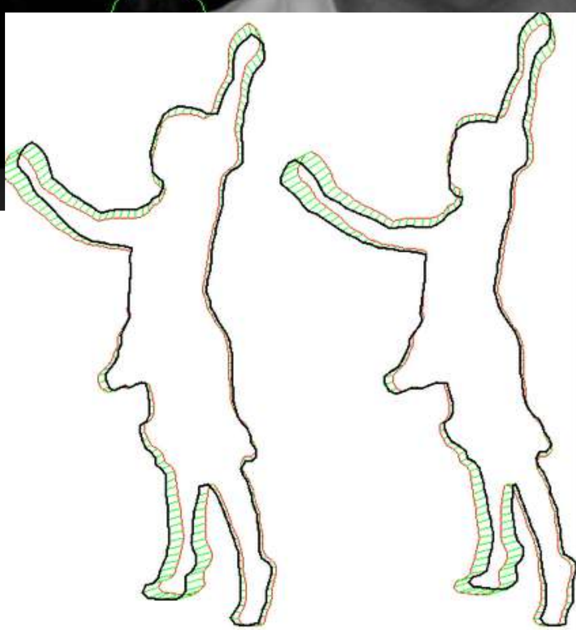
HLLE on Sym_3^+

Shape Manifolds

- Captures what is invariant in a set of k points in $\mathbb{R}^n$, when you take away rotation, translation and scale.
- Formally, a shape is an equivalence class of k points, where two sets of k points are equivalent if they are related by rotation, translation and scaling.
- Jayasumana et al, (ICCV 2013)



Atul Kanaujia



Guillaume Charpiat

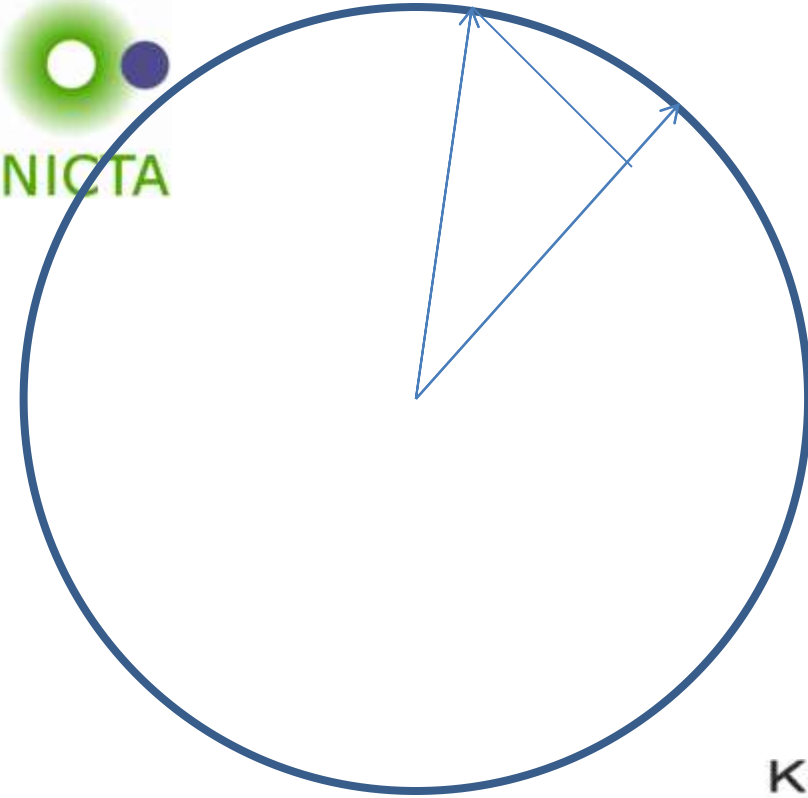
Shape Manifold.
Captures the configuration of a set of points, allowing for rotation, translation and scaling.

2D Shape manifolds

- Represent each point as a complex number.
- Vector of n complex numbers represents a shape.
- Normalize this vector to unit length and scale to length 1.
- "Preshape manifold" is equal to the complex n-dimensional sphere.

$$S = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{pmatrix}$$

- Multiplication by a non-zero complex unit complex number $z = e^{i\theta}$ rotates all the points.
- Shape manifold is equal to the complex projective space.



Kernels on the shape manifold

- Define $\cos\theta = \| \langle X, Y \rangle \|$
- $\sin(\theta)$ is the "full-Procrustes" distance – yields a positive definite RBF kernel
- Other possible distance
  - $d_P(X, Y) = 2\sin(\theta/2)$ does not
  - Geodesic distance θ does not

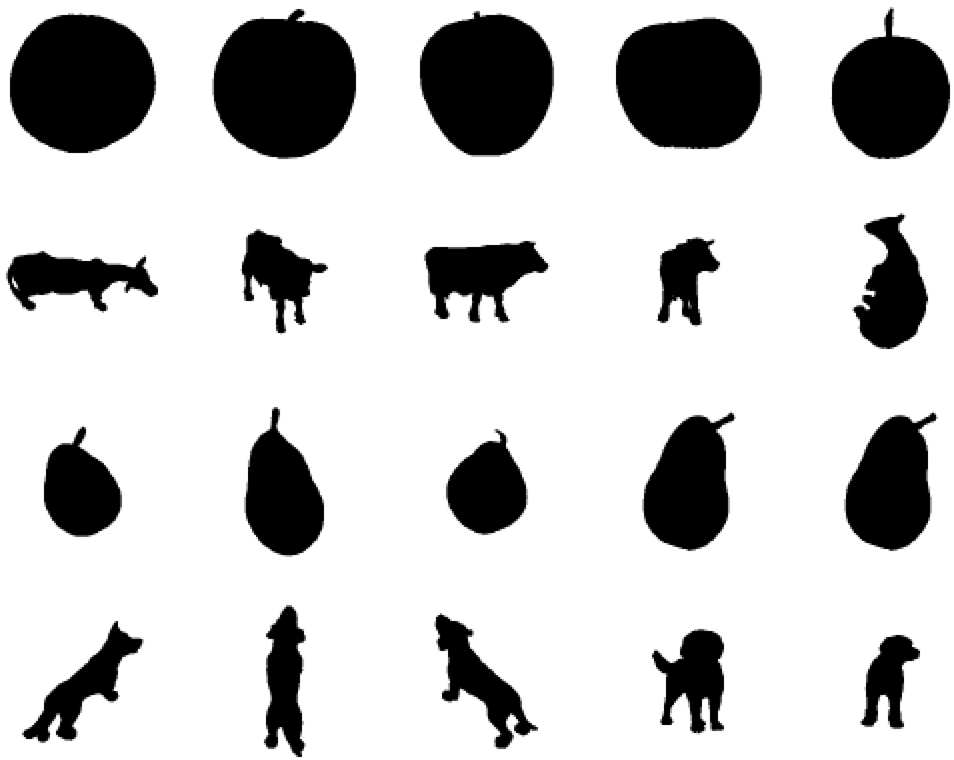
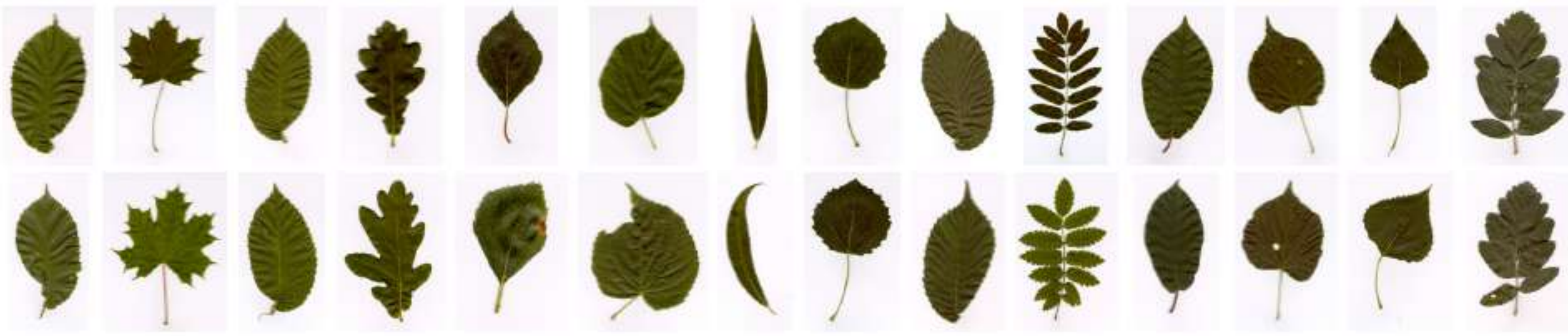
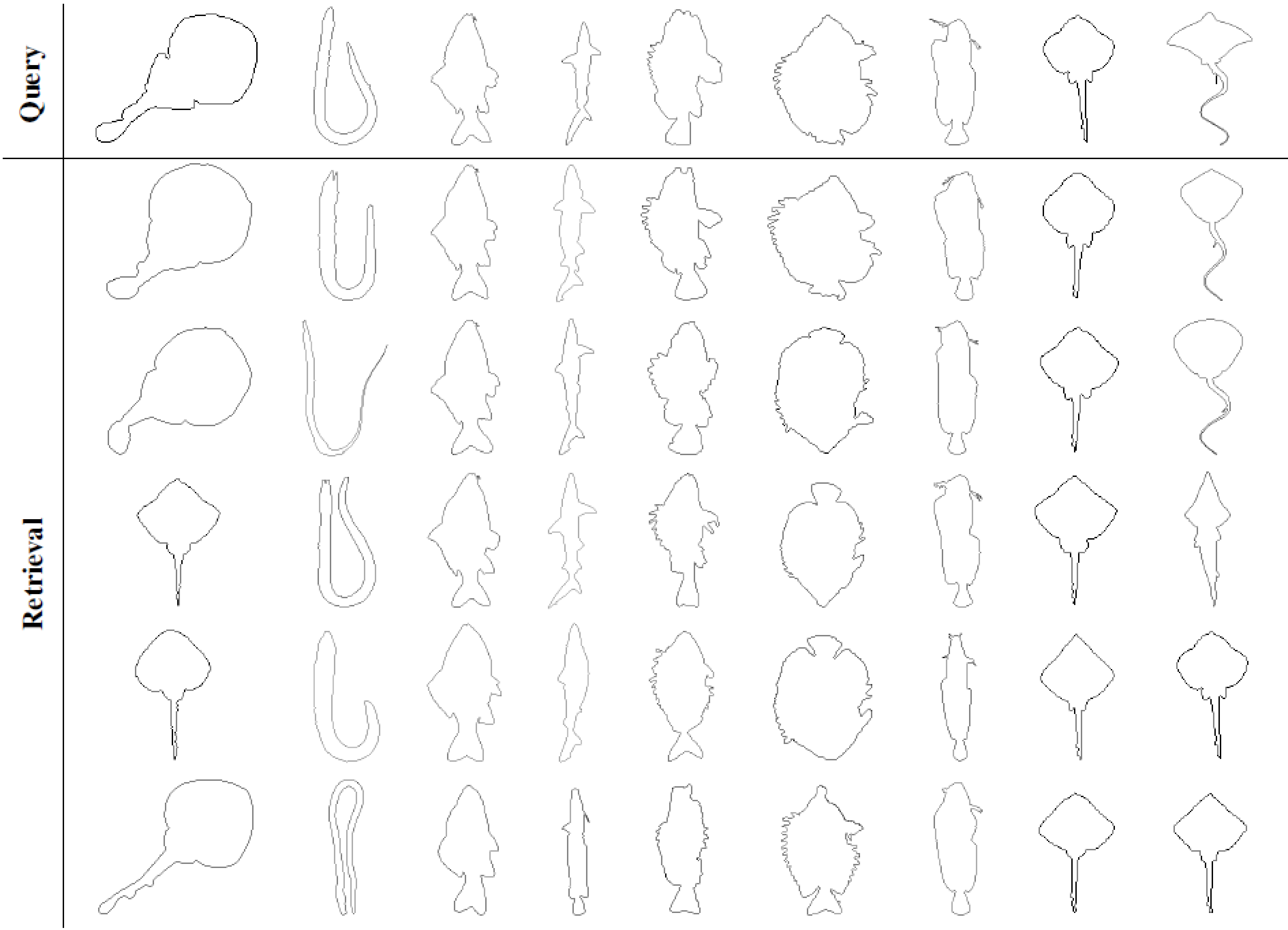


Figure 1: **The ETH-80 dataset.** Sample images from different objects and classes in the ETH-80 dataset.



Leaf Database


Query
Retrieval

Radial kernels on n-sphere

$$k_i(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle^i, \text{ for } i \in \mathbb{N}^0,$$

$$k_{-1}(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{if } \mathbf{x} = \mathbf{y}, \\ -1 & \text{if } \mathbf{x} = -\mathbf{y}, \\ 0 & \text{otherwise}, \end{cases}$$

$$k_{-2}(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{if } \mathbf{x} = \pm\mathbf{y}, \\ 0 & \text{otherwise}. \end{cases}$$

Schoenberg's result

Theorem 4.3. *A kernel $k : S_{\mathcal{H}} \times S_{\mathcal{H}} \to \mathbb{R}$ is radial with respect to the geodesic distance and is p.d. if and only if it admits the form*

$$k(\mathbf{x}, \mathbf{y}) = \sum_{i=-2}^{\infty} a_i k_i(\mathbf{x}, \mathbf{y})$$

where $\sum_i a_i < \infty$ and $a_i \geq 0$ for all i. Furthermore, k is continuous if and only if $a_{-1} = a_{-2} = 0$.

Radial kernels on n-sphere

$$k_i(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle^i, \text{ for } i \in \mathbb{N}^0,$$

$$k_{-1}(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{if } \mathbf{x} = \mathbf{y}, \\ -1 & \text{if } \mathbf{x} = -\mathbf{y}, \\ 0 & \text{otherwise,} \end{cases}$$

$$k_{-2}(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{if } \mathbf{x} = \pm\mathbf{y}, \\ 0 & \text{otherwise.} \end{cases}$$

- As i increases k_i rapidly approaches either k_{-1} or k_{-2}.
- Therefore, the infinite series can be closely approximated with a finite sum.
- Readily fits in to a Multiple Kernel Learning (MKL) framework!

Extending to other manifolds

- Grassmann manifold with Projection distance

$$d_P([Y_1],[Y_2]) = \|Y_1Y_1^T - Y_2Y_2^T\|_F$$

$$k_i([Y_1],[Y_2]) = \langle Y_1Y_1^T - Y_2Y_2^T\rangle_F^i$$

- Shape manifold with full Procrustes distance

$$d_{FP}([\mathbf{z}_1],[\mathbf{z}_2]) = \sqrt{1 - |\langle \mathbf{z}_1, \mathbf{z}_2\rangle|^2}$$

$$k_i([\mathbf{z}_1],[\mathbf{z}_2]) = |\langle \mathbf{z}_1, \mathbf{z}_2\rangle|^{2i}$$

Hand sketch recognition

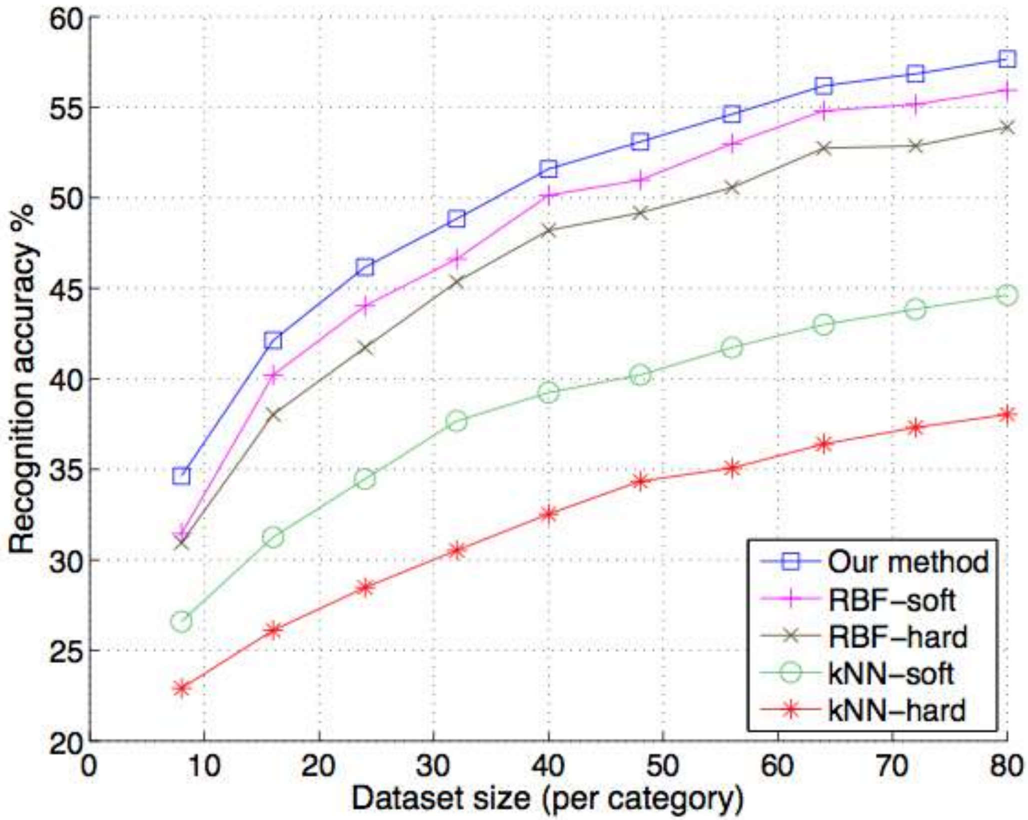


Figure 1: **Hand sketch recognition.** Recognition accuracies for different dataset sizes. The curves for the baselines were reproduced from [7].

Face & action recognition

Method	YT-Celebrity dataset	Ballet dataset
GDA [10]	58.72 ± 3.0	67.33 ± 1.1
GGDA [11]	61.06 ± 2.2	73.54 ± 2.0
Projection kernel k_P [10]	64.76 ± 2.1	74.66 ± 1.2
Proj. Gaussian kernel k_{PG} [12]	71.78 ± 2.4	76.95 ± 0.9
Our method	$\mathbf{72.00 \pm 1.9}$	$\mathbf{78.05 \pm 1.0}$

Table 2: **Face and action recognition.** Average recognition accuracies of our method compared to other kernel methods on $\mathcal{G}_n^r$.

Shape recognition

Method	Butterfly dataset	Pet dataset
Procrustes kernel k_{FP}	57.75 ± 2.0	67.48
Proc. Gaussian kernel k_{FPG} [13]	60.37 ± 1.6	77.34
Tangent Gaussian kernel [13]	58.96 ± 1.8	75.77
Our method	$\mathbf{63.98} \pm \mathbf{1.6}$	**80.87**

Table 3: **Shape recognition.** Average recognition accuracies of our method compared to other kernel methods on $\mathcal{SP}^n$. Note that the train/test partition on the Pet dataset is fixed and given by [17].

NICTA

Acknowledgements.

This talk deals with work done by myself and my collaborators, particularly

- Mehrtash Harandi
- Sadeep Jayasumana
- Mathieu Salzmann
- Hongdong Li
- Brian Lovell
- Fatih Porikli

Papers in the last 3 years at top 3 vision conferences on Riemannian manifolds.

NICTA : 9
Maryland : 7
Florida state : 6
University of Florida : 6
John Hopkins : 2
INRIA : 1

The End