Data with low intrinsic dimension.

Yoav Freund
UCSD

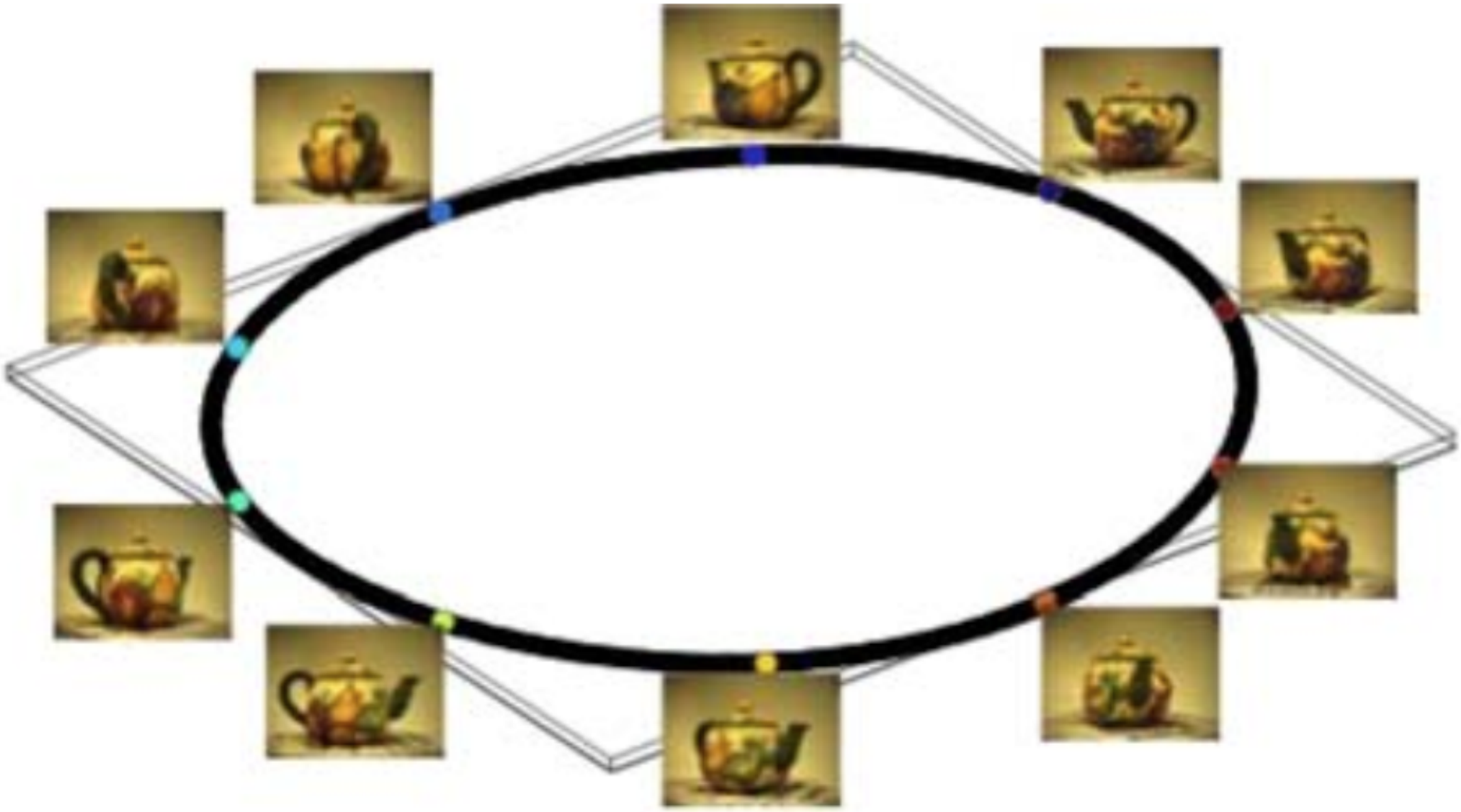
Goals of this course

- Where does low dimensionality of data come from?
- mathematical models of low dimensional data.
- Efficient algorithms for learning low dimensional models.

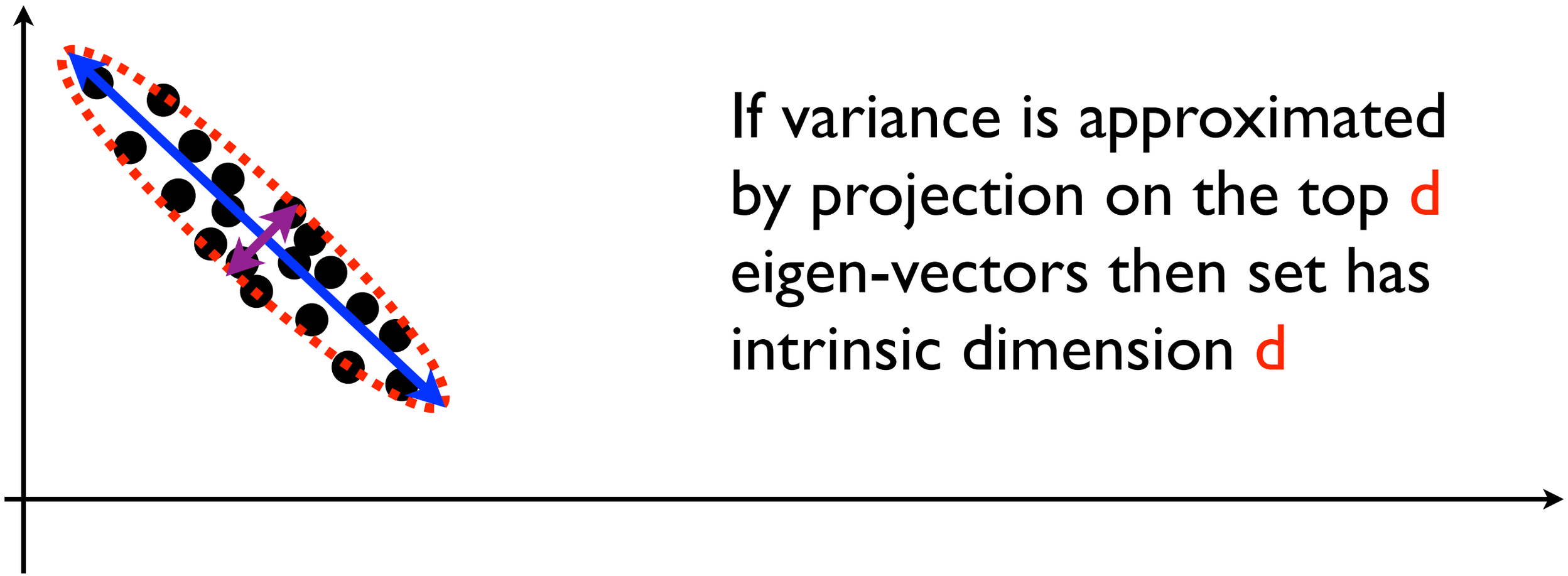
What is low intrinsic dimension

- A video frame is high dimensional.
  - A 1000X1000X3 image is defined by 3 million parameters.
  - However, the scene captured is typically parametrized by a much smaller number of parameters = Small number of degrees of freedom.

The turning tea-pot



PCA



If variance is approximated by projection on the top d eigen-vectors then set has intrinsic dimension d

Eigen-Faces

Beyond Eigenfaces: Probabilistic Matching for Face Recognition

Baback Moghaddam
Mitsubishi Electric Research Laboratory

Wasiuddin Wahid and Alex Pentland
MIT Media Laboratory

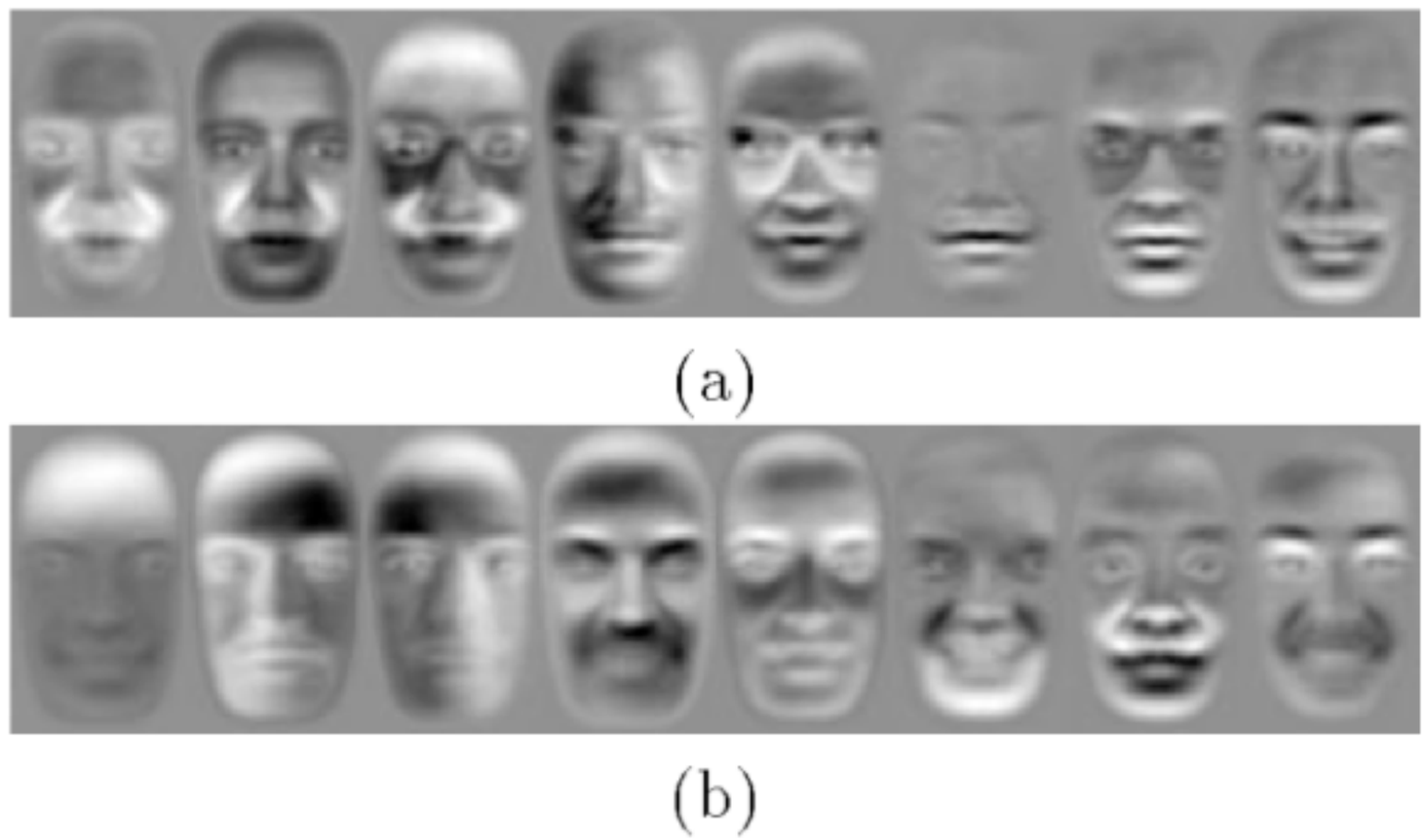
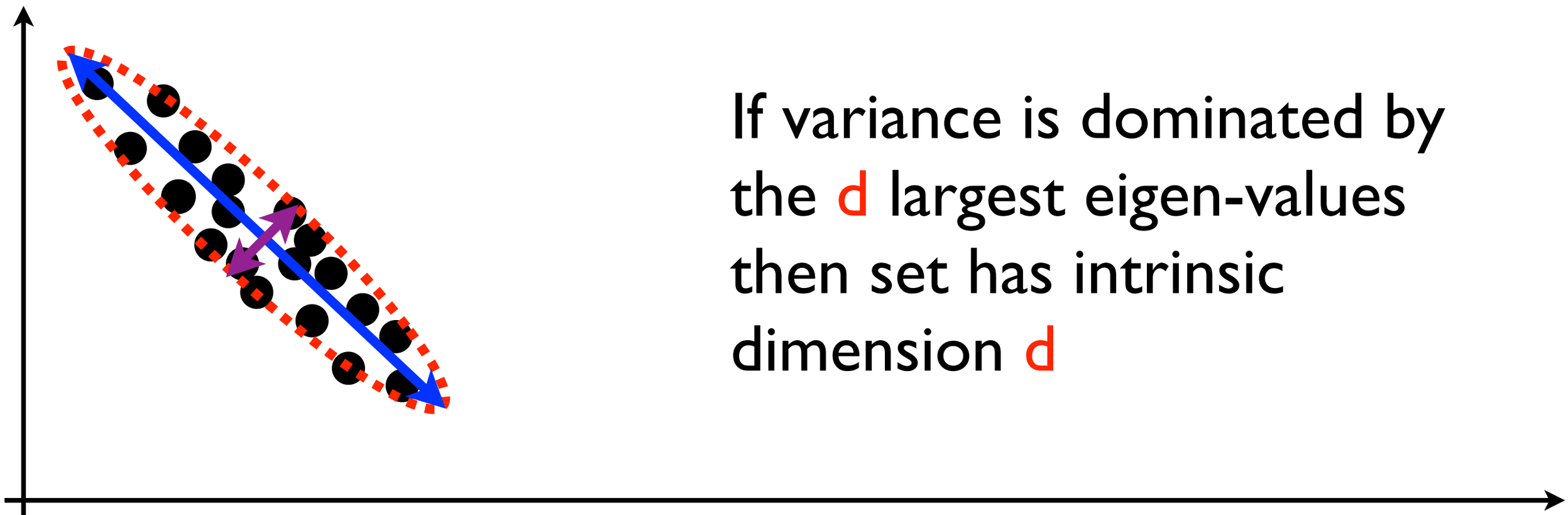


Figure 6: "Dual" Eigenfaces: (a) Intrapersonal, (b) Extrapersonal

PCA

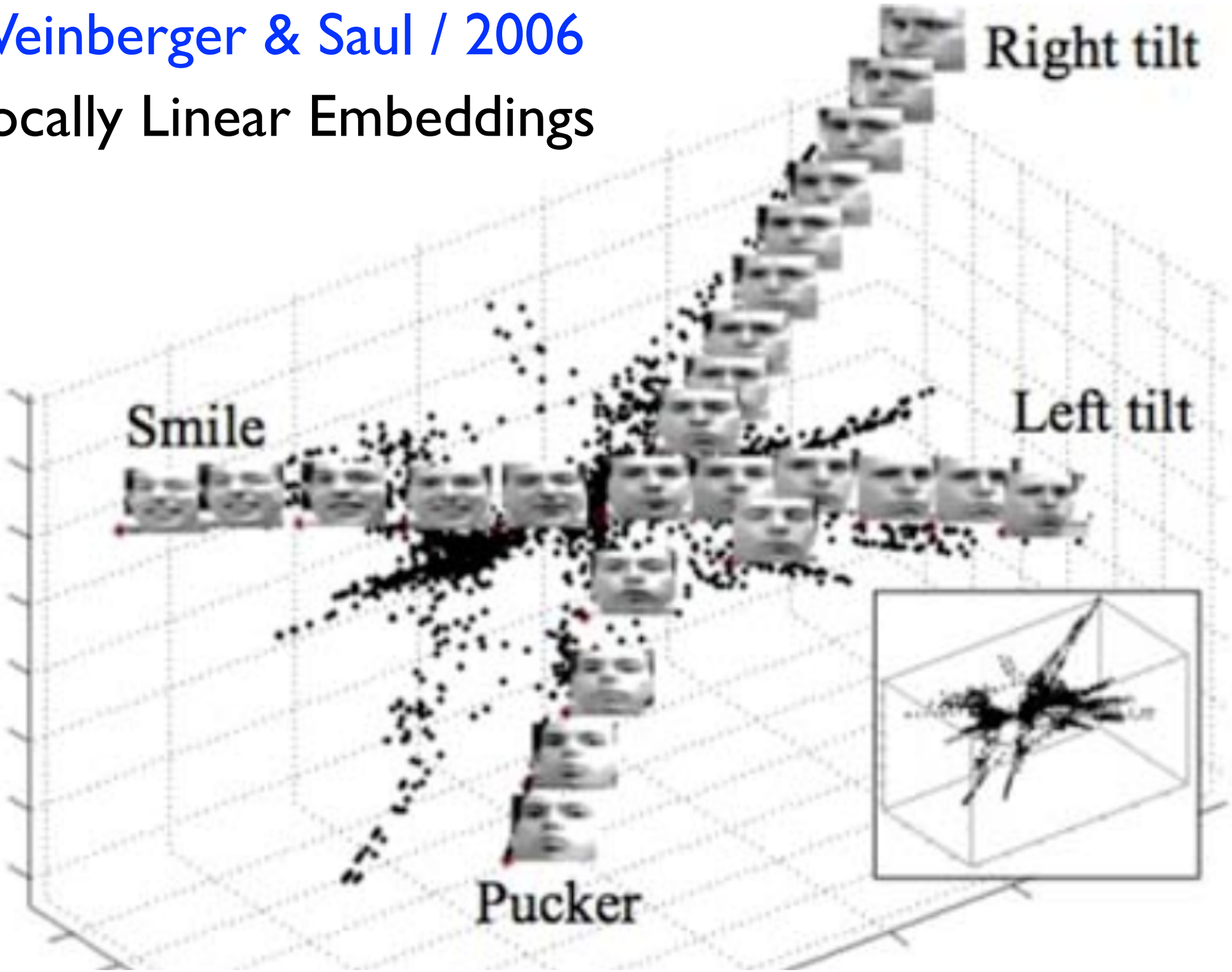


If variance is dominated by the d largest eigen-values then set has intrinsic dimension d

What can we do if set is on d-dim manifold that is not affine?

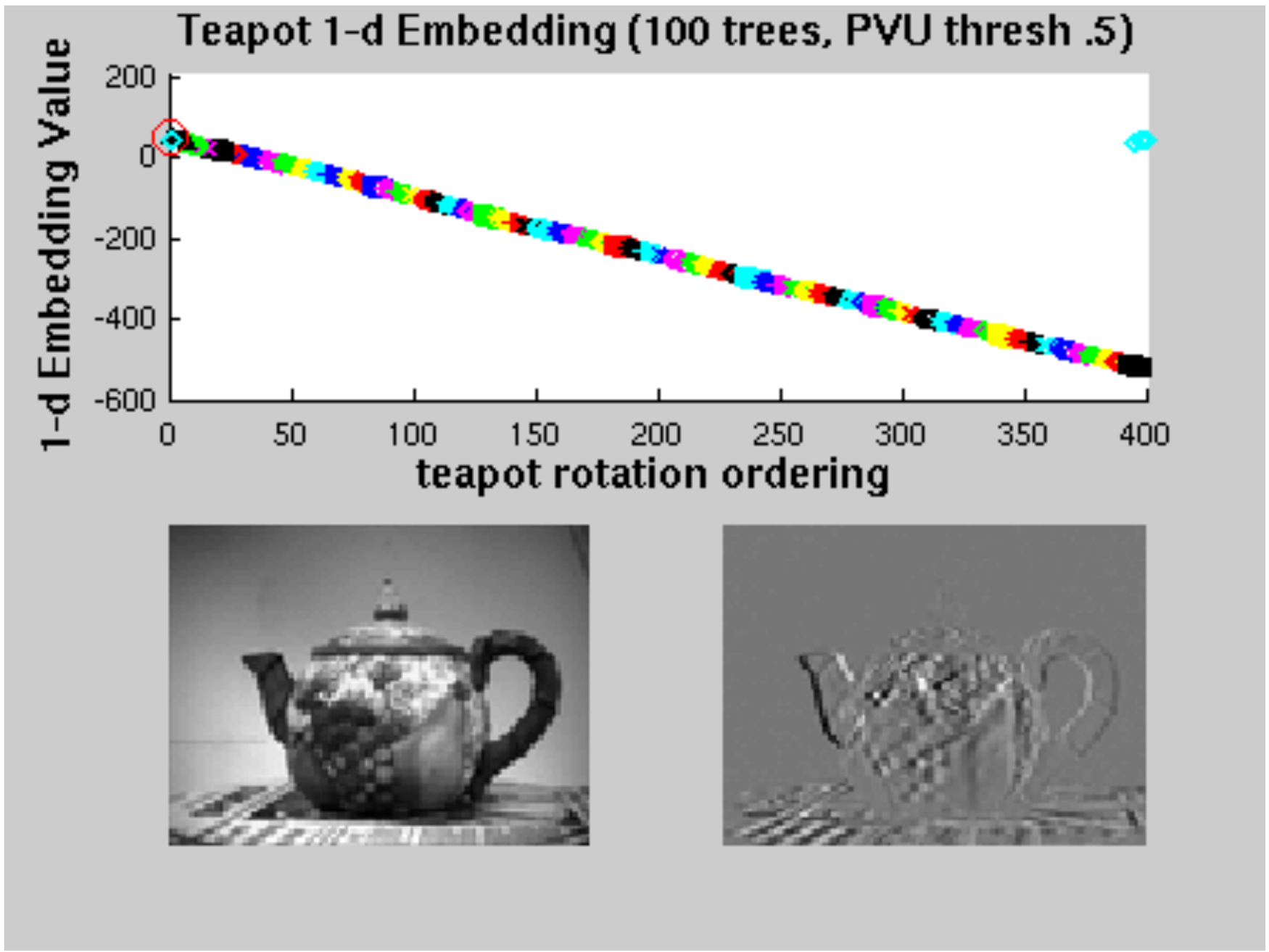
Weinberger & Saul / 2006

Locally Linear Embeddings



Charting turning teapot manifold

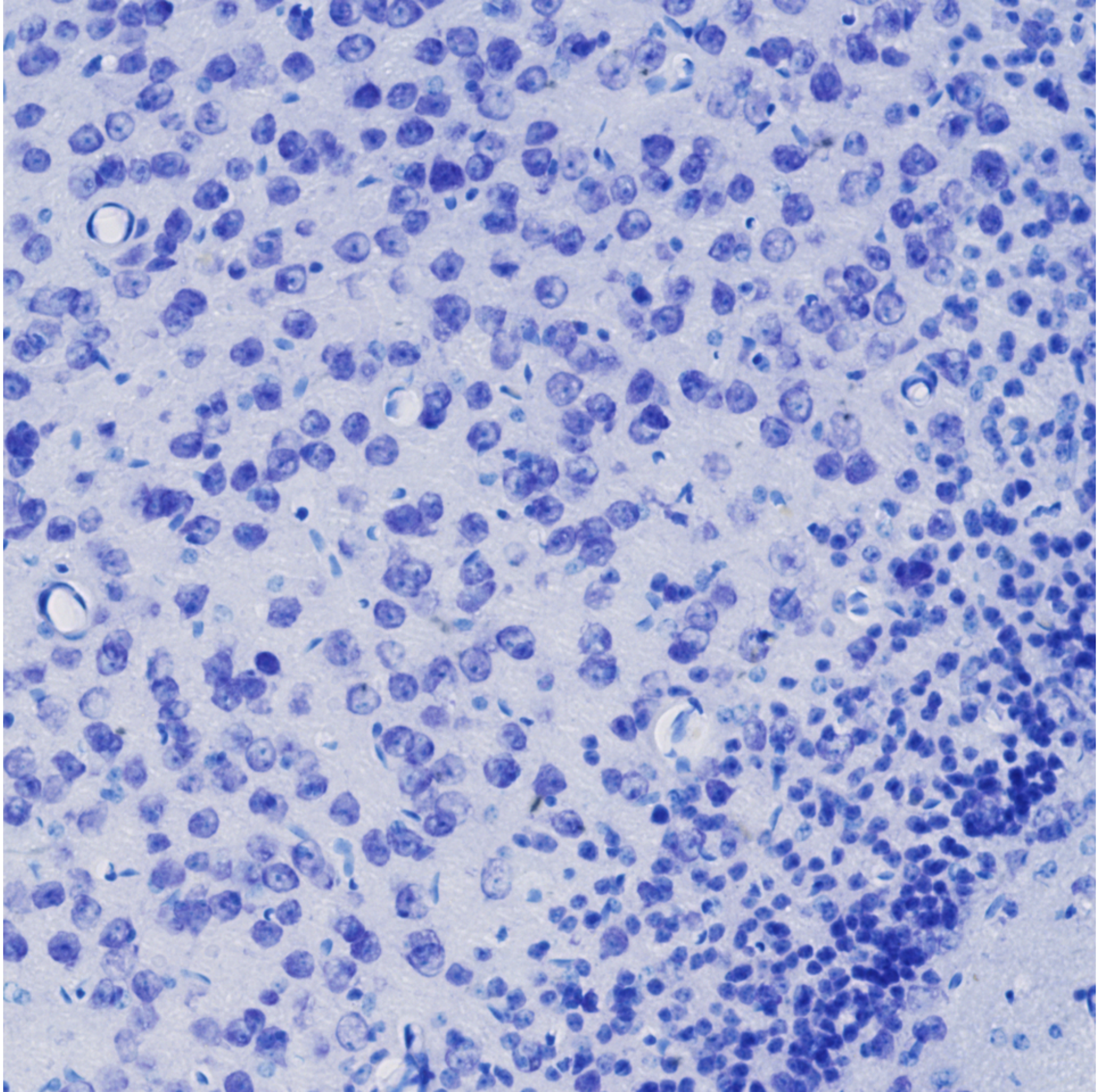
Given an unordered set of frames, organize them according to their rotation angle.

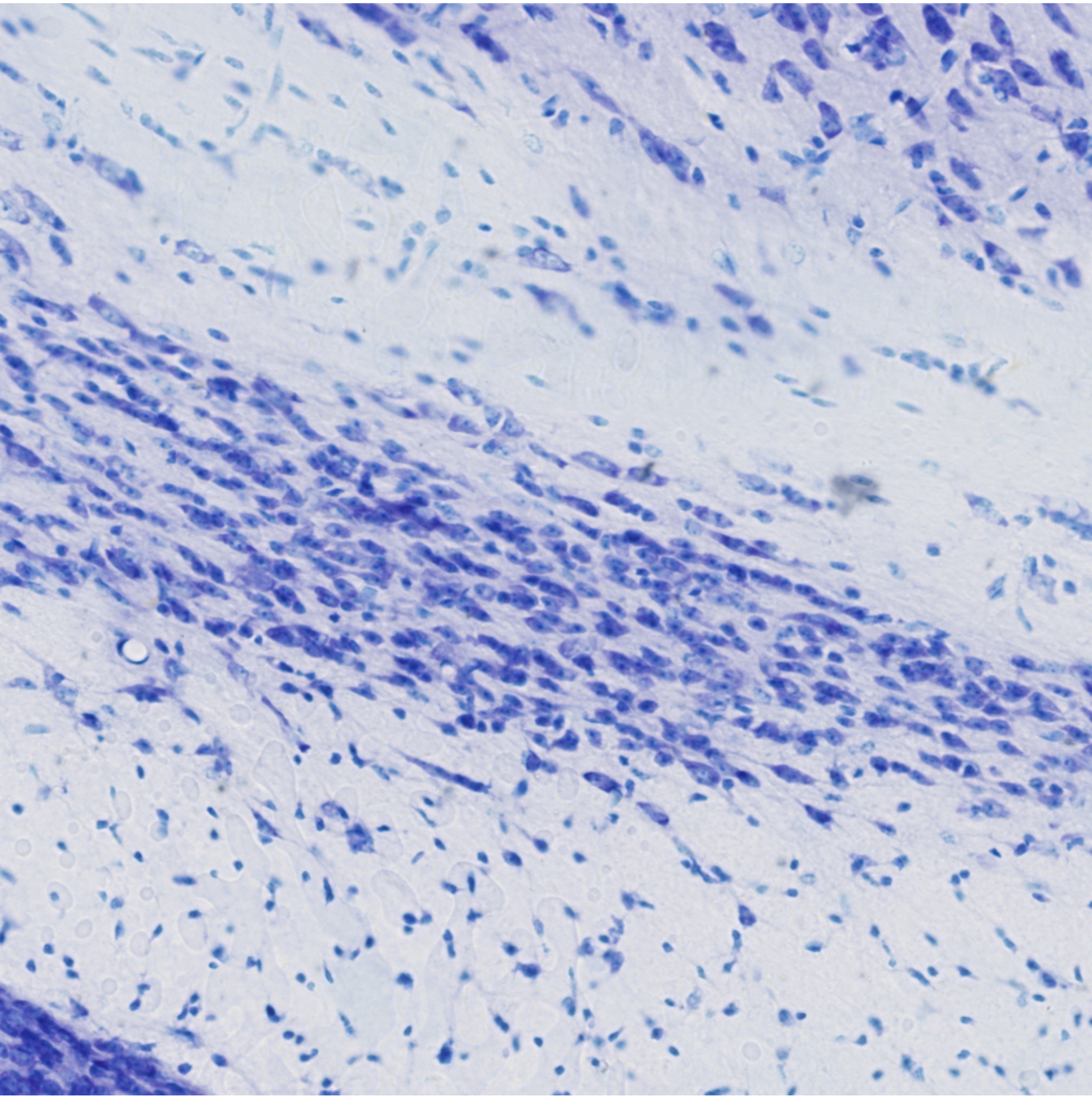


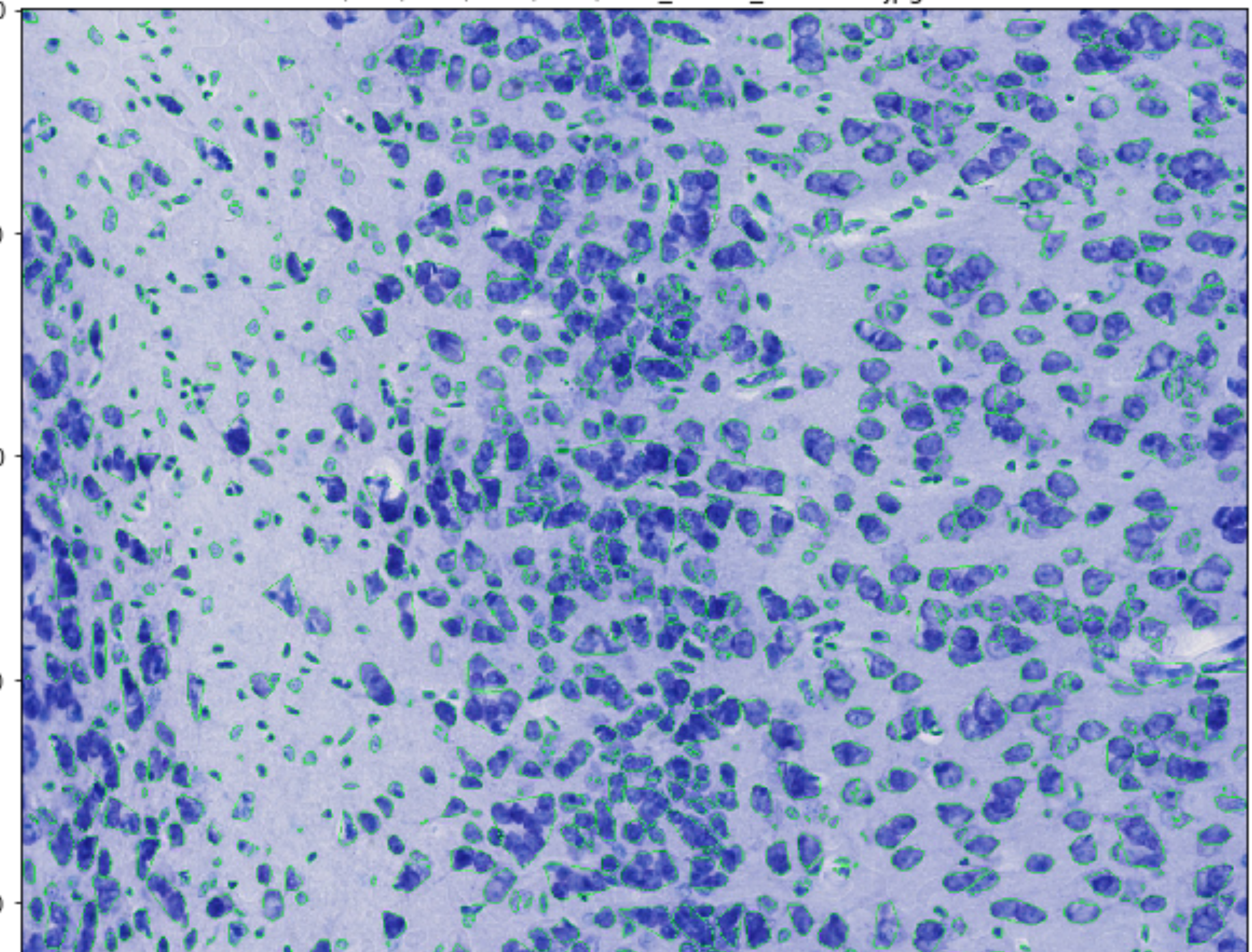
How does this fit with the rest of ML ?

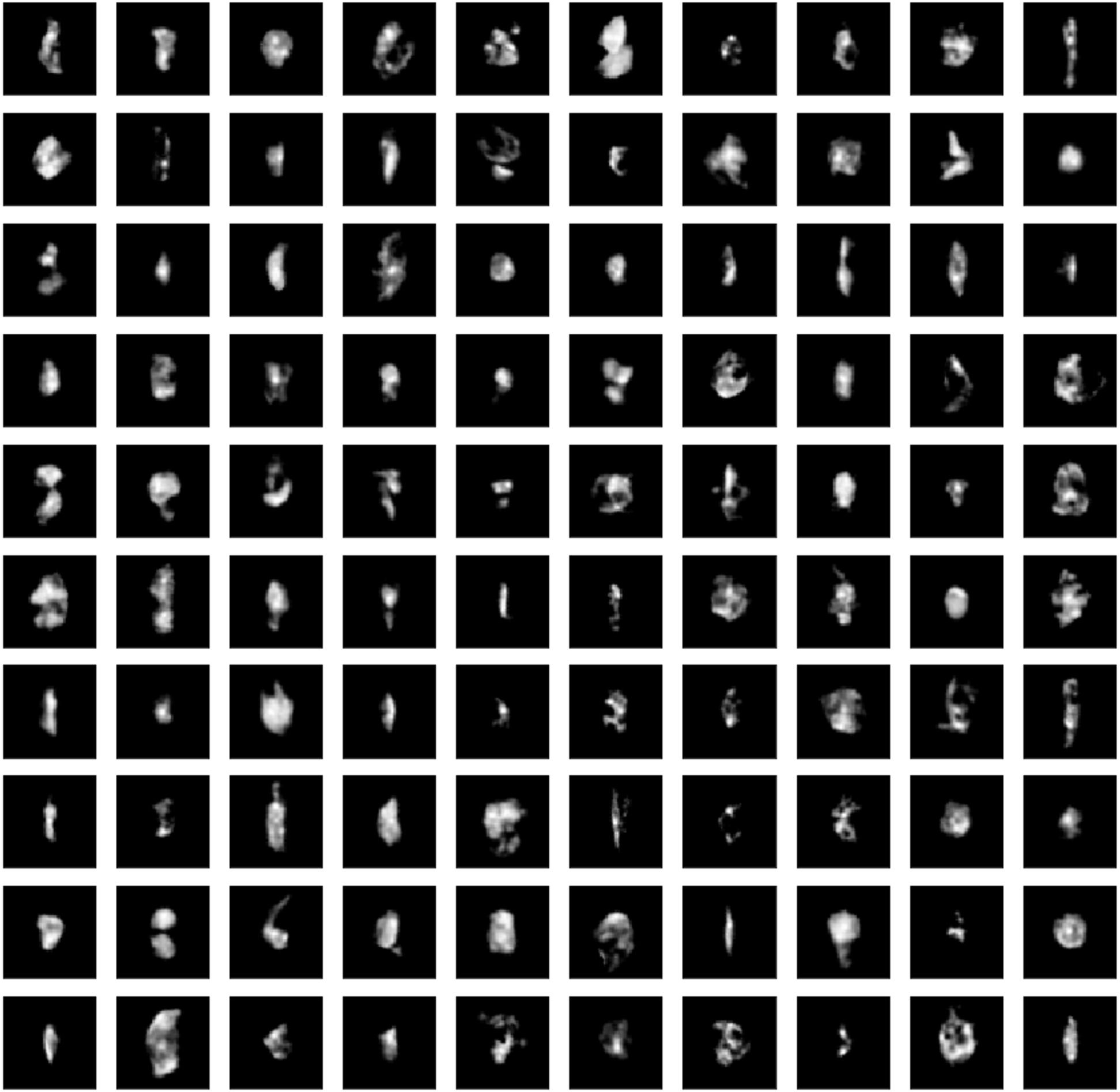
- Neural Networks very good for **supervised** learning complex non-linear models, not very effective for **unsupervised** learning.
- Low-dim manifolds - **unsupervised** learning of low dimensional models - dimensionality reduction.
- Easy to get lots of unlabeled data.
- **Challenge**: can we perform the learning in IO-efficient ways: without keeping the whole dataset in memory

Parametrizing the shapes of neurons



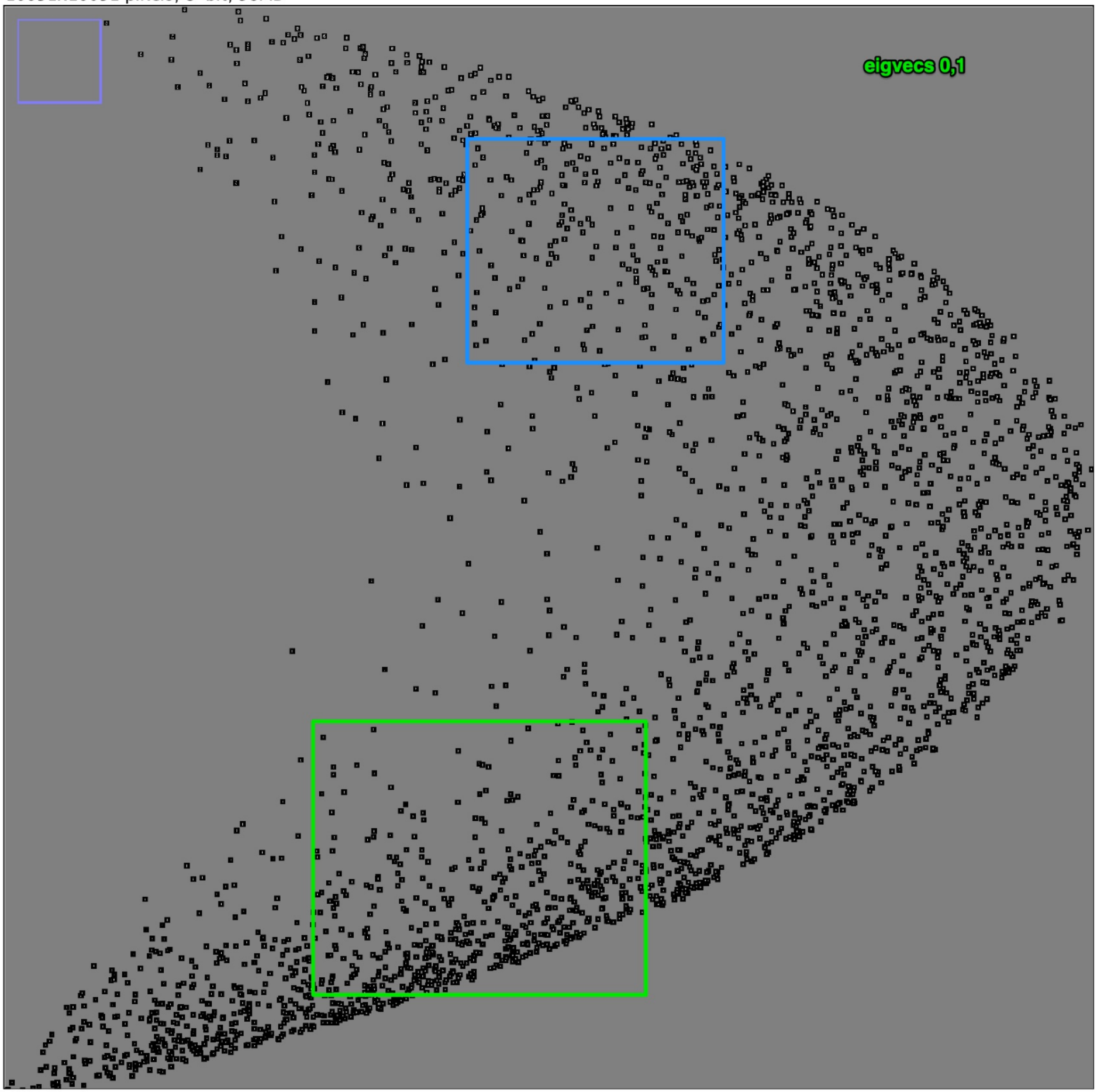


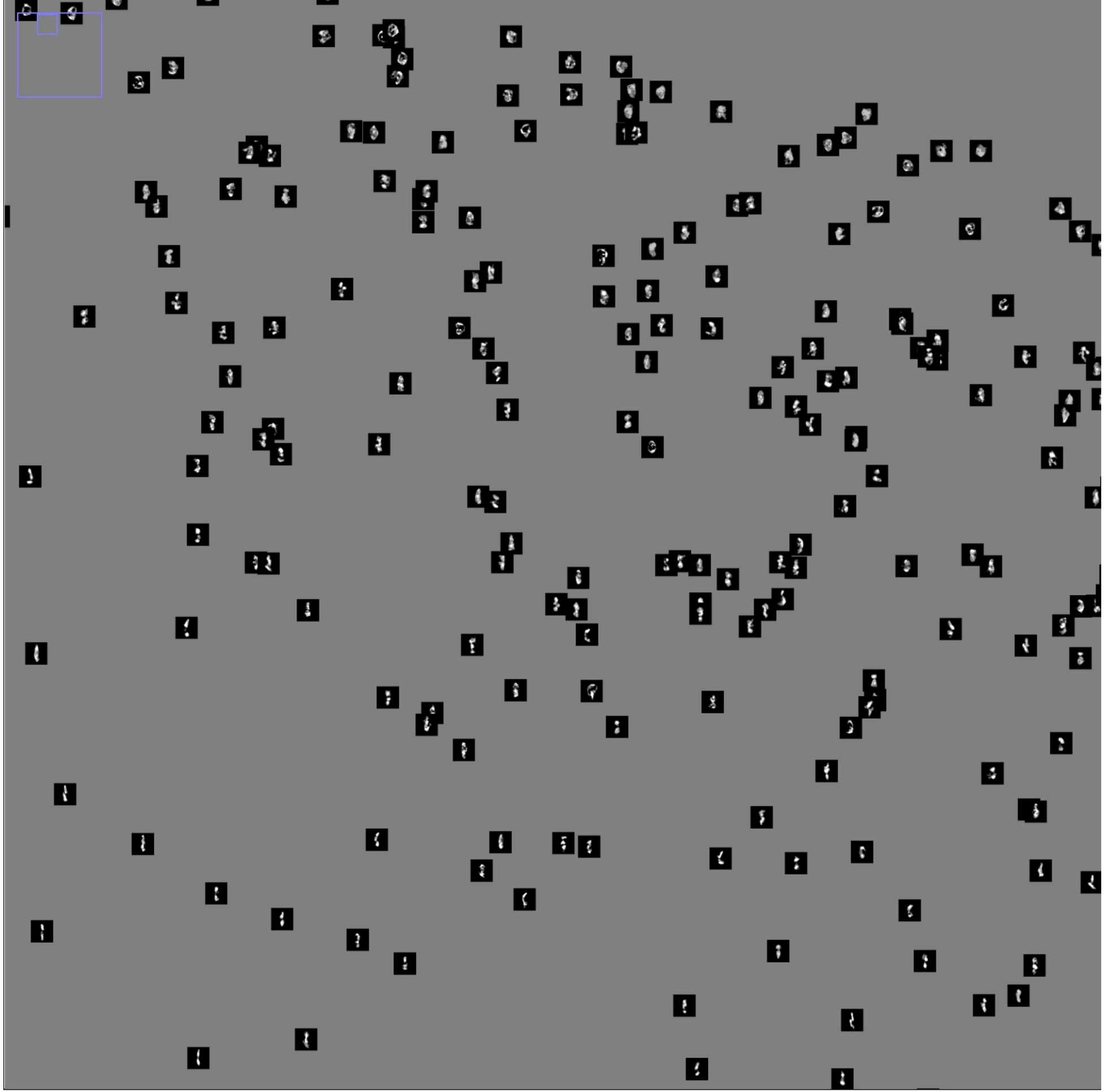


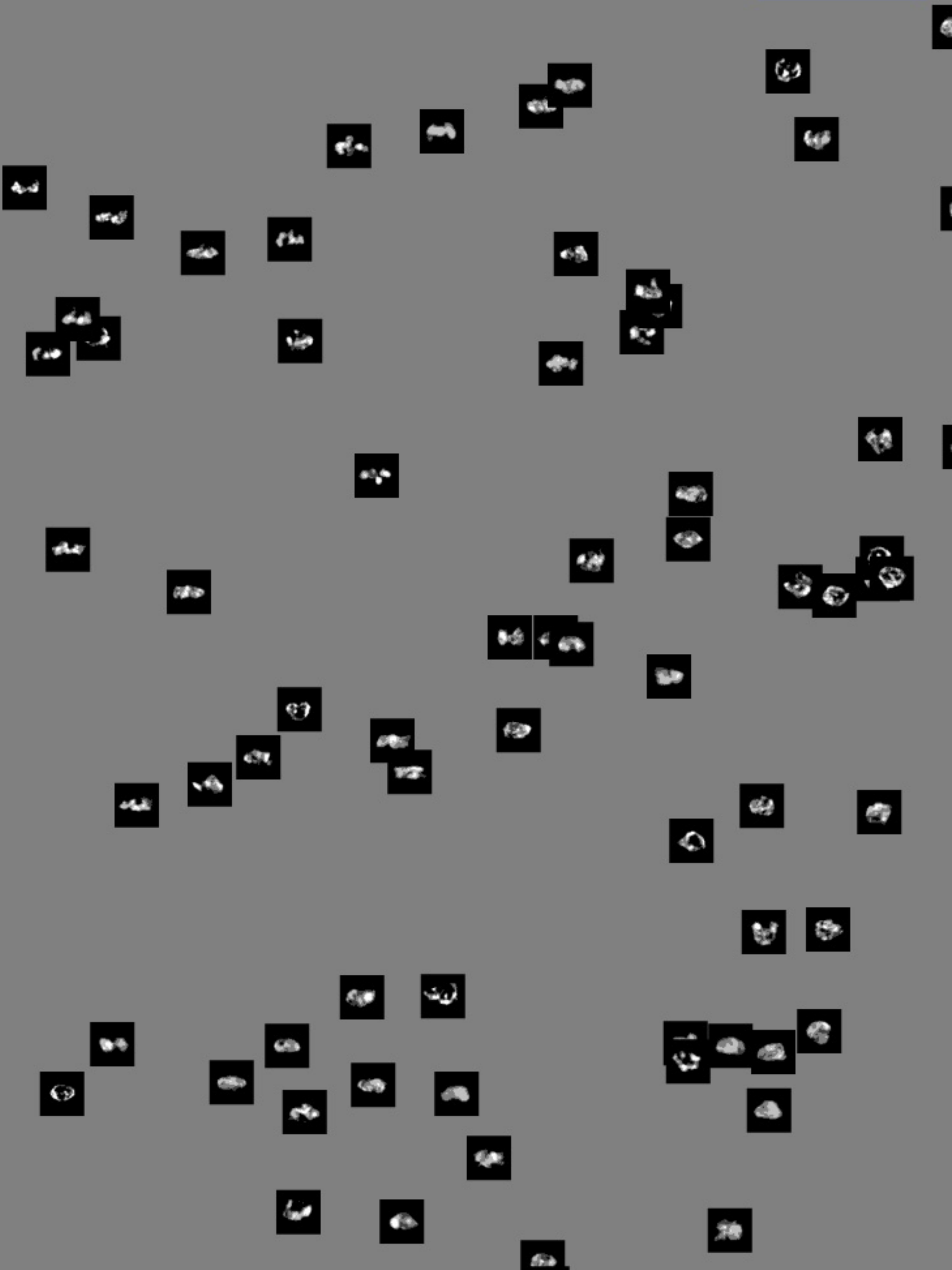


scatter01.tiff (8.3%)

10051x10051 pixels; 8-bit; 96MB

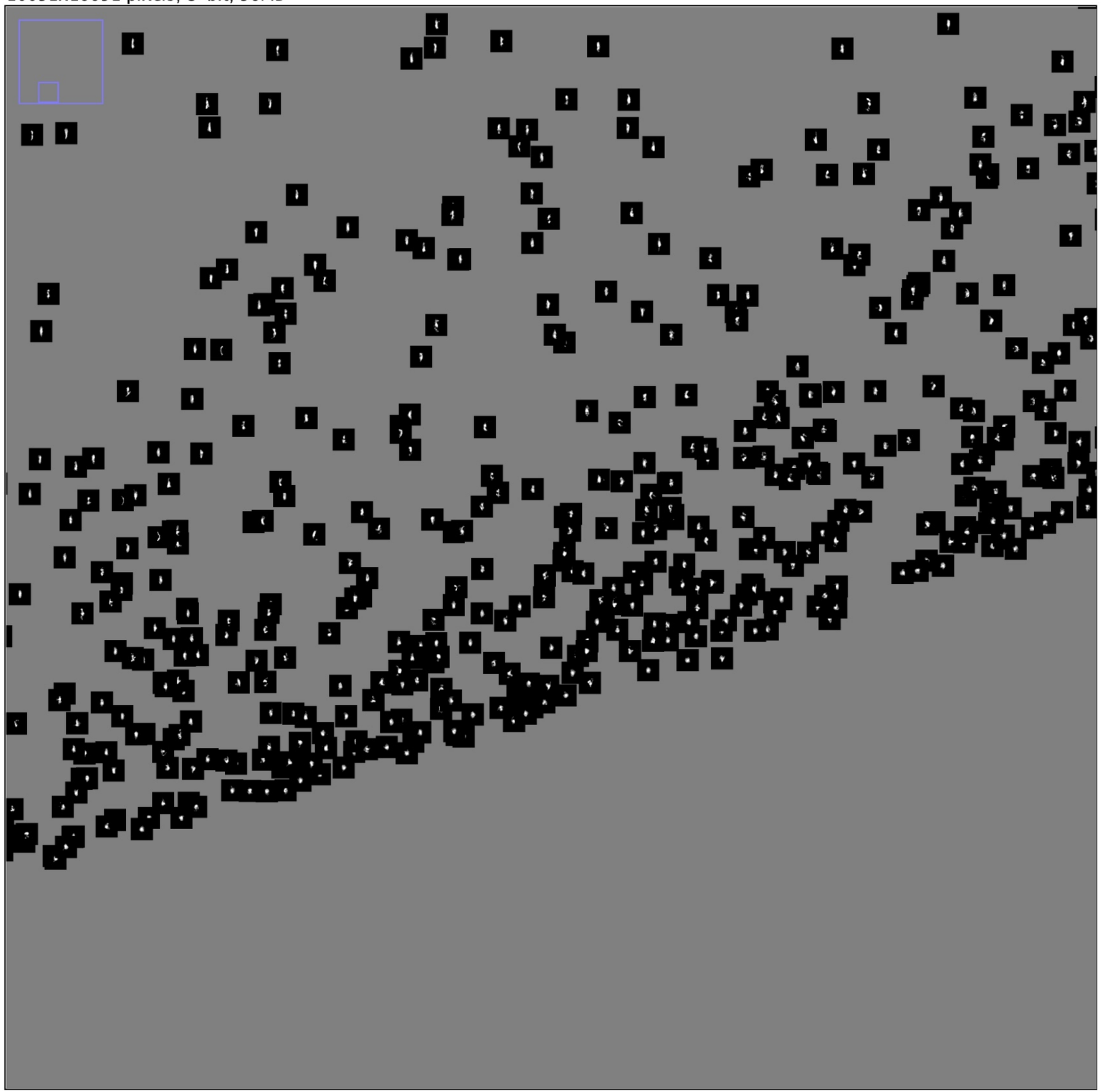


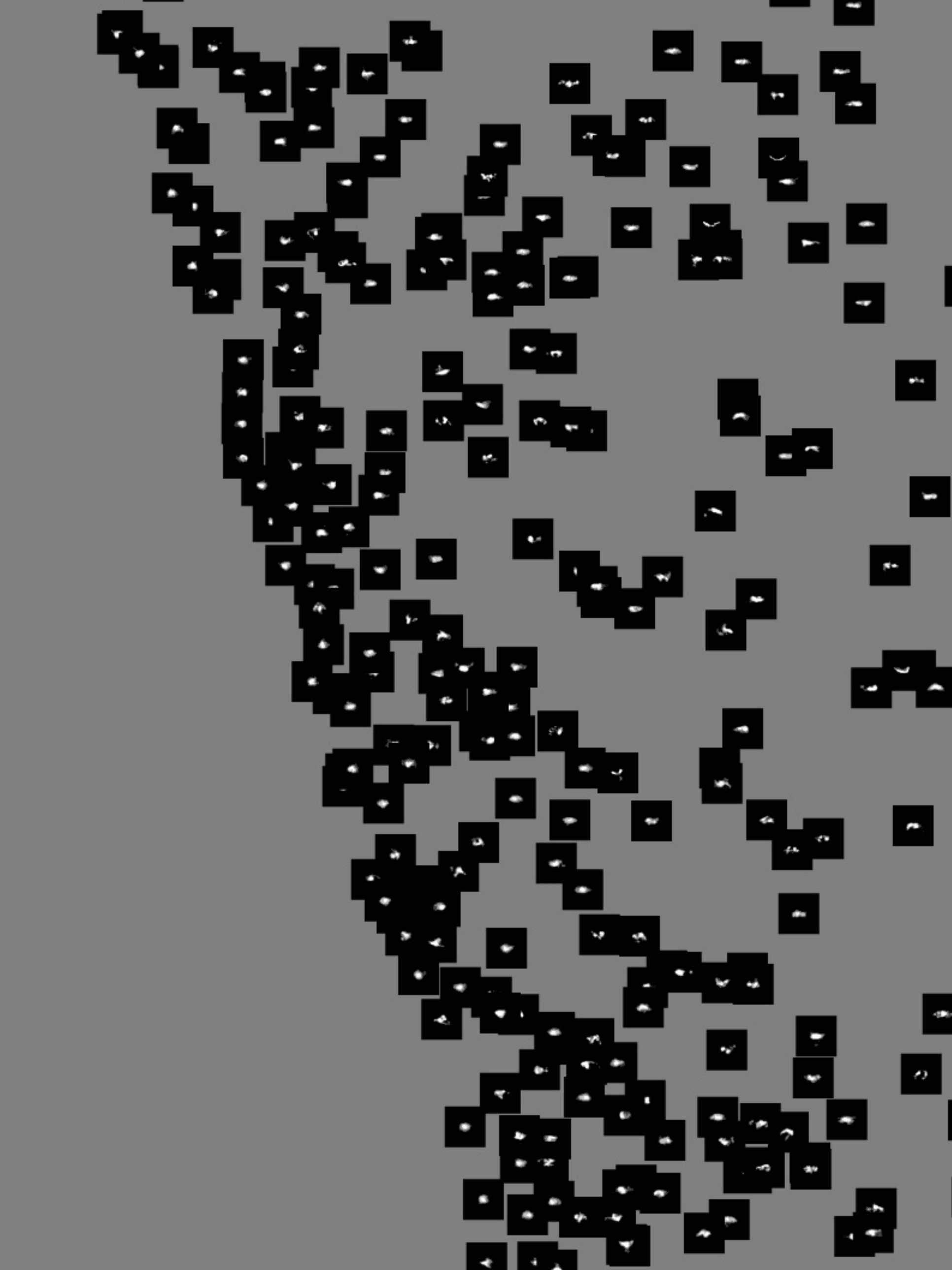




scatter01.tiff (33.3%)

10051x10051 pixels; 8-bit; 96MB





Some challenges

- Unsupervised learning from videos

Motion Capture

The Colbert Manifold

Turning mug

Turning and Swinging mug

The tools

Methods for identifying the dimension:

- Haussdorff dimension
- Doubling dimension
- epsilon-covers
- …

Methods for modeling the low dimensional structure

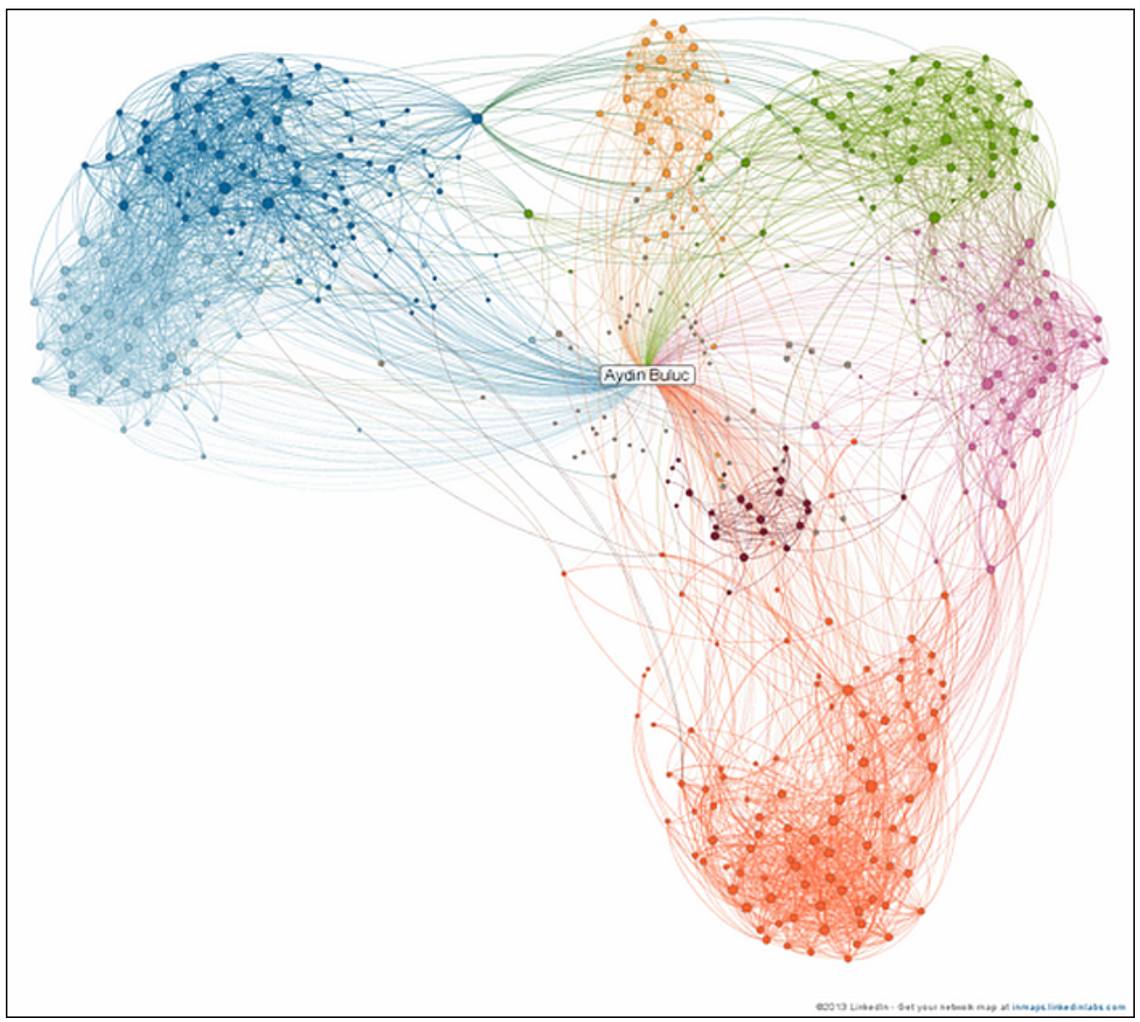
- Spectral analysis of graphs
- Differential geometry
- …

spectral graph theory

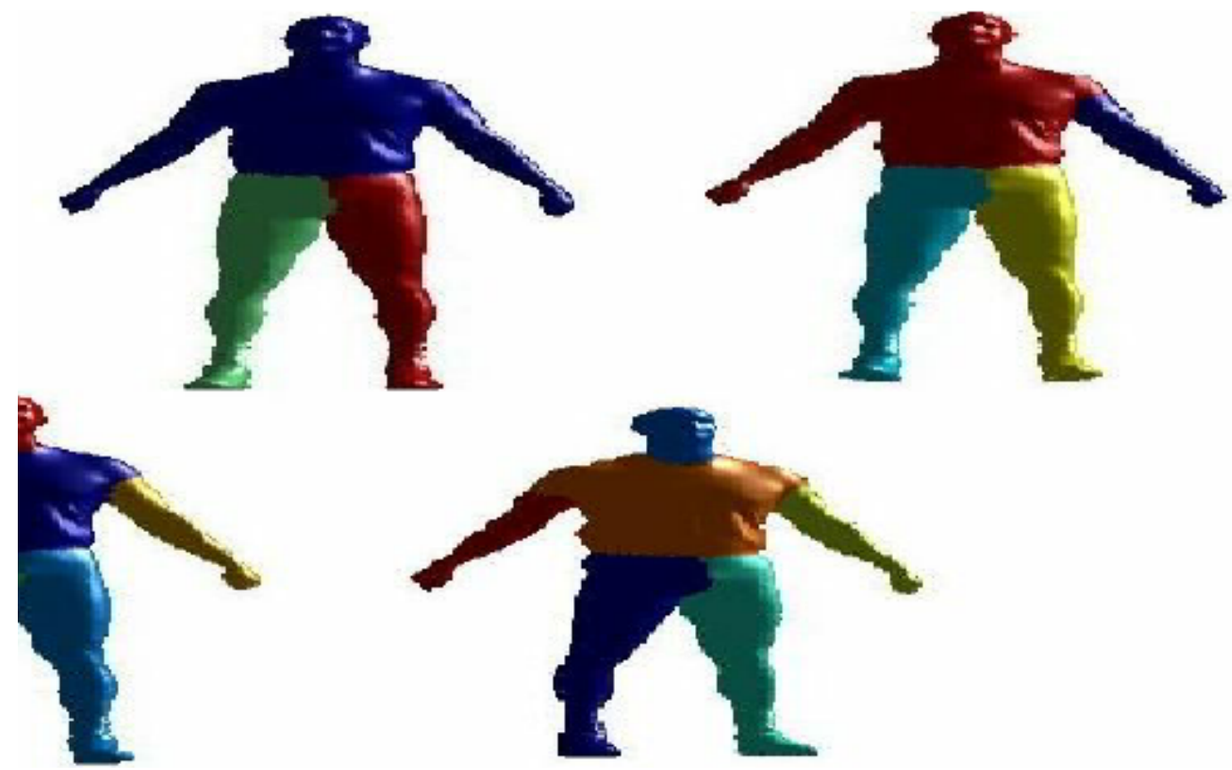
- data as an undirected graph: only small distances can be trusted.
- The adjacency matrix
- The Laplacian
- Random walk and the meaning of the eigen-vectors.
- The eigen-vectors of the laplacian.
- dimensionality reduction

Spectral clustering

Using eigen-vectors to find small cuts

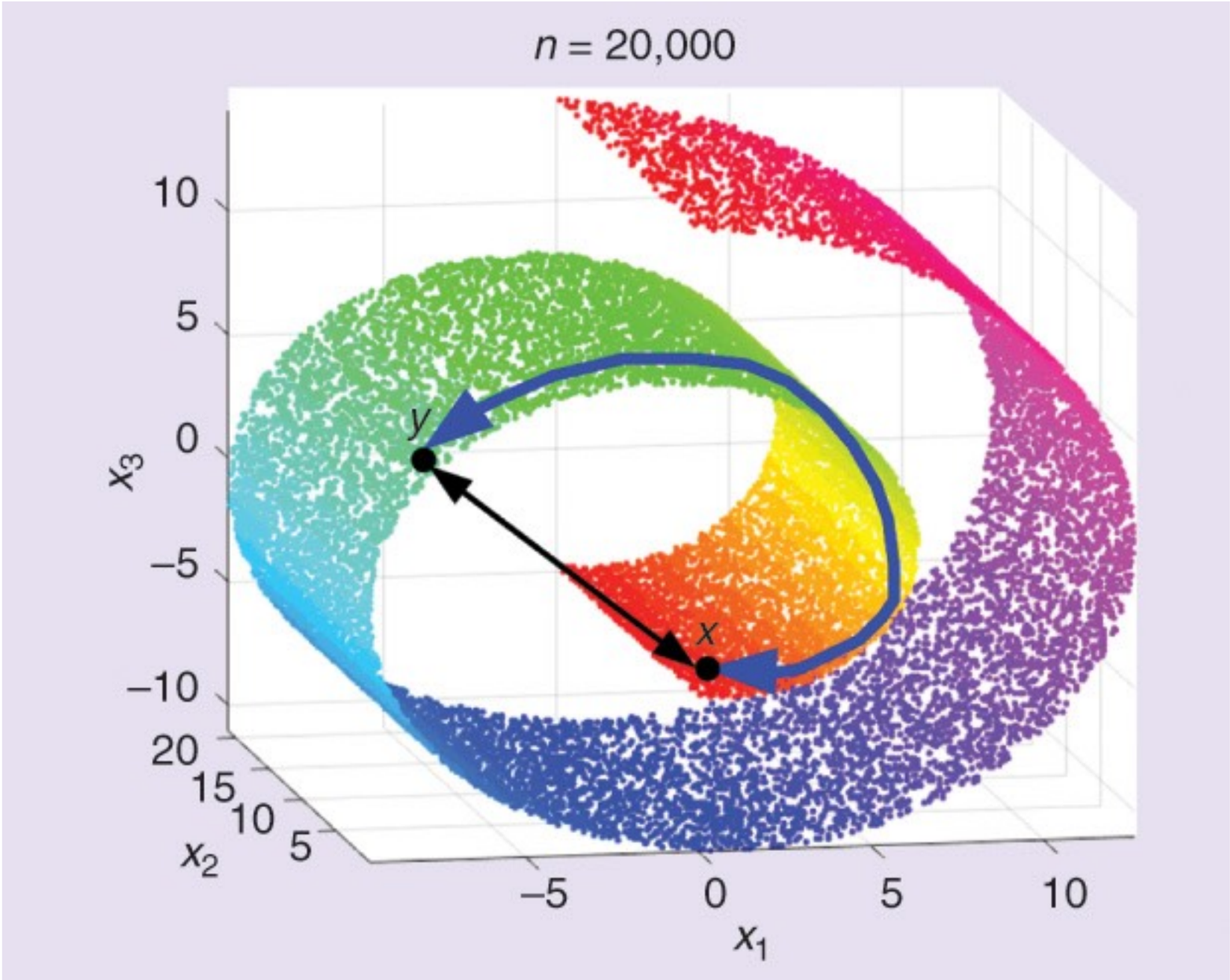


Popular method for image segmentation



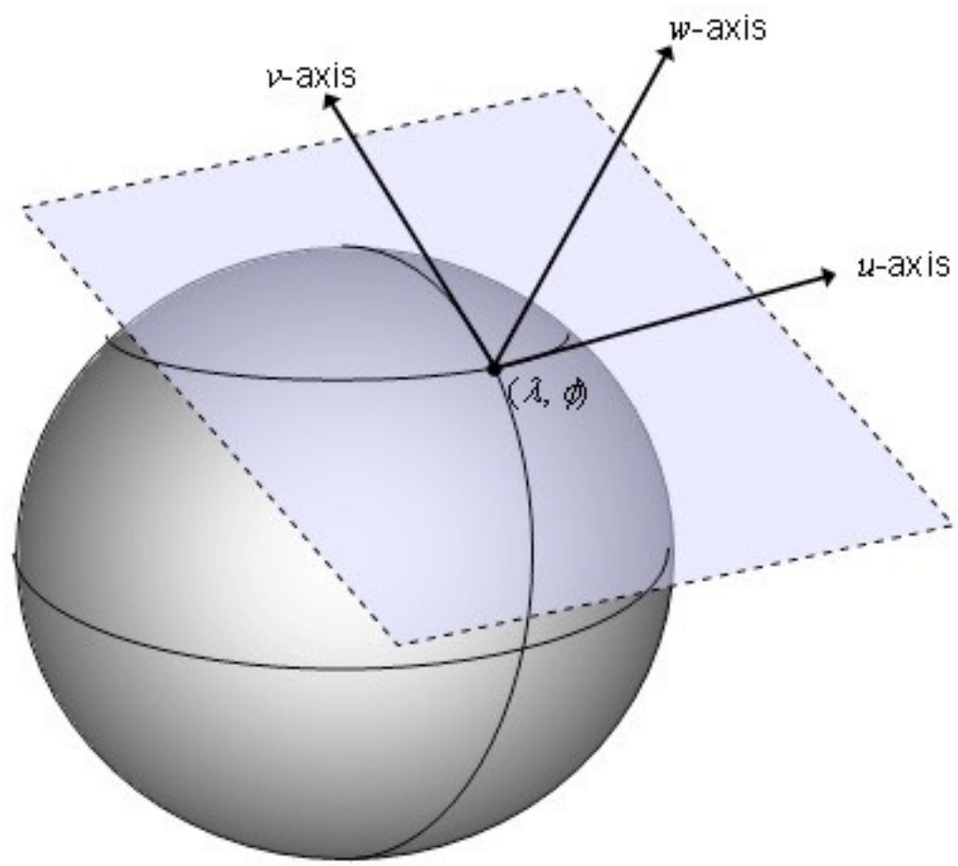
Diffusion maps

- Using the eigen-vectors to define coord. system



Differential geometry

- Differentiable manifold: a set of points in Rn which is locally homomorphic to Rd.



Differential Geometry vs. Graphs

- Discrete
- Edges
- Matrices
- Laplace operator
- Eigenvectors
- Computational

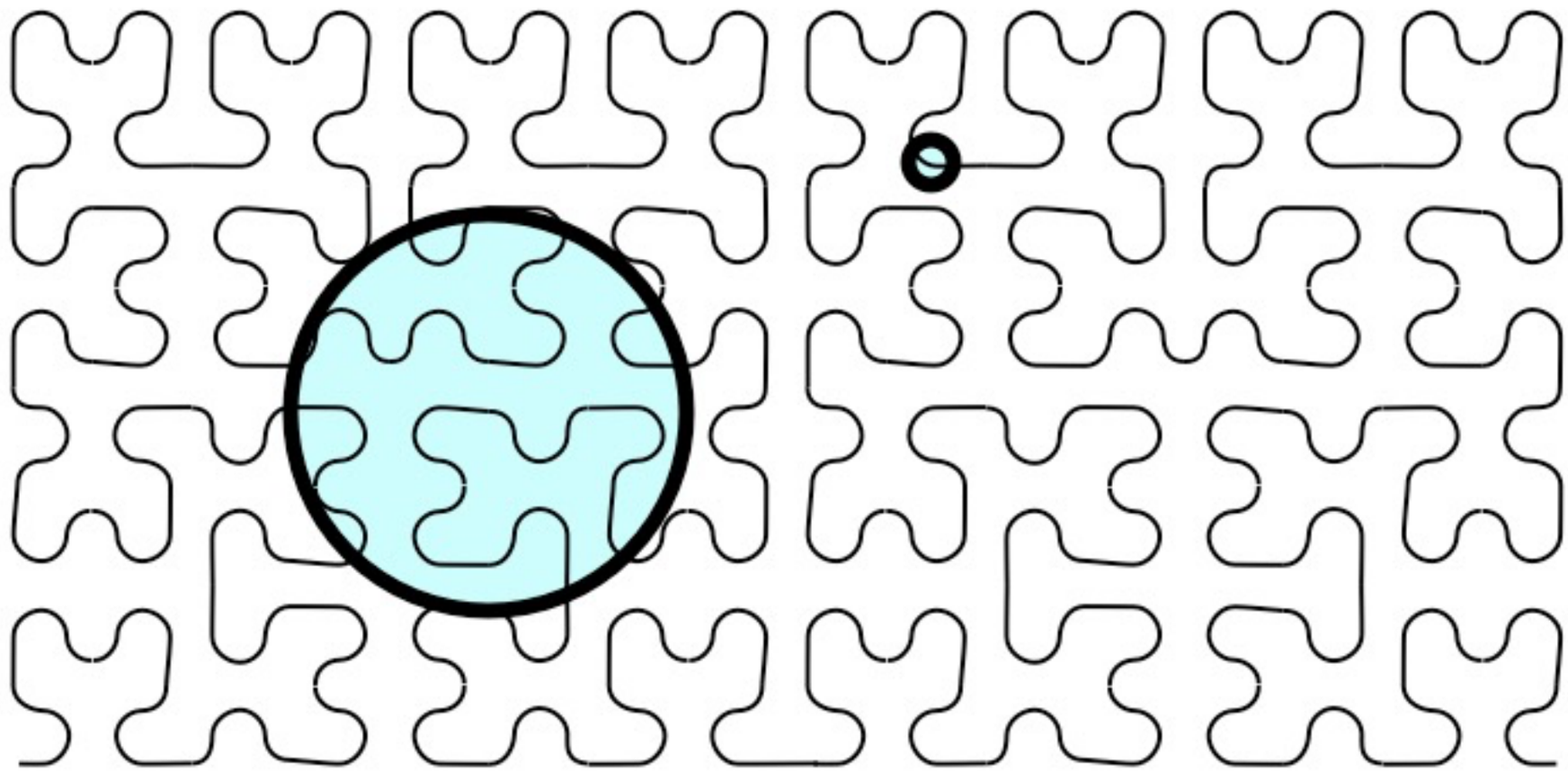
- Continuous
- Neigborhoods
- Differential operators
- Laplace Beltrami
- Analytical

Next class

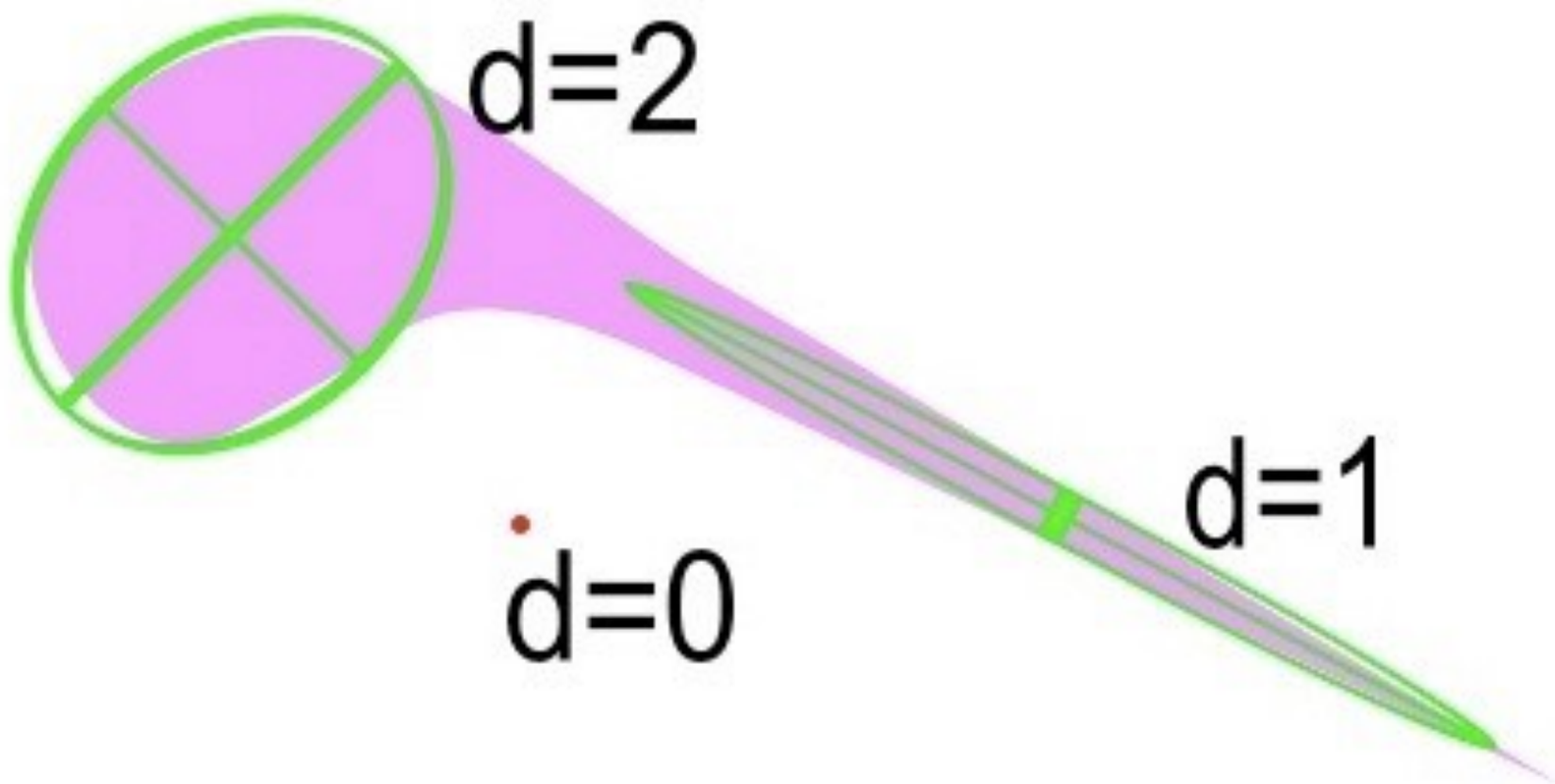
- Next Tuesday, no class Thursday
- Spielman's notes 1,2

End

Dimension can depend on scale



Dimension can depend on location



Doubling dimension

- Doubling dimension of set S is d if:
  - For any ball B of radius r
  - Intersection of set S and B can be covered by at most 2^d balls of radius r/2.
- Global, all scales, does not require smoothness.
- More general than manifold dimension.
- No clear connection to physics.

Local covariance dimension

- $S\{x_i\}_{i=1}^N$ is a finite set in R^D (a sample).
- Mean vector: $\mu = \frac{1}{N}\sum_{i=1}^N x_i$. Assume wlog $\mu = 0$
- Covariance matrix: $C = \frac{1}{N}\sum_{i=1}^N x_i^T x_i$
- $\{v_i\}_{i=1}^D$ are eigen-vectors of C with eigen-values $\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_D^2$
- S has covariance dimension (d, ϵ) if

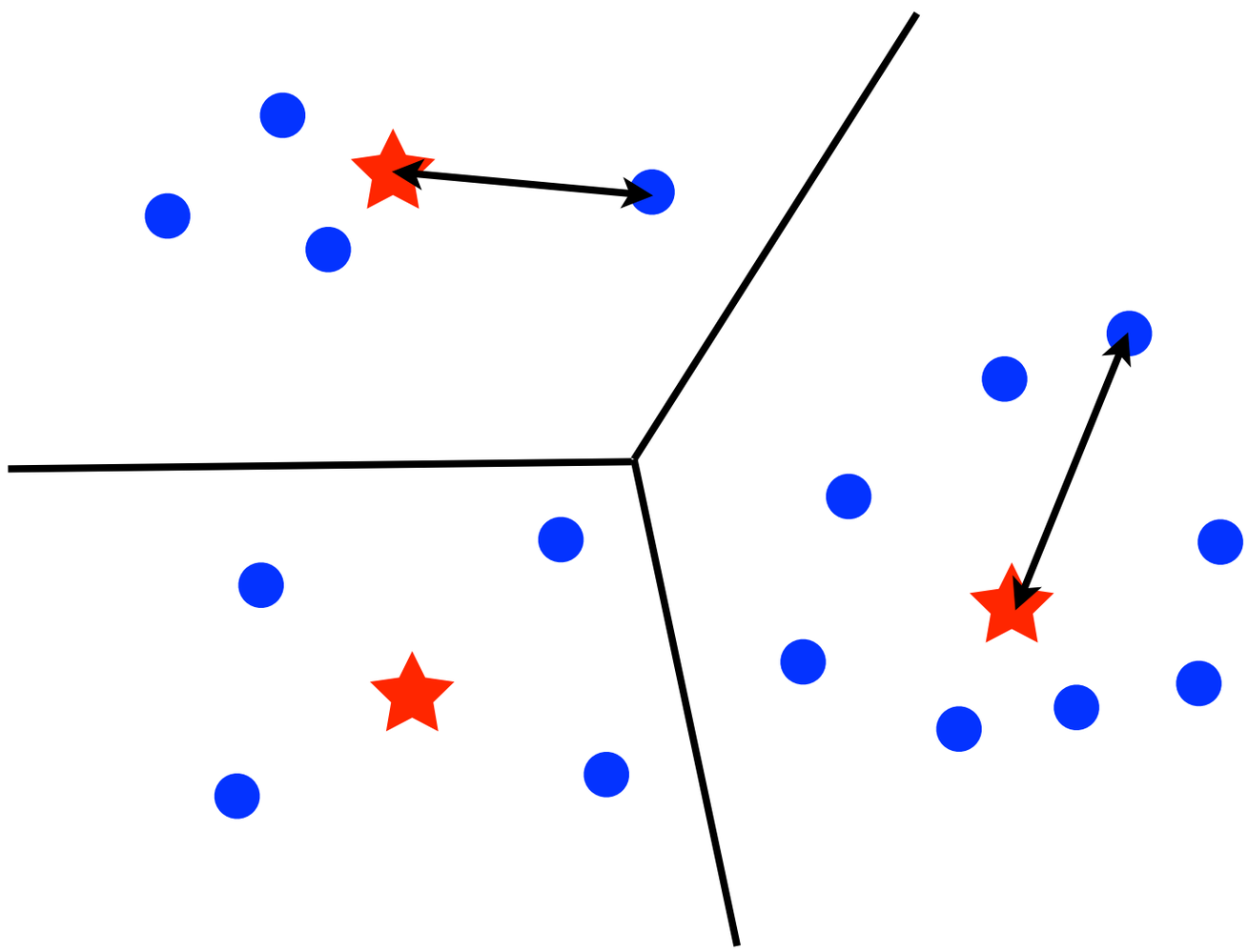
$$\sum_{i=1}^{d} \sigma_i^2 \geq (1-\epsilon)\sum_{i=1}^{D} \sigma_i^2$$

- S has local covariance dimension (d, ϵ) in the ball $B(x, r)$ if $S \cap B(x, r)$ has covariance dimension (d, ϵ).

Plan of talk

- Low dimensional representations for high dimensional data.
- Notions of intrinsic dimension.
- Vector Quantization using the RP-tree algorithm.
- Fast online PCA
- Charting the manifold
- Using manifolds for system calibration
- Future directions.

The vector quantization problem



The vector quantization problem

- **Given:**

$$\{x_i\}_{i=1}^{N}, x_i \in R^D$$

- **Find:**

$$\{r_i\}_{i=1}^{K}, r_i \in R^D$$

- **To minimize:**

$$(1/N) \sum_{i=1}^{N} \min_j \|x_i - r_j\|_2^2$$

The K-means algorithm

- **Initialize:** Pick $\{r_j\}_{j=1}^{K}$ at random from $\{x_i\}_{i=1}^{N}$
- **Repeat until convergence:**
  1. Assign each x_i to the closest r_j.
  2. Set each r_j to the average of the x_i's assigned to it.

Guaranteed to converge,
but not neccessarily to global minimum.

Computational complexity of K-means

- The **2-means** problem is **NP-complete**.
- **Decision problem**: Given $\{x_i\}_{i=1}^N$ and $T > 0$, is there a partitioning of x_i's into two sets S_1, S_2 and two vectors r_1, r_2 such that

$$\sum_{i \in S_1} \|x_i - r_1\|_2^2 + \sum_{i \in S_2} \|x_i - r_2\|_2^2 \leq NT$$

- **Caveat:** The reduction requires $D = 2N$. Not true for constant D because number of separating hyperplanes is polynomial.

VQ with square euclidean distance

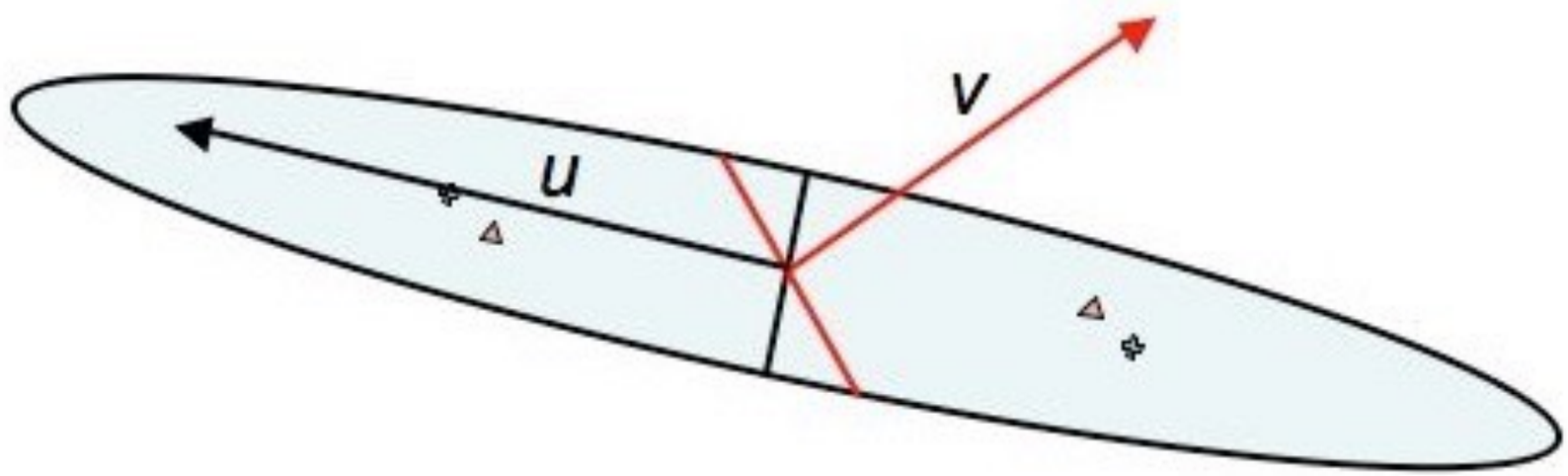
- $r = \mu$: The best representative vector is the mean.
- Assume wlog $\mu = 0$

$$\frac{1}{N^2}\sum_{i,j}\|x_i - x_j\|_2^2 = \frac{1}{N^2}\sum_{i,j}\|x_i\|_2^2 + 2x_i \cdot x_j + \|x_j\|_2^2 = \frac{2}{N}\sum_i \|x_i\|_2^2$$

- Partition $1, \ldots, N$ into disjoint sets S_1, S_2 with means μ_1, μ_2:

$$\frac{1}{N}\left(\sum_{i=1}^{N}\|x_i - \mu\|_2^2 - \sum_{i \in S_1}\|x_i - \mu_1\|_2^2 - \sum_{i \in S_2}\|x_i - \mu_2\|_2^2\right) = \|\mu_1 - \mu_2\|_2^2$$

Splitting a set with low covariance dimension



- “optimal” split - orthogonal to largest eigen-vector.
- Split on random direction - almost optimal with constant probability.

Random projection trees

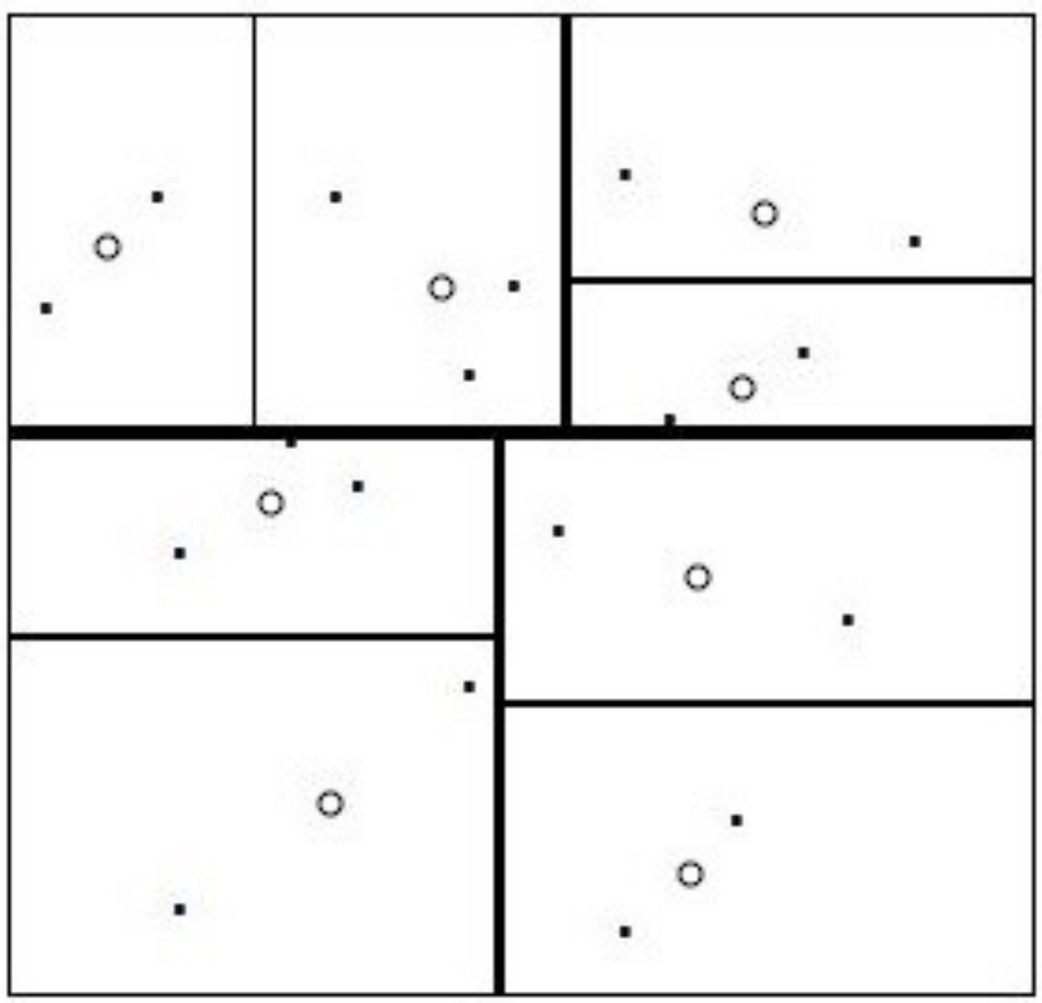
- Apply recursively:
  - project on random direction
  - Split at median
- Similar to KD-trees, but using random projections instead of coordinates.

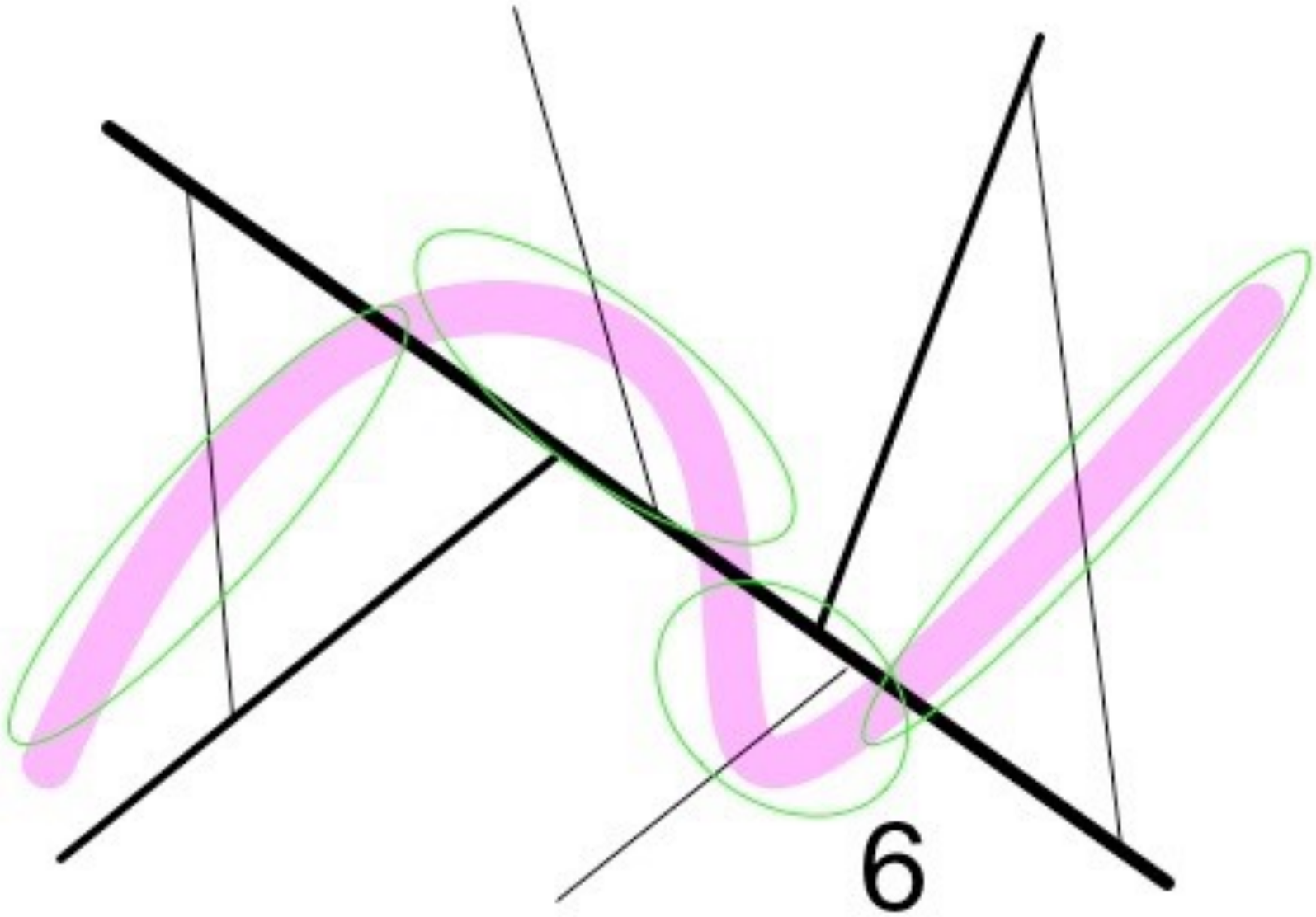
theoretical properties of RP-trees.

Dasgupta & Freund, STOC08

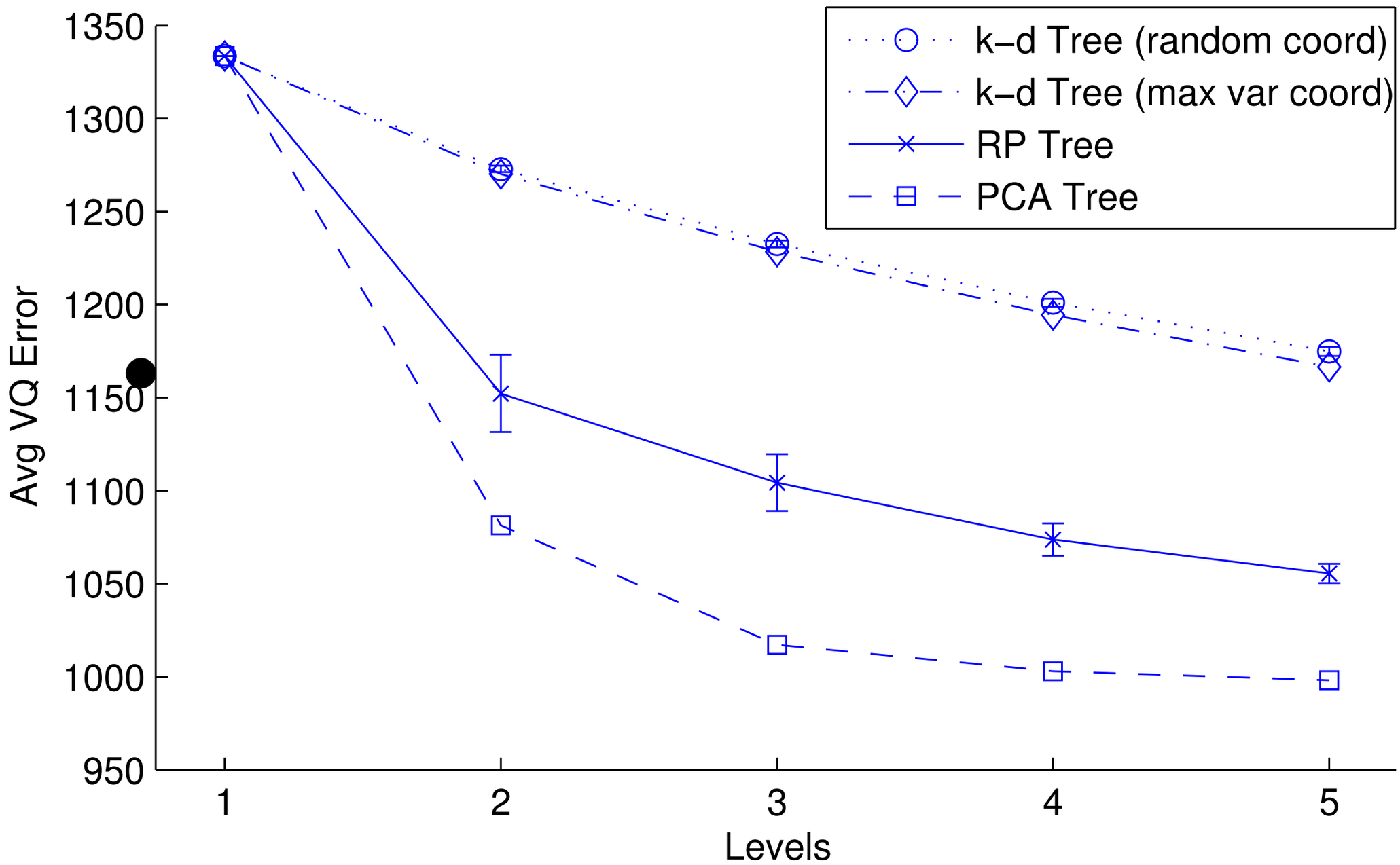
- Space: R^D
- Measure of progress: average cell diameter
- Tree-structured VQ: average diameter halved every D tree levels
- Data of intrinsic dimension d<<D
- RP-tree: average diameter halved every d tree levels (with constant probability)

KD-trees vs RP-trees




6

KD-tree vs. RP-tree performance



Plan of talk

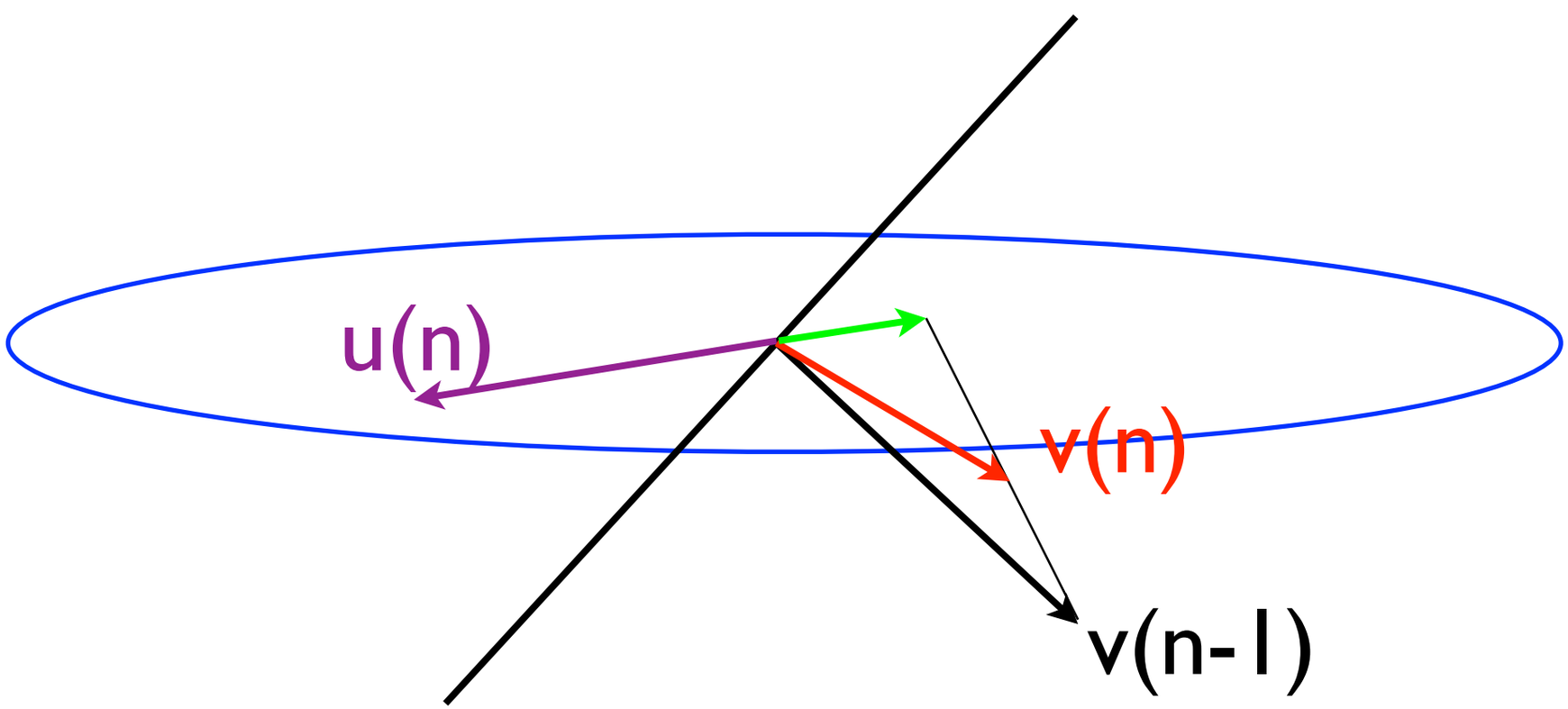
- Low dimensional representations for high dimensional data.
- Notions of intrinsic dimension.
- Vector Quantization using the RP-tree algorithm.
- Fast online PCA
- Charting the manifold
- Using manifolds for system calibration
- Future directions.

Candid Covariance-Free Incremental Principal Component Analysis

Juyang Weng, *Member*, *IEEE*,
Yilu Zhang, *Student Member*, *IEEE*, and
Wey-Shiuan Hwang, *Member*, *IEEE*

$$v(n) = \frac{n-1}{n} v(n-1) + \frac{1}{n} u(n) u^T(n) \frac{v(n-1)}{||v(n-1)||}$$

Scalar



Online PCA for low dimensional data

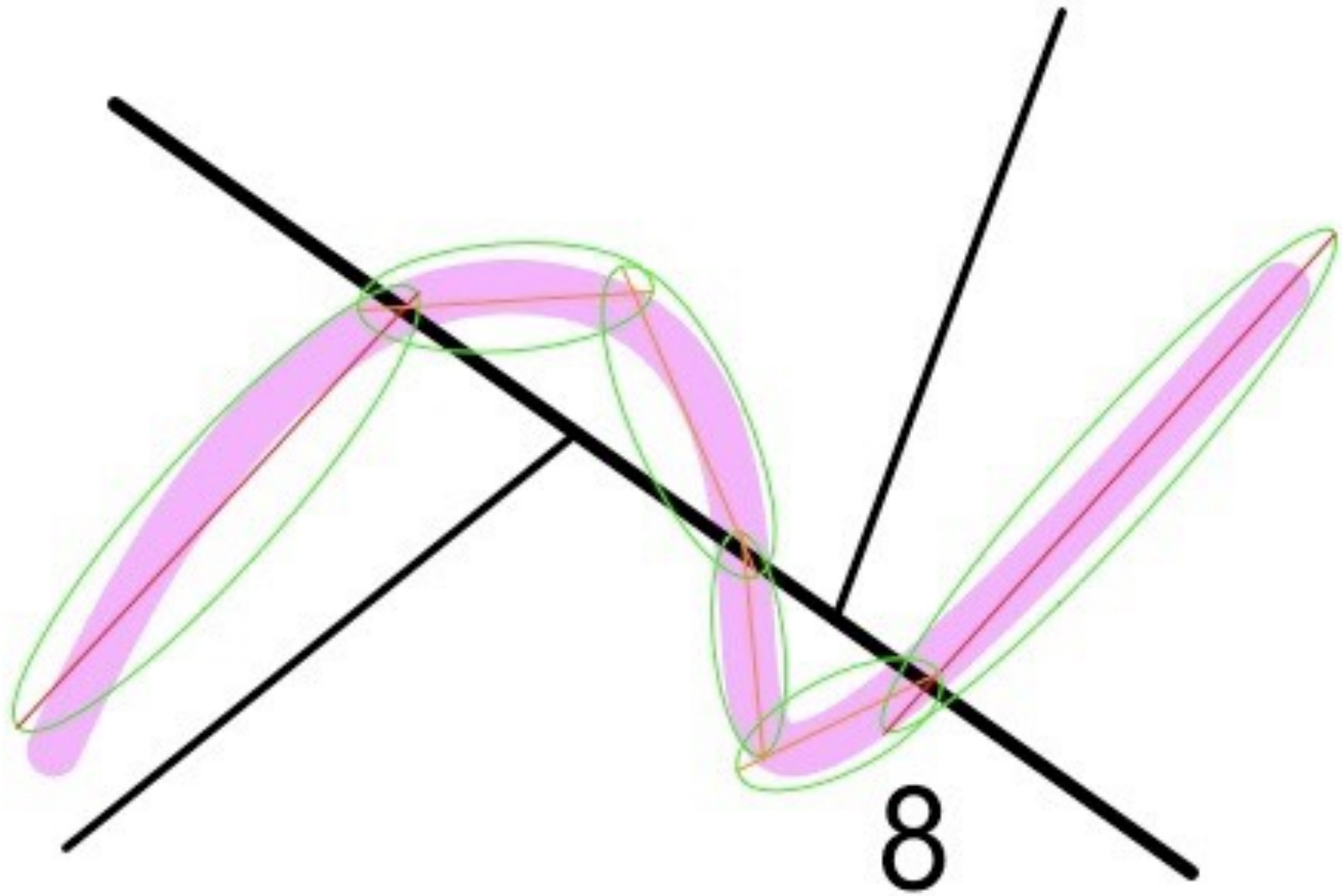
- D dimensional data
- Covariance matrix has DxD entries
- Accurate estimation of all entries requires O(D) samples
- If d<<D eigenvectors dominate, then online PCA converges after O(d) samples.

Plan of talk

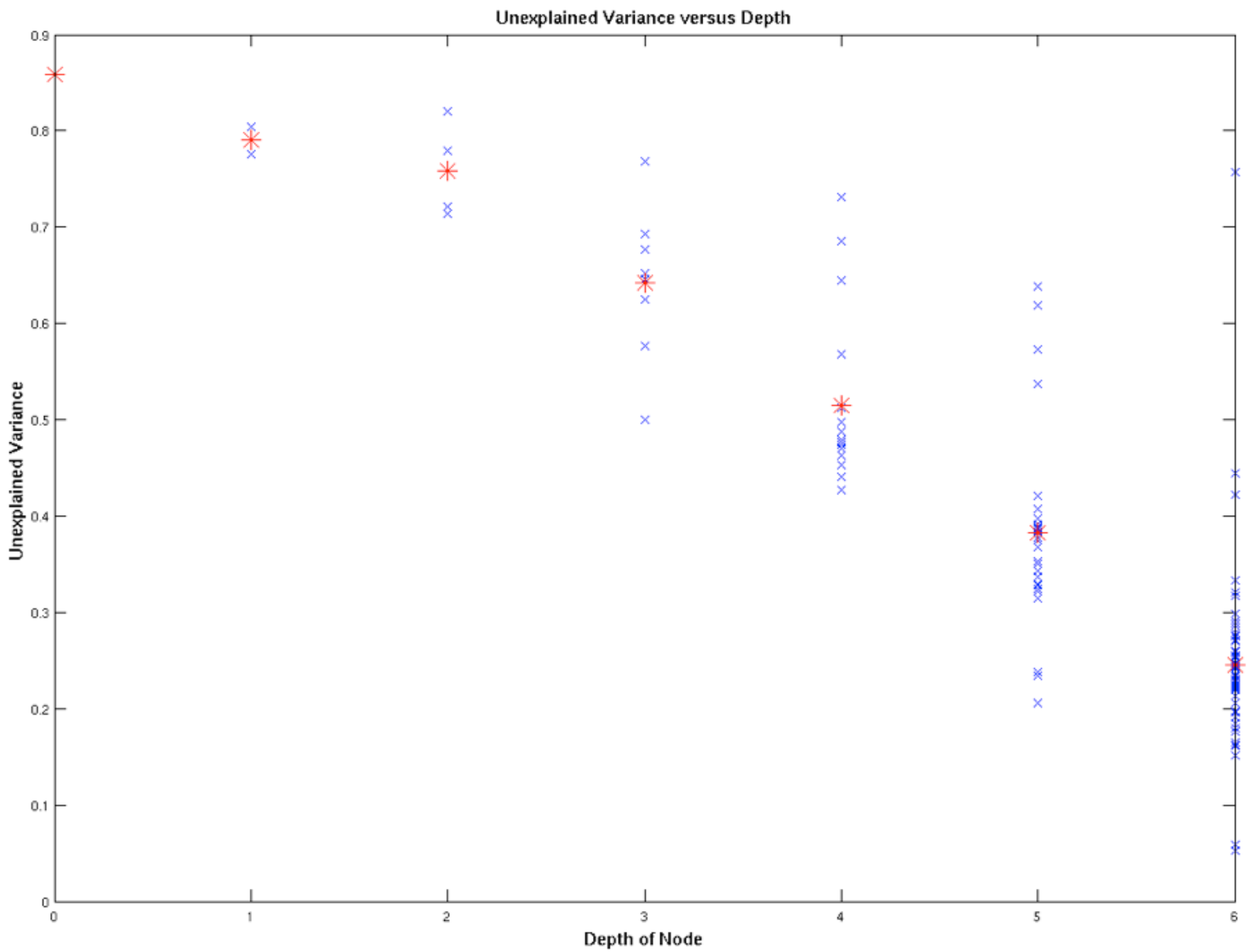
- Low dimensional representations for high dimensional data.
- Notions of intrinsic dimension.
- Vector Quantization using the RP-tree algorithm.
- Fast online PCA
- Charting the manifold
- Using manifolds for system calibration
- Future directions.

The charting problem

- Goal: A differentiable mapping of D dimensional input to R^d
- Identify the sufficiently linear pieces. (percent variance explained)
- Glue the pieces consistently.


8

Unexplained variance vs. tree depth



Plan of talk

- Low dimensional representations for high dimensional data.
- Notions of intrinsic dimension.
- Vector Quantization using the RP-tree algorithm.
- Fast online PCA
- Charting the manifold
- Using manifolds for system calibration
- Future directions.

Plan of talk

- Low dimensional representations for high dimensional data.
- Notions of intrinsic dimension.
- Vector Quantization using the RP-tree algorithm.
- Fast online PCA
- Charting the manifold
- Using manifolds for system calibration
- Future directions.

Future direction
learning piecewise-linear control

Tedrake et al. "Learning to walk in 20 minutes" Science 2004