Common Manifold Learning using Alternating-Diffusion for Multimodal Signal Processing

Roy R. Lederman[1], Ronen Talmon[2] and Ronald R. Coifman[1]

[1]Yale University, New Haven, CT
[2]Technion-IIT, Haifa, Israel

SIAM CSE, March 2015

Multimodal data analysis: sleep stage identification

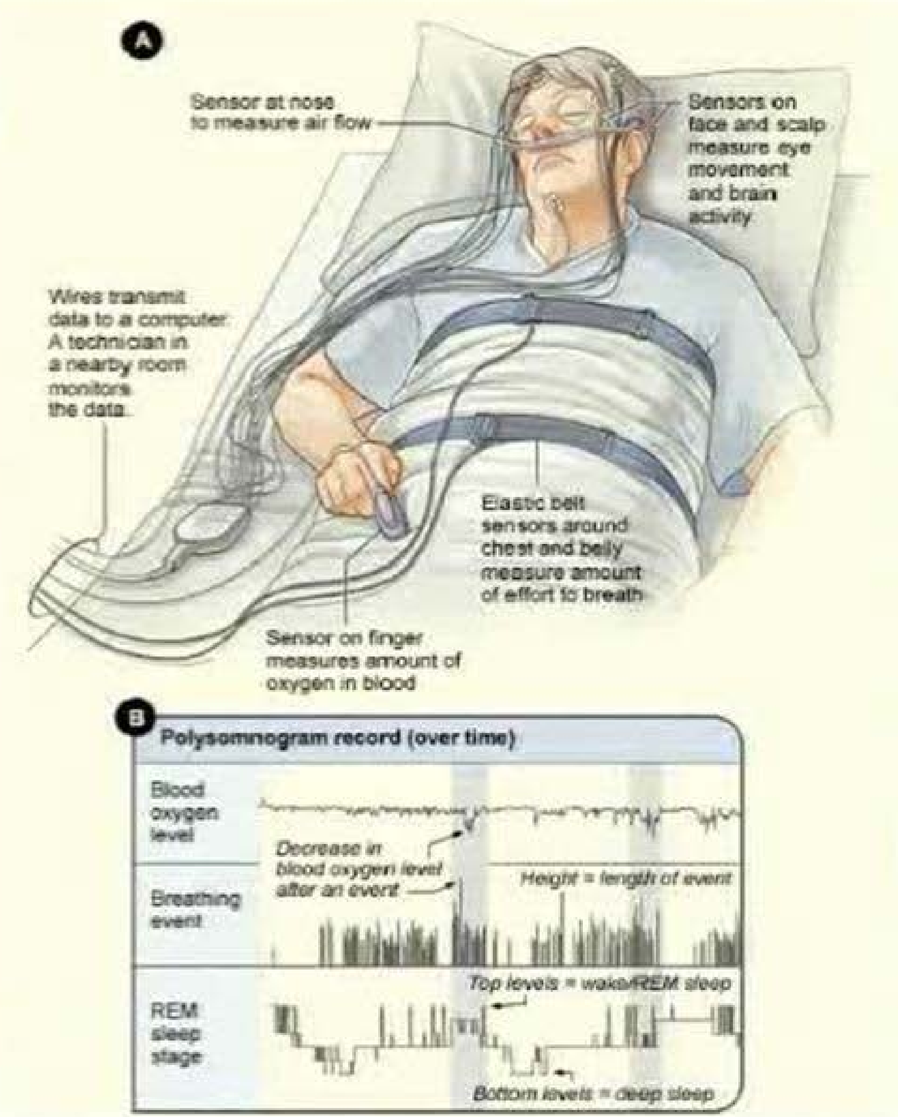

Figure: Sleep polysomnography [Source: NIH]

Problem definition

Setting:
Three *hidden* random variables, X, Y and Z,
such that given X, the variables Y and Z are independent.

$$(X, Y, Z) \sim \pi_{x,y,z}(X, Y, Z)$$

$$\pi_{x,y,z}(X, Y, Z) = \pi_x(X)\pi_{y|x}(Y|X)\pi_{z|x}(Z|X),$$

Two measured *observable* random variables: $S^{(1)} = g(X, Y)$ and $S^{(2)} = h(X, Z)$, where g and h are bilipschitz functions.

Dataset:
n pairs of measurements $\left\{(s_i^{(1)}, s_i^{(2)})\right\}_{i=1}^{n}$,
for n realizations of the hidden variables $\{(x_i, y_i, z_i)\}_{i=1}^{n}$.

Goal: Recover a parametrization of the common variable X.

A Toy Example

No pun intended

Figure: Experimental setup

Organizing data with one variable

Organizing data with one variable

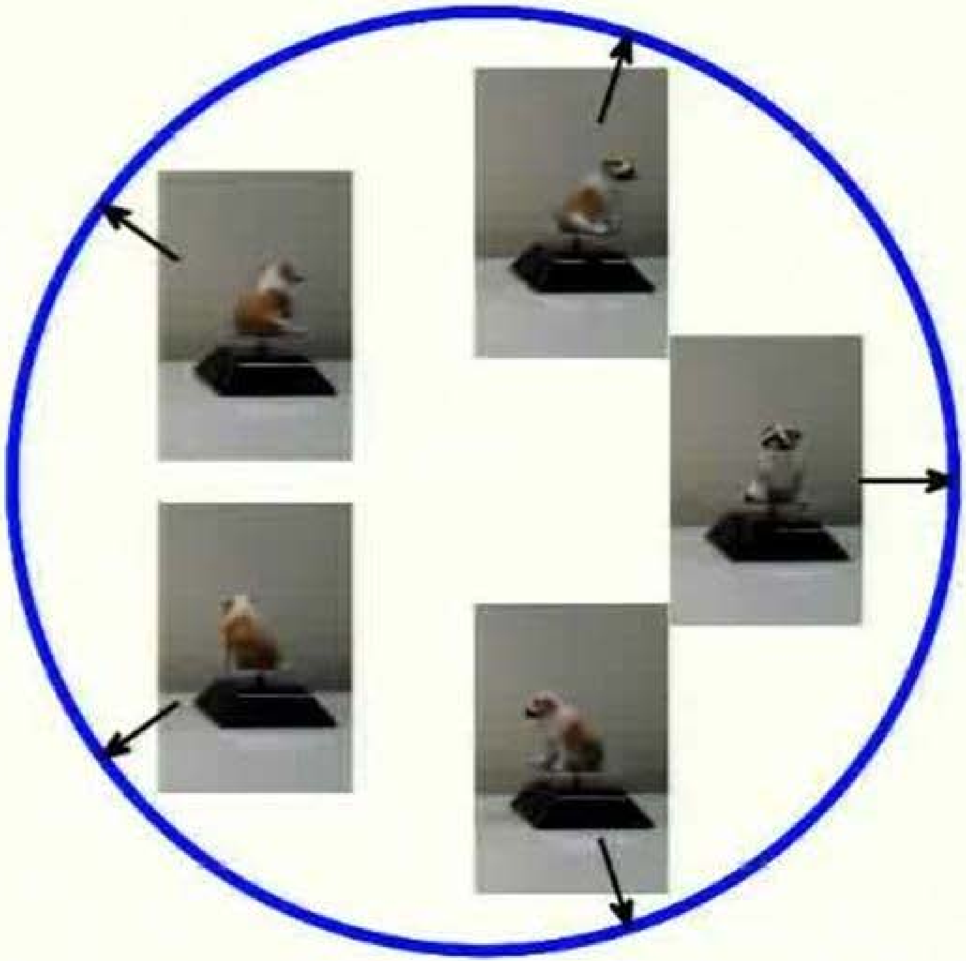

Figure: each sample (snapshot) is a point on the circle, representing the rotation angle.

Multiple variables

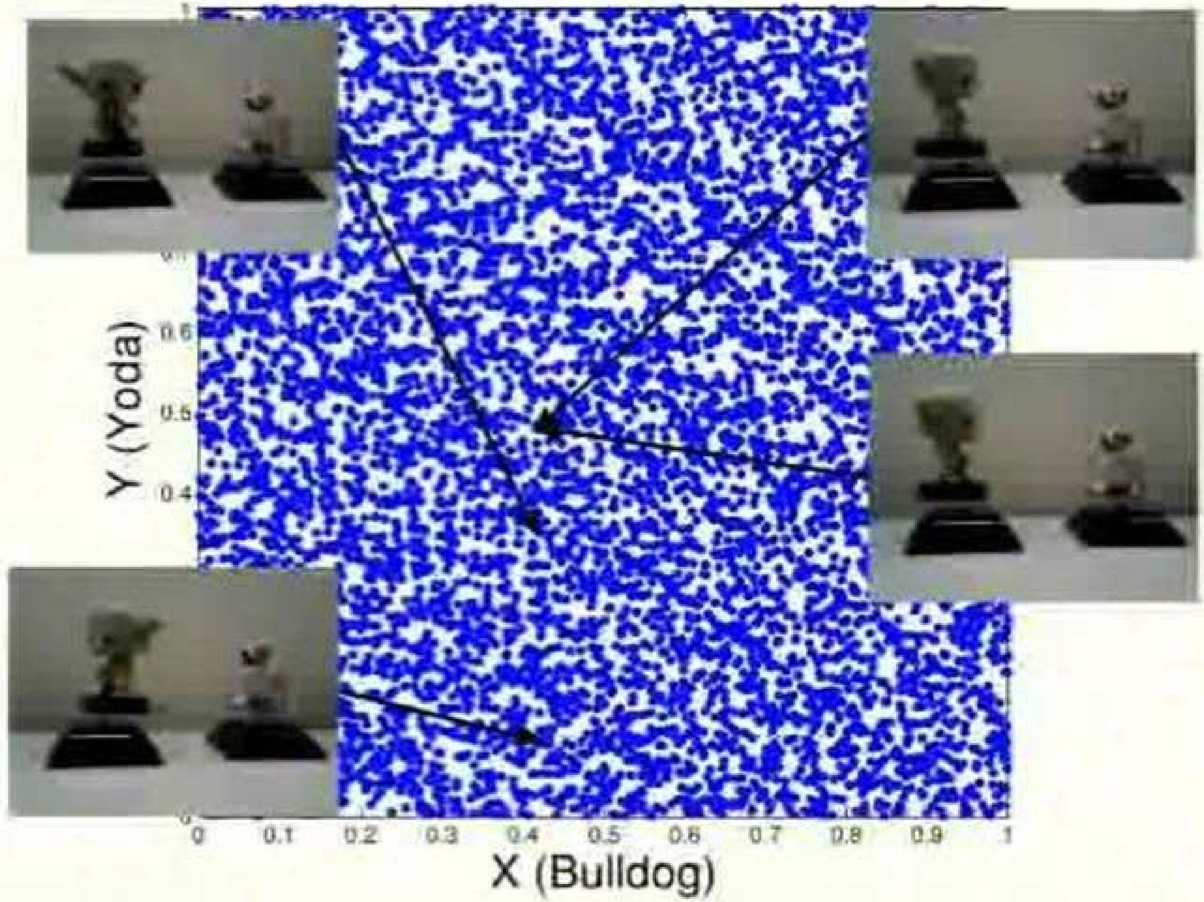

Figure: Each variable is the rotation angle of one of the objects.

Diffusion maps: multiple variables

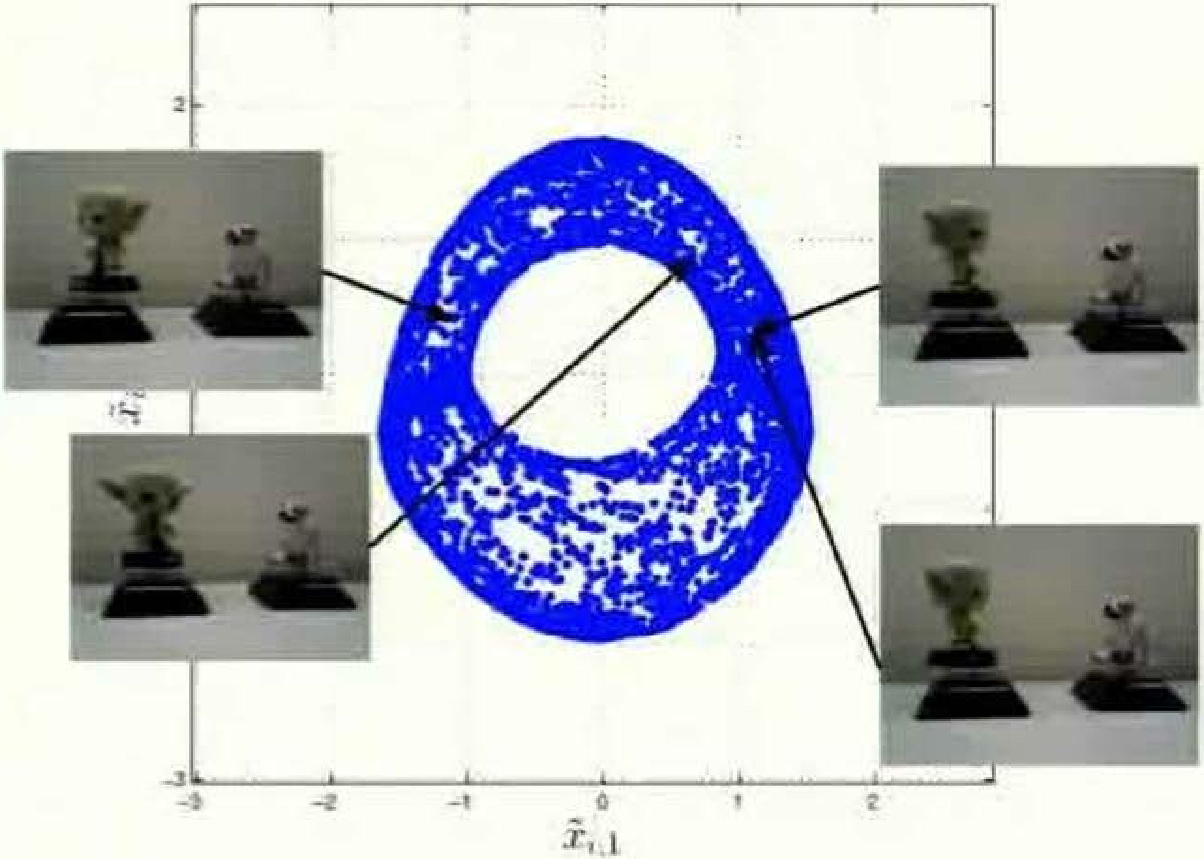

Figure: The embedding captures the two sources of variability.

Diffusion maps

Construct a random walk kernel W on the graph of samples.
Construct a sequence of probability distributions $p_{i,t} = W^t p_{i,0}$.

Figure: diffusion sequence for two variables

Diffusion maps

Construct a random walk kernel W on the graph of samples.
Construct a sequence of probability distributions $p_{i,t} = W^t p_{i,0}$.

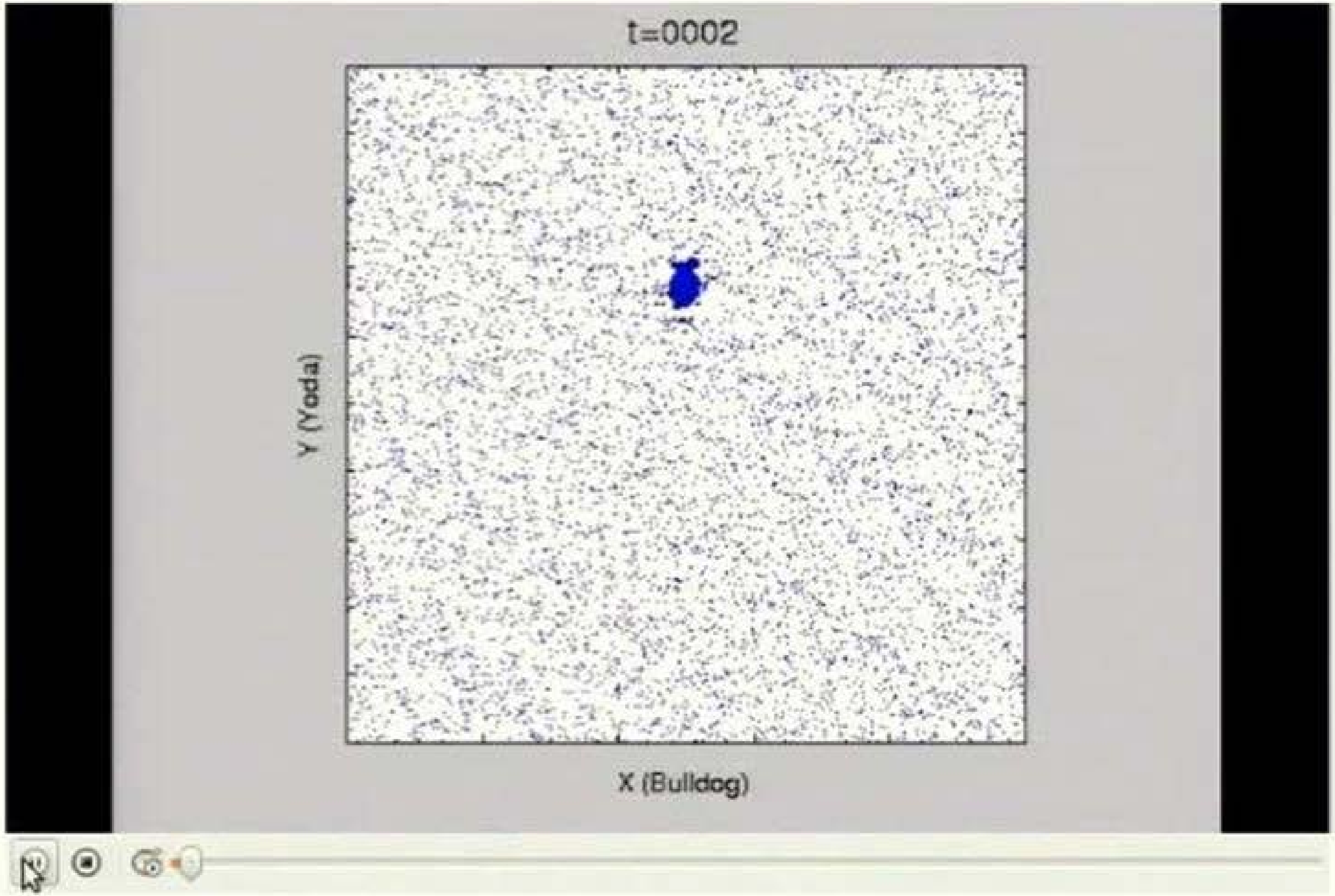

Figure: diffusion sequence for two variables

Diffusion maps: multiple variables

Compute the diffusion distance: $d_t(i, j) = \|p_{i,t} - p_{j,t}\|_M$,
and a low dimensional embedding consistent with this distance:

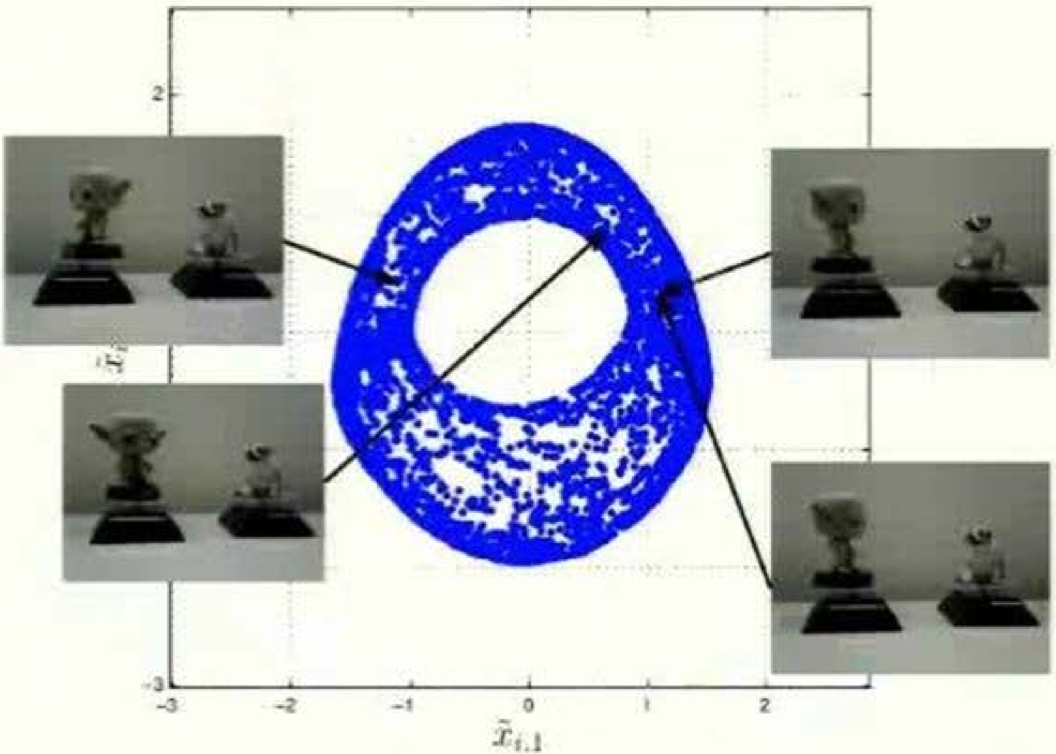

Figure: The embedding captures the two sources of variability, but does not separate the sources of variability.

Our approach : alternating diffusion

Construct the diffusion operators $K^{(1)}$ and $K^{(2)}$ for Sensor 1 and Sensor 2, respectively.
$K^{(1)}$ alone would generate the diffusion for Sensor 1,
$K^{(2)}$ alone would generate the diffusion for Sensor 2.

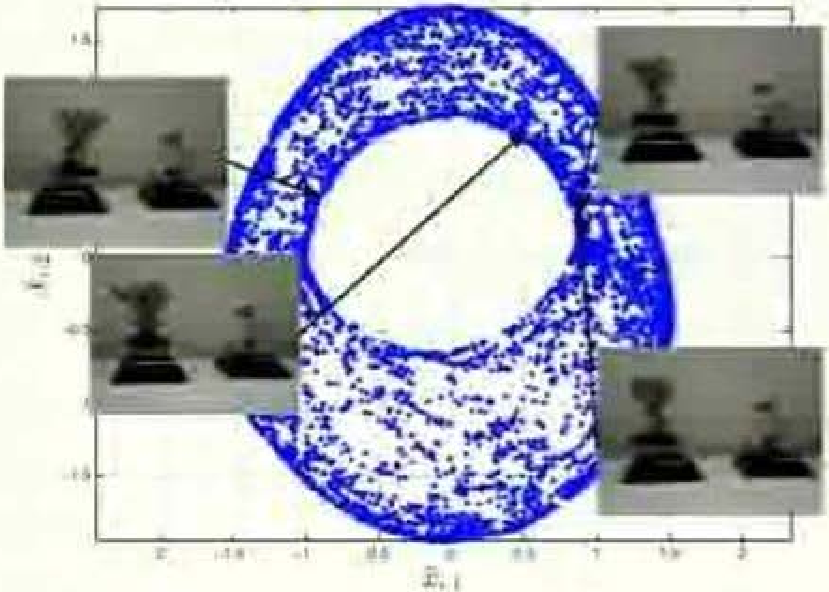

Figure: Embedding of samples from Sensor 1

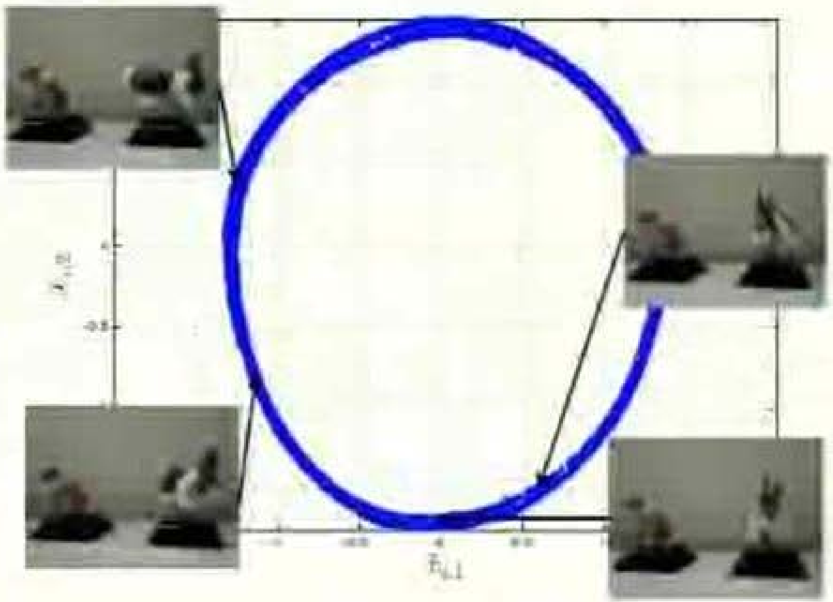

Figure: Embedding of samples from Sensor 2

Alternating diffusion

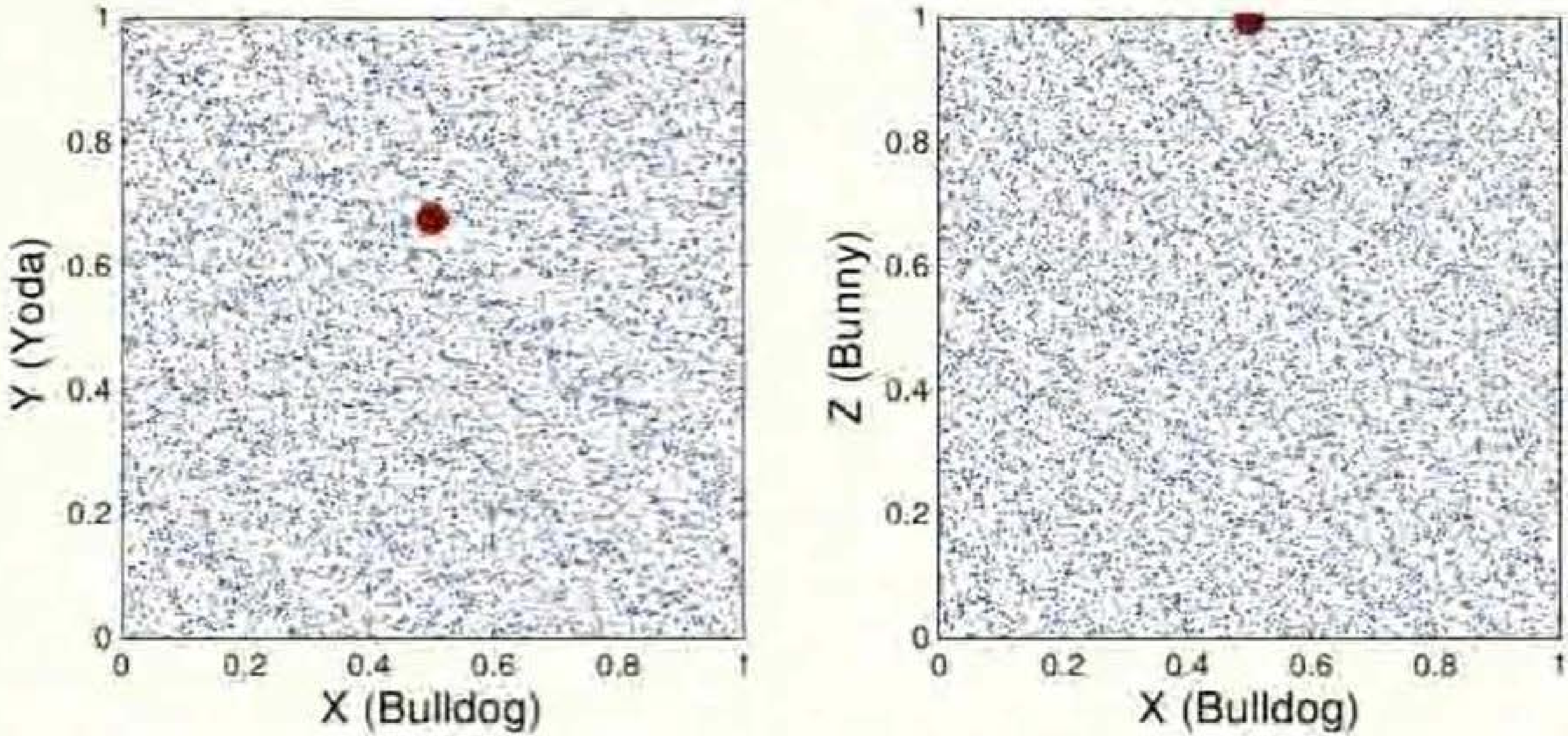

Figure: Diffusion: $p_{i,0}$

Alternating diffusion

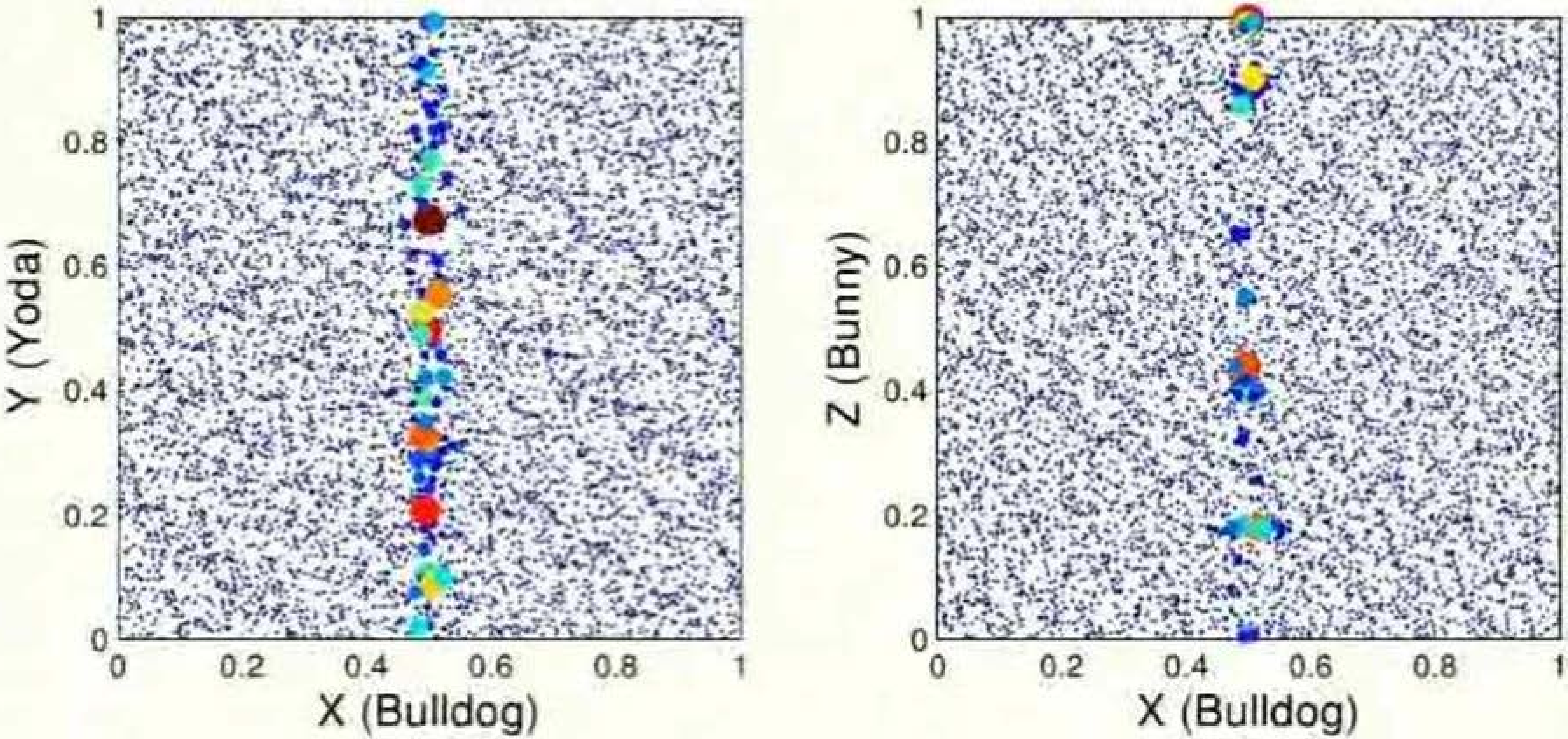

Figure: Diffusion: $p_{i,2} = K^{(2)} p_{i,1}$

Alternating diffusion

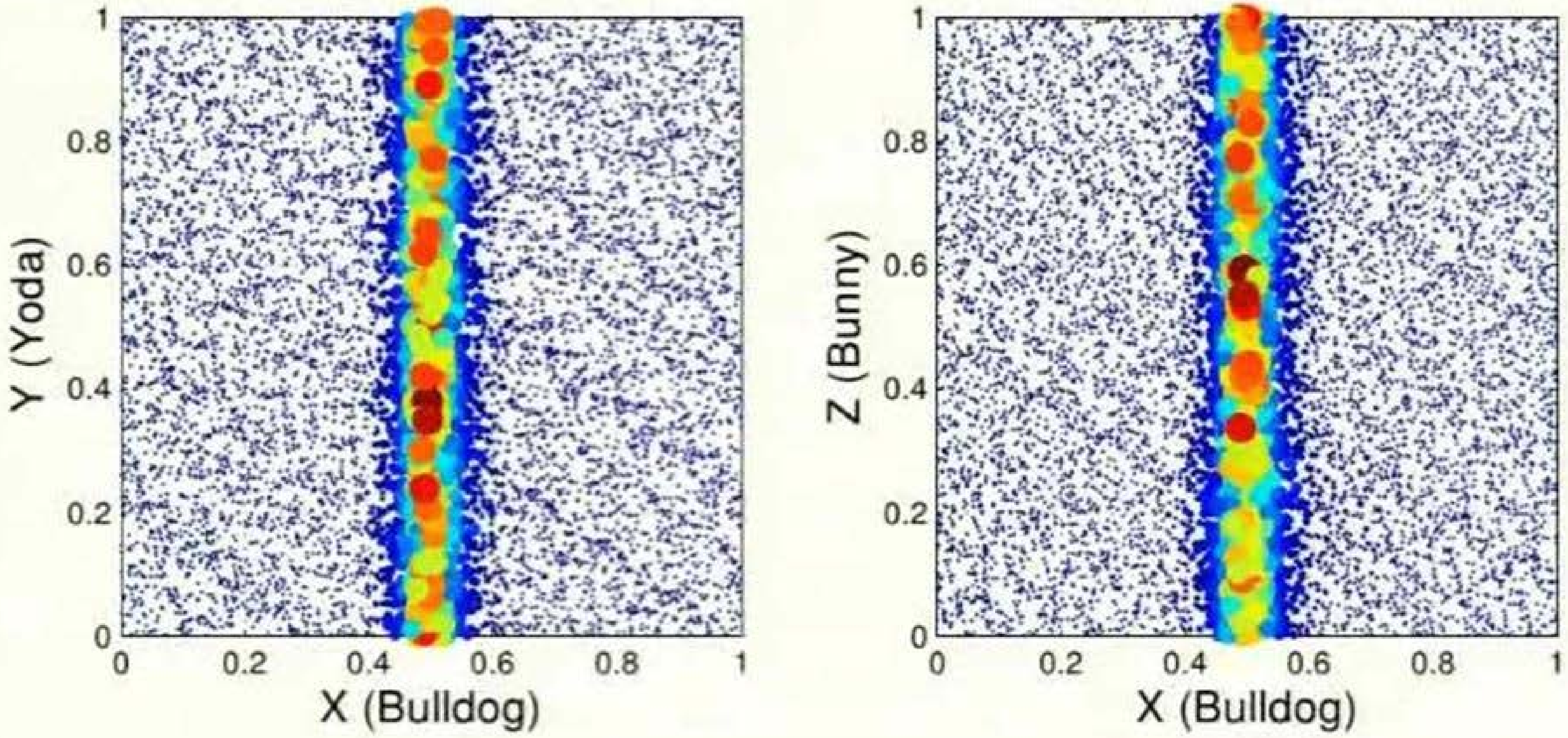

Figure: Diffusion: $p_{i,6} = K^{(2)} p_{i,5}$

Alternating diffusion

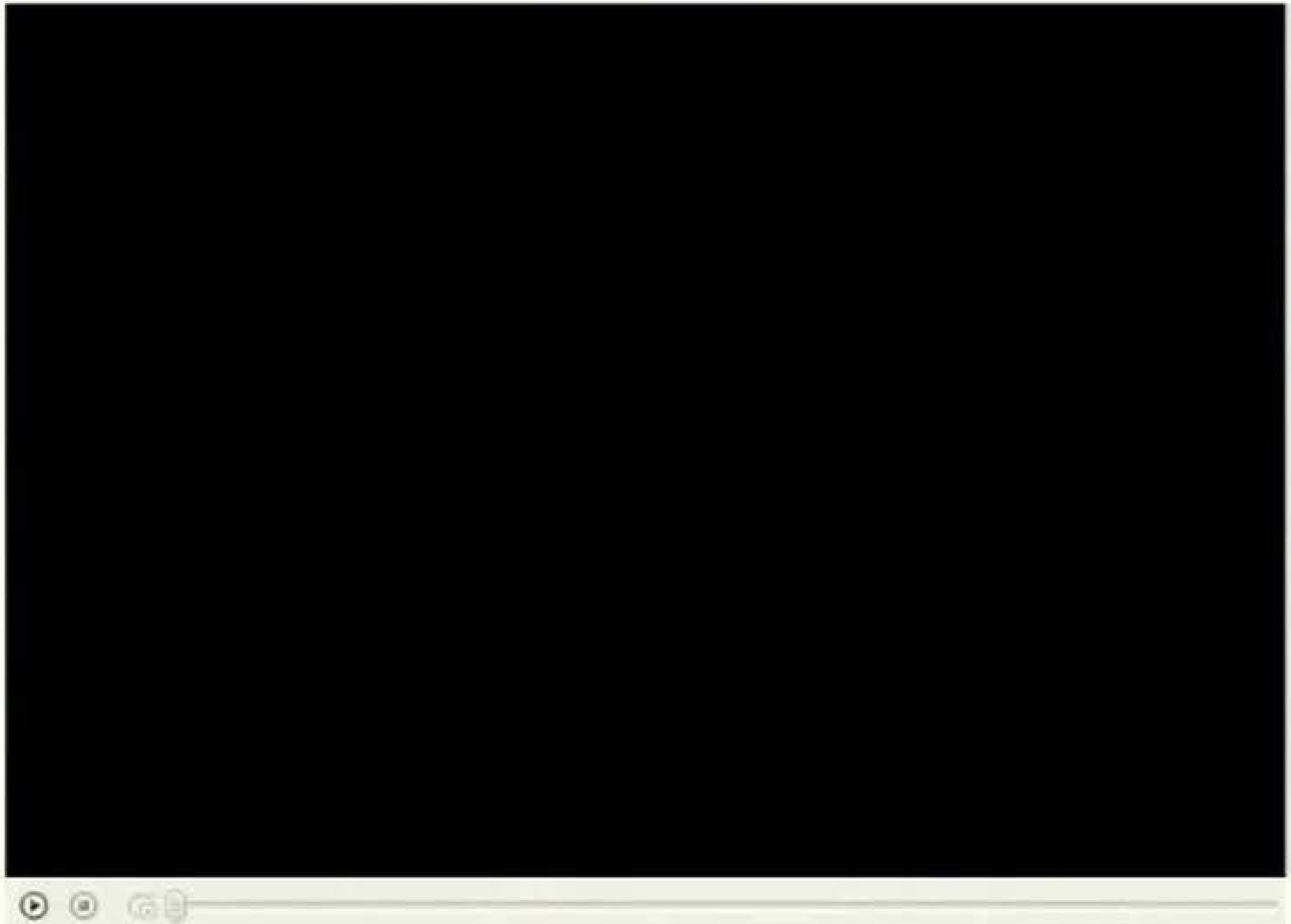

Figure: diffusion sequence $p_{i,t}(x, y, z)$, projected on $X \times Y$ and $X \times Z$

Alternating diffusion

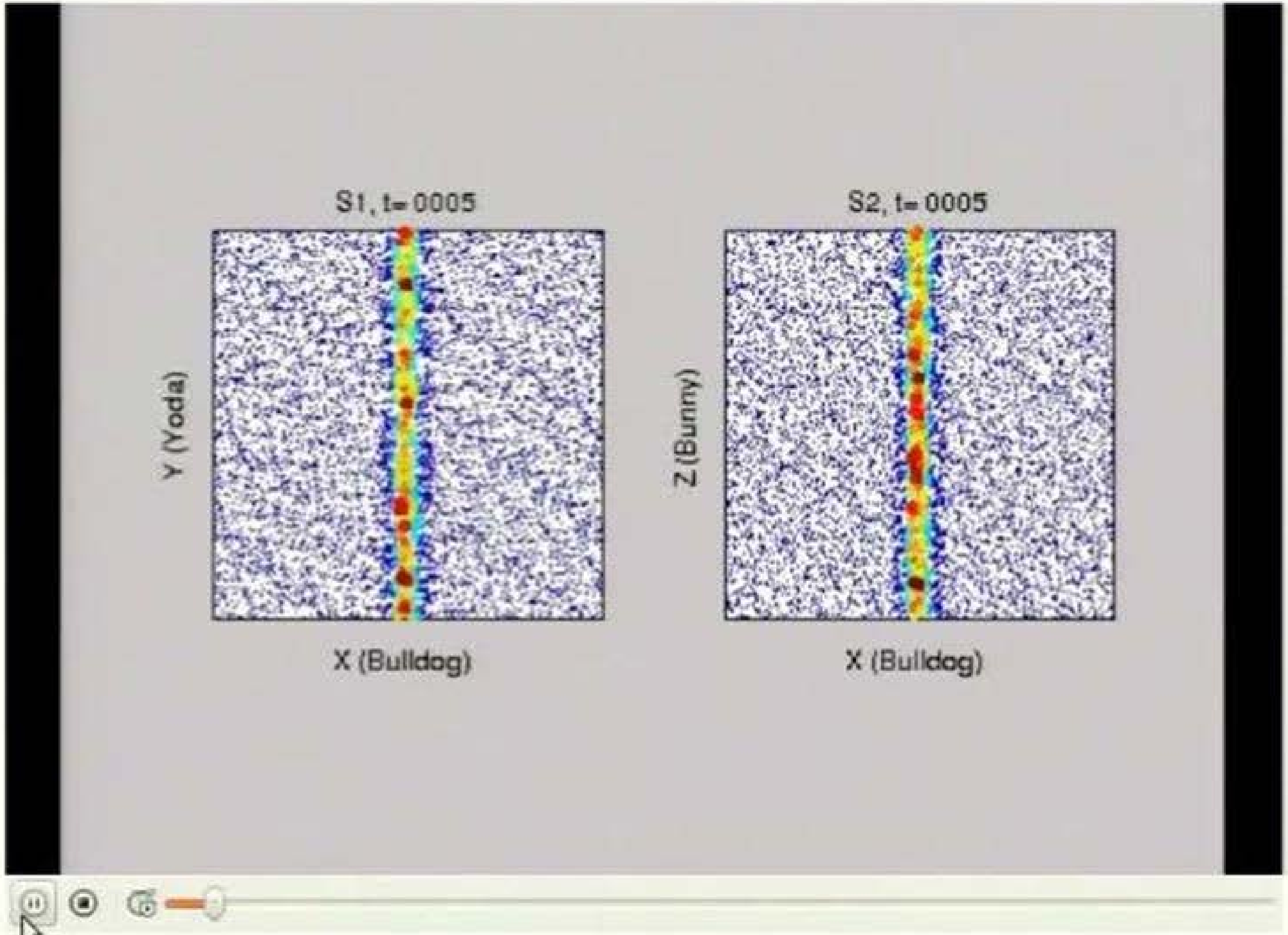

Figure: diffusion sequence $p_{i,t}(x, y, z)$, projected on $X \times Y$ and $X \times Z$

Alternating diffusion embedding

Compute the diffusion distance: $d_t(i,j) = \|p_{i,t} - p_{j,t}\|_\pi$,
and a low dimensional embedding consistent with this distance:

Figure: alternating diffusion captures the geometry of the *common* variable and ignores the sensor-specific variables.

Alternating diffusion embedding

Compute the diffusion distance: $d_t(i,j) = \|p_{i,t} - p_{j,t}\|_\pi$,
and a low dimensional embedding consistent with this distance:

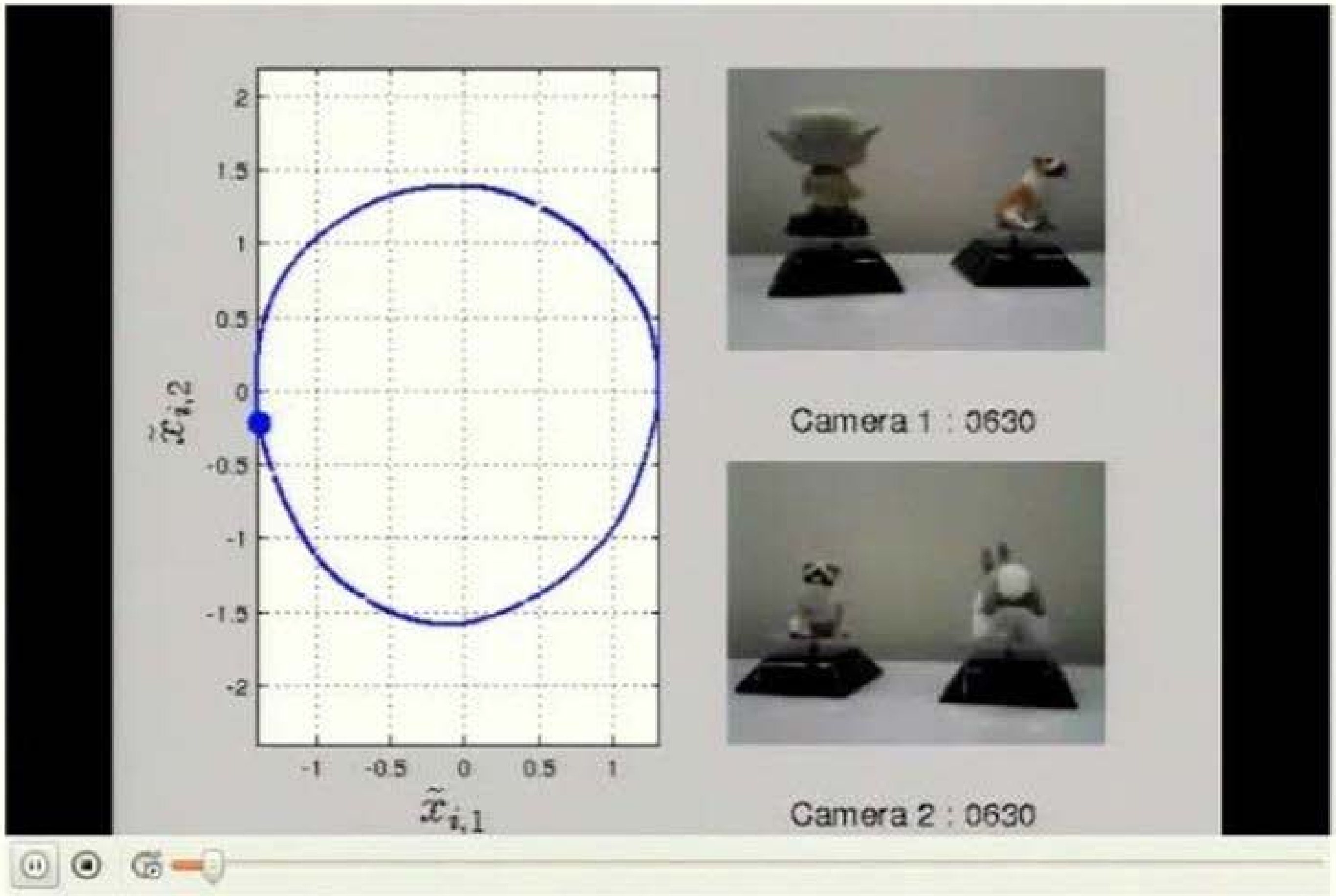

Figure: alternating diffusion captures the geometry of the *common* variable and ignores the sensor-specific variables.

Algorithm 1 Alternating-diffusion

1. Calculate two affinity matrices
$W_{ij}^{(1)} = \exp\left(-\frac{\|s_i^{(1)} - s_j^{(1)}\|^2}{\varepsilon^{(1)}}\right);\ W_{ij}^{(2)} = \exp\left(-\frac{\|s_i^{(2)} - s_j^{(2)}\|^2}{\varepsilon^{(2)}}\right).$
2. Compute diffusion operators $K_{ij}^{(1)} = \frac{W_{ij}^{(1)}}{\sum_{l=1}^{n} W_{lj}^{(1)}}; K_{ij}^{(2)} = \frac{W_{ij}^{(2)}}{\sum_{l=1}^{n} W_{lj}^{(2)}}.$
3. Compute the alternating diffusion operator $K = K^{(2)} K^{(1)}$.
4. Compute the diffusion distance at time $2m$ between each two points:
$$d_{2m}(i,j) = \|(K^m)_{\cdot,i} - (K^m)_{\cdot,j}\|_2.$$
5. (Optionally:) Refine using a standard diffusion maps algorithm.

Main results

We define the effective ("marginal") functions $p^{(e)}_{i,t}(x)$,

$$p^{(e)}_{i,t}(x) = \int p_{i,t}(x, y, z)\pi_{y,z|x}(y, z)dydz$$

Theorem 1 (Alternating diffusion sequences)

The sequence of effective functions $p^{(e)}_{i,2m+1}(x)$ is a diffusion sequence with the appropriate diffusion operator $D^{(e)}$:

$$p^{(e)}_{i,2m+3}(x) = \left(D^{(e)}\left(p^{(e)}_{i,2m+1}\right)\right)(x)$$

Ideally, we would now like to define an alternating diffusion distance of the form

$$\left\|p^{(e)}_{i,2m+1}(x) - p^{(e)}_{j,2m+1}(x)\right\|_M .$$

Main results

The effective functions $p_{i,t}^{(e)}(x)$ cannot be measured directly, but the alternating diffusion distance can be computed from the alternating diffusion sequences $p_{i,t}(x,y,z)$.

Theorem 2 (Computing the alternating diffusion distance)

The alternating diffusion distance is related to the diffusion sequence $p_{i,2m+2}(x,y,z)$ *by*

$$d_{2m+1}(i,j) = \left\| p_{i,2m+1}^{(e)}(x) - p_{j,2m+1}^{(e)}(x) \right\|_M =$$
$$= \| p_{i,2m+2}(x,y,z) - p_{j,2m+2}(x,y,z) \|_\pi ,$$

with the appropriate norms $\|f(x)\|_M$ *and* $\|f(x,y,z)\|_\pi$.

Sleep stage identification

[Lederman, Talmon, Wu, Lo, Coifman, to appear in ICASSP 2015]

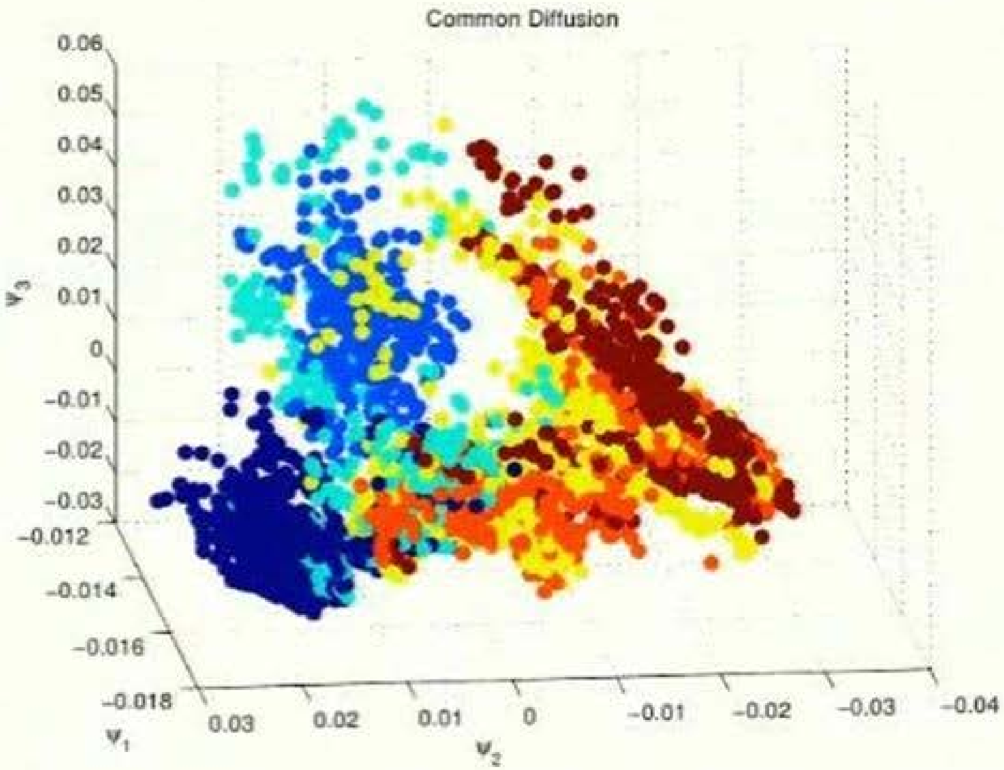

Figure: common variable in respiration signals (airflow, chest belt, and abdominal belt).

Sleep stage identification

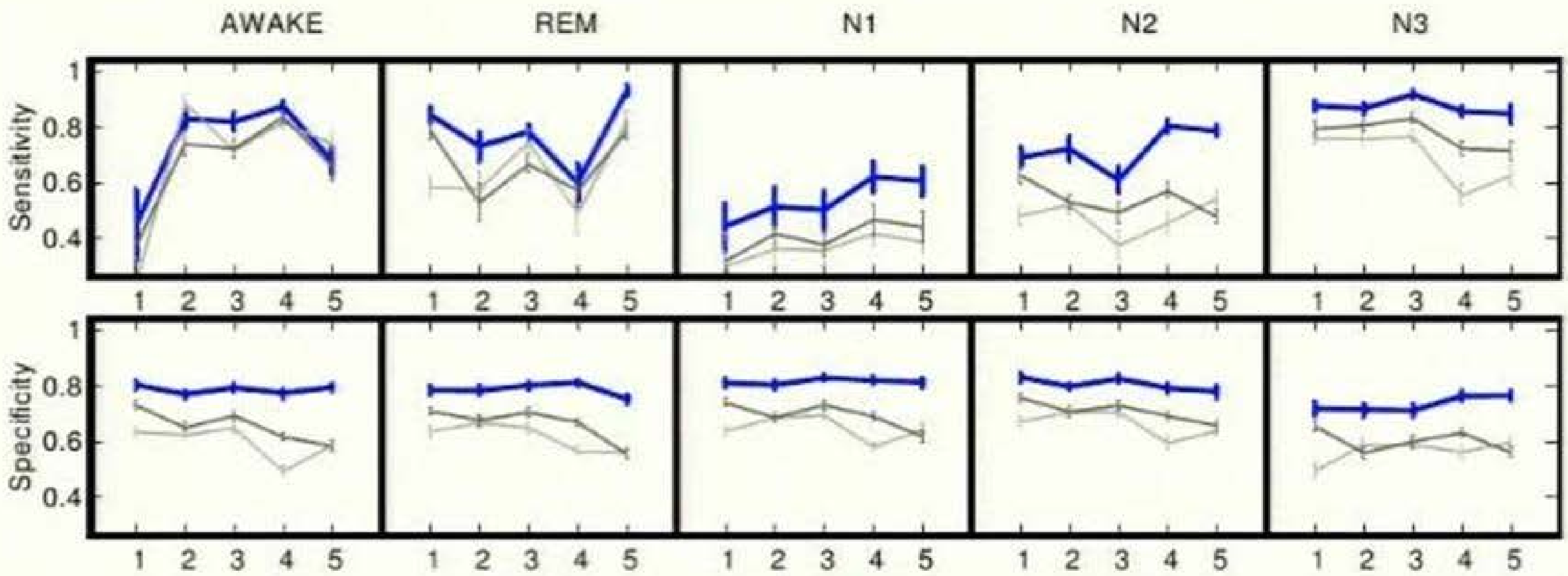

Figure: the sensitivity and the specificity of the sleep stage identification per individual.
Gray: single channels identification (light gray - airflow, dark gray - abdominal motion).
Blue: alternating diffusion identification.

Related work

- Canonical Correlation Analysis (CCA) [Hotelling, 1936]
- Kernel CCA [Lai, 2000 ; Bach, Jordan, 2003]
- Two-Manifold Problems [Boots, Gordon, 2012]
- Cross-diffusion [Wang, et al. 2012; Lindenbaum, 2015]

Alternating diffusion - random projection of images

Figure: alternating diffusion captures the geometry of the *common* variable and ignores the sensor-specific variables.

Alternating diffusion - random projection of images

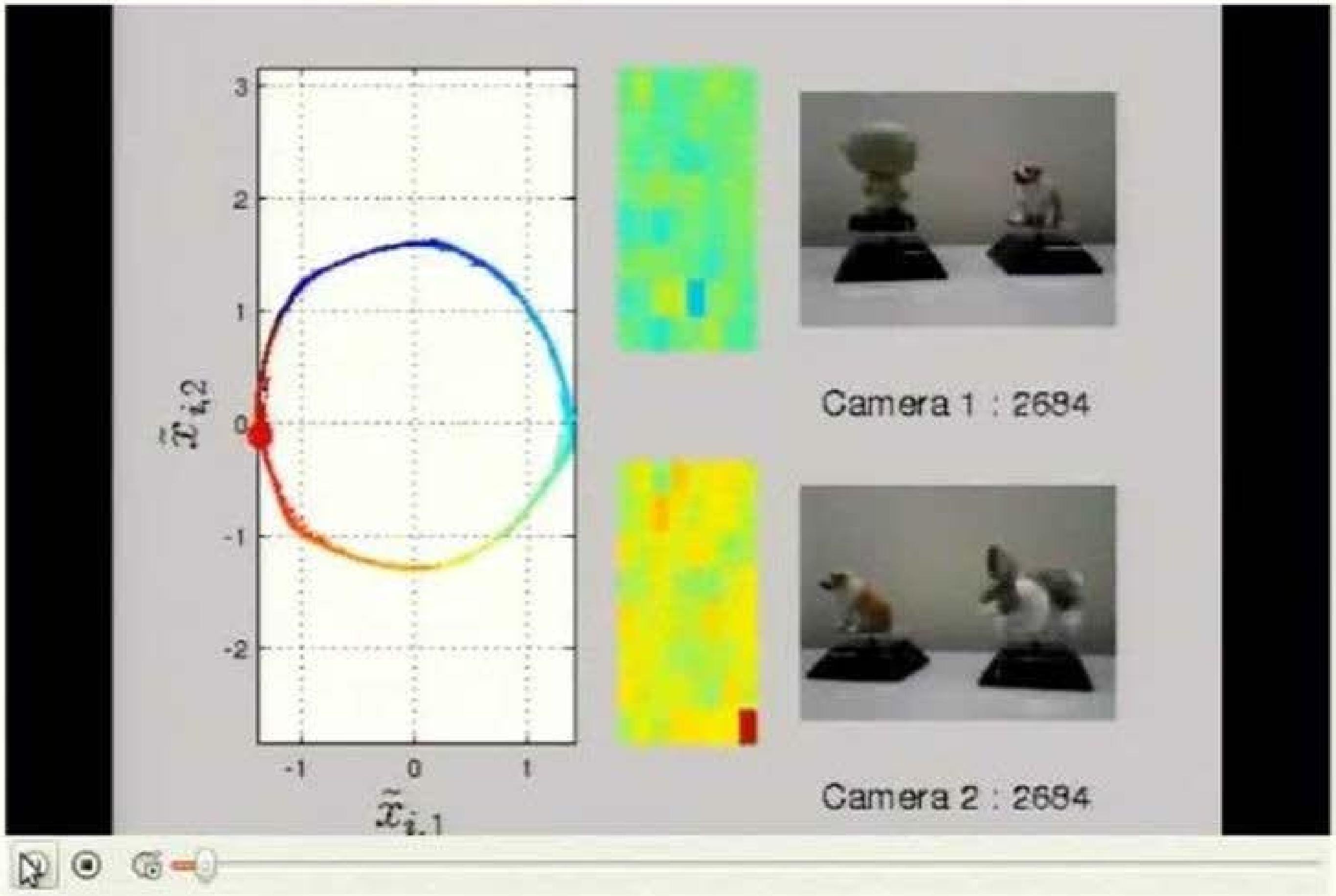

Figure: alternating diffusion captures the geometry of the *common* variable and ignores the sensor-specific variables.

Thank you!

More information:

http://roy.lederman.name/alternating-diffusion/

Technical report: YALEU/DCS/TR1497

roy.lederman@yale.edu