Non-parametric Density Estimation on a Transformation Group for Vision

Erik G. Miller, UC Berkeley

Christophe Chefd’hotel, INRIA

Goal

- Develop a simple, practical non-parametric density estimator for linear shape change.

Previous Work

- Probabilities on non-Euclidean group structures:
  - Grenander ('63)
- Parameter estimation on groups:
  - Grenander, M. Miller, and Srivastava ('98)
- Theoretical results (convergence) for non-parametric density estimators on groups:
  - Hendriks ('90)
- Diffeomorphisms:
  - Grenander, Younes, M. Miller, Mumford, others

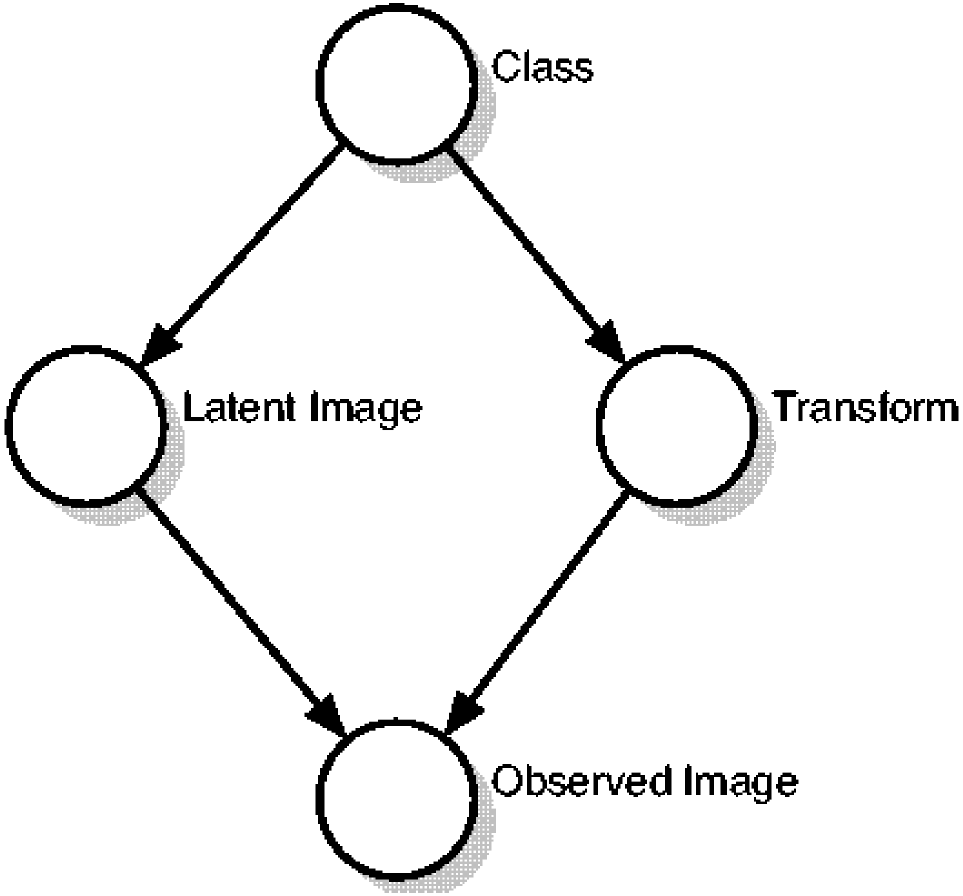
Outline

- Latent image-transform factorized image models.
  - Focus on transform density.
- Justification of matrix group structure for transformations.
- A natural inheritance structure:
  - The group difference.
  - An equivariant distance metric.
  - An equivariant kernel function.
  - An equivariant density estimator.
- Experiments:
  - Comparison of Euclidean transform density to equivariant estimator.

Latent Image-Transform Modeling

- Grenander
- Vetter, Jones, Poggio ('97)
- Jojic and Frey ('99)
- E. Miller, Matsakis, Viola ('00)

A Generative Image Model

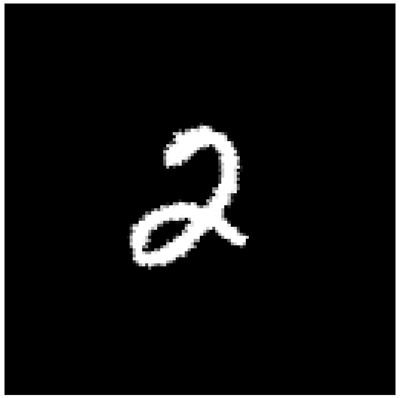
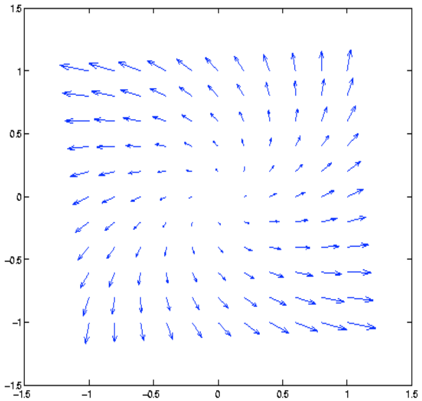
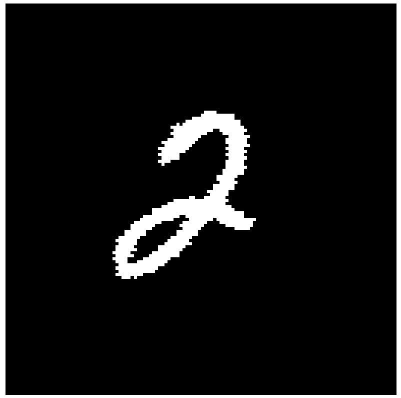


- A factored model:

Prob(Observed Image) =
Prob(Latent Image) *Prob(Transform)

An Image Decomposition

Latent Image * **Transform** = **Observed Image**

 *  = 

Estimating a Factored Image Model

- Step 1
  - Estimate latent images and linear transforms from observed images.
- Step 2
  - Build densities on sets of latent images and transforms.

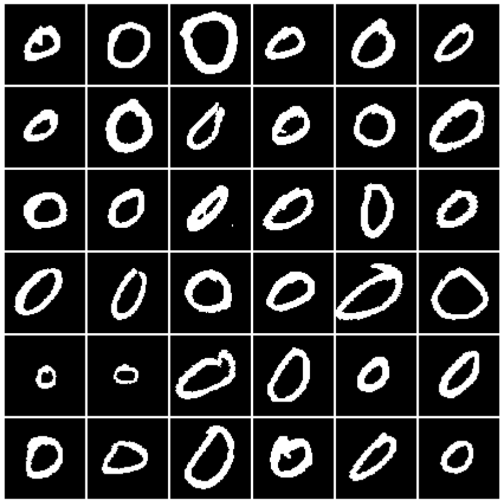
Miller and Chefd'hotel

Congealing: Automatic Factorization

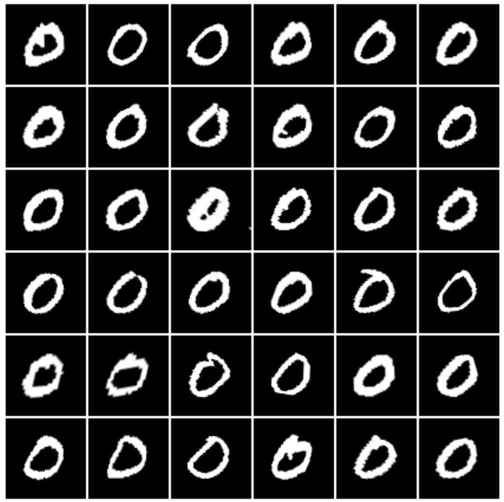
Observed Images → Latent image estimates

Observed Images → Transform estimates

Before



After



See Miller et al, CVPR 2000 for details.

A Set of Transforms From Congealing

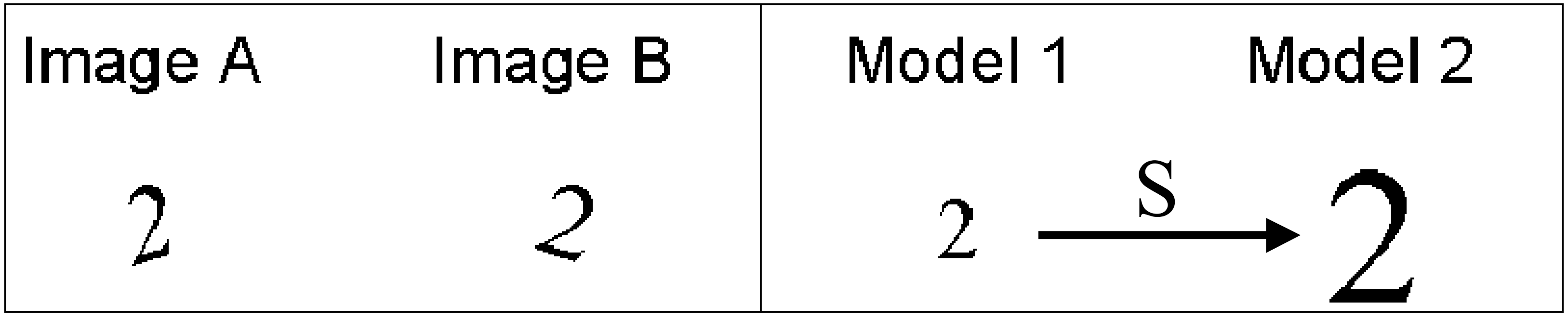
$$\begin{bmatrix} 1.1 & 0.1 \\ 0.2 & 1.3 \end{bmatrix} \quad \begin{bmatrix} 0.6 & -0.1 \\ 0.0 & 1.3 \end{bmatrix}$$

$$\begin{bmatrix} 1.1 & 0.1 \\ 0.2 & 1.0 \end{bmatrix} \quad \begin{bmatrix} 1.4 & 0.4 \\ -0.2 & 1.3 \end{bmatrix}$$

$$\begin{bmatrix} 1.0 & -0.1 \\ 0.1 & 1.1 \end{bmatrix} \quad \begin{bmatrix} 1.2 & -0.1 \\ 0.2 & 0.9 \end{bmatrix}$$

Why not just treat them as 4-vectors?

Desired Invariance



- The difference between A and B should be invariant to the choice of model:

$$
\begin{aligned}
D(T_A^1, T_B^1) &= D(T_A^2, T_B^2) \\
&= D(S * T_A^1, S * T_B^1),
\end{aligned}
$$

Equivariance of Group Difference

$$D_G(T, A) = T^{-1} \circ A,$$

$$\begin{aligned} D_G(B \circ T, B \circ A) &= [B \circ T]^{-1} \circ B \circ A \\ &= T^{-1} \circ B^{-1} \circ B \circ A \\ &= T^{-1} \circ A \\ &= D_G(T, A). \end{aligned}$$

General Linear Group

- GL(2): 2x2 non-singular matrices with matrix multiplication as group operator.
- $GL^+(2)$: 2x2 matrices with positive determinant.
- Equivariant difference is

$$D(T, A) = T^{-1}A$$

An Equivariant Distance

$$d(\mathbf{A},\mathbf{B}) = ||\log(\mathbf{A}^{-1}\mathbf{B})||_{\mathcal{F}}$$

- Matrix logarithm: inverse of $e^{\mathbf{A}} = \mathbf{X}$
  Not necessarily unique.
- Generalization of geodesic distance on SO(N).

An Equivariant Kernel

$$K(T;A) = \frac{1}{C(\sigma^2)} e^{-\frac{1}{2\sigma^2}||\log(T^{-1}A)||_F^2},$$

- Generalization of log-normal density to multiple dimensions.

A Subtlety

- Kernel *function* is equivariant, but is integral of kernel function? Not necessarily!

? ? ?

$$\begin{aligned} Prob(E) &= \int_{T \in E} \frac{1}{C} e^{-\frac{1}{2\sigma^2} \|\log(T^{-1}A)\|_F^2} d\mu \\ &= \int_{T \in B*E} \frac{1}{C} e^{-\frac{1}{2\sigma^2} \|\log(T^{-1}BA)\|_F^2} d\mu \\ &= Prob(B*E), \end{aligned}$$

  – Must use group invariant measure for integration:

$$d\square = \frac{1}{|T|^2} d\overline{x}$$

An Equivariant Estimator

$$f(U;T_1,T_2,...,T_N) = \frac{1}{N}\sum_{i=1}^{N} K(U;T_i).$$

$$\begin{aligned} f(BU;BT_1,BT_2,...,BT_N) &= \frac{1}{N}\sum_{i=1}^{N} K(BU;BT_i) \\ &= \frac{1}{N}\sum_{i=1}^{N} K(U;T_i) \\ &= f(U;T_1,T_2,...,T_N). \end{aligned}$$

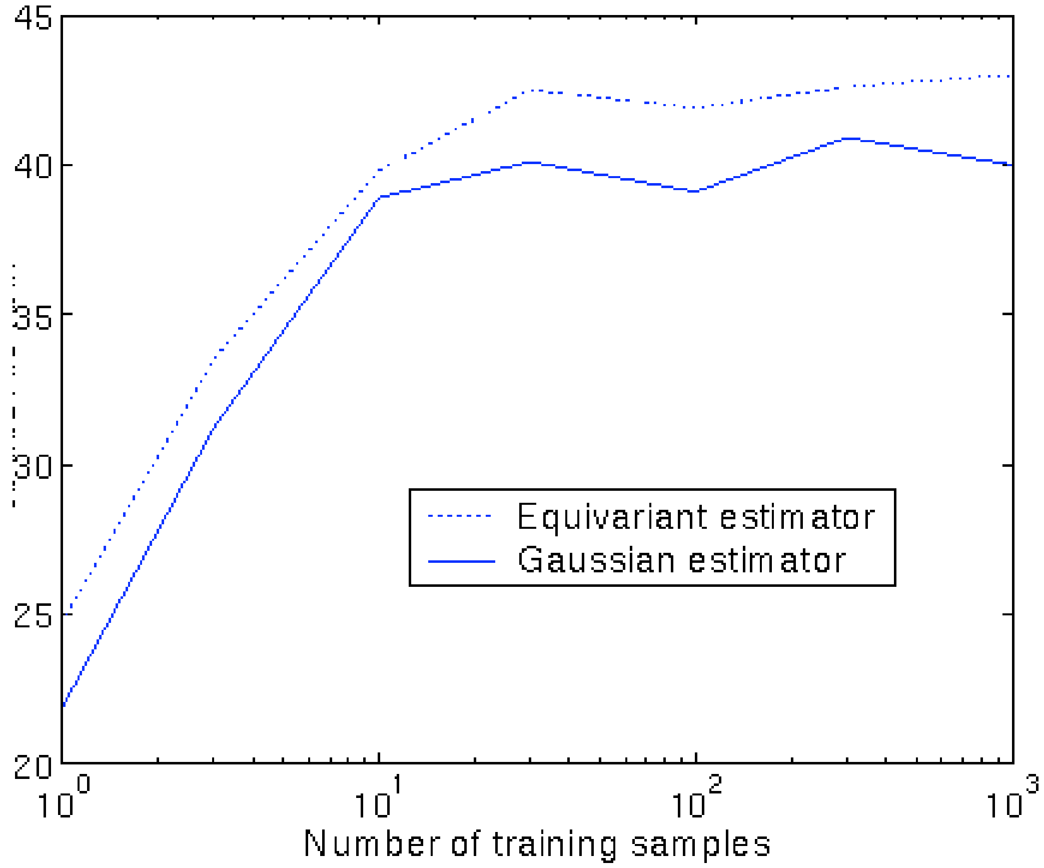
Choosing the Bandwidth

- Bandwidth parameter is not equal to variance.
- To maximize likelihood, must compute normalization constant.
  - Use Monte Carlo methods.
    - Slow, but doable.

Experiments

- Likelihood of held-out points
  - Cross-validated mean log-likelihood based on 100 examples: 1.7 vs. 0.2.
- One example classifier
  - 89.3% vs. 88.2%
- Transform-only classifier:
  - Align a test digit to each model:
  - Classify based only on transform.
  - 9-6 example.

Transform-only classifier



Summary

- A simple density estimator based on the group difference.
- Easy to implement.
- Improves performance over naïve Gaussian kernel estimate.

Equivariance of Euclidean Kernel

$$K(T;A) = K(B \circ T; B \circ A).$$

$$K(T;A) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(T-A)^2}$$

$$\begin{aligned} K(B+T;B+A) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(B+T-(B+A))^2} \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(T-A)^2} \\ &= K(T;A). \end{aligned}$$