Statistical Shape Segmentation

Ayres Fan

with

John Fisher, Sandy Wells, Steve Haker, Mujdat Cetin, Alan Willsky
Outline

1. Philosophy (Problem statement, background)
2. Math (General Framework)
3. Recess (Monge-Kantorovich)
4. Detention (Sampling and mixture algorithms)
5. Art (Results)
6. Final Bell (Summary)
Hard Segmentation Problems

- Segmentation: Process of dividing an image into coherent regions
- Can be hard for a number of reasons:
  - Occlusions
  - Missing data
  - Poor contrast
  - Missing edges
  - Poor image models
Extra Information

- Training examples
- Partial segmentations (e.g., slices)
- Relative objects
  - Use easy to segment objects to help locate hard to segment objects
- Can view all of these in a probabilistic framework
Previous Work

- Cootes and Taylor (PCA on marker points)
  - Simple and fast
  - Need correspondence (very hard in 3D)
- Leventon et al (PCA on SDF)
  - No correspondence problem
  - Linear operation on nonlinear manifold
- Tsai et al (multiple objects)
  - Problems with limited example space
- Paragios (mean SDF plus random field)
- Srivastava et al (geodesics on manifolds)
General MAP Model

\[ P(\Gamma|y; S) \propto P(y|\Gamma)P(\Gamma; S) \]

• \( \Gamma \) is a segmentation (can be a curve, indicator function, etc.)
• \( y \) is the observed image (can be vector)
• \( S \) is a shape model
• Data model usually IID given \( \Gamma \)
Traditional Curve Evolution

• Chan-Vese energy functional:

\[
E(\Gamma) = \int_{\Omega} (y - \mu(\Gamma(x)))^2 \, dx + \alpha \oint_{\partial \Gamma} ds
\]

• If we discretize the data term, equivalent to:

\[
P(\Gamma) \propto \left[ \prod_{i} \exp((y_i - \mu(\Gamma_i))^2) \right] \exp(\alpha \oint_{\partial \Gamma} ds)
\]
Shape Representations

• Parameterized curves (marker points)
• Implicit surfaces
  – signed distance functions
• Space conditioned probabilities (PERPS)
Shape Spaces

- A manifold embedded in an infinite dimensional Hilbert space (e.g., L_p)
- Oftentimes infinite codimension (e.g., for SDF, |∇Ψ|=1 for a.e. x)
- Mainly interested in local regions
- Curvature induced by metric, representation
- Generally large equivalence classes (pose)
Desired Characteristics

• Want to view probability as being related to shape distance
• Want locally-flat manifolds (adjust representation, distance metric)
• Capture co-variation (intra- and inter-object)
• Computationally feasible
Parzen Methods

• Lp exponential kernels

\[ K(x_1, x_2) = \frac{1}{Z} \exp(-d^p(x_1, x_2)/\alpha) \]

\[ p(x) = \frac{1}{N} \sum_{i=1}^{N} K(x, x_i; \alpha) \]

• Similar to a smoothed histogram
Distance Functions

• Now with this framework, we can view the problem of constructing pdfs on shape as choosing an appropriate distance function.

• Ideally, this distance would be computed along the manifold, but this is easier said than done.

• Most use distance function for Hilbert space (e.g., L2).
1. Philosophy (Problem statement, background)
2. Math (General Framework)
3. Recess (Monge-Kantorovich)
4. Detention (Sampling and mixture algorithms)
5. Art (Results)
6. Final Bell (Summary)
Monge-Kantorovich

- Two densities, $\mu_0$ and $\mu_1$. Want to reshape $\mu_0$ using a mass-preserving diffeomorphism (bijective, differentiable) $u$.

\[
\text{MP}(\mu_0, \mu_1) = \{u | \mu_1(x) = |Du| \mu_0 \circ u(x) \forall x\}
\]

- Define optimal MK map as:

\[
u^* = \arg \min_{u \in \text{MP}(\mu_0, \mu_1)} \int_{\Omega} \|u(x) - x\|_p^p \mu_0 dx\]
Fundamental Result

• Brenier 1987, 1991
  – $u^*$ is curl free (and hence $u^*$ is the gradient of a [convex] potential function)
  – for any mass preserving $u$, we can write as:
    \[ u = \nabla w \circ s \]
• Polar factorization ($s$ is the inefficiency)
• Intuition: curl is just wasted rotational energy
Quick Example

- Infinitely many choices (flexibility in zero areas)
- Only two possible diffeomorphisms if we add epsilon
- Note convexity of optimal solution implies $u$ is monotonically increasing
Features

- Not a metric in embedding space
- Topology changes are straightforward
- Well behaved with respect to translation and scaling (L2 is not)
- Gives a nice physical intuition behind shape distance
- Also gives dense correspondence
- With time formulation (viewing mass movement as a time evolution) on PERPS, intermediate steps should also lie on manifold
Feasible Gradient Descent

• Haker et al
• Simple initialization (composition)
• Stay in functions that are MP

\[ u_t = -\frac{1}{\mu_0} (Du) \xi, \quad \text{div} \xi = 0 \]

• Remove the curl (Helmholtz decomposition, \( u = \nabla v + \chi, \text{div} \chi = 0 \))

\[ u_t = -\frac{1}{\mu_0} (Du)(u - \nabla \Delta^{-1} \text{div}(u)) \]
Source and target
Deformation field
Hard to get gradient

• Computing $\frac{\partial d_{MK}^2}{\partial \mu_1}$ is tough because set MP changes with $\mu_1$

• Consider an unbalanced method:
  \[ \tilde{d}_{MK}^2(\mu_0, \mu_1) = \int_{\Omega} ||u(x) - x||^2 \mu_0 dx + \lambda \int_{\Omega} (\mu_1 - |Du| \mu_0 \circ u)^2 dx \]

• Can view first term as mass movement, second term as mass creation

• Why not use a simpler distance function?
Outline

1. Philosophy (Problem statement, background)
2. Math (General Framework)
3. Recess (Monge-Kantorovich)
4. Detention (Sampling and mixture algorithms)
5. Art (Results)
6. Final Bell (Summary)
Sampling methods

- Only need to be able to evaluate pdf
- Attractive empirical convergence results
- Explores configuration space
- Metropolis algorithm:
  1. Start with $x_0$
  2. Generate candidate $y_{t+1}$ (given $x_t$)
  3. Set $x_{t+1} = y_{t+1}$ with probability $\min(1, p(y_{t+1})/p(x_t))$, otherwise $x_{t+1} = x_t$
  4. Go back to 2
Sampling Algorithm

- Model:
  - L2 on data
  - Parzen windows using Lp kernels on PERPS
- Sample by adding smooth random fields
- Gradually-sloped edges are more likely to be moved than steeply-sloped edges
Initial Results

- Problems—way too few training examples
- Sampling method isn’t very good (too smooth)
Better Sampling Method

\[ \frac{d\vec{C}}{dt}(s) = f(s)\vec{N}(s) \]

\( \vec{C} : [0, 1] \rightarrow \mathcal{R}^2, \quad f \in L_2([0, 1]) \)

• Can do partially deterministic (mean), partially random

• This may allow faster convergence speed
Fast(er) MK Computation

• For these sampling methods, we are dealing with perturbations. Have map from $\mu_0$ to $\mu^{t+1}$, $\mu^t$ is close to $\mu^{t+1}$, want to compute map from $\mu_0$ to $\mu^{t+1}$.

• Very inefficient: for every sampling step, we compute an optimal diffeomorphism. In that computation, for every iteration, we solve Poisson’s equation.

• Using gradient descent, so initialization matters.

• Note that $u^t$ is close to $u^{t+1}$, but they are not in the same MP set.

• Compute $v$, any MP map between $\mu^t$ and $\mu^{t+1}$.

• Then $u^{t+1}$ is close to $v(u^t(x))$. 
Mixture Model

- Take convex combinations of PERPS
  \[ \phi^j(x) = \sum_{i=1}^{N} \alpha_i \Phi^j_i(x) \]
- View function values as prior space-conditioned marginal probabilities on labels
- We then have a parameterized prior model with unknown parameters
- Data term is L2
- Prior term...
EM Algorithm

• x observed, y missing/hidden/auxillary
• E-step: \( Q(\theta, \theta^{t-1}) = E[\log p(x, y|\theta)|x, \theta^{t-1}] \)
• M-step: \( \theta^t = \arg \max_{\theta} Q(\theta, \theta^{t-1}) \)

• Useful when the complete data likelihood is much easier to maximize than the observed data likelihood
• In mixture models, the E step often takes the form of expected weights (class probabilities)
Very Bad Assumption

• Our E-step looks like

\[ \alpha_i = p(i|\Gamma) = \frac{p(\Gamma|i)p(i)}{\sum_j p(\Gamma|j)p(j)} \]

• We need to convert our marginal prior probabilities into a joint density

• We use an IID assumption

\[ p(\Gamma|i) = \prod_x p(\Gamma(x)|i) \]

• Still should capture global features
Low SNR Data

- 0 dB (but “effective” SNR much higher)
Stochastic Systems Group

Missing Data

• 3 dB, missing data on part of the wing
Future Directions

• Make MK faster (multiresolution)
• Spatially varying sigma (encodes certainty of boundaries)
• Get MK flow for shape segmentation
• Better sampling methods
• Better sampling algorithm (use information from all samples)