Statistical Shape Models from A to Z
The Medical Imaging Chain

Acquisition

Analysis
Endless Opportunities…

... but automatic image analysis requires \textit{a-priori} knowledge!
Statistical Shape Models

Biological variation through mathematical transforms:

Training data
Statistical Shape Models from A to Z

I) Construction of SSMs
II) Appearance Modeling
III) Local Adaptation Strategy
IV) Initialization

Heimann & Meinzer 2009
I) Construction of Statistical Shape Models
Shape Representation

**Landmarks:**
- Cloud of points on object boundary
- Very versatile, arbitrary topology
- Simple to store as list of coordinates:
  \[ x = (x_1, y_1, z_1, \ldots, x_k, y_k, z_k)^T \]
- Also known as Point Distribution Model (PDM)

Landmarks on all training shapes have to be located at corresponding positions!
Iterative Closest Points (ICP) for Shape Model Construction?

Only delivers rigid transform, i.e. templates registered with ICP only not helpful for building statistical models

- Post-processing: project resulting template points to surface of target shape
- Only usable for objects with low shape variation

We require non-linear registration!
Gaussian Mixture Model (GMM)

**Key idea:** represent point sets by Gaussian mixture models
- Allows to interpret each point set as statistical sample
- Transforms discrete optimization problem into continuous one

For point matching:
- One Gaussian component for each point
- All weighted equally
- All share same spherical covariance

Jian & Vemuri 2011
Chose transformation model $T$, parameters $\theta$

- Rigid transform
- Thin-plate splines
- Gaussian radial basis functions

Minimize L2 distance between both distributions:

$$d_{L_2}(S, M, \theta) = \int (gmm(S) - gmm(T(M, \theta)))^2 dx.$$ 

Closed solution by fast Gauss transform: $O(n+m)$

with $n$, $m$ as size of the point sets
GMM Overview

- Initialize transform and scale (bandwidth)
- Set up objective function to minimize for current transform
- Optimize objective function using gradient-based optimizer (e.g. quasi-Newton or nonlinear conjugate gradient)
- Apply calculated transform
- Decrease scale (annealing)
- Check for convergence
Correspondence by Optimization

Occam’s Razor: The simplest model is the best one!

• MDL = Minimum Description Length
• Simplest Model has minimum stochastic complexity
• Based on information theory
• Cost function ~ Sum of logarithmic variances

Davies et al. 2002
Initial Landmark Generation

- Create a bijective mapping between each training shape and a common base domain.
- For two-manifolds without holes or self-intersections, suitable base domain is a sphere.
- Define common set of landmarks on sphere.
- Use inverse parameterization functions to generate individual training samples.
Population-based Correspondence

Spread landmarks equally on surface

Create shape model
Calculate cost function
Change landmark distribution

Until cost function converges at minimum

Cost function based on Minimum Description Length

\[ F = \sum_{m} \mathcal{L}_m \quad \text{with} \quad \mathcal{L}_m = \begin{cases} 1 + \log(\lambda_m/\lambda_{\text{cut}}) & \text{for } \lambda_m \geq \lambda_{\text{cut}} \\ \lambda_m/\lambda_{\text{cut}} & \text{for } \lambda_m < \lambda_{\text{cut}} \end{cases} \]
**Optimization Procedure**

- Calculate gradients of cost function for all landmarks

- Warp parameterizations in optimal direction using local Gaussian kernels

- Randomize kernel positions for each iteration

- Reduce kernel size in the course of optimization

Heimann et al. 2005
Dimensionality Reduction

1. Store each training sample as a (n x d)-dimensional vector
2. Build a landmark configuration matrix \( L \) from columns of all training samples
3. Calculate the mean landmark positions and build matrix \( M \) with all columns set to the mean
4. Calculate the covariance matrix:
   \[
   \Sigma = \frac{1}{n-1} (L-M) (L-M)^T
   \]
5. Perform a principal component analysis (PCA) on the covariance matrix \( \Sigma \)
The Statistical Shape Model

• PCA aligns a dataset along the axes of maximum variance

• Result from PCA:

\((n_s-1)\) modes of variation with a displacement vector and a variance each (eigenvector and eigenvalue)

• Each training sample can be described by a linear combination of the mean and the displacement vectors:

\[
x_i = \bar{x} + \sum_{m=1}^{n_s-1} y_i^m \cdot p^m
\]

• Modes with low standard deviation can be neglected to reduce the number of parameters
II) Appearance Modeling
Appearance Modeling by Profiles

- Fitting SSM to new data requires additional appearance model
- Common solution: sample a line perpendicular to model contour/surface (profile)
- Profiles typically consist of intensity values or gradients, sometimes normalized
- Sometimes true boundary not on strongest edge…
Training Appearance Models

- Use information from training data!
- Build statistical model of profiles:
  - Separate profile model for each landmark
  - Sample profiles across model contour in all training images

Assuming Gaussian distribution

- Use Mahalanobis distance as similarity metric, based on covariance matrix

No assumptions w.r.t. distribution

- Non-linear model based on moderated kNN-classifier estimates boundary probability as:

\[ p(b|g) = \frac{b_k(g) + 1}{k + 2} \]
Appearance Modeling by Steerable Features

- 24 features based on intensity and gradient
- Probabilistic boosting tree (PBT) for classification

Zheng et al. 2008
Appearance Modeling by Haar-like Features

- Arbitrary block sizes
- Generalize to arbitrary dimensions
- Constant evaluation time due to integral images:

- Integral image only valid for specific orientation
Benchmarking Local Appearance Models

- Test generated appearance models on known locations in training data
- Randomize position along boundary and normal vector
- Displace profile to inner and outer side of the boundary and estimate boundary probability
- Calculate performance index f:

\[ f_{ir} = \left( \sum_{k=-K}^{K} |k|^d (p_i(b|g_r0) - p_i(b|g_rk)) \right) / \left( \sum_{k=-K}^{K} |k|^d \right) \]

- Average over several runs
III) Local Adaptation Strategy
Active Shape Model Search

- Iterative approach with local landmark search
- Initialize model and repeat:
  1. Find optimal displacements for each landmark independently
  2. Retain shape and find best matching pose $T$
  3. Fix $T$ and find best matching shape parameters
  4. Update landmarks
- Typical extension: Multi-resolution search

Cootes et al. 1995
SSMs and Over-constrained Deformation

- In addition to systematic variation, biological objects feature essentially random perturbations.
- Constraining deformation to limited number of modes is too restrictive.

Green = model
Blue = best fit
Possible Solutions

• Add artificial degrees of freedom during model construction:
  • FEM simulation
  • Ad-hoc deformations

• Post-processing steps:
  • Level Sets
  • Freely deformable models

• Allow additional degrees of freedom during search:
  • Freely deformable model with statistical shape priors
Deformable Model

- Representation as discrete mesh (same topology as SSM)
- Evolution by forces equilibrium:
  - Internal forces for stabilization
  - External forces to adapt model to data

\[ p_i^{t+1} = p_i^t + F_{int}(p_i^t) + F_{ext}(p_i^t) \]

Heimann et al. 2007
Internal Forces: Tension

- Edge lengths should be similar as in SSM
- For each vertex in mesh:
  - Determine distances to neighbors
  - Calculate correction vectors
  - Add all correction vectors

\[
F_T(p, q) = \alpha \left( 1 - \frac{|\tilde{p} - \tilde{q}|}{|p - q|} \right) (p - q)
\]
Internal Forces: Rigidity

• Angles between neighboring faces should be similar as in SSM
• Rotation around edge yields $F_R$
• Position of a constellation may not be changed: Neutralizing force $F_N$

$$F_R(q, [p_1, p_2]) = T(q, [p_1, p_2], \beta \delta) - q$$

$$F_N([p_1, p_2]) = -\frac{1}{4} (F_R(q_1, [p_1, p_2]) + F_R(q_2, [p_1, p_2]))$$
External Forces

Standard method:
• Evaluate appearance model at different locations, pick best one
• Problem: outliers
Optimal Candidate Detection

Possibilities:

- Soft constraints by penalty costs
- Hard constraints by strict rules
Optimal Surface Detection (1/2)

Optimal surfaces with hard constraints efficiently solvable by transformation to graph problem:

1. Build column graph from mesh topology
2. Assign each node costs from corresponding appearance model
3. Add source and sink, connect systematically to all other nodes
4. Determine maximal flow / minimum cut
Optimal Surface Detection (2/2)

Direct comparison for the calculation of external forces on Liver CT:

Standard method

Optimal surface detection
Active Appearance Model Search

- Combine shape and appearance variation in one linear system
- Assume constant relationship between appearance model residuals and parameter updates:
  \[ dp = -Rr(p) \]
- Success depends largely on derivative matrix R
- Calculate R by multivariate linear regression or numeric differentiation

Cootes et al. 2001
Shape Regression Machine

Predict shape directly from image patch: \( p = F(I) \)

- Haar-like features
- Non-linear Image-based boosting ridge regression (IBRR)
- Single-shot approach, highly efficient
- For robustness, perturb rigid bounding box parameters \( n \) times and average results (weighted by parameter density)

Zhou 2010
IV) Initialization
Global Search with Evolutionary Algorithm

- Initialize population
- Repeat:
  - Evaluation of fitness
  - Selection by random sampling
  - Gaussian mutation

Heimann et al. 2007
Encoding and Evaluation

Individuals in the populations:
• Transform (3x translation, z-rotation, scale)
• Reduced set of shape parameters
• Stored as real-valued vector
• Initialized according to training data

Evaluation of Fitness:
• Evaluate boundary probabilities $p(b|g)$ at each landmark and multiply

$$w_s = \exp \left( \frac{1}{n} \sum_{i} \log p_i(b|g_i) \right)$$

• Statistical evidence for individual shape in the image
Selection and Mutation

Random sampling:
• Probability of reproduction proportional to individual fitness value
• Draw completely new population

Gaussian mutation:
For each element in vector $v$:
$$v_i = v_i + \sigma_t \cdot X, \quad X \sim N(0, \sigma_i^2)$$
• $\sigma_i^2$ variance for individual element
• $\sigma_t$ global perturbation factor, decreased each iteration:
$$\sigma_{t+1} = 0.95\sigma_t$$

1000 individuals over 40 generations sufficient for a reliable, fully automated initialization
Performance Considerations

- Global search runs in strongly down-sampled image
- Reduce SSM resolution by eliminating landmarks with inferior appearance models
- Assure adequate surface coverage by preventing gaps larger than a certain geodesic distance
Machine Learning for Object Detection

- Binary classifier can determine existence of object at specific position
- Exhausting search yields best candidate for entire image
- Also need to consider orientation and scale!

Explosion of hypotheses!
Marginal Space Learning

- Do not search all parameters at once!
- Search marginalized parameter space
- Keep best hypotheses for next stage and expand there

Zheng et al. 2008
Constrained Marginal Space Learning

To increase computational efficiency, limit hypothesis generation to distributions learned from training data.

- Also captures relations between different parameters, e.g. position and orientation

Zheng et al. 2009
MSL Initialization for Liver CTs
Shape Regression Machine

- Learn difference vector from sample to target
- Image-based boosting ridge regression (IBRR) for learning
- Haar-like features
- To increase robustness, predict position from multiple patches
- For each result, run binary classifier to estimate weight and remove outliers

Zhou 2010
### Training Images and Difference Vectors

<table>
<thead>
<tr>
<th>(-15, -12)</th>
<th>(-3, -8)</th>
<th>(-4, -6)</th>
<th>(-5, -17)</th>
<th>(-8, 25)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Image 1" /></td>
<td><img src="image2" alt="Image 2" /></td>
<td><img src="image3" alt="Image 3" /></td>
<td><img src="image4" alt="Image 4" /></td>
<td><img src="image5" alt="Image 5" /></td>
</tr>
<tr>
<td><img src="image6" alt="Image 6" /></td>
<td><img src="image7" alt="Image 7" /></td>
<td><img src="image8" alt="Image 8" /></td>
<td><img src="image9" alt="Image 9" /></td>
<td><img src="image10" alt="Image 10" /></td>
</tr>
<tr>
<td>(-7, -21)</td>
<td>(15, 16)</td>
<td>(15, -6)</td>
<td>(17, 6)</td>
<td>(16, -5)</td>
</tr>
</tbody>
</table>

The images above illustrate the training images and their corresponding difference vectors.
Results for LV Initialization in 2D US Data
Useful Software

Gaussian Mixture Model registration
• C++ and Python source code from:
  http://gmmreg.googlecode.com

MDL 3D correspondence optimization
• C++ source code (ITK style) from:
  http://hdl.handle.net/1926/224

Scikit-Learn Machine learning library
• Python package from:
  http://scikit-learn.org/
References

- Cootes et al. 1995: Active shape models - their training and application. Comput Vis Image Underst 61(1), 38-59
- Cootes et al. 2001: Active appearance models. IEEE PAMI 23(6), 681-685
- Davies et al. 2002: 3D statistical shape models using direct optimisation of description length. Proc ECCV, 1-17
- Heimann et al. 2007: A Shape-Guided Deformable Model with Evolutionary Algorithm Initialization for 3D Soft Tissue Segmentation. Proc IPMI, 1-12
- Jian & Vemuri 2011: Robust Point Set Registration Using Gaussian Mixture Models. IEEE PAMI 33(8), 1633-1645
- Zheng et al. 2008: Four-Chamber Heart Modeling and Automatic Segmentation for 3D Cardiac CT Volumes Using Marginal Space Learning and Steerable Features. IEEE TMI 27(11), 1668-1681
- Zhou 2010: Shape regression machine and efficient segmentation of left ventricle endocardium from 2D B-mode echocardiogram. MedIA 14, 563-581
Acknowledgements:

• German Cancer Research Center DKFZ, Heidelberg: my own work on SSMs originates there…

• My colleagues Yefeng Zheng and Kevin Zhou (Siemens CT ICV, Princeton) for providing additional material

Thank you for listening!

Questions?

Contact: tobias.heimann@siemens.com