Shape Matching & Object Recognition Using Shape Contexts

~ S. Belongie, J. Malik, J. Puzicha

Presented by Medha Bhargava

Introduction

- **Goals:**
  - Shape matching and isolated object recognition
  - General object categorization
- **Basic Idea:**
  - Find correspondences between points on shapes
  - Estimate aligning transformation
  - Measure dissimilarity
    - Σ(matching errors) + magnitude of aligning transform

Motivation

- "Related but not identical shapes can often be deformed into alignment"
  - D'Arcy Thompson, *On Growth and Form* (1917)
- Biological shapes

D'Arcy Thompson's Cartesian transformation of a *human* skull to the skull of a *chimpanzee* and a *baboon*

Related Work: Deformable Templates

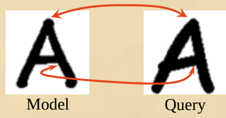
- Energy minimization in a mass-spring model
  - Fischler and Elschlager (1973)
- Probabilistic models
  - Grenader et al (1991)
- Model fitting using Gradient descent
  - Yuille (1991)
- Elastic graph matching
  - Lades et al (1993)

Shape Matching

- **Feature-based**
  - Utilize spatial relationships between features
    - Silhouettes, edges, keypoints
    - Distance Transforms
    - Geometric Hashing
    - Decision trees
- **Brightness-based**
  - Utilize pixel brightness directly
    - Registration methods based on grayscale values
    - Learning algorithms, including SVMs, neural networks, etc

Proposed Algorithm

Finding Correspondences



- Sample shape contours to extract a set of points
  - Need not be landmark points or curvature extrema
  - Uniform spacing
  - The more the merrier! (~100 used)
- Use **shape context** as shape descriptor
- Compare point sets

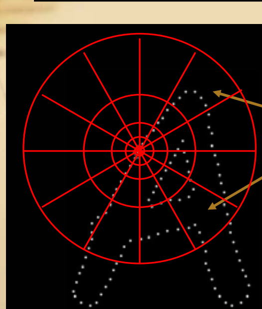
Shape Context

- Rich local descriptor
- Records distribution of relative positions of points
  - i.e. incorporates global shape information
- For each point p_i, a coarse histogram h_i of relative coordinates of remaining (n-1) points is computed

$$h_i(k) = \#\{q \neq p_i : (q - p_i) \in bin(k)\}$$

- Bins uniform in log-polar space

Computing Shape Context



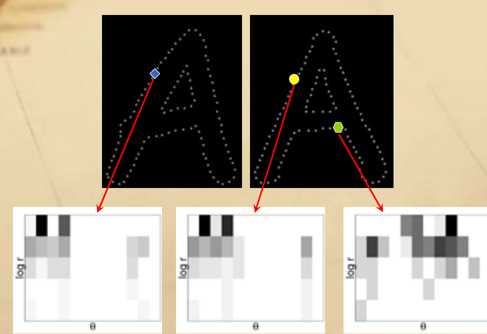
count = 6

count = 10

5 bins for $\log r$; 12 bins for θ

- Sample shape contours
- Count no. of points/bin
- Compute **log-polar** histogram

Shape Contexts



Comparing Shape Contexts

- Use χ^2 test statistic to compute matching costs

$$C_{ij} = C(p_i, q_j) = \frac{1}{2}\sum_{k=1}^{K}\frac{[h_i(k) - h_j(k)]^2}{h_i(k) + h_j(k)}$$

- May incorporate additional local appearance similarity
- Minimize total cost of matching

$$H(\pi) = \sum_i C(p_i, q_{\pi(i)})$$

- Weighted bipartite graph matching problem

Matching

- Recover correspondences by solving linear assignment problem with costs C_{ij}

Why Shape Context?

- Invariant under **scale** and **translation**
  - Normalize radial distances by mean distance between all n^2 point pairs
- Can be made **rotation**-invariant
  - Place reference frame with tangent as x-axis
- Robust to partial **occlusion** and **noise**
- Can handle **outliers**
  - Use 'dummy' nodes with fixed error ϵ_d
- Tolerant to small **locally affine distortions**

Estimation of Transformation

- Map one shape onto the other
- Degree of alignment provides a convincing measure of shape similarity
- Affine model standard among several families
- Regularized **thin plate splines** (TPS) transformation employed here
  - Popular representation of flexible coordinate transformations

Thin Plate Splines

- "Given a set of data points, a weighted combination of TPS centered about each point gives the interpolation function $f(x,y)$ that passes through the points exactly while minimizing the *bending energy*."

  \- S. Belongie. From *MathWorld*

- 2D analog to cubic spline, and has the form

$$U(r) = r^2 \log r \qquad r > 0$$

TPS – The Math

- **Bending energy:**

$$I_f = \iint_{\mathbb{R}^2} \left(\frac{\partial^2 f}{\partial x^2}\right)^2 + 2\left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 + \left(\frac{\partial^2 f}{\partial y^2}\right)^2 dxdy$$

- **Regularization:** to relax the requirement that $f(x,y)$ pass through the data points exactly

$$H[f] = \sum_{i=1}^{n} \left(v_i - f(x_i, y_i)\right)^2 + \lambda I_f$$

Regularization parameter

- Degenerates to the least-squares affine model when regularized

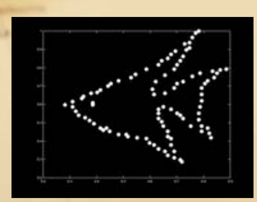
TPS Modeling

- Co-ordinate transformation:

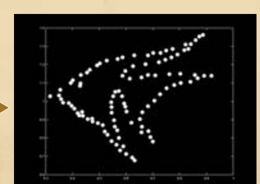
$$T(x,y) = \left(f_x(x,y), f_y(x,y)\right)$$

- Algorithm iterated for better performance
- Jitter noise in correspondences smoothed out during alignment

Evaluation of Method

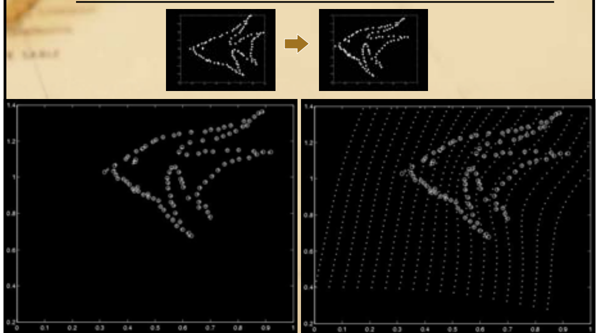


Model



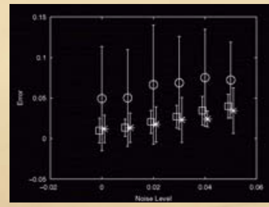
Synthetically designed target

Illustration



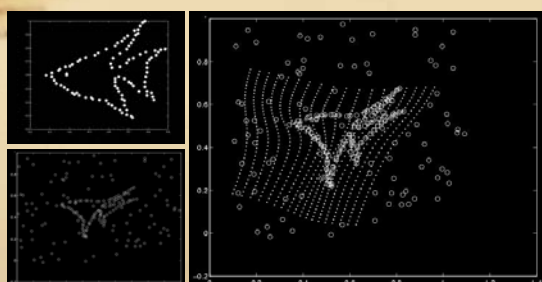
Synthetic Test Results

Fish - deformation + noise



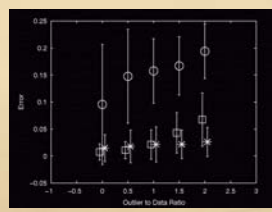
○ ICP □ Shape Context ✱ Chui & Rangarajan

Evaluation with Outliers



Outlier Test Results

Fish - deformation + outliers



○ ICP □ Shape Context ✱ Chui & Rangarajan

Object Recognition

- Prototype-based recognition
- (k-)Nearest-neighbor method
- **Shape Distance:**

$$D_{shape} = \underbrace{\alpha_1 D_{SC}}_{\text{Shape context distance}} + \underbrace{\alpha_2 D_{AC}}_{\text{Appearance cost}} + \underbrace{\alpha_3 I_f}_{\text{Bending Energy}}$$

- Shape Context matching cost:

$$D_{SC}(P,Q) = \frac{1}{n}\sum_{p \in P} \arg\min_{q \in Q} C\left(p, T(q)\right) + \frac{1}{m}\sum_{q \in Q} \arg\min_{p \in P} C\left(p, T(q)\right)$$

  – symmetric sum over best matches

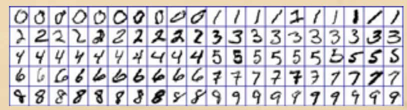
Experiments & Results

Experiments

- Handwritten digits - MNIST
- Silhouettes – MPEG-7
- 3D objects – COIL-20
- Trademarks
- Human Body Poses

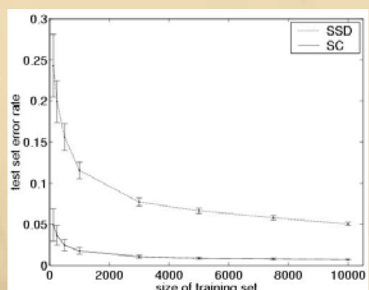
Digit Recognition

- **MNIST** dataset



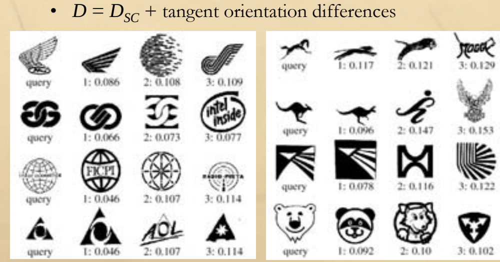
  - 60,000 training and 10,000 testing digits
  - Local tangent angle information incorporated
  - $D_{shape} = D_{SC} + 1.6 D_{AC} + 0.3 I_f$
  - Error-rate: 0.63% (63 errors out of 10000 samples)
    - Using 3-NN and 20000 training samples

Digit Recognition Results



Trademark Similarity

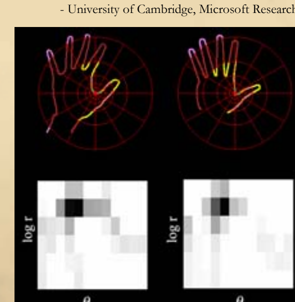
- Database of 300 trademarks
- Affine transformation model;
- $D = D_{SC}$ + tangent orientation differences



Other Experiments

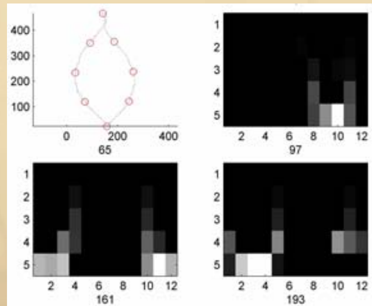
Hand Tracking using Shape Context

- University of Cambridge, Microsoft Research (CVPR 2003)

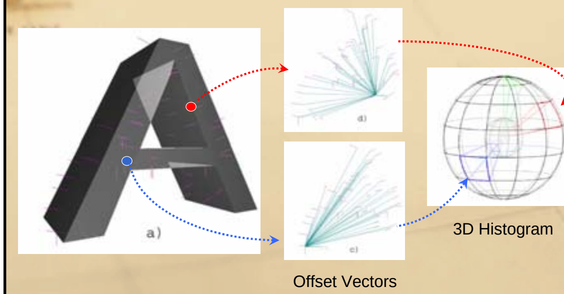


Leaf Classification using Shape Context

- University of Maryland



Extending Shape Contexts to 3D



Quick Summary

In Conclusion

- Invariant under translation, scale, rotation and local affine distortions
- Robust to noise, outliers and occlusion
- Applicable in a variety of domains
- Can handle large shape variations from templates
- Not invariant in clutter
- Not feasible for real-time applications

Questions & Comments