2D Shape Matching (and Object Recognition)

Raghuraman Gopalan

Center for Automation Research
University of Maryland, College Park
Outline

- What is a shape?
- Part 1: Matching/Recognition
  - Shape contexts [Belongie, Malik, Puzicha – TPAMI ’02]
  - Indexing [Biswas, Aggarwal, Chellappa – TMM ’10]
- Part 2: General discussion
Some slides were adapted from Prof. Grauman’s course at Texas Austin, and Prof. Malik’s presentation at MIT
Where have we encountered shape before?

Edges/ Contours

Silhouettes
A definition of shape

- **Defn 1:** A set of points that collectively represent the object
  - We are interested in their location information alone!!

- **Defn 2:** Mathematically, shape is an equivalence class under a group of transformations
  - Given a set of points $X$ representing an object $O$, and a set of transformations $T$, shape $S = \{t(X) \mid t \in T\}$
  - Issues? – Kendall ‘84
Applications of Shapes

Analysis of anatomical structures
Figure from Grimson & Golland

Recognition, detection
Fig from Opelt et al.

Pose

Morphology
http://usuarios.lycos.es/lawebdelosfosiles/i

Characteristic feature
Fig from Belongie et al.
Part 1: 2D shape matching
Recognition using shapes – (eg. – model fitting)

For example, the model could be a line, a circle, or an arbitrary shape.
Example: Deformable contours


Applications:
- Traffic monitoring
- Human-computer interaction
- Animation
- Surveillance
- Computer Assisted Diagnosis in medical imaging
Issues at stake

- Representation
  - Holistic
  - Part-based
- Matching
  - How to compute distance between shapes?
- Challenges in recognition
  - Information loss in 3D to 2D projection
  - Articulations
  - Occlusion…. Invariance???
  - Any other issue?
Representation [Veltkamp '00]

- Holistic
  - Moments
    
    The \((p, q)\)-moment of an object \(O \subseteq \mathbb{R}^2\) is given by
    
    \[
    m_{p,q} = \int_{(x,y) \in O} x^p y^q \, dx \, dy
    \]
  
    - Fourier descriptors
  
  - Computational geometry
  
  - Curvature scale-space
Figure 4: Contour evolution reducing curvature changes, see
Part-based

- medial axis transform – shock graphs
Fig. 1. Examples of two handwritten digits. In terms of pixel-to-pixel comparisons, these two images are quite different, but to the human observer, the shapes appear to be similar.
Discussion for a set of points

- Hausdorff distance

Hausdorff distance (HD)

For two sets of points $A = \{a_1, \ldots, a_m\}$ and $B = \{b_1, \ldots, b_n\}$

$$H(A, B) = \max(h(A, B), h(B, A))$$

where

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|$$
Chamfer distance

- Average distance to nearest feature

\[ D_{\text{chamfer}}(T, I) \equiv \frac{1}{|T|} \sum_{t \in T} d_I(t) \]

- \( T \): template shape \( \rightarrow \) a set of points
- \( I \): image to search \( \rightarrow \) a set of points
- \( d_I(t) \): min distance for point \( t \) to some point in \( I \)
Chamfer distance

\[ D_{\text{chamfer}}(T, I) \equiv \frac{1}{|T|} \sum_{t \in T} d_I(t) \]

How is the measure different than just filtering with a mask having the shape points?

Edge image
Distance Transform is a function $D(\cdot)$ that for each image pixel $p$ assigns a non-negative number $D(p)$ corresponding to distance from $p$ to the nearest feature in the image $I$.

Features could be edge points, foreground points,…
Distance transform

Value at (x,y) tells how far that position is from the nearest edge point (or other binary image structure)

>> help bwdist
Chamfer distance

- Average distance to nearest feature

\[ D_{\text{chamfer}}(T, I) \equiv \frac{1}{|T|} \sum_{t \in T} d_I(t) \]
Chamfer distance

Fig from D. Gavrila, DAGM 1999
A limitation of active contours

- External energy: snake does not really “see” object boundaries in the image unless it gets very close to it.

![Diagram of image gradients](image.png)

\[ \nabla I \]

image gradients \( \nabla I \) are large only directly on the boundary
What limitations might we have using only edge points to represent a shape?

How descriptive is a point?
Comparing shapes

What points on these two sampled contours are most similar? How do you know?
Shape context descriptor [Belongie et al ’02]

Count the number of points inside each bin, e.g.:

- Count = 4
- Count = 10

Compact representation of distribution of points relative to each point.
Shape context descriptor
Comparing shape contexts

Compute matching costs using Chi Squared distance:

\[ C_{ij} = \frac{1}{2} \sum_{k=1}^{K} \left[ \frac{h_i(k) - h_j(k)}{h_i(k) + h_j(k)} \right]^2 \]

(Then use a deformable template match, given the correspondences.)
Invariance/ Robustness

- Translation
- Scaling
- Rotation

- Modeling transformations – thin plate splines (TPS)
  - Generalization of cubic splines to 2D

- Matching cost = f(Shape context distances, bending energy of thin plate splines)
  - Can add appearance information too
  - Outliers?
An example of shape context-based matching

Fig. 1. Examples of two handwritten digits. In terms of pixel-to-pixel comparisons, these two images are quite different, but to the human observer, the shapes appear to be similar.

Fig. 4. Illustration of the matching process applied to the example of Fig. 1. Top row: 1st iteration. Bottom row: 5th iteration. Left column: estimated correspondences shown relative to the transformed model, with tangent vectors shown. Middle column: estimated correspondences shown relative to the untransformed model. Right column: result of transforming the model based on the current correspondences; this is the input to the next iteration. The grid points illustrate the interpolated transformation over \( \mathbb{R}^2 \). Here, we have used a regularized TPS model with \( \lambda_v = 1 \).
Some retrieval results
Efficient matching of shape contexts
[Mori et al ’05]

Fig. 2. Matching individual shape contexts. Three points on the query shape (left) are connected with their best matches on two known shapes. $L^2$ distances are given with each matching.
Efficient matching of shape contexts

[Mori et al ’05] – cont’d

- Vector:

Fig. 3. (a) Line drawings. (b) Sampled points with shape label. \(k = 100\) shapes were extracted from a known set of 260 shapes (26,000 generalized shape contexts). Note the similarities in shape label (2, 41 on the left side, 24, 86, 97 on the right side) between similar portions of the shapes.
Are things clear so far?
Fig. 1. Three objects. The dashed lines denote shortest paths within the shape boundary that connect landmark points.
Inner distance vs. (2D) geodesic distance

Fig. 2. Geodesic distances on 2D shapes. Using the geodesic distances along the contours, the two shapes are indistinguishable.
The problem of junctions

- Is inner distance truly invariant to articulations?

Fig. 5. Articulated objects. (a) An articulated shape. (b) Overlapping junctions (the five dark areas). (c) Ideal articulation.
Theorem. Let $O$ be an articulated object and $f$ be an articulation of $O$ as defined above. $\forall x, y \in O$, suppose the shortest path $\Gamma(x, y; O)$ goes through $m$ different junctions in $O$ and $\Gamma(x', y'; O')$ goes through $m'$ different junctions in $O'$, then

$$|d(x, y; O) - d(x', y'; O')| \leq \max\{m, m'\} \epsilon. \quad (5)$$
Non-planar articulations? [Gopalan, Turaga, Chellappa ’10]

![Diagram showing non-planar articulations]

(a) (b)

\[ \| \mathbf{d}_{21} - \mathbf{d}_{22} \|_2 < \| \mathbf{d}_{22} - \mathbf{d}_{23} \|_2. \]

(b) Keypoints with similar shape description obtained from our method. Points were picked in the first frame, and their ‘nearest neighbors’ are displayed in other two frames. No holistic shape matching was done, emphasizing the importance of a shape representation. *(All figures are best viewed in color).*
Indexing approach to shape matching
[Biswas et al ’10]

- Why?

Fig. 1. A few applications that can benefit from robust and efficient shape matching. (a) Matching and retrieval of 2-D shapes [1], like trademark retrieval 2], leaf recognition [3], etc. (b) Activity classification [4]. (c) Gesture recognition. (d) Pose estimation in sports clips [5].
Indexing - framework

Pair-wise Features:
Inner distance, Contour length, relative angles etc.

Fig. 2. Prototype of the proposed shape indexing framework. Each shape in the database is indexed to a hash table using a set of indexable feature vectors extracted from the shape.
Fig. 6. (Left) Retrieval algorithm. (Right) Post-retrieval rank refinement to improve accuracy.
Part 2 – Shapes as equivalence classes

- Kendall’s shape space
  - Shape is all the geometric information that remains when the location, scale and rotation effects are filtered out from the object
Pre-shape

- Kendall’s Statistical Shape Theory used for the characterization of shape.
- Pre-shape accounts for location and scale invariance alone.
- $k$ landmark points ($X:k \times 2$)
- Translational Invariance: Subtract mean
- Scale Invariance: Normalize the scale

$$Z_c = \frac{CX}{\|CX\|}, \quad \text{where} \quad C = I_k - \frac{1}{k} 1_k 1_k^T$$

Some slides were adapted from Dr. Veeraraghavan’s website
Feature extraction

- Silhouette
  - Landmarks
  - Centered Landmarks
  - Pre-shape vector
Shape lies on a spherical manifold. Shape distance must incorporate the non-Euclidean nature of the shape space.
Affine subspaces [Turaga, Veeraraghavan, Chellappa '08]

Figure 1. Synthetic data generated from the MPEG database. The first column shows base-shapes from the original MPEG dataset for 5 objects. The remaining columns show random affine warps for the base shapes with increasing levels of additive noise.
Other examples

- Space of Blur
- Modeling group trajectories etc..
Conclusion

- Shape as a set of points configuring the geometry of the object
  - Representation, matching, recognition
  - Shape contexts, Indexing, Articulation
- Shape as equivalence class under a group of transformations
Announcements

- HW 5 will be posted today;
  - On Stereo, and Shape matching
  - Due Nov. 30 (Tuesday after Thanksgiving)

- Turn in HW 4

- Midterms solutions by weekend
Questions?