CSE 252C: Computer Vision III

Lecturer: Serge Belongie
Scribes: Andrew Rabinovich and Vincent Rabaud
Edited by: Catherine Wah

LECTURE 12
Shape-Based Recognition

12.1. Introduction

What is a shape? That is one of many what-is-X questions in vision (see also X = texture), and they are not easy to answer.

One informal definition of shape is what's left after factoring out color, texture, motion, shading, blur, *etc*. For example, line drawings are often "pure shape"; other instances include MNIST digits, trademarks, and Attneave's cat.

Formally, *shape* can be thought of as characteristics of a point set that do not depend on transformations (such as similarity or affine). More flexible transformations can also be considered, as in D'Arcy Thompson examples; this motivated Bookstein to propose the use of TPS for biological shape studies (morphometrics). We can still use the bending energy, *etc*. as a penalty term, but the key is that it is separated out.

In this discussion, we make some assumptions:

[1]Department of Computer Science and Engineering, University of California, San Diego.

October 11, 2009

- Known correspondences: points and/or edge contours are extracted and the correspondences are known (not "meaningful" features or parts, just samples)
- Family of transformation is specified (*e.g.* similarity, affine, regularized TPS).

As we've seen, the two aforementioned tasks are nontrivial in themselves and are very likely linked, but in this lecture we focus on what follows from these assumptions.

12.2. Without correspondences

First let's see what we can do if correspondences are not known, *i.e.* the input is two point sets (let's assume 2D) and we'd like to measure their similarity. Here we see a familiar progression, with simpler techniques based on moments (*e.g.* mean, covariance, *etc.*) and more advanced techniques based on nonparametric density estimates (*e.g.* histograms of various measurements on the point sets).

12.2.1. Moment-based shape matching

As we saw when doing least squares affine transform estimation, we generally start by eliminating relative translation: we subtract centroids, computed as the *first moment* (the center of mass or center of gravity):

$$\boldsymbol{\mu} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i, \text{ where } \boldsymbol{x} = \begin{bmatrix} x_i \\ y_i \end{bmatrix}. \tag{12.1}$$

Successive moments are computed as:

$$m_{k,l} = \frac{1}{N}\sum_{i=1}^{N} x_i^k y_i^l \tag{12.2}$$

(for a centered version, or "central moments," the mean needs to be subtracted). For example, the 2×2 symmetric *scatter matrix*, or centered second moment matrix, is:

$$C = \frac{1}{N}\sum_{i=1}^{N} (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\top. \tag{12.3}$$

The first moment is generally discarded; the second moment allows affine alignment without correspondences. Then one can use a variety of packagings of higher order moments such as Hu, Legendre, or Zernike. For example, Hu (1962) proposed a set of seven invariant moments up to third order. While elegant, these approaches have serious problems with clutter and occlusion.

12.2.2. Global shape distribution-based methods

To address these issues, several approaches have been developed based on distributional shape descriptors, *e.g.* Ankerst *et al.* (1999), shape context from Belongie *et al.* (2000), spin image from Johnson & Hebert (1997), and Funkhouser *et al.* Several of these were proposed as local shape descriptors, for use in jointly estimating correspondences and shape similarities, but one can also consider their global versions.

The histograms are often computed on derived measurements, for instance, Funkhouser's D2 histogram estimates density over Euclidean distances between point pairs. In addition, as another example, his D4 measure is the cube root of the volume of the tetrahedrom between four points.

12.2.3. "Shape context" matching

These approaches are cheap, powerful (in a bag of features kind of way), and more robust to occlusion and clutter, but naturally limited compared to shape comparison methods that exploit correspondence.

One such method that incororates correspondences is "shape context" matching. It serves as a cost for estimating correspondences, as well as an overall shape similarity score when averaged over all point pairs for two aligned shapes. Generally, it is applied iteratively with an affine or regularized TPS transform. Some speedup was achieved with later work with shapemes and random representations, while Ling & Jacobs (2005) addressed articulation with inner distance shape context.

Note that in addressing shape in this manner (just assuming a sampled distribution of points in $\mathbb{R}^2$, with no figural continuity or closed curve assumption in general), we've taken a decidedly statistical approach; if in some way you are able to extract clean silhouette contours (especially closed external boundaries), a completely different (and elegant) set of approaches can be applied. Examples in this vein include curvature scale space (CSS) (Mokhtarian, 1998) and a variety of methods based on dynamic programming, Fourier/wavelet descriptors, edit distances, *etc.*

Unfortunately, this type of clean contour is somewhat elusive in natural images, but in certain constrained environments (*e.g.* http://www.like.com) one can make great use of such methods.

Speaking of natural images, a few approaches have been proposed that migrate the distributional shape matching concept away from abstract point sets and closer to raw images:

- Geometric blur (Berg & Malik, 2001)
- TextonBoost (Shotton *et al.* , 2007)

When meaningful parts of an image (*e.g.* patch features or contour framents) can be identified, one moves into the territory known as "constellation based matching," which we'll discuss next time.