Computer Vision
ETH Zürich
Fall 2020

Lab Assignment 8 - Shape Context

Mokhles Bouzaien

1 Shape Context Descriptors

I started by loading and plotting the data. As we can see in figure 1, there are 3 object classes (heart, fork and watch) and 5 instances for each class. The next step

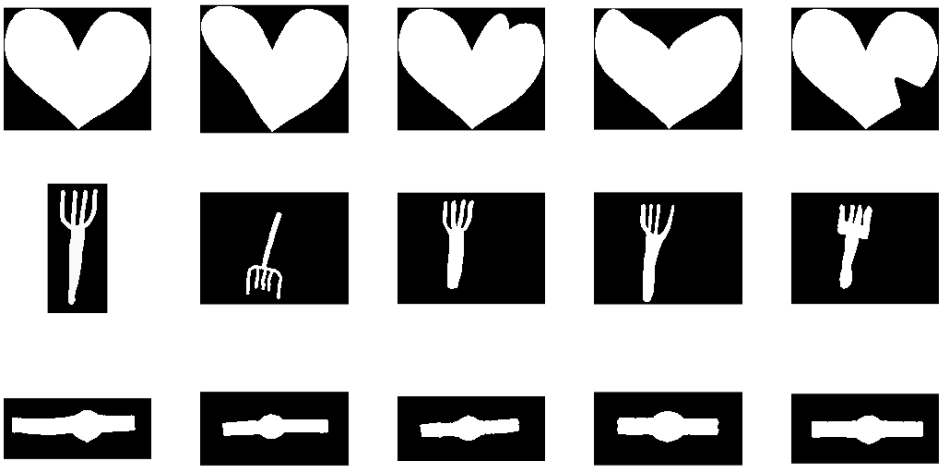

Figure 1: Object classes and instances.

is to calculate the shape context descriptors for each pixel of the object edge using the `sc_compute` function. At a high level, this function computes the distance r and orientation θ between all pair of points (x_i, x_j) where $i \neq j$. And then create, for each point, the histogram of relative coordinates using predefined bins as described in the paper.

2 Cost Matrix

The `chi2cost` function is used to calculate the cost matrix C between 2 shapes. The element C_{ij} is the cost between the point p_i of shape 1 and point q_j of shape 2. This is done after sampling the same number of points (100 samples) from the source and target shapes (figure 2).

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

where $h_i(k)$ is the element k of histogram of relative coordinates of point p_i. ϵ is added to avoid division by zero in case of $h_i(k) = h_j(k) = 0$.

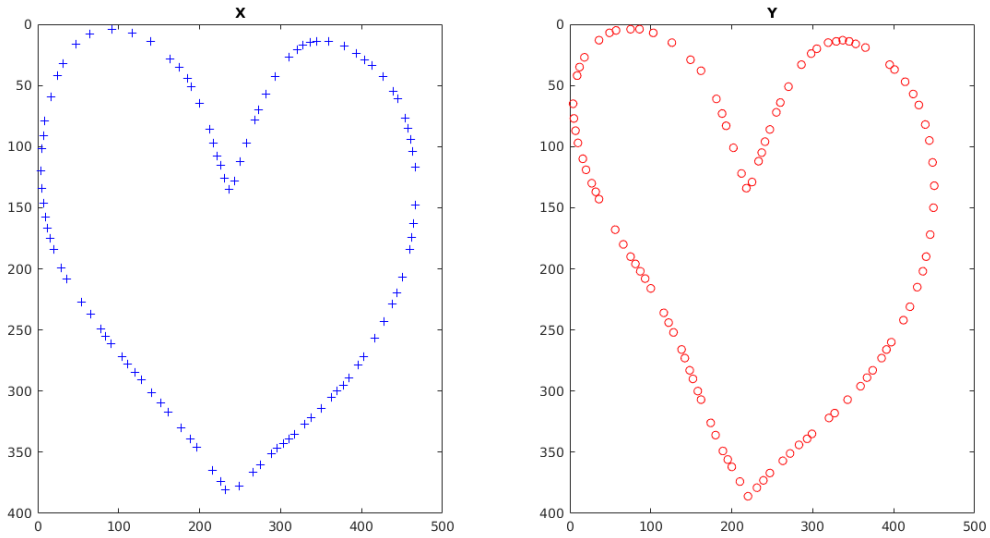

Figure 2: Sampled points of the source and target shapes.

3 Thin Plate Splines

In this part, I will try to estimate the transformation to map points from the source to the target object. This is done by:

- Defining $U(t) = t^2 \log(t^2)$
- Calculating matrix K such that $K_{ij} = U(\|p_i - p_j\|)$
- Defining $P = [1, X]$
- Solving

$$\begin{pmatrix} K + \lambda I & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} w_x \\ a \end{pmatrix} = \begin{pmatrix} v_x \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} K + \lambda I & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} w_y \\ a \end{pmatrix} = \begin{pmatrix} v_y \\ 0 \end{pmatrix}$$

  where v_x and v_y are the target x and y coordinates.

- Calculating the energy $E = w_x^T K w_x + w_y^T K w_y$

As we can see in the example figure 4, mismatched pointed can affect the calculated transformation.

Is the shape context descriptor scale-invariant? The relatives distances are normalized before calculating the histogram of the distance distribution which makes the descriptor scale invariant.

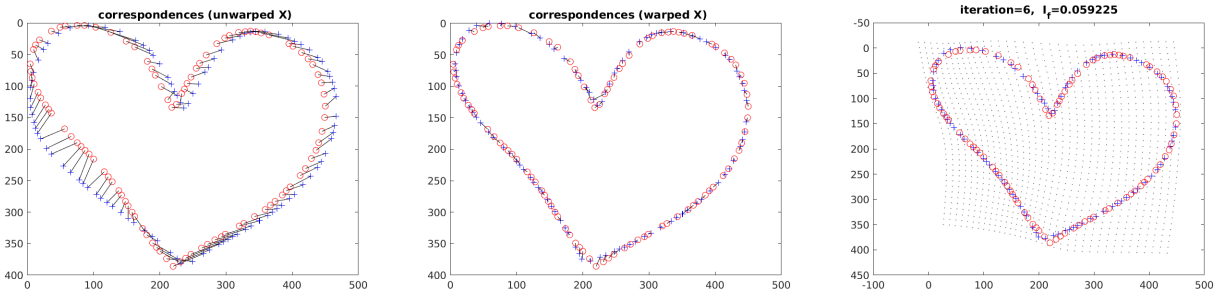

Figure 3: Matched points and the estimated transformation.

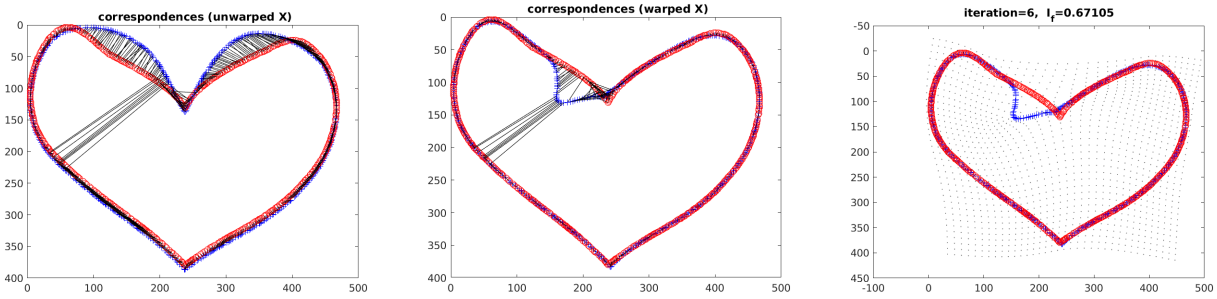

Figure 4: Mismatched points and the resulted transformation.