Report of Anton Maksimov (antonma, 16-952-137), Task 8 "Shape Context"
on ETHZ course "Computer Vision".

---

1. To compute shape descriptors we tried different strategies varying on method to assign angular component of the descriptor (flags to change mode are in the section of parameters in `shape_matching.m`). Suppose we have current point with position p and we processing at the moment point at the distance between $r =$ `smallest_r` and $R =$ `biggest_r` distances. We calculate squared distance and place it in correspondent bin in between

1. «local» — angle as arctan of angle between $p1-p$ and X–axis. Problem is that it isn't rotationally invariant, so works well only when sets of points are not rotated initially.
2. «global» — we find difference of angles of $p1-c$ and $p-c$, where c is the center of the point cloud $(\max + \min)/2$ for both coordinates. Also isn't rotationally invariant and could work better for complex structures that the «local».
3. «tangent» — we find the closest point p_c to p, so get tangent angle τ as angle between $p_c - p$ and X axis. Then we calculate angles as in «local» mode, but subtracting τ. As sampling is more or less uniform (when sample many enough points), the tangent direction could be well approximated with $p_c - p$.

Problems with «tangent» mode. In paper attached to the task it's mentioned that for rotation–invariant descriptor it's better to use local system, such that calculate angles relatively to tangent direction. But it isn't said how to choose direction from two. We decided to calculate cost in the «tangent» mode such that we take the minimum value of two: when one descriptor is normal and when it's inverted with respect to angle values (if number of angular bins is even, it will mean that we changed direction tangent direction by $\pi/2$).
Another problem is that many figures are symmetric (hearts, forks are also), so points which are symmetric could be easily assigned to the opposite side and the figure will bend much, especially when number of iteration is big. So we used 3 iterations. After that it's clearly seen that parts which were assigned to the right side are fit well, though parts with incorrect side assignment try to bend more and more with every new iteration (fig. 1).
Also because of this bending we have to assign bigger $\lambda = 10 \cdot \text{mean\_distance}^2$ for smoothing. We don't change number of bins and rescaled r and R, because they shouldn't depend on image in reality a priori (or we should use some automatic assignment of them).
Conclusion In general, without removing outliers «tangent» mode works worse (+ because of the symmetry), but if remove outliers it should be better. It could be possible to use «dummy» points in template, which have some threshold level of cost for all points in image which we are trying to warp.

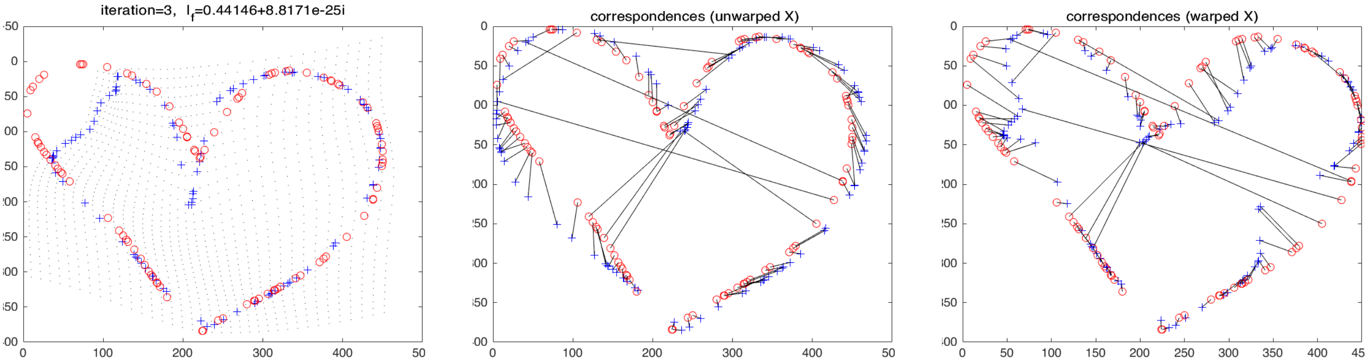

Figure 1: «Tangent» mode, $\lambda = 10 \cdot \text{mean\_distance}^2$, 3 iterations.

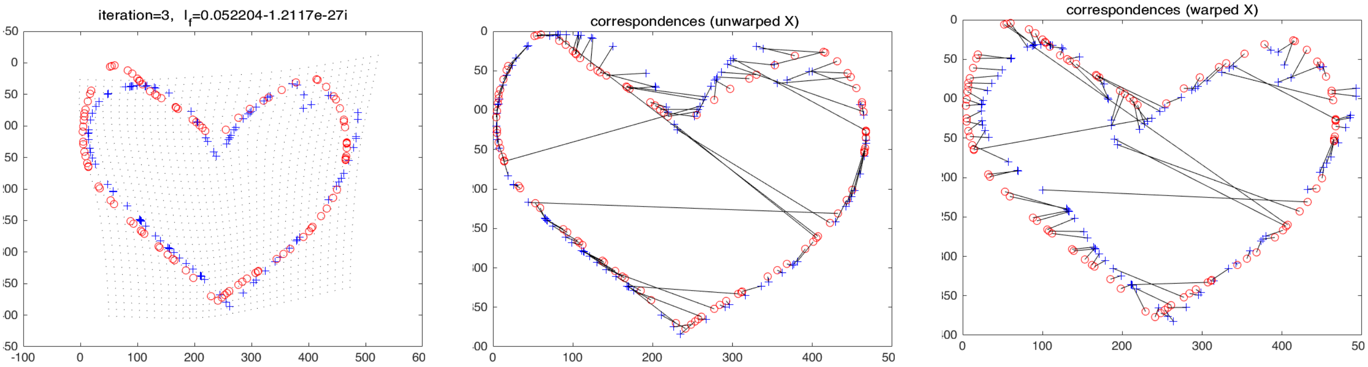

Figure 2: «Local» mode, $\lambda = 10 \cdot \text{mean\_distance}^2$, 3 iterations.

Answer to question. Yes, shape–context descriptors are scale–invariant (all three modes we mentioned in the beginning), because they use points as the origin of local coordinate system and distances are normalized using mean distance between points, which scales equally with distances between selected point and others.

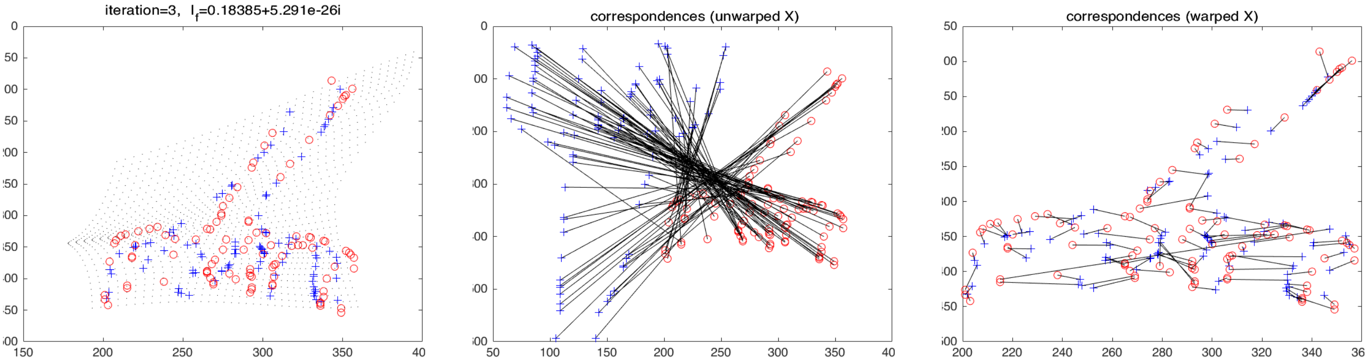

Figure 3: «Local» mode, $\lambda = 10 \cdot \text{mean\_distance}^2$, 3 iterations.

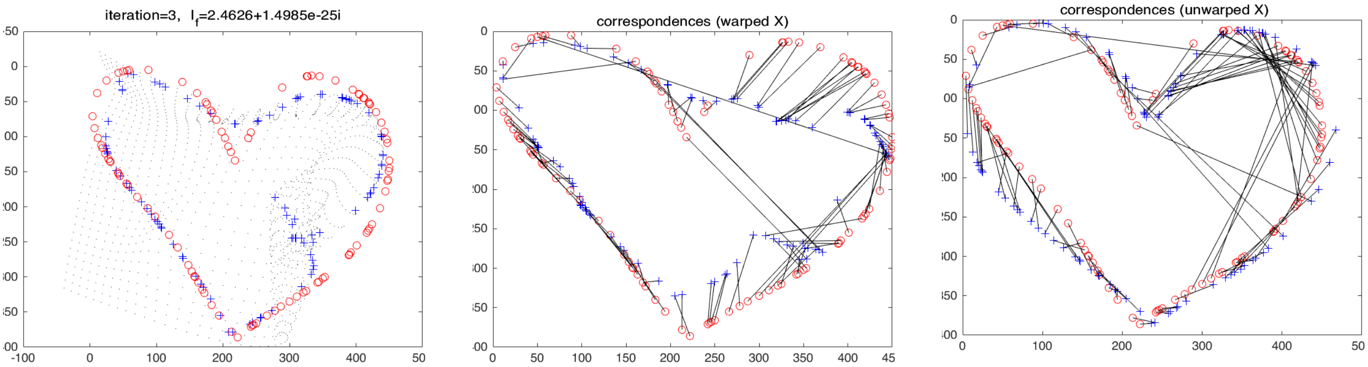

Figure 4: «Global» mode, $\lambda = 10 \cdot \text{mean\_distance}^2$, 3 iterations.