

SOFT SHAPE CONTEXT FOR ITERATIVE CLOSEST POINT REGISTRATION

David Liu and Tsuhan Chen

dliu@cmu.edu, tsuhan@cmu.edu

Department of Electrical and Computer Engineering
Carnegie Mellon University

ABSTRACT

This paper introduces a shape descriptor, the soft shape context, motivated by the shape context method. Unlike the original shape context method, where each image point was hard assigned into a single histogram bin, we instead allow each image point to contribute to multiple bins, hence more robust to distortions. The soft shape context can easily be integrated into the iterative closest point (ICP) method as an auxiliary feature vector, enriching the representation of an image point from spatial information only, to spatial and shape information. This yields a registration method more robust than the original ICP method. The method is general for 2D shapes. It does not calculate derivatives, hence being able to handle shapes with junctions and discontinuities. We present experimental results to demonstrate the robustness compared with the standard ICP method.

1. INTRODUCTION

Iterative closest point (ICP) [1] registration is a popular method for aligning two point sets. The registration process iterates two steps repeatedly: find the best correspondence between the two point sets based on spatial distance, and update the transformation based on the found correspondence. Assuming the closest points provide an estimate of the correct correspondence is often not a valid assumption, thus making ICP sensitive to the initial alignment. To address this problem, methods have been proposed to improve the step of finding correspondence by using extra cues like color [2], surface normal, curvature, gradient information [3], and global features like moment invariants and spherical harmonics invariants [4].

In this paper we consider the registration of 2D point sets in general, as opposed to single closed curves [5] or simple polygons (polygons without holes)[6]. A 2D point set often has 3-way or 4-way junctions, and it is difficult

to define the gradient or curve normal at a junction, hence ICP variations for 3D surfaces [3] do not directly apply to the 2D case. Also, while moment invariants and spherical harmonics invariants provide invariant features [4], they are global features and hence do not apply to situations where one point set is a subset of the other, as happens in partial matching. We propose to incorporate into ICP a shape descriptor that does not rely on derivatives, hence being able to handle shapes with junctions and discontinuities.

The paper is organized as follows. Section 2 defines the problem and notations. In Section 3 and 4 we introduce two shape descriptors. Section 5 incorporates our shape descriptor into the ICP framework. Finally we present experimental results in Section 6.

2. THE ICP FRAMEWORK

Given two sets of 2-dimensional points, called the model set \mathbf{M} and data set \mathbf{D} , with their elements denoted by a sequence of position vectors $\{\mathbf{m}_i\}_{i=1}^{N_M}$ and $\{\mathbf{d}_i\}_{i=1}^{N_D}$, $\mathbf{m}_i, \mathbf{d}_i \in \mathcal{R}^{2 \times 1}$, our task is to find the parameter \mathbf{a} of a transformation T that best aligns the two point sets in the presence of outliers in set \mathbf{M} . For 2D Euclidean transformation,

$$T(\mathbf{d}_i; \mathbf{a}) = T(\mathbf{d}_i; \theta, t_x, t_y) = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \mathbf{d}_i + \begin{pmatrix} t_x \\ t_y \end{pmatrix}$$

so that $\mathbf{a} = [\theta, t_x, t_y]$. For 2D Affine or Projective transformations, the parameter \mathbf{a} becomes a 6-element vector or 8-element vector, respectively.

In standard ICP, the following function is minimized with respect to \mathbf{a} to achieve alignment

$$E(\mathbf{a}) = \sum_{i=1}^{N_d} \min_j \varepsilon_{dist} \left(\left| \mathbf{m}_j - T(\mathbf{d}_i; \mathbf{a}) \right| \right)$$

with the typical error function $\varepsilon_{dist}(|\mathbf{x}|) = \|\mathbf{x}\|^2$, and the goal is to find \mathbf{a} to minimize $E(\mathbf{a})$. To do this, ICP runs two steps iteratively:

Step 1. Finding the correspondences, $\phi(\cdot)$:

$$\phi(i) = \arg \min_{j \in \{1, \dots, N_m\}} \varepsilon_{dist} \left(\left| \mathbf{m}_j - T(\mathbf{d}_i; \mathbf{a}_k) \right| \right), \quad i = 1, \dots, N_d$$

Step 2. Updating the transformation, \mathbf{a} :

$$\mathbf{a}_{k+1} = \arg \min_{\mathbf{a}} \sum_{i=1}^{N_d} \varepsilon_{dist} \left(\left| \mathbf{m}_{\phi(i)} - T(\mathbf{d}_i; \mathbf{a}) \right| \right)$$

Instead of using $\varepsilon_{dist}(\|\mathbf{x}\|) = \|\mathbf{x}\|^2$, other variations use the Huber robust estimator in Step 2 [7], or exclude points with large errors at Step 1 [8]. Termination can be determined when the correspondences found in Step 1 no longer change. Convergence to local (not global) minimum is guaranteed, since both steps reduce the error.

A successful alignment process based on ICP relies on the assumption that the closest points provide a reasonable estimate of the correct correspondence. This assumption, however, is rarely met and might lead to unpredictable results when the background clutter (outliers) is too heavy, or when the initial alignment is not good enough. An example of ICP registration is shown in Table 2. It is natural to think that incorporating information about the shape should benefit the correspondence step. In the following sections we show how to do that.

3. SHAPE CONTEXT DESCRIPTOR

The work in [9] introduced a pure shape based matching technique based on a shape descriptor, the *shape context*. Intuitively, consider any point $\mathbf{d}_i \in \mathbf{D}$ as a reference point, and all the other $N_d - 1$ points as vectors originating from that reference point. Then these vectors express the shape of the entire point set \mathbf{D} relative to the reference point.

More precisely, the shape context of an image point \mathbf{d}_i is a histogram which describes the relative position of the remaining points as follows,

$$h_i(k) = \# \{ \mathbf{d}_j \neq \mathbf{d}_i : (\mathbf{d}_j - \mathbf{d}_i) \in \text{bin}(k) \}, \quad k=1 \sim K.$$

The bins are uniform in log-polar space as shown in Fig. 1. The absolute size of the shape context is determined by the system designer. It is a trade-off: If too small we lose global information; if too large we introduce more outliers.

Now that every image point in set \mathbf{D} and set \mathbf{M} has a shape context, a cost can be assigned for matching two points:

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

where K is the number of bins. Given the cost between all pairs of points i in set \mathbf{D} and j in set \mathbf{M} , [9] used graph

theory methods to find the minimum cost correspondence between the two point sets.

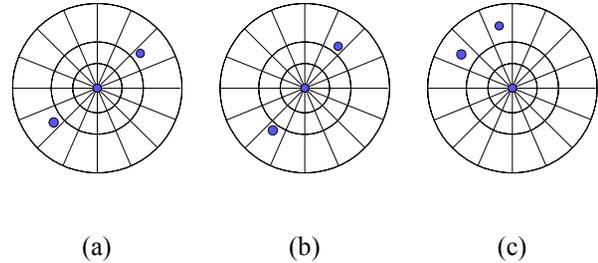


Figure 1. The original shape context method does not differentiate between (a), (b), and (c) because each point contributes to only one bin.

4. THE PROPOSED SOFT SHAPE CONTEXT

4.1. Solving the problem of background clutter

The shape context method achieved top recognition rate in the MNIST [10][9] handwritten digit database, where digits are segmented, i.e., each image contains a single digit. However, for aligning an isolated object to a bunch of non-segmented objects, shape context becomes unreliable [11]. The reason is that the shape context will include outliers, and shape information becomes contaminated.

We propose to use shape context at a small scale, with radius about 9 pixels for images of size 50×50 to 100×100 pixels. The shape context has K bins in the tangential direction and one bin in the radial direction as shown in Fig. 2. We used $K=12$ in the experiments. This size and partition is appropriate for capturing shape features at a local scale, e.g. 3-way or 4-way joints, large curvature, straight segments, etc. This shape descriptor avoids the explicit calculation of first or second order derivatives, which is difficult or unreliable, if not impossible, for images with joints or discontinuities.

If we were to rely only on this local shape information to do alignment, results would be unreliable, because such a small shape context loses global shape information and the ‘‘aperture problem’’ similar to solving optical flow arises. However, since this shape information is to be used as an auxiliary feature vector in the proposed ICP modification, there is no such concern.

4.2. Solving the problem of histogram bins

The original shape context shape descriptor is sensitive to image distortions, as shown in Figure 1. In Fig. 1(a), two image points lie close to the boundary of their bins. Fig.

1(b) is a slightly distorted version of Fig. 1(a), where image points are assigned to neighbor bins. Fig. 1(c) is a point set with completely different configuration. Although visually (a) and (b) are more similar, the cost between the histograms of (a), (b) and (c), however, are all the same, according to Eq. (1). To solve this problem, we make an attempt to diffuse, or spread, the contribution of each image point as follows.

4.3. Label assignment for histogram bins

Bins are assigned labels in a counter-clockwise order, from 0 to $K-1$. It does not matter from which bin we start labeling, as long as we are consistent. This kind of labeling of bins facilitates *diffusion*: If an image point falls within a bin, not only the count at that particular bin will be incremented by 1, but also all bins with approximate labels will be incremented by, say, 0.5, 0.25, etc., according to how much the labels differ. This process essentially softens the boundary of the bins, as shown in Figure 2.

4.4. Bin assignment for image points

To create the soft shape context of an image point $\mathbf{d}_i = [x_i, y_i]$, we center the K -bin diagram at \mathbf{d}_i . All the other image points \mathbf{d}_j in set \mathbf{D} that are inside the K -bin diagram are assigned a label according to which bin they fall in.

4.5. Diffusion and counting

If image point \mathbf{d}_j falls in a bin with label k , we increment the bins $l \in \{(k+i) \bmod K\}_{i=-w}^w$ according to a symmetric triangular function $Triang_k(l)$ with peak value one, center at k , and support $2w+1$. We use $w=1$ in the experiments. In this way, the count is spread to approximate labels with different weightings. This process is done for all image points \mathbf{d}_j in set \mathbf{D} . The final count for all bins is a histogram of real numbers, and captures the information of how the point set is locally distributed around \mathbf{d}_i . This histogram is called the soft shape context, SSC , of image point \mathbf{d}_i :

$$SSC_{\mathbf{d}_i}(l) = \sum_{\substack{k=\text{label}(\mathbf{d}_j) \\ \mathbf{d}_j \in \mathcal{S}(\mathbf{d}_i)}} Triang_k(l), \quad l \in \{1, \dots, K\}$$

where $\mathcal{S}(\mathbf{d}_i)$ is the K -bin diagram \mathcal{S} centered at \mathbf{d}_i .

It is worth noting that the soft shape context is essentially equivalent to applying a low-pass filter to the original shape context.

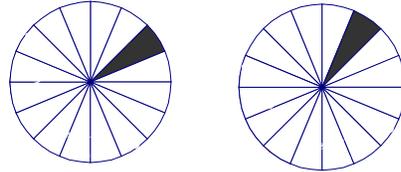


Figure 2(a)

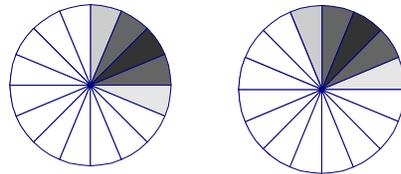


Figure 2(b)

Figure 2. (a) The original shape context. (b) The proposed soft shape context. Bins are diffused.

5. ICP WITH SOFT SHAPE CONTEXT

As stated in section 2, ICP would benefit if shape information is incorporated into the correspondence step. That is, the *closest* point is closest not in the sense of spatial distance, but closest in both spatial and shape distances. Hence the following modification:

Modified Step 1. Finding the correspondences, $\phi(\cdot)$:

$$\phi(i) = \arg \min_{j \in \{1, \dots, N_m\}} \left\{ \varepsilon_{\text{dist}}(|\mathbf{m}_j - T(\mathbf{d}_i; \mathbf{a}_k)|) + \alpha \cdot \varepsilon_{\text{shape}}(\mathbf{m}_j, \mathbf{d}_i) \right\}$$

where

$$\varepsilon_{\text{shape}}(\mathbf{m}_j, \mathbf{d}_i) = \sum_{k=1}^K (SSC_{\mathbf{m}_j}(k) - SSC_{\mathbf{d}_i}(k))^2,$$

and α is a constant which we set at 6.

6. EXPERIMENT

We compare ICP with and without soft shape context on 80 trademark logo images from the database in [13]. The data set and model set are created as follows. The data set is the set of 80 images, resized to about 70×70 with original aspect ratio. All pixels from the Canny edge filtered image were used as image points. We then added outliers to these images and save them as the model set, as shown in Fig. 3. Hence we know the ground truth location and orientation for aligning each pair of the model set and the data set. For each image pair, we conducted 200 experiments, with starting position of the data set being

initialized randomly within -10 to 10 pixels displacement from the ground truth position, and random initial orientation within -20 to 20 degrees. Hence both algorithms are run $80 \times 200 = 16000$ times over all these different translation-orientation combinations. The final average registration error for both translation and rotation are shown in Table 1. Two of the experiments are shown in Table 2.

	Final translation error (pixel)	Final rotation error (degree)
ICP	2.2075	7.1906
ICP with shape context	0.4261	3.8159
ICP with soft shape context	0.2134	2.8648

Table 1. Comparison of registration error between ICP with and without (soft) shape context.

7. CONCLUSION AND FUTURE WORK

In this paper we proposed the concept of soft shape context and incorporate it into the ICP method. The shape descriptor we use is robust to distortions, and does not need the calculation of first or second order derivatives. This shape information is shown to improve the registration performance of the standard ICP method. Future work includes using different kind of shape descriptors, for example, a grid-like rectangular shape context shape descriptor instead of the original circular shape context shape descriptor. Also of interest is to use different cost functions in Eq. (1), e.g., the Earth Mover's Distance [12].

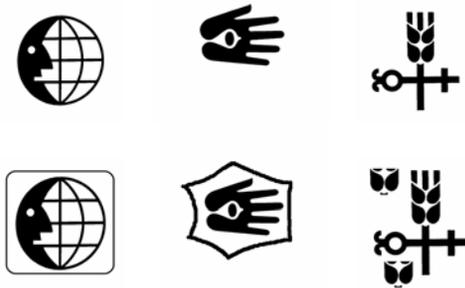


Figure 3. Part of the image pairs used in the experiment. The top row shows three data point sets, the bottom row shows the model point sets to be aligned to.

8. REFERENCES

[1] Paul J. Besl and Neil D. McKay, "A method for registration of 3-D shapes," *IEEE Trans. PAMI*, 14(2), pp. 239-255, 1992.

[2] A. Johnson and S. Kang, "Registration and integration of textured 3D data," *Proc. 3DIM '97*, pp. 234-241, 1997.

[3] A. Gourdon, N. Ayache, "Registration of a curve on a surface using differential properties," Research report RR-2145, INRIA, 1993.

[4] G. Sharp, S. Lee, and D. Wehe, "ICP registration using invariant features," *IEEE Trans. PAMI*, 24(1), pp. 90-102, 2002.

[5] T. B. Sebastian, P. N. Klein, and B. B. Kimia, "Recognition of shapes by editing shock graphs," *IEEE Int. Conf. on Computer Vision*, pp. 755-762, 2001.

[6] P. F. Felzenszwalb, "Representation and detection of deformable shapes," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 102-108, 2003.

[7] A. W. Fitzgibbon, "Robust registration of 2D and 3D point sets," *British Machine Vision Conference*, pp. 411-420, 2000.

[8] G. Champleboux, S. Lavalée, R. Szeliski, and L. Brunie, "From accurate range imagings sensor calibration to accurate model-based 3-d object localization," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 83-89, 1992.

[9] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. PAMI*, 24(4), pp. 509-522, 2002.

[10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, 86(11), pp.2278-2324, 1998.

[11] A. Thayananthan, B. Stenger, P. H. S. Torr, and R. Cipolla, "Shape context and chamfer matching in cluttered scenes," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 127-133, 2003.

[12] Y. Rubner, C. Tomasi, and L. J. Guibas, "A metric for distributions with applications to image databases," *Proc. IEEE Conf. Computer Vision*, pp. 59-66, 1998.

[13] <ftp://ftp.cps.msu.edu/pub/prip/database/>

		ICP Final translation error 6.3 pixels Final rotation error 7.5 degrees
		ICP with soft shape context Final translation error 1.6 pixels Final rotation error 1.8 degrees
		ICP Final translation error 8.2 pixels Final rotation error 6.1 degrees
		ICP with soft shape context Final translation error 0.2 pixels Final rotation error 5.7 degrees

Table 2. Initial alignment and final registration results.