Recognizing hand-drawn images using shape context

Gyozo Gidofalvi
Department of Computer Science and Engineering
University of California, San Diego
La Jolla, CA 92037
gyozo@cs.ucsd.edu

Abstract

The objective of this paper is twofold: to gather real world samples for a subset of the standardized set of 260 line drawings introduced by Snodgrass and Vanderwart [4] and to test the performance of the representative shape context method for rapid retrieval of similar shapes introduced by Mori et al. [1]. To experiment with the expressive power of the shape context at different location in the image, we introduce a modification to the representative shape context method, which draws representatives from a distribution based on the pixel density of the image in a given area. Furthermore, we test the performance of the representative shape context method for detection of these objects when embedded in an arbitrary environment. We find that the performance of the representative shape context method is slightly worse on hand–drawn images than on synthetic data presented in [1]. We also find that the density based sampling methods perform worse than the original method. Finally, we find that the representative shape context in its original form is highly affected by the presence of clutter and is not appropriate to recognize objects when embedded in an environment. However, results suggest that a sampling method that incorporates both spatial considerations and density measures may improve query performance for embedded objects.

1 Introduction

The shape context, a recently introduced shape descriptor by Belongie et al. [3], has proven to be accurate at matching similar shapes. In [1], Mori et al. divide the shape matching task into two stages: *fast pruning*, during which a small set of candidate shapes are retrieved, and *detailed matching*, during which a more expensive and accurate matching procedure is applied to only the candidate shapes. They describe two methods for the fast retrieval of similar shapes: *representative shape context* and *shames*. Both of these methods are based on matching shape contexts between the query image and the known images in the dataset using the nearest neighbor algorithm. The following a basic description of the representative shape context method, which we have used to conduct our experiments.

A shape is represented as set of points $P = \{p_1, p_2, \ldots, p_n\}$ sampled from the internal or external outline of the image. At each of these points, the *shape context* is the histogram of the relative locations of the remaining points of the shape. To ensure that the shape context is more sensitive to nearby points than to points further away, bins that are uniform in log-polar space are used to obtain this histogram corresponding to the shape context.

Using the X^2 distance as a metric to represent the cost of matching two shape contexts the problem of shape matching can be reduced to the weighted bipartite matching problem, which can be done in $O(n^3)$ time using the Hungarian method, where n is the number of points to be matched. Although this method gives highly accurate matching results, it is computationally slow, hence should be applied to only a small set of candidate shapes.

To obtain this small set of candidate shapes Mori et al. [1] present the following simple method.

```
1. Represent the query image by a small number of shape context descriptors
2. Calculate the cost of a match between the query shape and a known shape as the sum of costs between a query shape context and the closest shape context of the known shape by performing nearest neighbors search
3. Return a short list of known shapes of the fist K best matches
```

It is important to note that as a result of the coarse matching, the matching does not obey the one-to-one mapping between query and known shape contexts.

2 Real life data set of line drawings

In [1], Mori at al. propose that although the shape context descriptor is not invariant under arbitrary affine transforms, due to the log-polar coordinate system of the shape context, small, local changes in the image result in correspondingly small changes in the shape context. These small local changes would include distortions due to pose change and intra-category variations. Due of the lack of multiple instances within the categories in case of the Snodgrass and Vanderwart line drawings, to test the performance of these pruning methods Belongie et al. used the thin plate spline (TPS) model to create a synthetic distorted set of images for querying. To verify the performance of the representative shape context method on real world situations we have gathered a real world dataset.

2.1 Snodgrass and Vanderwart line drawings

In [4] Snodgrass et al. presented a standardized set of 260 pictures to investigate the differences and similarities in the processing of pictures and words among human subjects. Pictures in this set were selected based on three criteria: "first, that they be unambiguously picturable; second, that they include exemplars from […] the category norms of Batting and Montague; and third, that they represent concepts at the basic level of categorization." [4]. The following is the 15 categories defined by the Batting and Montague: four-footed animal, kitchen utensil, article of furniture, part of the human body, fruit, weapon, carpenter's tool, article of clothing, part of building, musical instrument, bird, type of vehicle, toy, and insect. The selected pictures were standardized based on four variables that are relevant to human memory and cognitive processing: name agreement, image agreement, familiarity, and visual complexity [4].

2.2 Procedure for gathering real life data

We gather 6 samples for a subset of size 50 of the original Snodgrass and Vanderwart line drawings. When selecting the 50 objects, we tried to obey the second criteria used by [4] and selected objects from most of the 15 categories. A list of the objects selected can be found in the appendix. Since the ambiguity in verbal description of objects could possibly lead to too large variation within classes of shapes and in pose we adopt the following method for acquiring our samples. Our subjects are presented with a sample shape for a short interval of 1 second and are allowed an arbitrary amount of time to draw their image. Figure 1 shows an example of a known shape and some hand-drawn shapes that correspond to it. The slideshow used to acquire our samples can be viewed at http://rick.ucsd.edu/~gyozo/images.htm. Images gathered are scanned and cropped to 500 by 500 pixel binary images.

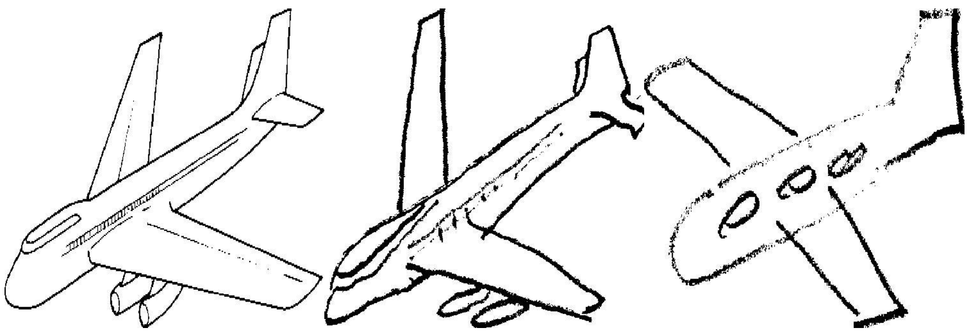

Figure 1: Image in the left shows an instance of a known image. Images to the right of the known image are samples taken for this object.

3 Performance of the representative shape context method

We test the original representative shape context method using the 300 hand-drawn query shapes and the 50 original, known images. A query of a hand-drawn shape is successful if the corresponding known shape is included in the set of retrieved candidate shapes. From [1] we adopt the terminology and call this set the *short list*. Figure 2 shows the results in terms of error rate for varying pruning factors.

In fugure 2, as well as in rest of graphs in the paper, error rates are shown for various pruning factors on a logarithmic scale. For a retrieval of k candidate shapes from a set of known shapes of size D, the pruning factor is defines as $P = D / k$. In our case, when the $P = 1$, the length of the short list k is 50 (i.e.: all shapes from the database are on the short list), and hence the error rate is 0. When $P = 50$, the length of the short list contains 1 candidate shape only and the error rate is 0.47.

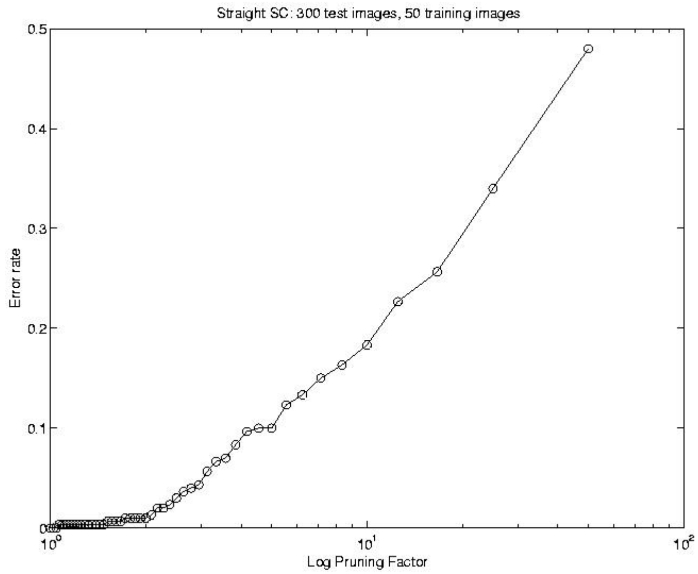

Figure 3: Retrieval results of the representative shape context method using hand-drawn images

3.1 Sampling representative shape contexts

As discussed earlier the shape context used by [1] and in this project uses bins that are uniform in log-polar space. By definition such a descriptor is more sensitive to points that are near by than to points that are farther away from the center of the sampled shape context.

In the original representative shape context algorithm introduced in [1], as discussed earlier, shape context samples are uniformly drawn from the query shape and are used to match possible candidate solutions for the retrieval. Since uniformly sampled shape context of the shape may contain redundant information this method sampling of representative shape context is questionable. As an example consider a straight line. This line can easily be defined in terms of its end points; hence any additional points would be redundant and would not make the description more precise. Similar argument suggests that a uniform sampling of shape context may result in redundancy in the representation of a shape. Is there a different sampling method that yields better performance due to the higher expressiveness of the representative shape context used for the matching? The following section introduces a method that samples representatives from a distribution, which is based on the pixel density of the image in a given area.

We define the pixel density $pd(i)$ of an image point i in a window w centered at i as the normalized number of image points that are inside w. To perform pixel density based selection of representative shape contexts, we assign a probability to each image point i proportional to $pd(i)$, and perform roulette wheel selection based on these probabilities. We conduct two sets of experiments: one in which image points with higher pixel densities-, and one in which image points with lower pixel densities receive higher chance to represent a shape. Figure 3 visualizes an instance of these sampling methods.

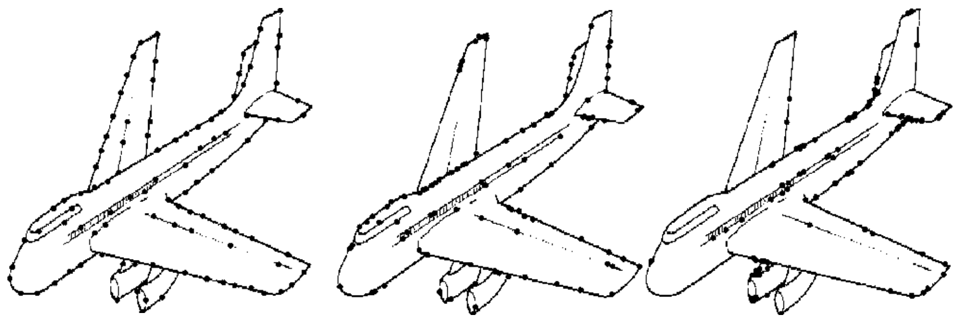

Figure 3: Visualization of the results of the different sampling methods. From left to right: uniform sampling, promoting lower density sampling, promoting higher density sampling.

Since the discriminative power of the representative shape contexts obtained with these different sampling methods is not intuitively clear, we test them on the on our hand-drawn samples. Figure 4 shows the retrieval results for the different sampling methods. Results show that neither of the density-based methods gives better results than the original uniform sampling method.

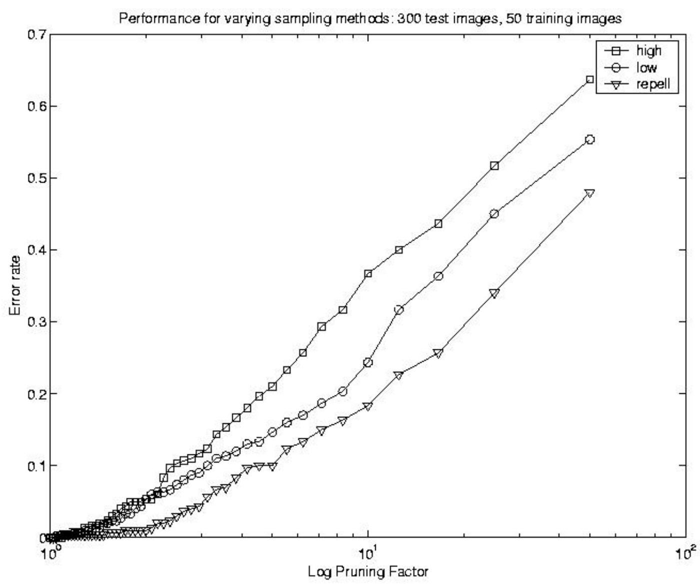

Figure 4: Retrieval results for different methods of sampling

Figure 5 and 6 shows some query examples and the short list returned for each query. It shows results for two sets of hand-drawn samples: one, in which query objects are visually very similar to known objects (figure 5), and one in which queries objects are visually very different from the known objects (figure 6). Interestingly, there are no obvious differences between the results for these two sets, which justifies the generalization ability of the shape context descriptor.

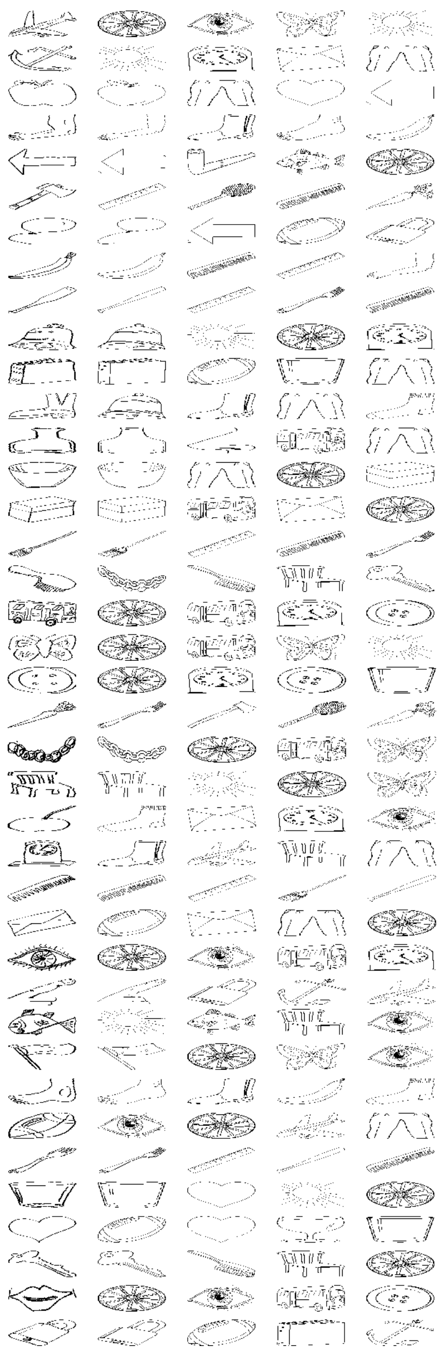

Figure 5: Sample query results for hand-drawn images that are very similar to the known objects. First column shows the query object, remaining objects in a row are the first few objects on the short list.

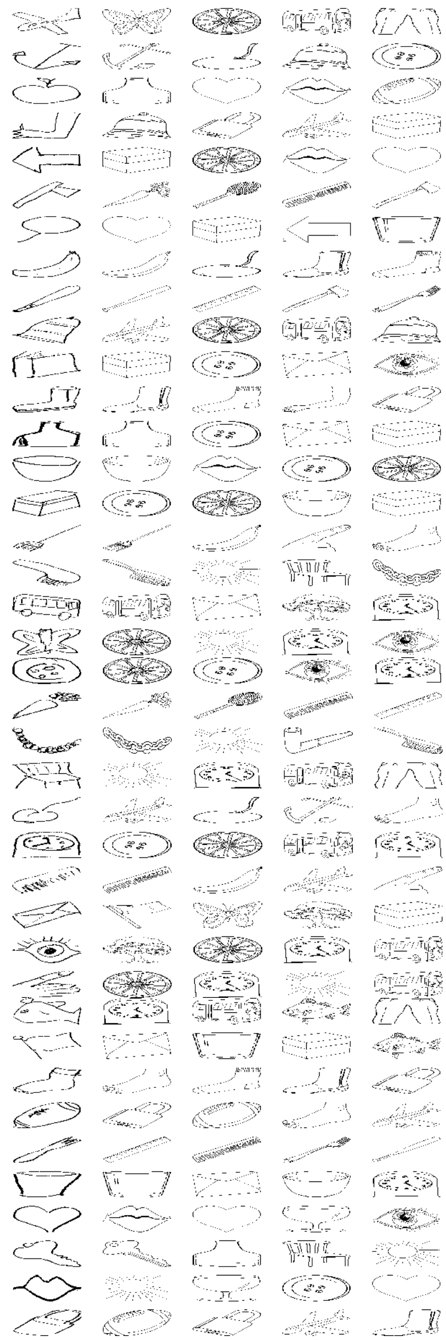

Figure 6: Sample query results for hand-drawn images that are very different from the known objects. First column shows the query object, remaining objects in a row are the first few objects on the short list.

4 Finding embedded objects using shape context

To test the performance of the representative shape context method in cases when the query object is embedded in an arbitrary environment, we place the cropped and normalized objects into an arbitrary environment. This is achieved by first finding the outline of the object and by constructing a mask for it. Using binary operations on the mask, the image containing the object, and the image containing the environment we copy the non-overlapping parts of the environment to obtain and embedded version of the query. Figure 7 shows some of instances of embedded objects.

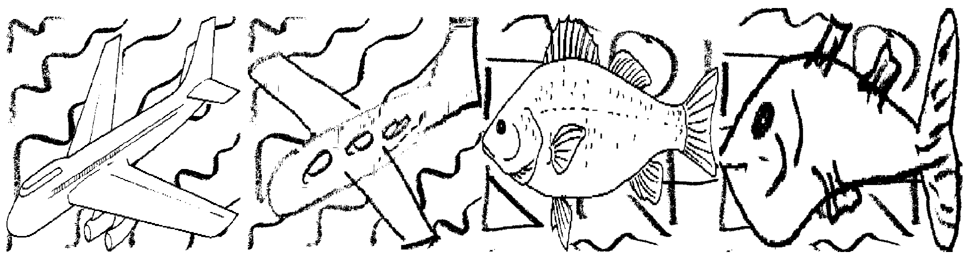

Figure 7: Instances of embeddings. Left images in the pairs of objects in the same categories are embedded known objects, while right images are embedded hand-drawn objects.

To test the performance of the representative shape context algorithm, we take both the hand-drawn and the known objects and place them into 8 different environments, and use these embedded objects as queries. Performance results are shown in figure 8.

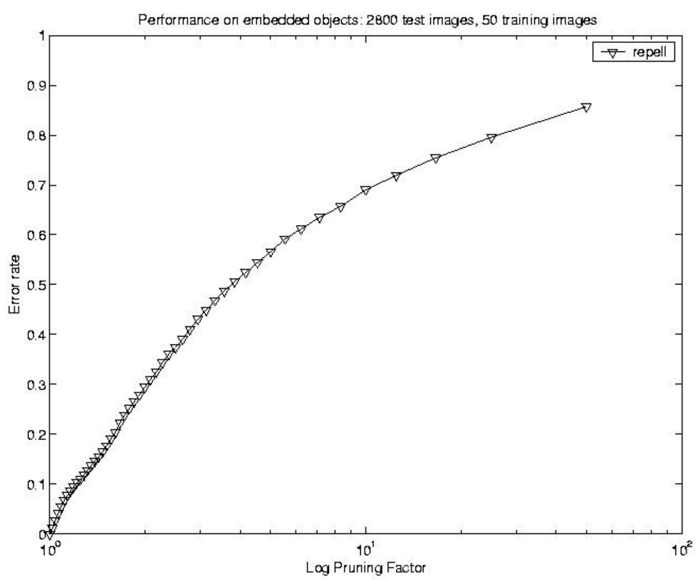

Figure 8: Retrieval results on embedded objects

The results for embedded objects significantly worse than for standalone hand-drawn images. To investigate the possible reasons for the degradation in performance, we conduct a test where queries are embedded versions of the known objects. Figure 9 shows performance results, while figure 10 shows some examples for embedded queries.

We find that objects that are large and occupy most of the space are more often classified correctly, since they are less affected by the environment or clutter around them. Conversely, objects that occupy only a small amount of the space are classified with lower accuracy and are mostly confused with larger objects that exhibit similar spatial distribution around the peripherals of the image as the clutter. Finally, objects that have areas that have particularly low (i.e.: heart) or high pixel densities (i.e.: eye) are classified correctly more often.

These findings suggest that, after all, a modification to the sampling used in the representative shape context method may improve overall performance. Such a sampling should possibly incorporate density measures presented in this paper and other spatial measures that sample representatives with higher probability towards the center of the image.

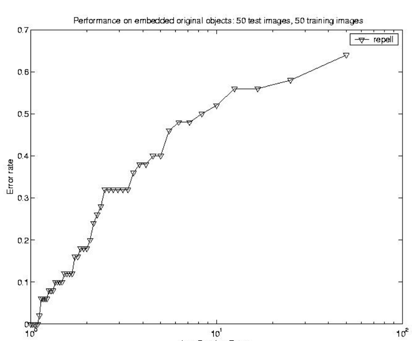

Figure 9: Retrieval results on embedded known objects

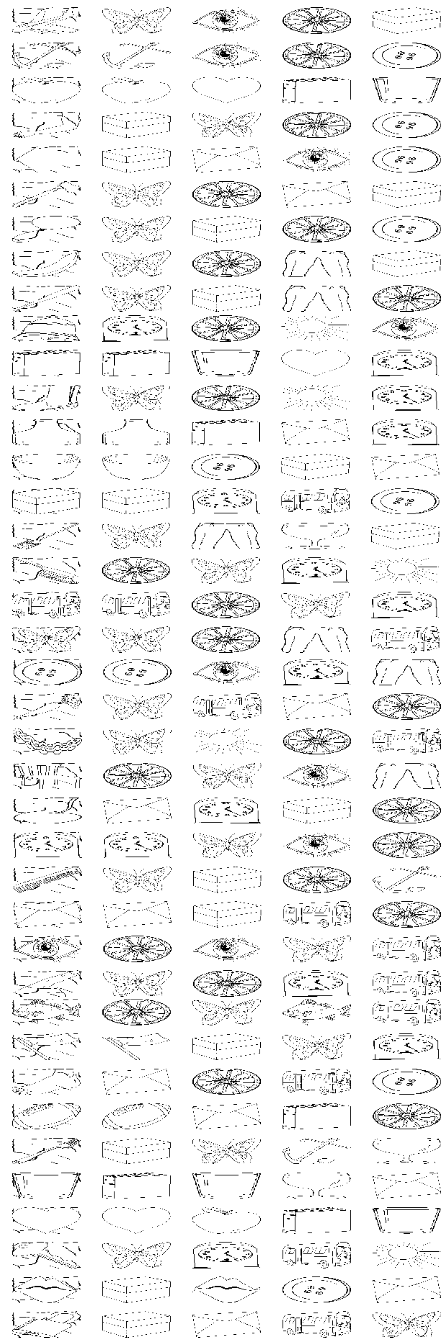

Figure 10: Sample query results for embedded known objects. First column shows the query object, remaining objects in a row are the first few objects on the short list.

5 Conclusions

We have gathered 6 sets of hand-drawn samples for a subset of size 50 of the Snodgrass and Vanderwart line drawings. We tested the performance of representative shape context method on these images and have found that the results are only somewhat worse on hand-drawn objects than on synthetically obtained objects used in [1]. We modify the original algorithm to bias the sampling of representatives based on pixel densities, and we find that this biased sampling only worsens the performance. Finally we test the performance of the original method on embedded object, and find that the representative shape context in its original form is highly affected by the presence of clutter and is not appropriate to recognize objects when embedded in an environment. Finally results suggest that a sampling method that incorporates both spatial considerations and density measures may improve query performance for embedded objects.

Acknowledgments

I would like to thank for Mori et al. for allowing for me to use an implementation of their pruning method, and providing me with the original Snodgrass and Vanderwart line drawing dataset. I would also like to thank Belongie for providing me with helpful ideas and directing me in this project. Last but not least, I would like to thank my family members, my girlfriend, and Hector for taking the time to draw the sample images.

References

[1] Mori, G., Belongie, S., & Malik, J., (2001) Shape Contexts Enable Efficient Retrieval of Similar Shapes, *CVPR*.

[2] Belongie, S., Malik, J., & Puzicha, J., (2001) Shape Matching and Object Recognition Using Shape Context *accepted for publication in PAMI*.

[3] Belongie, S., Malik, J., & Puzicha, J., (2001) Shape Matching, ICCV.

[4] Snodgrass, J. D., & Vanderwart, M., (1980) A standardized set of 260 pictures: Norms for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6:174-215

Appendix

Image numbers of selected drawings:

2, 4, 6, 7, 8, 12, 15, 16, 19, 25, 30, 31, 32, 34, 35, 37, 38, 39, 40, 41, 48, 52, 53, 54, 60, 65, 85, 86, 88, 89, 90, 94, 95, 97, 104, 119, 128, 141, 143,155, 162, 174, 192, 211, 217, 222, 229, 241, 254, 258

Hand-drawn data set gathered in this project can be obtained from the author.