POINT PATTERN RECOGNITION

J. Rošický, P. Studenovský

Czech Technical University in Prague, Faculty of Mechanical Engineering, Department of Instrumentation and Control Engineering

ADR-SYS, Sartoriova 31, 169 00 Prague

Abstract

Presented here is a method for pattern recognition. The pattern in this context is a set of points represented as a set of pairs of coordinates. Given a register of template patterns and an image pattern the method assigns a quantity to each template pattern expressing the similarity between that template pattern and the image pattern. The method uses an invariant description of the patterns unaffected by pattern transformations (translation, rotation, noise perturbations in coordinates or point deletion/addition). We performed a test to estimate some properties of the method, i.e. the probability that a wrong template pattern could be assigned to a given image pattern. The method can be used for object recognition in cases where an object can be represented as a point pattern, i.e. stars in the sky. Included is a brief note about the implementation in Matlab.

1 Introduction

Point pattern recognition techniques are useful for object recognition tasks when an object can be represented as a set of points. Such tasks can be encountered in fields like machine vision, image registration, biometry, biology [2], gel electrophoresis or astronomy to name a few.

2 Point pattern recognition

The formulation of the task of point pattern recognition is as follows. Given two point patterns in 2D space (Figure 1 and Figure 2) we want to calculate a measure of their similarity.

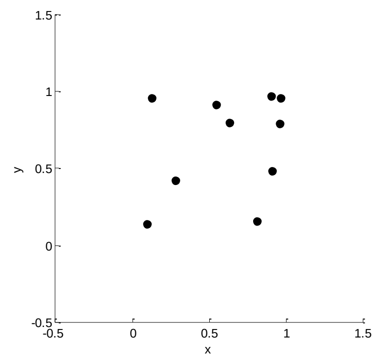

Figure 1: Point pattern *A*.

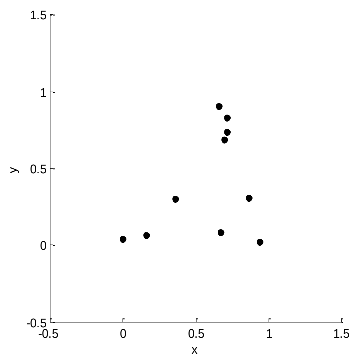

Figure 2: Point pattern *B*.

The positions of points in patterns can be transformed – translated, rotated (Figure 3), perturbed by noise (**Figure 5**) and some points could be deleted or added (Figure 4). We do not consider the change of the scale.

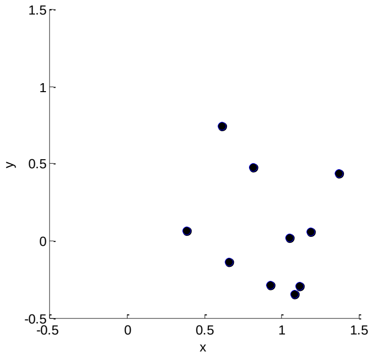

Figure 3: Point pattern *A*, translated and rotated.

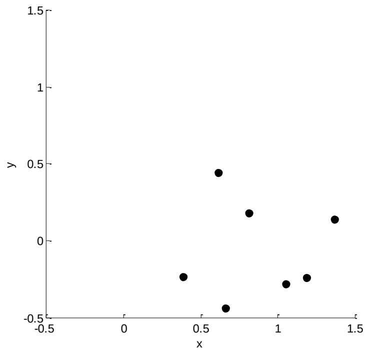

Figure 4: Point pattern *A*, translated, rotated and 3 points are missing.

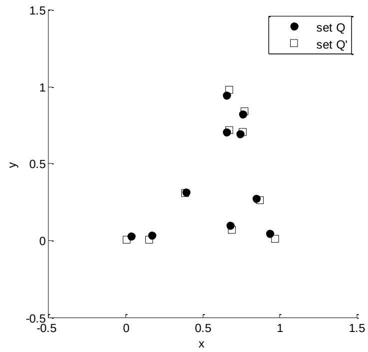

Figure 5: Two instances of point pattern B. Points are perturbed by noise.

3 Review of approaches

According to Cox et al. [1], techniques for point pattern recognition fall into three groups: clustering, interpoint distance algorithms and relaxation methods, plus there are some methods outside these three groups. Clustering methods use (quantised) parameter space and search it for strong clusters. Interpoint distance algorithms exploit several structures – sorted vectors of interpoint distances, sorted nearest neighbour vector, minimal spanning tree or graph representation of point pattern. Two relaxation approaches are mentioned in [1] – a translation invariant technique more tolerant of global distortions and a fuzzy relaxation technique for labeled patterns. Baumach et al. [4] describes signal based point pattern matching. This approach deals with the image signal instead of the point coordinates. Mo et al. [3] proposes point pattern method for real time application – object tracking. Matula et al. [2] developed a method based on point pattern matching for aligment of live cells. There is an implementation of procrustes method in Matlab [6] which can be used as a point pattern recognition approach. Murtagh [5] proposed so called „world-view vector“ for the point pattern description used for matching star lists.

4 Point pattern recognition

Current methods usually try to find the best transformation which maps one pattern to another and then evaluate the degree of similarity. Our approach is focused on the discriminative properties of the recognition process. We are interested in how possible is the failure of the recognition. Presented method for point pattern recognition uses a description of the point pattern that is invariant to rotation and translation transformations, noise perturbation in coordinates and deletion or addition of a low number of points. We call this description the *identification graph*. An example of identification graph is in Figure 6. Each point (node) conects to a defined number of its neighbours. We call this number the *locality* (*L*). The locality was set to 4 in the graphs in Figure 6. Some points (nodes) have more than 4 connections, because they have just 4 „outging“ connections and might have some „ingoing“ connections from other points than their „outgoing neighbours“. The topological structure of a graph is expressed as a set of *vectors* expressing „fragments“ of the graph. Fragments are portions of the identification graph. In our case fragments are „triplets“ – three nodes connected by two branches with one common node. The oriented angle between the two branches and the lengths of the branches form the „triplet“. The measure of similarity (T score) between two point patterns is calculated as a sum of vector „twins“ found among their associated graphs. The T score can be considered as a composite of two contributors which can be expressed as:

$$T = r + q \tag{1}$$

The expression r represents the random part and q represents the non-random part. The random part expresses accidental similarity between any two point patterns and the non-random part expresses significant similarity between matching point patterns. The distribution of r should be close to normal and the distribution of q should not be normal as its value is limited.

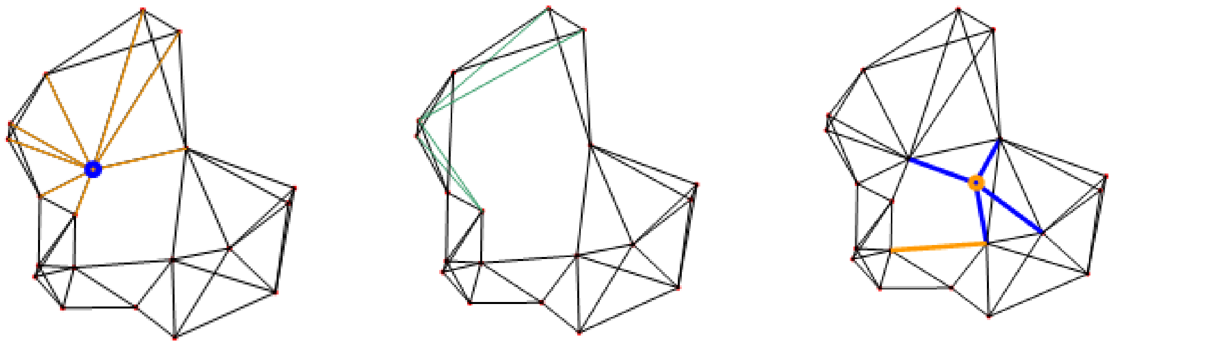

Figure 6. Identification graph for a point pattern: 20 points (left), 19 points (middle, one point deleted), 21 points (right, one point added).

5 Experiments and results

We carried out a simulation with the aim to estimate and evaluate the probability that the method will provide correct results. Actually we formulated the task as a hypothesis test with the null hypothesis: „The T scores of the first two candidates do not differ (the given pattern was not found in the database)“, and the alternate hypothesis: „The T scores of the first two candidates do differ (the given pattern was found in the database)“. In other words - for a chosen threshold (the critical score T_{crit}) we asked what is the probability of the type I error (α) and type II error (β) and what is the power of the test ($1-\beta$). The simulation was arranged in three blocks with four repetitions in each block. The first, second and third block included patterns with 30, 40 and 50 points respectively. In each block there were four patterns in the database and each of four repetitions inside a block compared a set of 1000 transformed patterns with the patterns in the database. The result of a block were 4x1000=4000 scores of the first candidate, another 4000 scores of the second candidate, and so on. The properties (i.e. the distribution of the point addition/deletion and noise) of the transformed patterns are depicted in **Figure 7** (for brevity only patterns with 30 points are shown; 40 and 50 point patterns share similar properties).

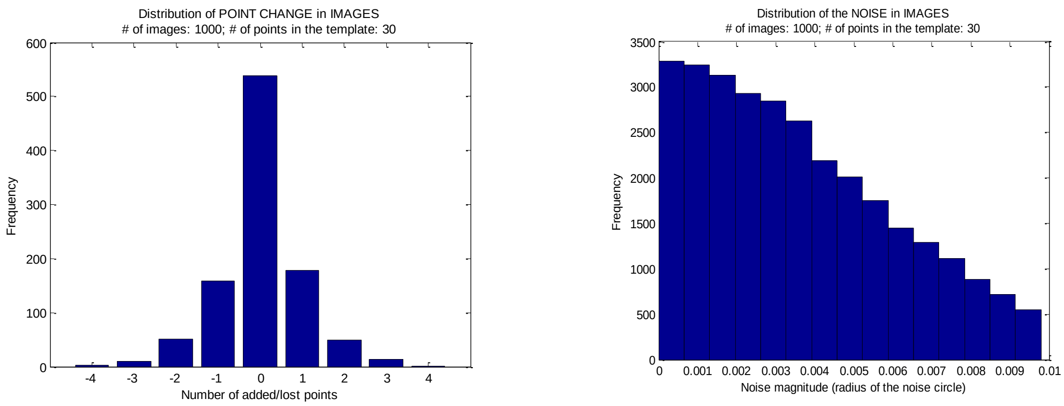

Figure 7. Properties of transformed patterns – distribution of point addition/deletion (left) and distribution of the noise magnitude in coordinates (right).

The distributions of the T score are shown in **Figure 8**. Further we will focus only on the results obtained from tests with 30 point patterns. The results for 40 and 50 point patterns are similar to the 30 point case.

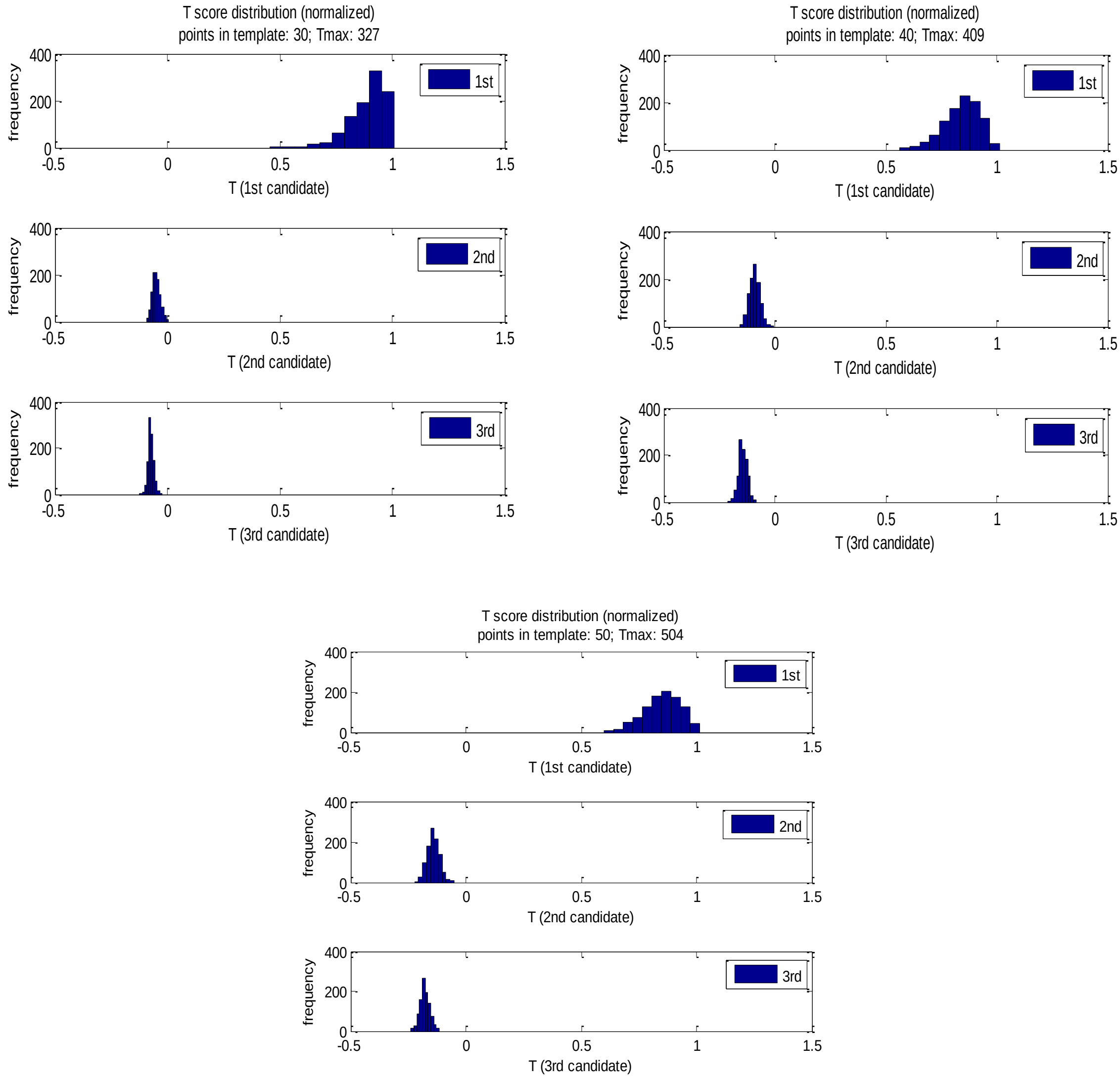

Figure 8. *T* score distributions.

The fitted distributions of the first and the second candidate are show in **Figure 9**. The distribution of the second candidate can be censidered close to normal with parameters $\mu = -0.03692$ and $\sigma = 0.01820$. The distribution of the first candidate cannot be censidered as normal so we used weibull distribution. To fit the weibul distribution the T score of the first candidate was flipped from left to right and translated to the origin. Weibul parameters are: *shape* = *1.221*, *scale* = *0.1142*.

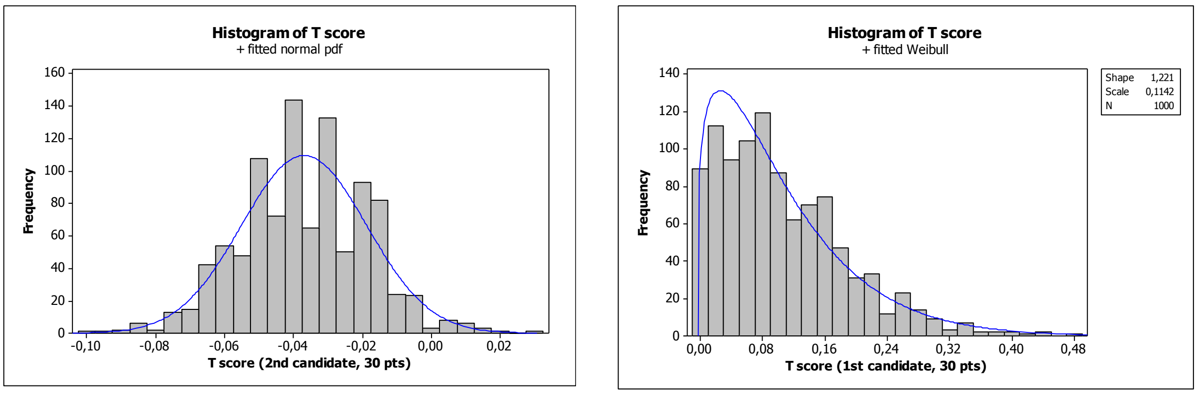

Figure 9. Fitted *T* score distribution – 2nd candidate (left) and 1st candidate (right, flipped and translated).

To calculate the type I (α) and II (β) errors for the T_{crit} = 0.05 value, the normcdf and wblcdf Matlab functions were used:

```
>> alfa = 1-normcdf(0.05,-0.03692,0.01820)

>> alfa = 8.948e-007

>> beta = 1-wblcdf(0.95,0.1142,1.221)

>> beta = 1.698e-006
```

Power of the test is *1 – beta = 9.99998e-001*. The interpretation is that for chosen criteria T_{crit}= 0.05 the T_1 score of the first candidate $T_1 > T_{crit}$ indicates that the first candidate matches the involved pattern from the database and there is a probability lower than alfa = 8.9e-005 % that it is a wrong decision. If the $T_1 < T_{crit}$ than it is an indication that the first candidate does not match any of the patterns in the database and there is a probability lower than beta = 1.698e-004 % that it is a wrong decision.

The algorithm was implemented in Visual Basic for Applications. Therefore we did not perform any time performance tests. Matlab was very useful for test data preparation and inspection. Its support for structured variables was invaluable for dealing with large heterogeneous data.

6 Conclusion

Next work will be aimed to improving the calculation speed. We plan to implement the algorithm using a faster platfoorm because we would like to achieve a real-time performance. Other possible areas to investigate are deeper analysis of statistical properties of the method.

Acknowledgement

This work was kindly supported by grant ČVUT SGS10/051/OHK2/1T/12.

References

[1] Cox, G.S., de Jager, G.: A survey of point pattern matching techniques and a new approach to point pattern recognition. *Communications and Signal Processing*, 1992. COMSIG '92., Proceedings of the 1992 South African Symposium on (0-7803-0807-7), p.248-248.

[2] Matula P., Kozubek M., Dvorak V.: Fast point-based 3-D alignment of live cells. *Image Processing*, IEEE Transactions on (1057-7149), 2006. Vol.15,Iss.8;p.2396-2396.

[3] Fan Mo, Haodong Pei, Shuying Li, Dan Li, Baorong Xie, Wei Zhou: A Novel Object Tracking Algorithm Based on Point Pattern Matching. *Information Science and Engineering* (ISISE), 2009 Second International Symposium on (978-1-4244-6325-1), 2009. p.459-459.

[4] Baumbach, T.; Ortmann, W.: Shift detection by restoration-demonstrated by signal based point pattern matching. *Image Analysis and Processing*, 1999. Proceedings. International Conference on (0-7695-0040-4), 1999, p.315-315.

[5] Murtagh F.: A New Approach to Point Pattern Matching. *Publications of the Astronomical Society of the Pacific*, 1992, in press.

[6] Mathworks: Procrustes. *Matlab Documentation*. Available on-line: [http://www.mathworks.com/access/helpdesk/help/toolbox/stats/procrustes.html]

Jiří Rošický
jiri.rosicky@fs.cvut.cz

Pavel Studenovský
pavel@studenovsky.cz