

UNIVERSITY OF LJUBLJANA
INSTITUTE OF MATHEMATICS, PHYSICS AND MECHANICS
DEPARTMENT OF MATHEMATICS
JADRANSKA 19, 1000 LJUBLJANA, SLOVENIA

Preprint series, Vol. 40 (2002), 830

PATTERN MATCHING OF
COLORED POINT SETS IN 3D

Marko Boben Alen Orbanić
Tomaž Pisanski Ante Graovac

ISSN 1318-4865

July 31, 2002

Ljubljana, July 31, 2002

Pattern Matching of Colored Point Sets in 3D

Marko Boben*, Alen Orbanic†, Tomaž Pisanski‡

University of Ljubljana, IMFM
Jadranska 19, SI-1000 Ljubljana, Slovenia

Ante Graovac§

Rudjer Bošković Institute
POB 180, HR-10002 Zagreb, Croatia

Abstract

The following problems are considered here:

Problem 1. Given a large collection of points S in 3D space and a small pattern set P find all occurrences of a subset P' in S such that P' is obtained from P as a result of rotation, translation, reflection, and scaling.

Problem 2. Same as Problem 1, except that points in both sets are labeled (colored).

Efficient algorithms for both problems are presented. Both problems could be diversified into a series of problems by modifying the set of allowed transformations like dropping the scaling transformation or introducing the mirror transformations.

The algorithms developed here should find applications in correcting the approximate geometries of fullerenes and other cages to achieve presumed symmetry.

1 Introduction

Recognition of the presence of a geometric pattern in a large set of points is a fundamental problem of computational geometry. It arises in such diverse applications as chemistry (recognition of substructures in molecules), astronomy (recognition of constellations), etc.

*Marko.Boben@fmf.uni-lj.si

†Alen.Orbanic@fmf.uni-lj.si

‡Tomaz.Pisanski@fmf.uni-lj.si supported in part by Ministrstvo za šolstvo, znanost in šport Republike Slovenije, grant J1-6161, J2-6193.

§graovac@rudjer.irb.hr

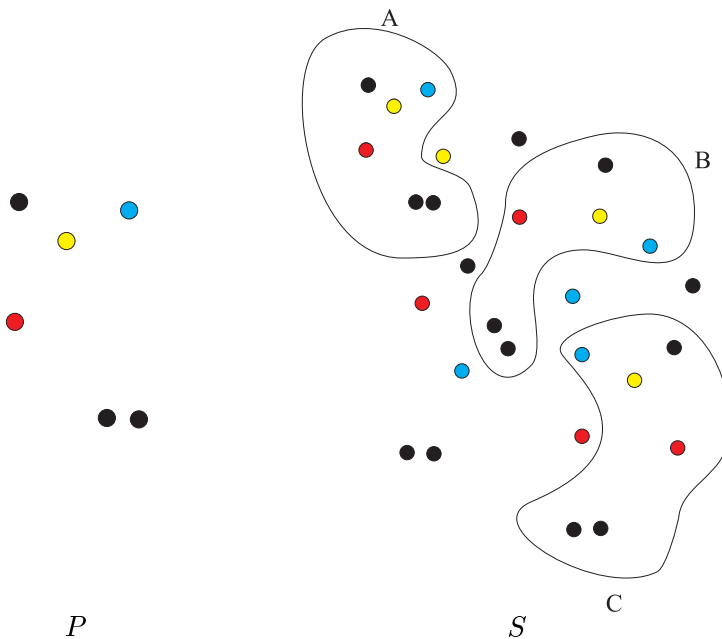


Figure 1: Pattern set P and sample set S with occurrences of P as a result of translation and scaling (A), translation and rotation (B), translation and reflection (C).

The problem can be described as follows. Given a *pattern set* P and a *sample set* S in \mathbb{R}^3 with $0 < k = |P| \leq |S| = n$, identify all subsets of S which are

- P itself, as a subset of S ,
- a translate of P ,
- a rotation of P ,
- a scaled set P ,
- a reflection of P .

Moreover, if P and S are considered to be colored we search only for occurrences of P in S which respect the coloring. See Figure 1.

Among the papers in which variants of the *Point Set Pattern Matching Problem* have been studied are Boxer 1992 [2], de Redzende and Lee [4], Boxer 1996 [3], Boxer 1998 [1].

2 The Algorithm

Let $S = \{s_1, s_2, \dots, s_n\} \subseteq \mathbb{R}^3$ be the sample set and $P = \{p_1, p_2, \dots, p_k\} \subseteq \mathbb{R}^3$ be the pattern set. We try to find translated, rotated, scaled, or reflected

occurrences of P in S .

We give the following algorithm.

1. Sort P by lexicographic order. This takes $O(k \log k)$ time (counting comparisons of real numbers).
2. Sort S by lexicographic order. This takes $O(n \log n)$ time.
3. If P is collinear then perform a simpler algorithm, else select from P three non-collinear points t_1, t_2, t_3 , i.e. points that form a triangle T .
4. For each $s_i \in S$ construct a sorted set $S_i = \{d_{ij_1}, d_{ij_2}, \dots, d_{ij_{n-1}}\}$ where d_{ik} is a distance between s_i and s_k . This takes $O(n^2 \log n)$ time.
5. Consider a pair (s_i, s_j) and try to match (t_1, t_2) to it – now we try to find an occurrence of T in S scaled for the factor $f = |s_i - s_j|/|t_1 - t_2|$. The candidates for s_k (the image of t_3) are points on the intersection of the sphere of radius $r_1 = f|t_1 - t_3|$ centered in s_i and the sphere of radius $r_2 = f|t_2 - t_3|$ centered in s_j . If the number of points at equal distance from s_i can be bounded by a constant A for each i , we have to verify at most A candidates for s_k . These candidates can be efficiently selected from the sets S_i and S_j in $O(\log n)$ time.
6. For each s_k found in the previous step a mapping from (t_1, t_2, t_3) to (s_i, s_j, s_k) uniquely determines an image of P in S . We must verify the existence of (transformed) P in S for the remaining $k - 3$ points. Since S is sorted, this can be done in $O(k \log n)$ time.
7. If there exists an unchecked pair (s_i, s_j) then go to step 5, else stop.

Theorem 2.1. *Let S be the sample set and P be the pattern set. If the number of points at distance d from s can be bounded by a constant for each $s \in S$ and $d \in \mathbb{R}$, then the algorithm described above solves the Pattern Matching Problem in $O(kn^2 \log n)$ time, where $n = |S|$ and $k = |P|$.*

Proof. This follows from the greater of two estimations in steps 5 and 6 above and taking into account all $O(n^2)$ pairs (s_i, s_j) . These steps clearly exceed the preparation which is done in steps 1 to 4. \square

Remarks.

1. To extend the algorithm to colored sets P and S we must additionally pay attention to the colors of elements when matching them in steps 5 and 6. This does not affect the estimation of the running time.
2. The space complexity of the algorithm is $O(n^2)$ since we must build n sets S_i of length $n - 1$ in step 4. In practice, the storage of these sets in memory limits the size of the input which can be processed more than the actual time complexity.

3. In practice (see the next section), the comparisons of points must be performed with a desired tolerance to find patterns that “approximately” match P . Formally, this can be realized by introducing the function $e : \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \{\mathbf{false}, \mathbf{true}\}$ where $e(s, t) = \mathbf{true}$ precisely when the points s and t should be treated as equal. It should be used by the algorithm at every test for equality of two points. For “approximate” matching we can, for instance, use $e(s, t) = |s - t| \leq \varepsilon$ introducing an additional parameter ε (tolerance) to the algorithm.

In the papers mentioned in the introduction we find several other algorithms.

The paper [4] gives point set pattern matching algorithms for exactly matching a pattern P of cardinality k in a sampling set of cardinality n in \mathbb{R}^d , $d \geq 2$, with running time of $O(kn^d)$.

The paper [3] presents an algorithm for $d = 3$ with running time $O(kn^{5/2} [\lambda_6(n)/n]^{1/4} \log n)$ (where $\lambda_6(n)/n$ is “nearly” a constant). This algorithm, however, does not feature searching for scaled patterns.

The upper bound for time complexity was further improved in [1] where the algorithm is given, for $d = 3$, with running time $O(kn^2 [\lambda_6(n)/n]^{1/2} \log n)$. Here, scaling is also omitted from the list of allowed operations.

3 Two Examples from Chemistry

We present two examples from chemistry which show possible applications of the algorithm in finding substructures in molecules.

In the first example we searched for (approximately) regular hexagons (pattern set P) in a set of 772 points (sample set S) which represent a part of a DNA molecule. The result is shown in Figure 2.

In the second example the sample set S consists of atoms of C_{94} fullerene and the pattern P is regular pentagon. As expected, 12 pentagons were found which (approximately) match P , see Figure 3.

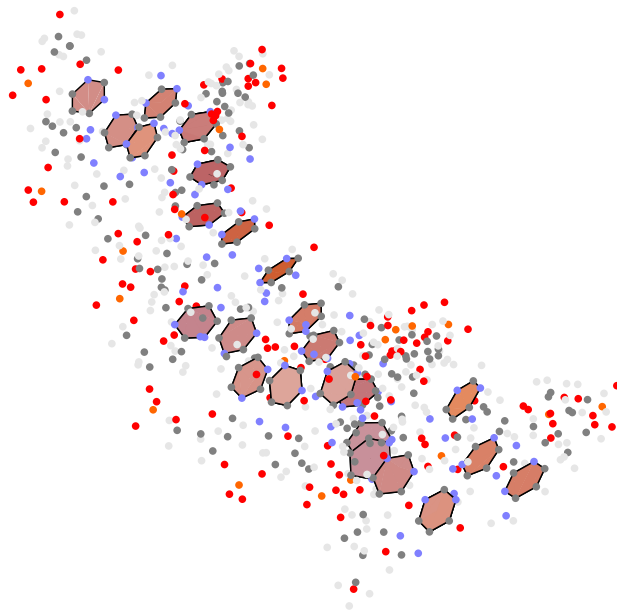


Figure 2: In this case, our algorithm was used to find all (colored) regular hexagons. The sample set is a part of a DNA molecule.

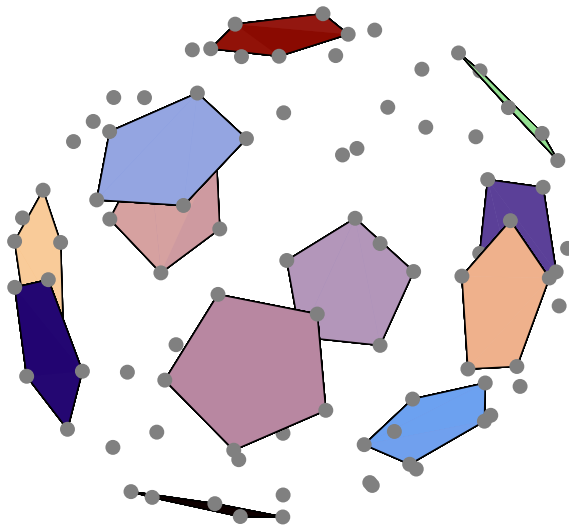


Figure 3: The algorithm was used to find all regular pentagons in a C_{94} fullerene.

References

- [1] L. Boxer, *Faster point set pattern matching in 3-D*, Pattern Recognition Letters 19 (1998), 1235-1240.
- [2] L. Boxer, *Finding congruent regions in parallel*, Parallel computing 18 (1992), 807-810.
- [3] L. Boxer, *Point set pattern matching in 3-D*, Pattern Recognition Letters 17 (1996), 1293-1297.
- [4] P. J. de Redzende and D. T. Lee, *Point set pattern matching in d-dimensions*, Algorithmica 13 (1995), 387-404.