Clustering methods: Part 5

Fast search methods

Pasi Fränti

5.5.2014

Speech and Image Processing Unit
School of Computing
University of Eastern Finland

UNIVERSITY OF EASTERN FINLAND

Methods considered

Classical speed-up techniques

- Partial distortion search (PDS)
- Mean distance ordered partial search (MPS)

Speed-up of k-means

- Reduced-search based on centroid activity

External search data structures

- Nearest neighbor graph
- Kd-tree

Partial distortion search (PDS)

[Bei and Gray, 1985: *IEEE Trans. Communications*]

- Current best candidate gives upper limit.
- Distances calculated cumulatively.
- After each addition, check if the partial distortion exceeds the smallest distance found so far.
- If it exceeds, then terminate the search.

$$e_{a,j} = \sum_{k=1}^{K} \left(c_{ak} - c_{jk} \right)^2$$

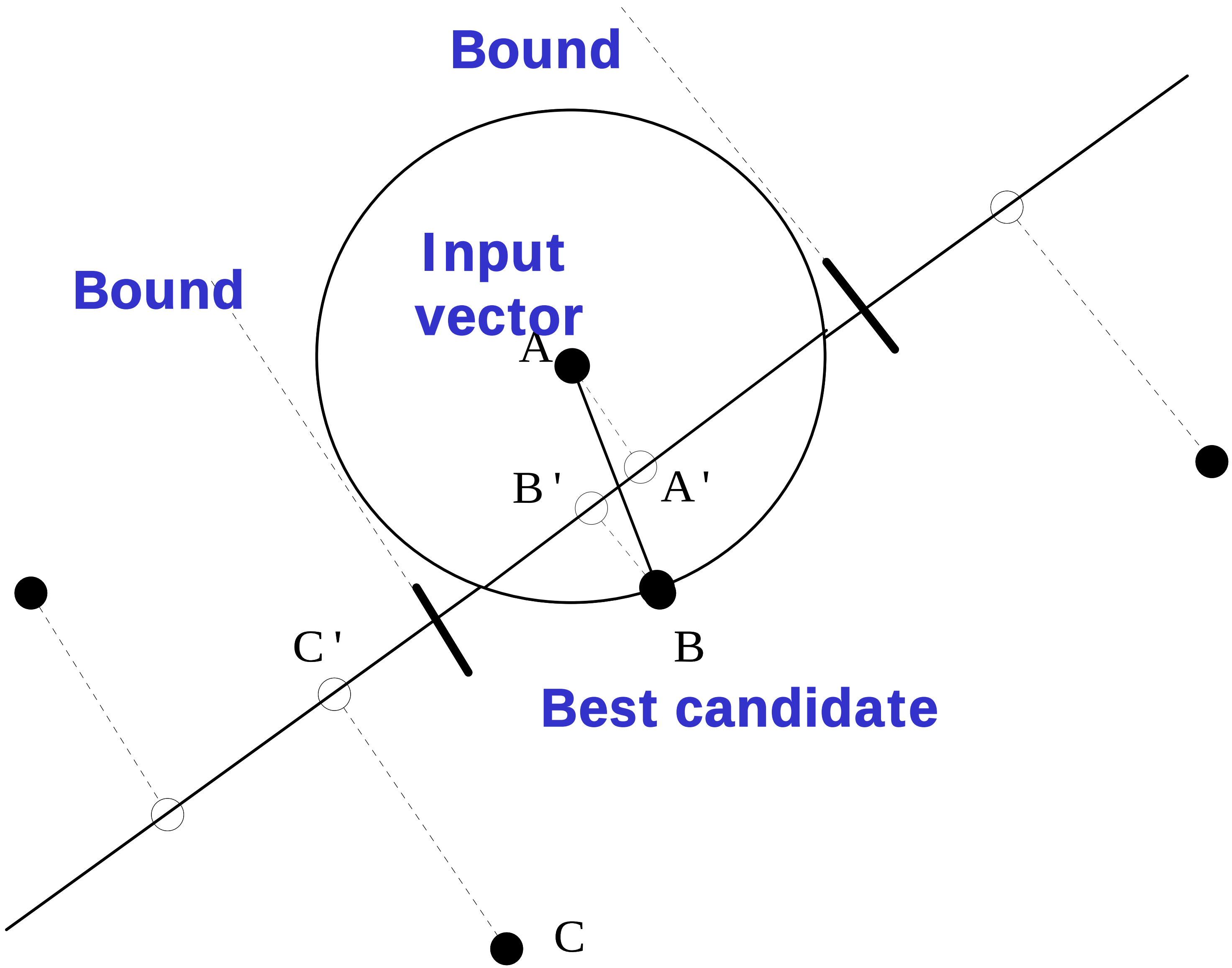
Mean-distance ordered partial search (MPS)

[Ra and Kim, 1993: *IEEE Trans. Circuits and Systems*]

- Calculate distance along projection axis.
- If distance is outside bounding circle defined by the best candidate, drop the vector.

$$\left(\sum_{k=1}^{K} x_i^k - \sum_{k=1}^{K} c_b^k \right)^2 > K\, d(x_i, c_a)$$

Bounds of the MPS method



Pseudo code of MPS search

SearchNearestNeighborUsingMPS(c_a, c_j, d_{min}) $\rightarrow$ nn_a, d_a;

$d_{min} \leftarrow \infty$;
up $\leftarrow$ TRUE;
down $\leftarrow$ TRUE;
$j_1 \leftarrow a$;
$j_2 \leftarrow a$;

WHILE (up OR down) DO

IF up THEN
$j_1 \leftarrow j_1 + 1$;
IF $j_1 > N$ THEN up $\leftarrow$ FALSE
ELSE CheckCandidate(s_a, s_{j1}, n_a, d_{min}, nn, up);

IF down THEN
$j_2 \leftarrow j_2 - 1$;
IF $j_2 < 1$ THEN down $\leftarrow$ FALSE
ELSE CheckCandidate(s_a, s_{j2}, n_a, d_{min}, nn, down);

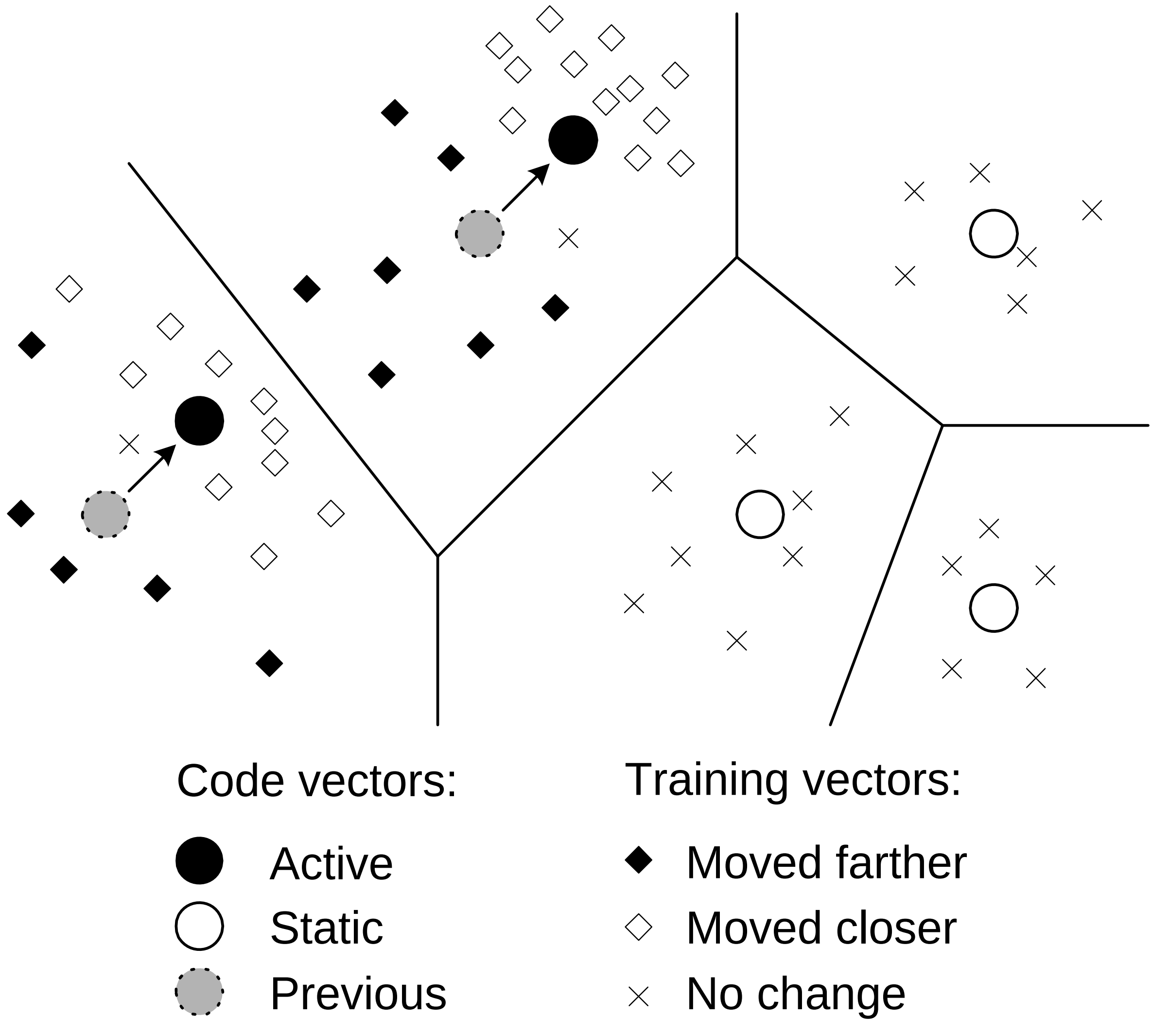
END-WHILE;

RETURN nn, d_{min};

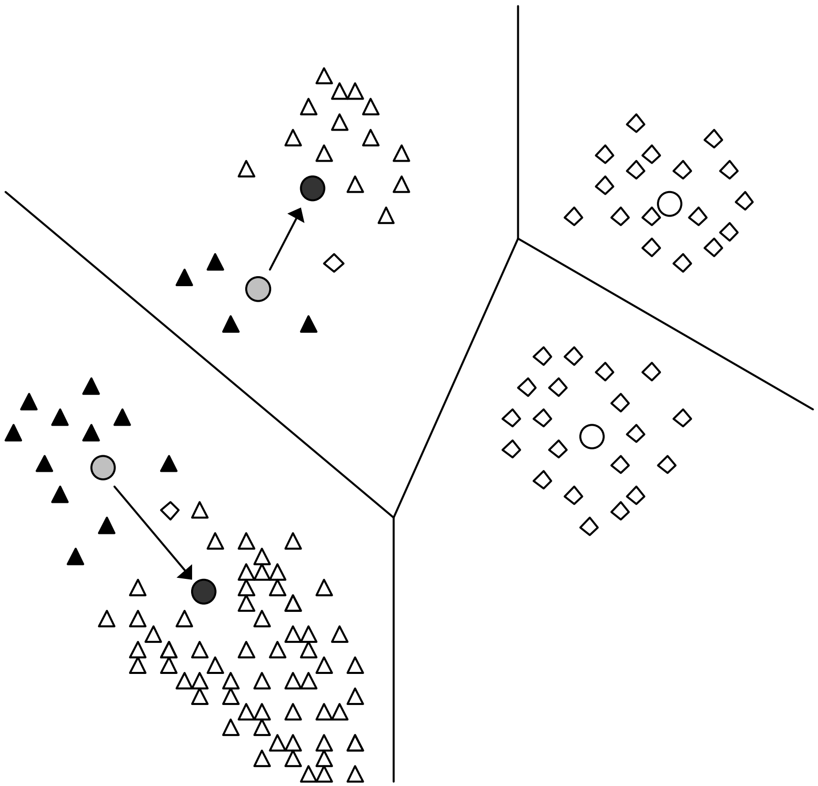
Needs updates

Activity classification

[Kaukoranta et al., 2000: *IEEE Trans. Image Processing*]



Reduced search based on activity classification



Lähimmän sentroidin haku määräytyy seuraavasti:

	▲	△	◇
●	T	O	O
○	-	-	O

T = täysi haku
O = osittainen haku

Etäisyyslaskujen määrä:

	▲	△	◇
●	100%	4%	4%
○	0%	0%	4%

Osuus alkioista:

	▲	△	◇
●	3,6%	3,7%	0,1%
○	0%	0%	92,6%

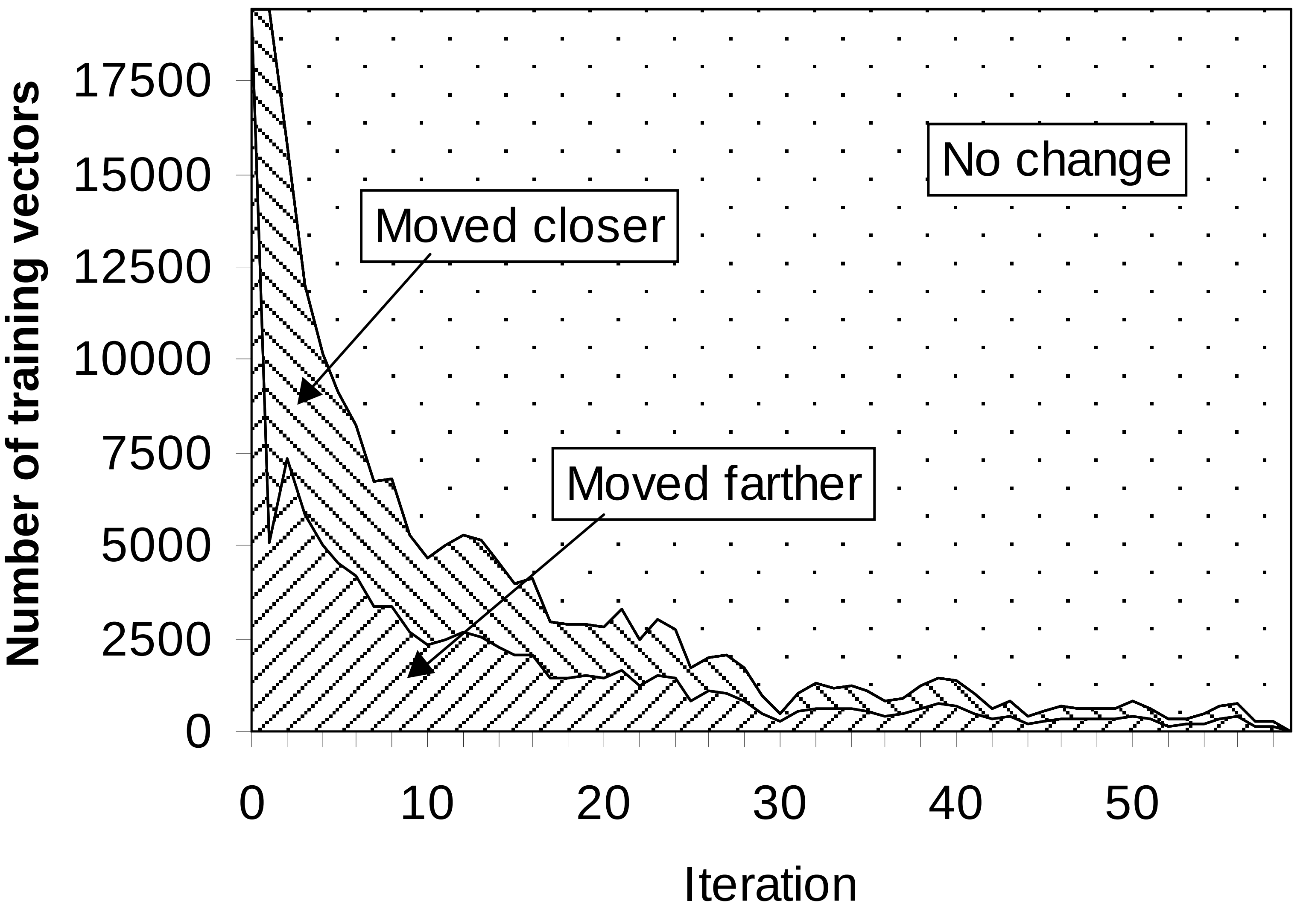
Sentroidit:

- ● Aktiivinen, uusi sijainti
- ◉ Aktiivinen, vanha sijainti
- ○ Staattinen

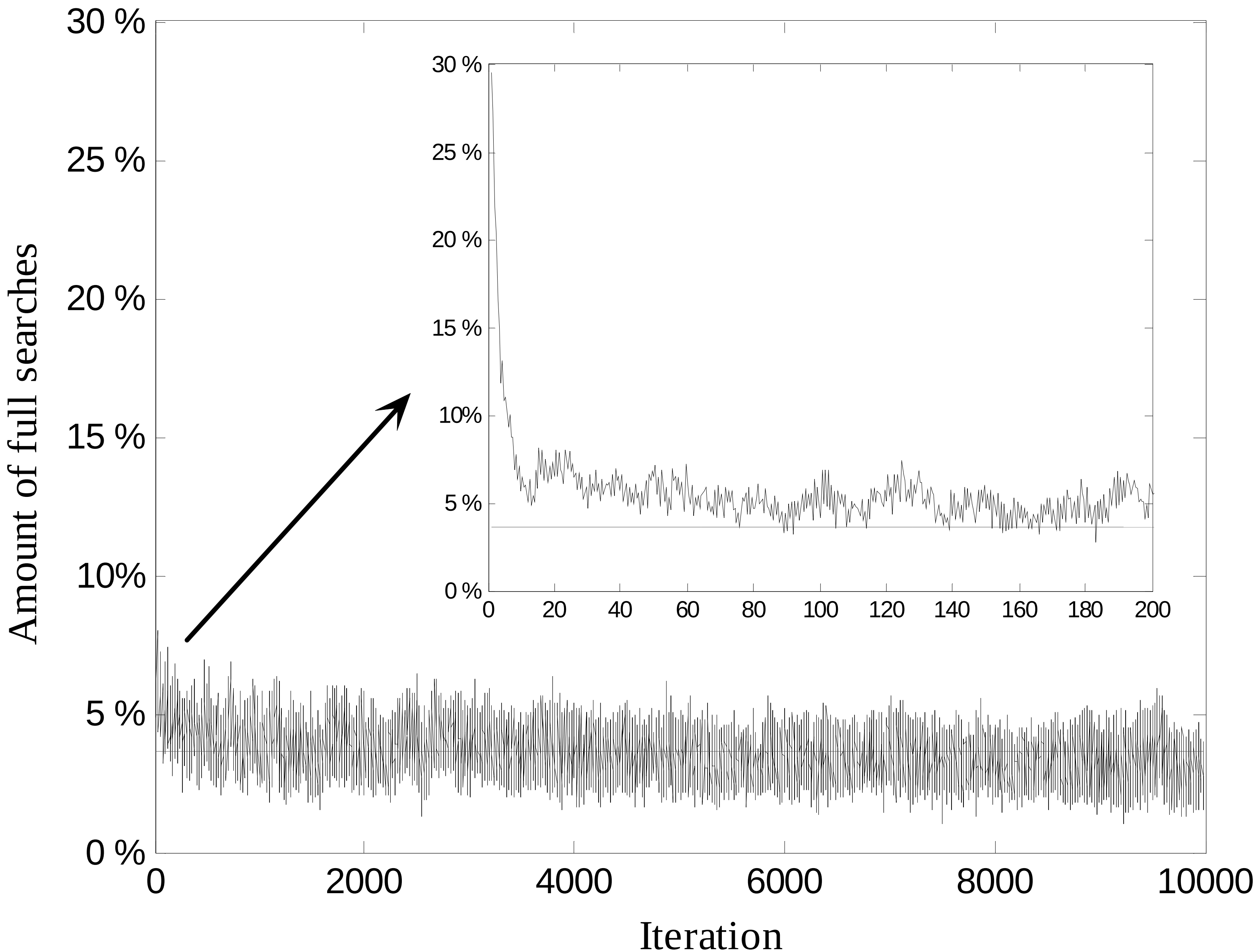
Alkiot:

- △ Sentroidi siirtynyt lähemmäksi alkiota
- ▲ Sentroidi siirtynyt kauemmaksi alkiosta
- ◇ Etäisyydessä sentroidiin ei muutosta

Classification due to iterations



Activity of vectors in Random Swap



Effect on distance calculations

K-means

	Distance calculations / search	Dimensions / distance calculation	Dimensions / search
Full	255.97	16.00	4095.48
PDS	255.97	2.34	598.96
MPS+PDS	26.97	8.07	217.60

K-means with activity classification

	Distance calculations / search	Dimensions / distance calculation	Dimensions / search
Full	61.44	16.00	983.07
PDS	61.44	2.08	127.98
MPS+PDS	5.35	6.72	35.97

Effect on processing time

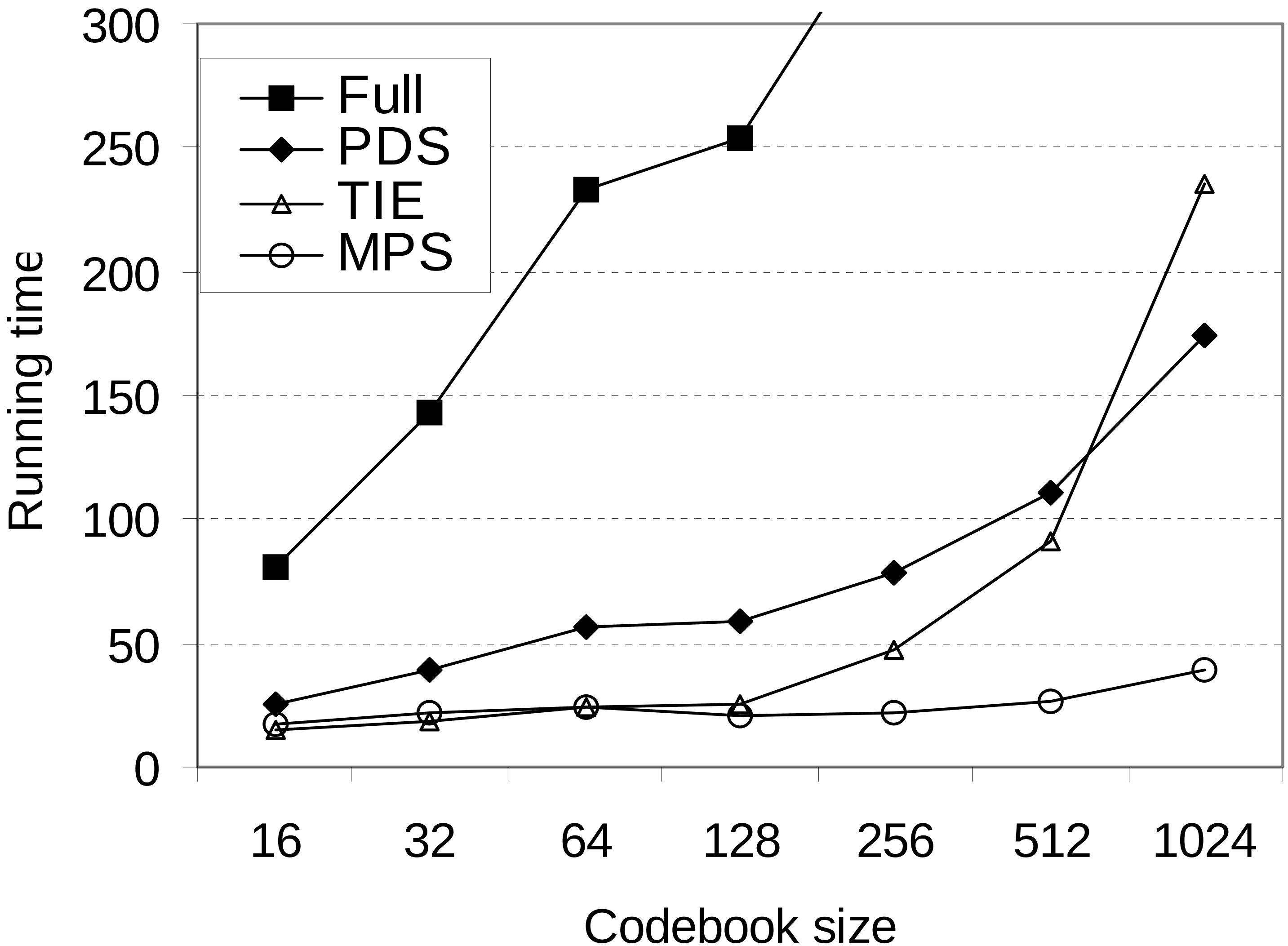
For improving K-means algorithm

	Bridge		Miss America	
	Without grouping	With grouping	Without grouping	With grouping
Full	127.6	46.1	1344.5	336.2
PDS	33.4	13.0	311.1	75.8
MPS+PDS	12.4	4.8	97.3	21.5

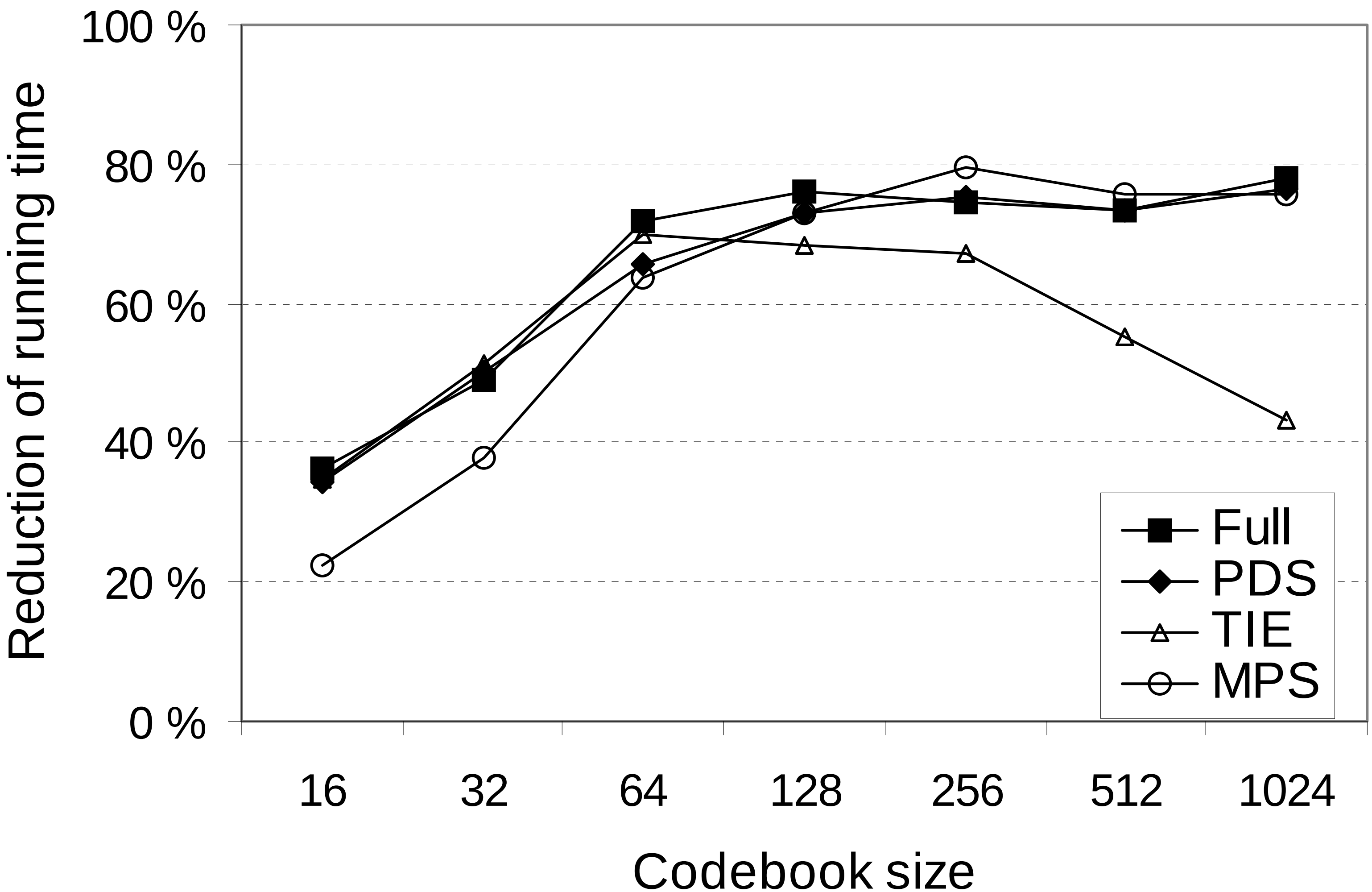
3.8 %

1.6 %

Comparison of speed-up methods

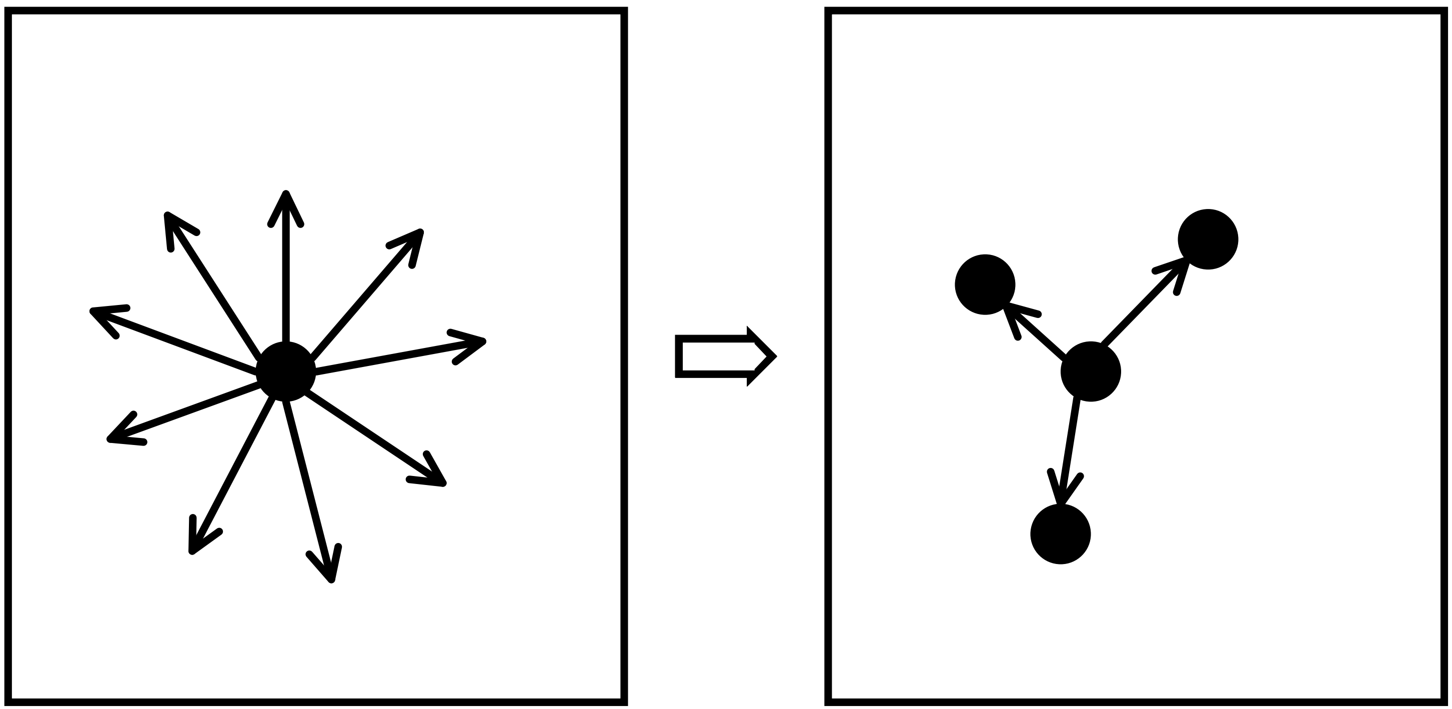


Improvement of reduced search



Neighborhood graph

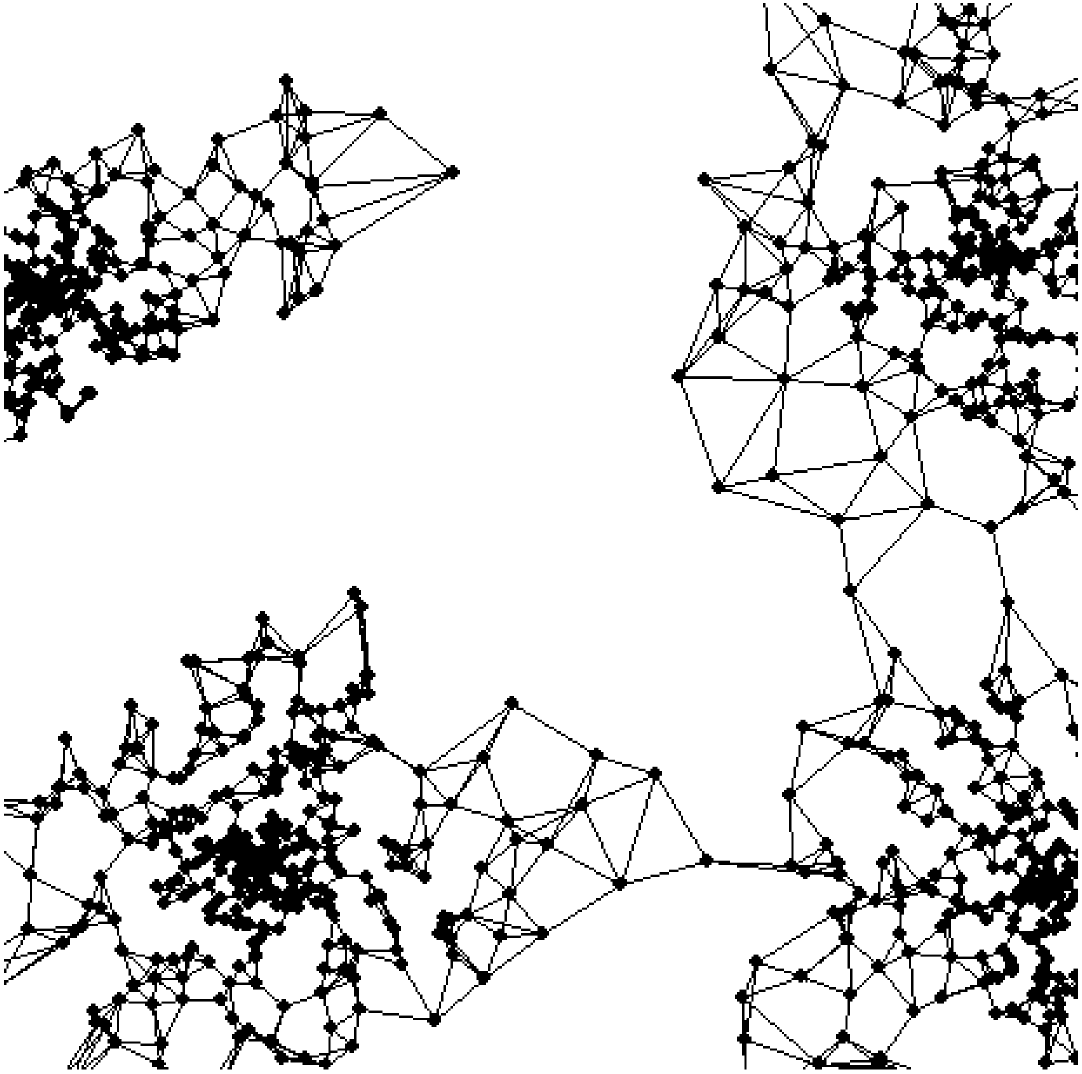
Full search: **Graph structure:**



Full search:
O(*N*) distance calculations.

Graph structure:
O(*k*) distance calculations.

Sample graph structure



Literature

1. T. Kaukoranta, P. Fränti and O. Nevalainen, "A fast exact GLA based on code vector activity detection", *IEEE Trans. on Image Processing*, 9 (8), 1337-1342, August 2000.
2. C.-D. Bei and R.M. Gray, "An improvement of the minimum distortion encoding algorithm for vector quantization", IEEE Transactions on Communications, 33 (10), 1132-1133, October 1985.
3. S.-W. Ra and J.-K. Kim, "A Fast Mean-Distance-Ordered Partial Codebook Search Algorithm for Image Vector Quantization", *IEEE Transactions on Circuits and Systems*, 40 (9), 576-579, Sebtember 1993.
4. J.Z.C. Lai, Y.-C.Liaw, J.Liu, "Fast k-nearest-neighbor search based on projection and triangular inequality", *Pattern Recognition*, 40, 351-359, 2007.
5. C. Elkan. Using the Triangle Inequality to Accelerate k-Means. *Int. Conf. on Machine Learning*, (ICML'03), pp. 147-153.

Literature

6. James McNames, "A fast nearest neighbor algorithm based on a principal axis search tree", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(9):964-976, September 2001.
7. J.H. Friedman, J.L. Bentley and R.A. Finkel, "An algorithm for finding best matches in logarithmic expected time," *ACM Trans. on Mathematical Software*, 3 (3), pp. 209-226, September 1977.
8. R. Sproull, "Refinements to nearest-neighbor searching in *K*-d tree," *Algorithmica*, 6, pp. 579-589, 1991.