Modified DBSCAN Algorithm on Oculomotor Fixation Identification

Beibin Li, Quan Wang, Erin Barney, Logan Hart, Carla Wall, Katarzyna Chawarska,
Irati Saez de Urabain, Timothy J. Smith, Frederick Shic Child Study Center, School of Medicine, Yale University

Motivation: While dispersion based fixation identification algorithms can hardly deal smooth pursuit, velocity based ones are too noisy.

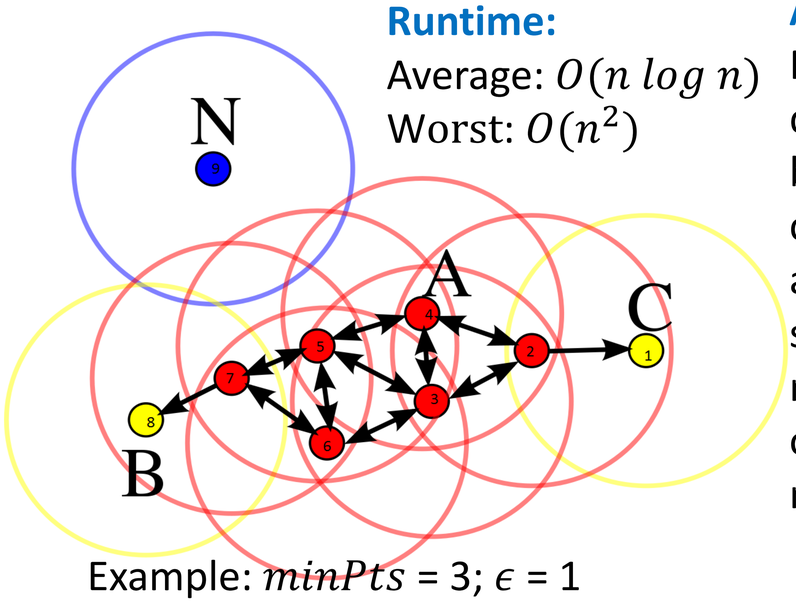
Definition

Point p is a **core point** if:

1. at least $minPts$ points are within distance ϵ to point p; and
2. these points form a consecutive subsequence $p_0, p_1, \dots, p_k$ of the dataset, where p_i and p_{i+1} are adjacent in time.
3. Those points are called **directly reachable** from p.

A point q is **reachable** from q if there is a path $p_1 = p, p_2, \dots, p_n = q$ where each p_{i+1} is directly reachable from p_i .

If p is a core point, then it forms a **cluster (fixation)** together with all points that are reachable from it.



Example: $minPts$ = 3; ϵ = 1

Runtime:
Average: $O(n\ log\ n)$
Worst: $O(n^2)$

Advantages:
Fixations are density based and can be arbitrary shaped. It's robust to outliers and noises.

Fixation Data	Temporal fixation overlap	SP Trajectory Coverage	# of fixations	coverage per fixation
MDBSCAN	4.1(1.5)	77%(25%)	7.5(2.6)	12.7%(6.7%)
I-DD	4(1.5)	73%(26%)	21.2(10.4)	4.2%(1.3%)
I-VT	2.2(1.9)	63%(32%)	10.9(6.8)	7.1%(2.4)

Performance

Combined Advantages of:
- Velocity based algorithm
- Dispersion based algorithm

github.com/BeibinLi/MDBSCAN

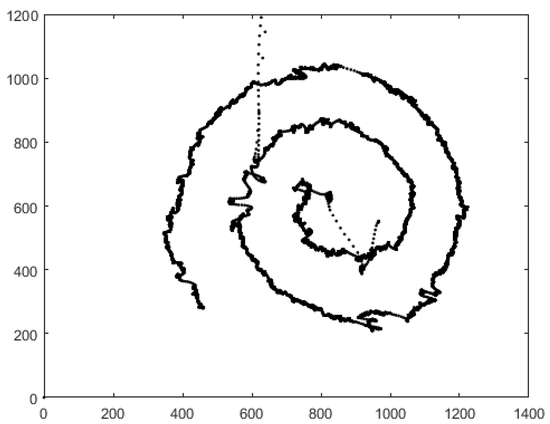
Comparison

Using GraFix, a semi-manual fixation identification tool, as ground truth, we compared the performance of I-DD, I-VT, and MDBSCAN. (Note: GraFix is based on I-VT). Linear Mix Model is applied to the comparison.
In Terms of Mean fixation time, I-VT is the closest to ground truth ($F(2487.9)=129.0$, $p< 0.001$); in terms of number of fixations per second, MDBSCAN is the closest ($F=18$, $p<0.001$); in terms of fixation coverage time, I-DD is the closest ($F=10.3$, $p<0.001$). All these algorithms have their own advantages. Further studies can research in their differences.

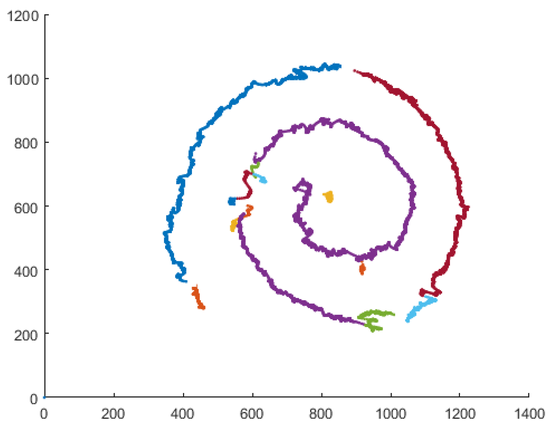
Acknowledgement: K01 MH104739, R21 MH102572, CTSAUL1 RR024139, R03 MH092618, NIH R01 MH100182, R01 MH087554, U19 MH108206; NSF #1139078, #0835767, DOD W81XWH-12-ARP-IDA, The Nancy Taylor Foundation for Chronic Diseases , and the Associates of the Child Study Center.

Beibin.li@yale.edu Frederick.shic@yale.edu

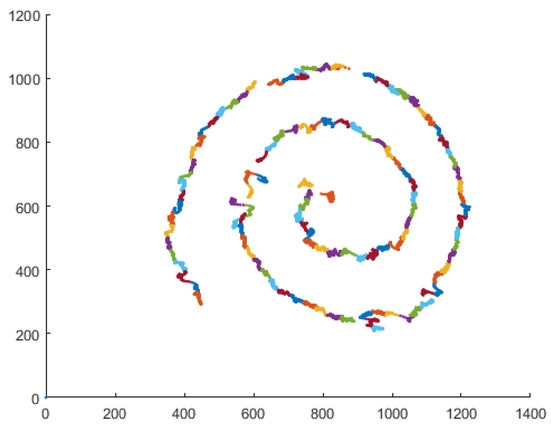
Original



MDBSCAN



I-DD



I-VT