# SMOOTH, REAL-TIME COMPETITOR CLUSTERING

BY CHRISTIAAN BOTHA SUPERVISOR: MARK WINTER RES. PROF.: ELMAR EISEMANN

## 1. INTRODUCTION

Clustering data allows for a visual analysis [1].
Current algorithms cluster once per day [2].

To analyse sailing regattas, the solution needs to:
- Cluster competitors 60 times per second.
- Produce a smooth timeline of results
- Cluster hierarchically

**"What are efficient methods for smooth, hierarchical clustering for multi-scale race visualisation?"**

## 2. METHODOLOGY

**Clustering competitors**

A distance function assigns clusters their meaning, because it calculates the similarity between competitors $c_i$, $c_j$

$$d_{euc}(c_i, c_j) + \alpha \cdot d_{dir}(c_i, c_j) + \beta \cdot d_{dtl}(c_i, c_j)$$

$d_{euc}(c_i, c_j)$ is the Euclidean distance
$d_{dir}(c_i, c_j)$ calculates the difference in tack
$d_{dtl}(c_i, c_j)$ calculates if competitors are catching up
$\alpha$ $\beta$ are weighting parameters

These weights were used to determine a stable distance function.

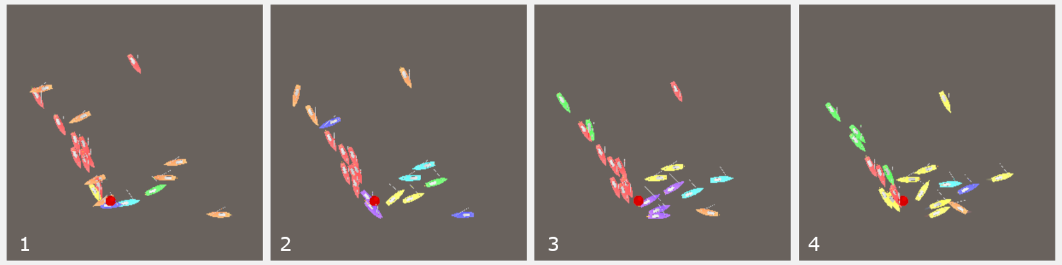
**Smoothing timeline**

Continuously clustering dynamic data is inherently unstable [3].
The timeline is stabilised by the smoothing distance matrix:

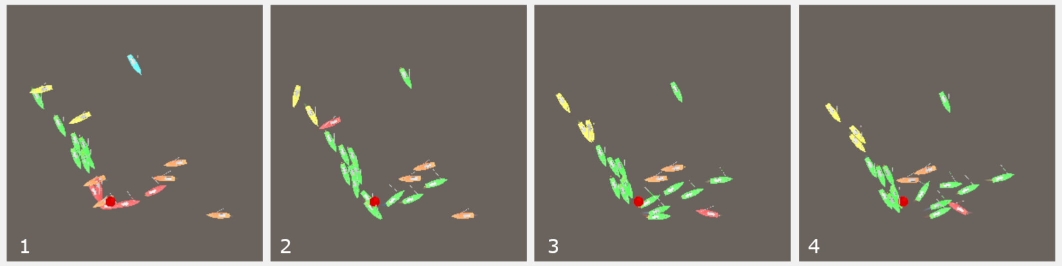
$$(1 - \delta) \cdot M_t + \delta \cdot M_{t-1}$$

$M_t$ is the distance matrix at time $t$
$\delta$ is the smoothing value

An adaptive δ can be chosen based on the average tree distance between two frames.

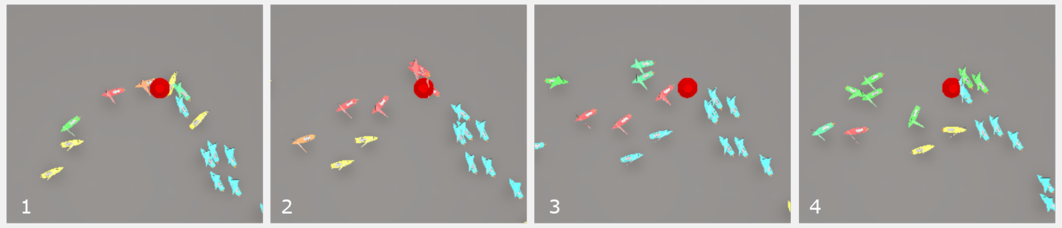



(a) Unstable distance function with weights α=0.5 and β=0.25

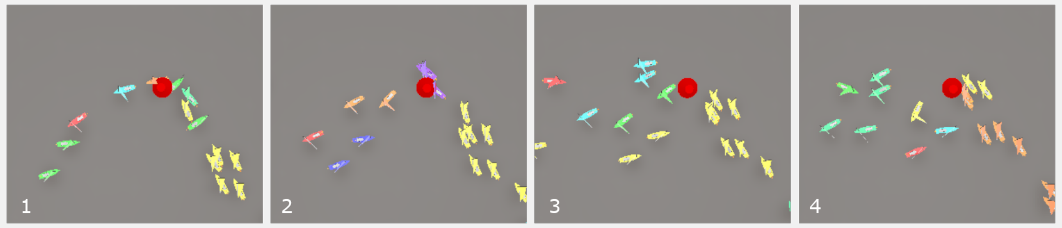



(b) Stable distance function with weights α=0.5 and β=0

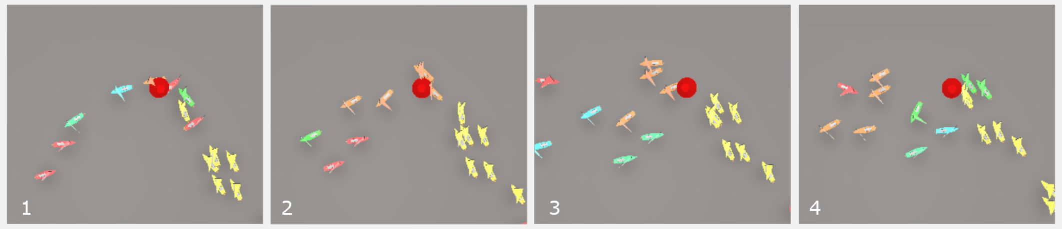
Figure 1: The differences between a stable and unstable distance function over time.




(a) No smoothing




(b) A smoothing value δ=0.6




(c) An adaptive smoothing value

Figure 2: The effects of different smoothing values over time.

The colours represent different clusters of competitors.
The red sphere represents a mark competitors have to pass.

## 3. RESULTS

Figure 1a shows that clusters change often for unstable distance functions.

With a stable distance function, clusters remain largely the same, as in Figure 1b.

A static smoothing value either produces an unstable clustering or is not accurate.

Yellow competitors in Figure 2a change cluster for one frame.

Clustering in Figure 2b is delayed three frames.

The adaptive smoothing value of Figure 2c is both accurate and stable.

## 4. CONCLUSION

Combining Euclidean distance and tack results in the most stable distance function.

An adaptive smoothing value further increases stability and remains faithful to the data.

## 5. LIMITATIONS

The expected performance is $O(n^3)$, but could be improved to $O(n^2)$.

Uncertain competitor direction causes the first few seconds of the regatta to be unstable.

[1] C. Tominski, S. Gladisch, U. Kister, R. Dachselt, and H. Schumann, “Interactive lenses for visualization: An extended survey”, *Computer Graphics Forum*, vol. 36, no. 6, pp. 173–200, 2017. doi: https://doi.org/10.1111/cgf.12871.

[2] C. C. Aggarwal, P. S. Yu, J. Han, and J. Wang, “A framework for clustering evolving data streams”, Proceedings 2003 VLDB Conference, pp. 81–92, 2003. doi: 10.1016/b978-012722442-8/50016-1.

[3] D. Chakrabarti, R. Kumar, and A. Tomkins, “Evolutionary clustering”, in Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD ’06, Philadelphia, PA, USA: Association for Computing Machinery, 2006, pp. 554–560, isbn:1595933395. doi: 10.1145/1150402.1150467.

Delft University of Technology

J.C.Botha@student.tudelft.nl
4802993