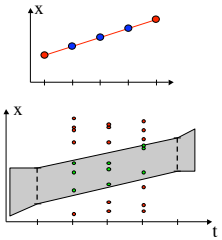
Variable Tree Algorithms

Key Intuition:

We only need to search over just enough points to define the model! But we still need to account for the support points (in order to use structure from other aspects of the problem).

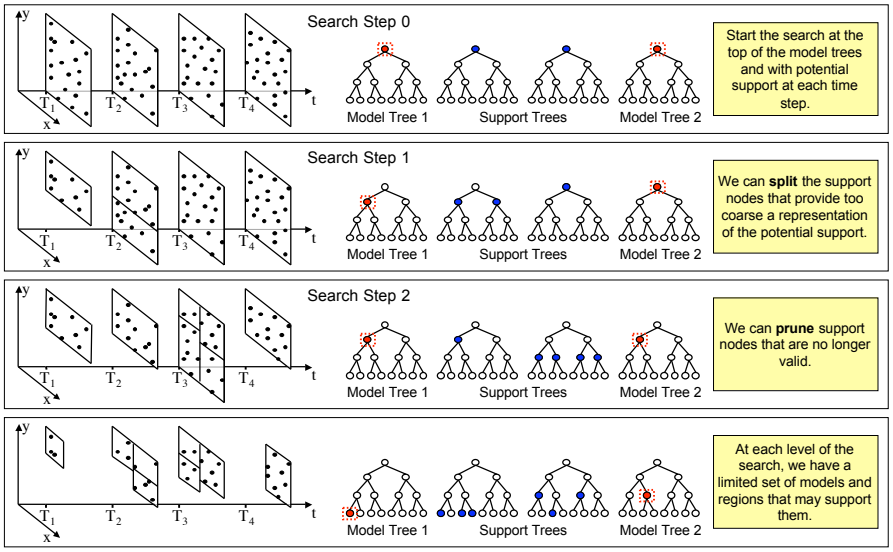
- We can break the points into two conceptual classes:
  - **Model Points** - points required to fit a model's free parameters.
  - **Support Points** - points that confirm or support the model.
- We can use subsets of model points to define regions of space through which a model (using these points) could pass. This information can be used to determine which points can or cannot support the model.



We can limit our search to just the model points and use the structure they provide (i.e. a subset of the models) to test the feasibility of the other supporting points.

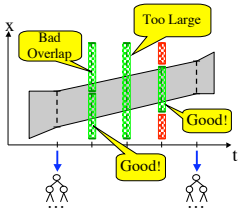
Variable Trees Algorithm (in short):

- Perform a multiple tree search over model points.
- Place the support points in one or more trees.
- Maintain a dynamic list of valid support tree nodes:
  - Remove nodes that are infeasible (given the model tree nodes).
  - Split nodes providing too coarse a representation of support space.
- Use the list of valid support nodes to help prune.
- Uses structure from all parts of the problem to facilitate pruning.
- Remove redundant work by pruning support points for many sets of model points at once.
- Adding more support points/evidence does not increase the branching factor.
- Use spatial structure from the support points to reduce computation.



How/when do we prune/split support nodes?

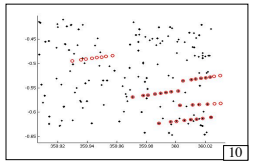
- We can prune whenever the support node cannot support any model defined by the model tree nodes
- We can split if the node is a "bad" representation (too coarse):
  - One of its children could be pruned immediately, or
  - It is too wide.

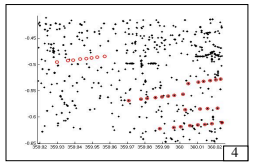


Experiments and Results

Asteroid Linkage:

- **Goal:** Find intra-night linkages, linear tracks that appear in at least 7 of 8 images.
- Images from 3.6-meter Canada-France-Hawaii Telescope (30 minute spacing).
- Limited preprocessing (e.g. no Halo removal)
- We can test the effect of *pushing into the noise* by varying minimum detection significance threshold.





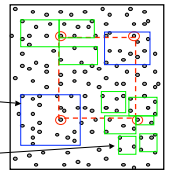
Sigma	10	8	6	5	4
N	3531	5818	12911	24068	48646
Sequential	2	7	61	488	2442
Multiple Tree	1	3	30	607	4306
Support List	4	10	64	498	2399
Variable Tree	< 1	1	4	40	205

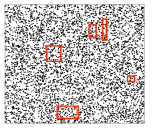
Algorithm running time (in seconds) vs. significance threshold (sigma) and number of points.

- Multiple tree algorithm takes a huge hit from the 8 time steps.
- Variable tree algorithm scales best (by 10x)!
- We would expect this advantage to **increase** with the number of time steps and the density.

Pattern/Template Matching:

- Consider the "simple" problem of finding rectangles.
- We can fully define the model with 2 opposite corners.
  - Search for rectangles using just 2 **model trees**.
- We can use the other corners for support.
  - Start with 2^D-2 support tree nodes.
  - Require at least one valid **support tree node** for each corner.





N	500	2500	10000	25000	50000
Brute Force	0.4	21.1	n/a	n/a	n/a
Single Tree	0.2	0.3	10.1	66.2	293.1
Multiple Tree	0.1	0.1	1.1	6.6	27.8
Variable Tree	0.1	0.1	0.8	4.3	16.3

Algorithm running time (in seconds) vs. the number of points.

- Again variable tree algorithm scales the best.
- The full multiple tree algorithm also performs well, but has trouble if we allow missing corner points.

Conclusions:

- Spatial structure can provide significant opportunities for computational savings.
- The use of multiple trees allows us to use structure from all aspects of the problem, but introduces a higher branching factor.
- The use of a (small) fixed number of model/search trees and a list of additional supporting tree nodes:
  - allows us to use structure from all aspects of the problem,
  - does not increase the branching factor,
  - allows us to change the representation of the supporting points on the fly so as to adapt it to the current search state.
- Thus we can address the trade-off between single tree and multiple tree approaches.

Acknowledgements

Jeremy Kubica is supported by a grant from the Fannie and John Hertz Foundation. Andrew Moore and Andrew Connolly are supported by a National Science Foundation ITR grant (CCF-0121671).