Efficiently Tracking Moving Sources in the LSST

J Kubica (CMU), T Axelrod (UA), K Barnard (UA), A Connolly (Pitt),
L Denneau (Hawaii), A Efrat (UA), J Heasley (Hawaii), R Jedicke (Hawaii),
B Moon (UA), A Moore (CMU), S Morris (UA), P Rao (UA),
and the LSST Collaboration

Large Synoptic Survey Telescope

The LSST will survey the sky with a cadence of several visits per month. This time sampling enables a detailed census of Solar System objects ranging from over a million Main Belt Asteroids, to 20,000 Trans Neptunian Objects and even potentially hazardous asteroids. The challenge in identifying all potential asteroid associations is that it can be very computationally expensive (due to the many potential tracks that must be tested in order to isolate true orbits). To address this issue, we have developed new approaches for tracking asteroids and for recovering orbits from isolated observations that use spatial structures in order to ignore ``obviously'' bad candidate sets. The goal here is to reduce the number of proposed associations that must be tested, not just to reduce the cost of a test. To this end, we have developed a new class of search algorithms based on spatial data structures (e.g. KD-trees) that use a variable number of tree nodes to capture information about the current state of a search. These algorithms are able to exploit simultaneously all available observations to limit the number of candidate asteroid tracks. Initial applications show that these approaches scale to the size of the LSST domain, can recover 95% of asteroids with a small false positive rate and have the potential to provide almost real time identification of moving sources. We validate and improve orbits by matching future and prior observations to the proposed orbits. Here we consider the probability that each observation should be paired with each hypothesized orbit in its vicinity, and maximize the overall association likelihood using bipartite graph matching. This gives a global solution that enforces the one-to-one constraint and provides for control of outliers. We propose strategies to allow mismatches consistent with estimated probabilities that the items do not have a match.

Linking Observations to Track the Orbits of Asteroids within the Solar System

Final Goals: Find ~90% of potentially hazardous asteroids with sizes >300m
Track over 10^6 main belt asteroids over a period of 10 years
Calculate the orbits for all linked observations

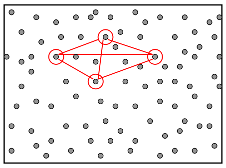
An artist's rendering of an asteroid impact (NASA).

Immediate Goal: to link together sources detected at different times that correspond to the same true *object* (asteroid).

Linkage or **Track Initiation** focuses on finding these associations *without* an initial track estimate.

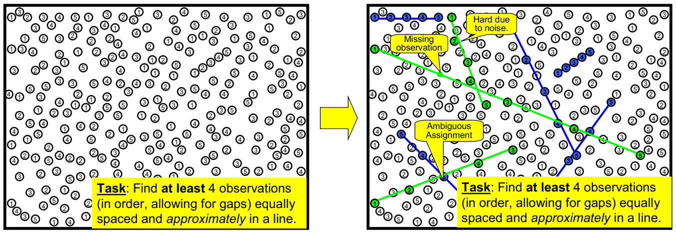
Problem Details:

- We are looking for known types of structure buried within the data.
- We are particularly interested in high-density, low-support domains.
- Unfortunately, this problem is very computationally expensive.
- We need to **efficiently find this structure in large noisy data sets**.



Examples of the Problem:

- **Template matching** ("Find all occurrences of this pattern.")
- **Spatial search** ("Find this type of structure in the data.")
- **Track initiation** ("Find all 'new' tracks in the data.")
- **'n'-point Correlation** ("How many times is this set of relations satisfied?")

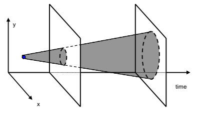


Task: Find **at least** 4 observations (in order, allowing for gaps) equally spaced and *approximately* in a line.

Task: Find **at least** 4 observations (in order, allowing for gaps) equally spaced and *approximately* in a line.

Why Is This Hard/Interesting?

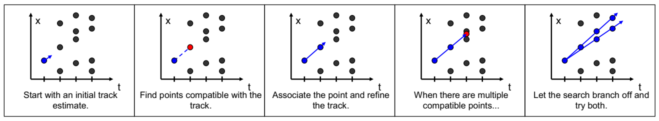
- Scale (many many detections):
  - LSST will be constantly scanning the sky.
  - Number of detections explodes as we look for fainter asteroids (both real and spurious noise).
- Lack of initial parameter information.
- Partial observations (we see projection on the sky, not distance to object).
- Sparse cadence where we see the sample field on the sky periodically.



It is easier to associate points that are **close** in time, because the point cannot move as much!

Constructive "Single Tree" Algorithms

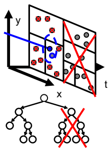
Key Intuition: Constructively build up the set of points (one point at a time), using the points already in the set and the model to determine the validity of the next point.



Start with an initial track estimate. | Find points compatible with the track. | Associate the point and refine the track. | When there are multiple compatible points... | Let the search branch off and try both.

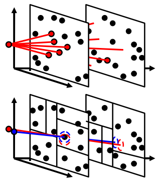
We can use spatial data structures to search efficiently (e.g. Uhlmann 1993):

- Create a tree on the points.
- Use the tree to efficiently find points compatible with the current model.
- For example in tracking, use the tree to find points near the predicted track position.



Drawbacks:

- We only use a structure from one aspect of the problem at a time. May miss "obvious" pruning opportunities if information from later time steps is considered.
- We may repeat work over similar initial sets.



"Multiple Tree" Algorithms

Key Intuition: Can we use structure from **all** of the time steps at once?

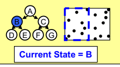
Yes! We can search over all sets/combinations of points using multiple tree nodes!

Use a **multiple tree algorithm** (Hjaltason and Samet 1998, Gray and Moore 2001):

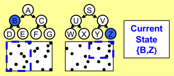
–Build *multiple* kd-trees, and

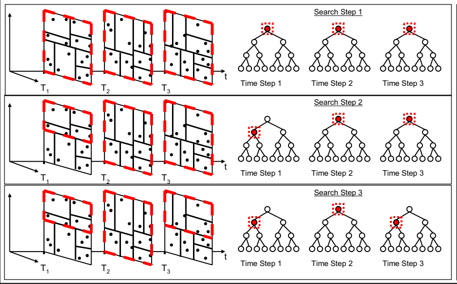
–Do a search of *combinations* of tree nodes.

Traditional "single tree" algorithms use a single node to capture the search state.



Multiple tree algorithms use a **combination** of tree nodes to capture the search state.

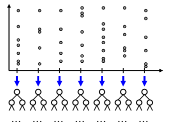




Drawbacks:

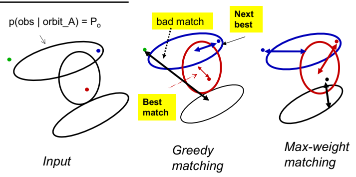
- By searching over all sets, the set is exponential in the number of time steps! $O(N^T)$
- Increased potential for taking deeper "wrong turns."

This is an additional challenge if we have many time steps or need many supporting points.



Matching Observations to Orbits

- Every orbit fit implies a probability distribution over its parameters, which implies a second distribution of expected observations
- For any observation, a, we estimate *P(a)*, the probability that a is a real object (i.e not noise).
- For any orbit A, we estimate *P(A)*, the probability that the object associated with A will be seen in the observation
- For any pair of an observation a and orbit, *A*, we estimate *P(a|A)*, the probability that the observation a is the object associated with *A*.
- Due to viewing conditions and orbital parameters every orbit has a probability that it is not observed in a given image (modulated by our certainty that it exists at all)



Generate a bipartite graph *G(Obs ∪ Obj, E),* and use min-weight algorithm. The weight of the edge *(a,A))* is -log*(p(a,A))*
Small circles: Observations
Ovals: Projected prediction for orbits at some threshold.

- Algorithmic tricks:
  - Introduce phantom observations with match probability set to the probability that the orbit being matched exists but will not be observed.
  - Introduce phantom orbits with match probabilities set to the probability that the observation being matched might be noise.
- Updated linkages give updated orbital parameters
- Poor matches/missed observations reduce our belief that the orbit is correct.
- Past observations must be evaluated against the evolving data sets (new orbits should be consistent with past observations).

Experiments and Results

Asteroid Linkage:

Goal: Find intra-night linkages that appear in at least 7 of 8 images.

Images from 3.6-meter Canada-France-Hawaii Telescope (30 min spacing).

Limited preprocessing (e.g. no bright star removal)

We can test the effect of *pushing into the noise* by varying minimum detection significance threshold.

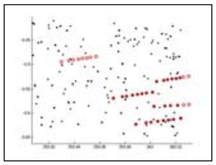
Sigma	10	8	6	5	4
N	3531	5818	12911	24068	48646
Sequential	2	7	61	488	2442
Multiple Tree	1	3	30	607	4306
Support List	4	10	64	498	2399
Variable Tree	< 1	1	4	40	205

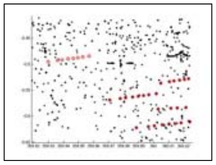
Algorithm run time (seconds) vs. significance threshold (sigma) and number of points.

Multiple tree algorithm takes a huge hit from the 8 time steps.

Variable tree algorithm scales best (by 10x)!

Advantage **increases** with number of time steps and density.





Conclusions:

Spatial structure provide significant opportunities for computational savings.

Scaling to the size, cadence and density of the data expected from the LSST, we can recover both the Main Belt Asteroids and the Potentially Hazardous Asteroids. Computational challenges remain due to missing data (i.e. sources dropping below the detection threshold or gaps due to weather), and when we link singleton observations.

The LSST research and development effort is funded in part by the National Science Foundation under Scientific Program Order No. 9 (AST-0551161) through Cooperative Agreement AST-0132798.
Additional funding comes from private donations, in-kind support at Department of Energy laboratories and other LSSTC Institutional Members.

National Optical Astronomy Observatory
Research Corporation
The University of Arizona
University of Washington
Brookhaven National Laboratory
Harvard-Smithsonian Center for Astrophysics
Johns Hopkins University
Las Cumbres Observatory, Inc.
Lawrence Livermore National Laboratory
Stanford Linear Accelerator Center
Stanford University
The Pennsylvania State University
University of California, Davis
University of Illinois at Urbana-Champaign