

Making the sky searchable

David W. Hogg (NYU)

people

- David W. Hogg (NYU) – *astro PI*
- **Dustin Lang (Toronto)**
- Keir Mierle (Toronto, Google)
- Sam Roweis (Toronto, Google) – *comp sci PI*
- (with help from Blanton, Finkbeiner, Barron, Stumm)

non-text queries

- need to search things that aren't text, with queries that aren't text
- even “image search” in Google requires accurate text meta-data
- multi-billion-dollar question: “Here's a picture, what is it a picture of?”
 - we have answered this in *one tiny domain*

our problem is easy

- the sky is just a set of points in 2-d
- excellent catalogs exist (esp USNO-B)

our problem is hard

- the sky is big!
- astronomical images are small
- bandpasses and sensitivities of images do not match those of the astrometric catalogs
- we don't necessarily know *anything* about the images we see

demo

- [this was a live demo of the system, plus a discussion of the different kinds of images that solve: amateur, professional, and archival.]

what we can solve

- digital camera snapshots
- amateur astrophotography
- archival photographic plates
- science imaging from modern equipment
 - LBT, GALEX, NOAO, IMACS, *etc*
 - SDSS: 9x12 arcmin, 55 sec, 2.5 m, 99.84% of 336,554 fields, zero false positives

what we can't

- x-ray, far-infrared, and radio images
 - hope to bootstrap to these bandpasses
- images smaller than a few arcmin
 - limitation is *fundamental* at present: *catalogs*
 - but we can build *new* catalogs
- ultra-deep images with terrible saturation
 - limitation is *not* fundamental

how it works

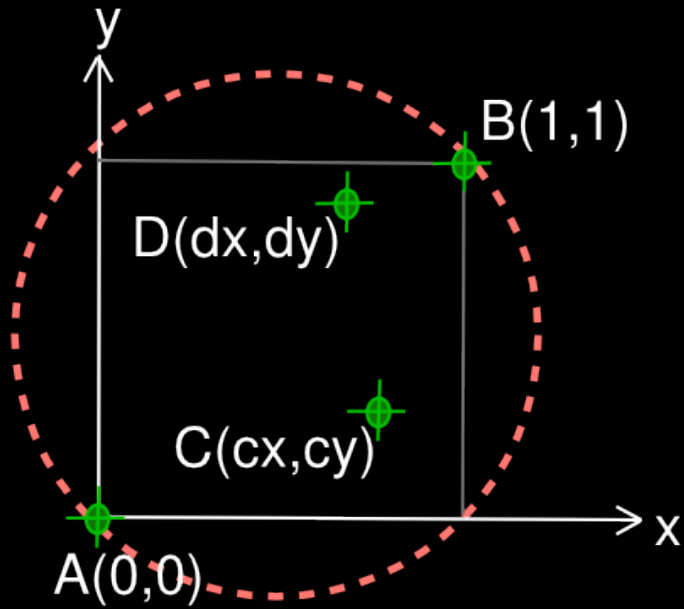
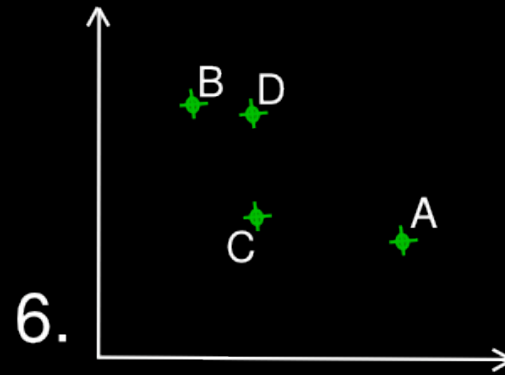
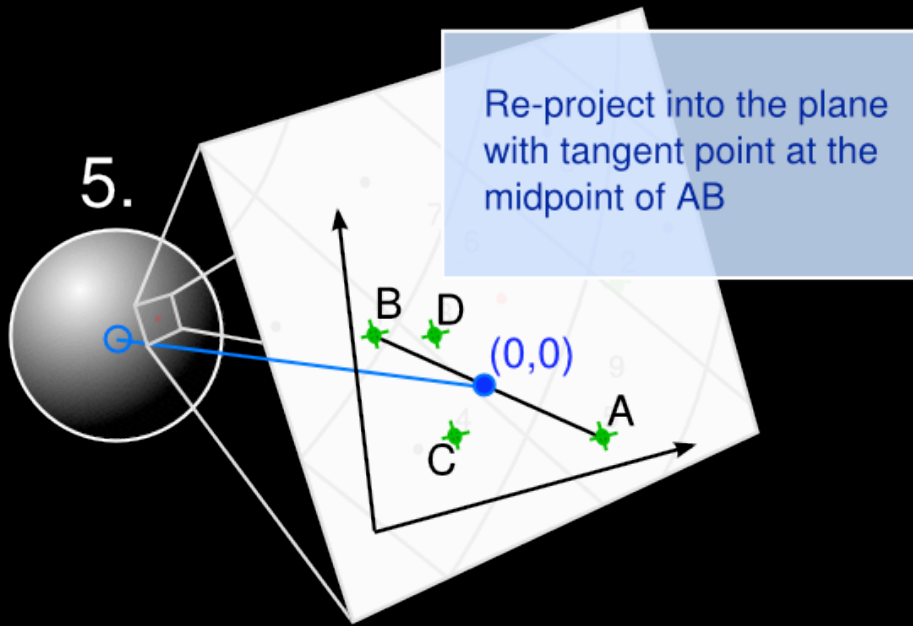
- use quads of stars to identify **hypotheses**
- test explanatory power of each hypothesis to **verify**
- typically try thousands of hypotheses per image
 - that's a lot, but a lot less than brute-force search
 - verify is fast

details: indexing

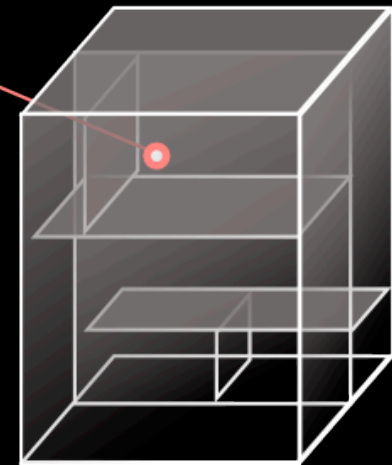
- build large indices of quads from USNO-B
 - each index specializes on one angular scale and one bandpass
- indices attempt not to over-use individual stars (robustness to missing & extra stars)
- indices attempt to have contiguous, overlapping coverage of sky (solve rate)
- quads are parameterized cleverly

details: quads

- choose 4 stars; use the largest-separation pair to define a coordinate system.
- quad “hash” is just the two two-d positions of the other two stars.
 - description is scale- and rotation-free
 - hashes uniformly fill four-d space
 - excellent continuity and error properties



CODE:
 (cx,cy,dx,dy)



8. 4D kdtree

details: fast lookup

- a typical index has nearly a billion quads
- build a balanced 4-d KD tree in quad space
- tree is stored on disk exactly as it is arranged in RAM
 - can be loaded in a mmap operation
 - no pointers to build
 - very fast response to queries

details: verify

- each quad match sets a candidate pointing, rotation, and scale for the image
- did we find, at that pointing, rotation, and scale, many more matches to *other* stars than we would expect by chance?
- verify is statistically exact if the input catalog has no strong clustering
 - *not* true; working to make this true

plans

- we are currently alpha (invitation only)
- go beta this winter
- *all code is open source* (vanilla c)
- implement precision tools
 - camera calibration (what instrument teams do)
 - deepen and correct standards catalogs
 - overlapping images
 - FITS WCS implementation issues

evil plans

- aggregate and make interoperable all amateur data
- same for *plate archives* and historical data
- coordinate activities of amateurs
 - time resolution all-sky
 - interesting objects
- build sky maps and new catalogs
- photometric calibration and mosaicing

the end

- <http://astrometry.net/>
- Lang *et al*, *Science*, submitted
- suggested questions:
 - *what do you do with the input data?*
 - *aren't you just de-skilling young astronomers?*
 - *why not triples or quintuples?*