Astrometry.net:

Blind astrometric calibration of arbitrary astronomical images

Dustin Lang\textsuperscript{1,2,7}, David W. Hogg\textsuperscript{3,4}, Keir Mierle\textsuperscript{1,5}, Michael Blanton\textsuperscript{3}, and Sam Roweis\textsuperscript{1,5,6}

AJ \textbf{139}, 1782L (2010) - \url{http://dx.doi.org/10.1088/0004-6256/139/5/1782}

\textsuperscript{1} Department of Computer Science, University of Toronto
\textsuperscript{2} Princeton University Observatory
\textsuperscript{3} Center for Cosmology & Particle Physics, Department of Physics, New York University
\textsuperscript{4} Max-Planck-Institut f"ur Astronomie
\textsuperscript{5} Google Inc.
\textsuperscript{6} Computer Science Department, New York University

Presentation by Paulo Penteado (pp.penteado@gmail.com)
\url{http://www.ppenteado.net/ast/pp_astrometry_201110.pdf}
Outline

• The problem: Lost in space

• The solution: Astrometry.net
  ➔ Diamonds in the sky
  ➔ Hashing the quads
  ➔ Indexing the sky
  ➔ Picking the diamonds

• Tests

• Using Astrometry.net

http://www.astrometry.net
AJ 139, 1782L (2010) - http://dx.doi.org/10.1088/0004-6256/139/5/1782

http://www.ppenteado.net/ast/pp_astrometry_201110.pdf
The problem: Lost in space

Which way is my spaceship going?
The problem: Lost in space

• Which way is my spaceship going?
• What is in the image?
• Are my targets in the image?
• Who are these stars?
• Where am I looking at?
The problem: Lost in space

Can I not just look at the FITS header?

SIMPLE = T / file does conform to FITS standard
BITPIX = 16 / number of bits per data pixel
NAXIS = 2 / number of data axes
NAXIS1 = 3072 / length of data axis 1
NAXIS2 = 2048 / length of data axis 2
EXTEND = T / FITS dataset may contain extensions
COMMENT FITS (Flexible Image Transport System) format is defined in 'Astronomy and Astrophysics', volume 376, page 359; bibcode: 2001A&A...376..359H
BZERO = 0 / offset data range to that of unsigned short
BScale = 1 / default scaling factor
INSTRUME = 'SBIG STL' / Image CCD
EXPTIME = 20. / exposure time (seconds)
DATE = '02:54:30 2011-03-17'
FILTER = ' (5) CLEAR ' / Filter
IMAGETYP = 'Light Frame'
O_BZERO = 32768 / Original BZERO Value
END
The problem: Lost in space

Can I not just look at the FITS header?

```
SIMPLE = T / file does conform to FITS standard
BITPIX = 16 / number of bits per data pixel
NAXIS = 2 / number of data axes
NAXIS1 = 3072 / length of data axis 1
NAXIS2 = 2048 / length of data axis 2
EXTEND = T / FITS dataset may contain extensions
COMMENT FITS (Flexible Image Transport System) format is defined in 'Astronomy and Astrophysics', volume 376, page 359; bibcode: 2001A&A...376..359H
BZERO = 0 / offset data range to that of unsigned short
BSCALE = 1 / default scaling factor
INSTRUME= 'SBIG STL' / Image CCD
EXPTIME = 20. / exposure time (seconds)
DATE = '02:54:30 2011-03-17'
FILTER = ' (5) CLEAR ' / Filter
IMAGETYP= 'Light Frame'
O_BZERO = 32768 / Original BZERO Value
END
```

Often, not enough:
- No data
- No field orientation
- Low precision
- Unreliable data
The problem: Lost in space

Does not a chart solve everything?
The problem: Lost in space

Does not a chart solve everything?
The problem: Lost in space

Does not a chart solve everything?
The problem: Lost in space

Does not a chart solve everything?
The problem: Lost in space

Does not a chart solve everything?

Often, no:

- **Need to have a first guess** (field center, size and orientation).
- **Interactive: work-intensive and time-wasting.**
- Not feasible for many images.
- Unreliable.
- Does not work for crowded fields.
The problem: Lost in space

When does the problem arise?

- Making the observation (are my targets in the image?)

- Processing the data:
  - what are the coordinates of each source in the image?
  - what are the coordinates of each pixel in the image?

- Building databases (what is contained in each image?)

- Navigating (where is my ship going?)
The solution: Astrometry.net

**Blind** astrometric calibration of arbitrary astronomical images

- Uses only the pixels
- No need for any (possibly absent or unreliable) metadata: not even the scale of the image.

Given any image, it automatically and efficiently

- Finds the brightest sources.
- Matches the stars against an all-sky catalog.
- Calculates the pointing solution and puts it into the standard metadata (WCS in the FITS header).

“to help organize, annotate, and make searchable all the world’s astronomical information.”
How does Astrometry.net work?

The easy part:

Locate the sources ("stars") in the image.

Old, well-studied problem in astronomy.
How does Astrometry.net work?

1) The easy part:

Locate the sources ("stars") in the image.

Old, well-studied problem in astronomy.

Finds by SExtractor (http://sextractor.sourceforge.net/) or its own algorithm:
• Flatten image with a median-filtered subtraction
• Measure the noise level $\sigma$
• Find the objects (above $8\sigma$)
• Clean the object list (looking at neighbors)
• Find the centroids of the objects (Gaussian fit)
Diamonds in the sky

The hard part:

Where in the sky is there a field that matches the image?

**The only thing known from the image is the coordinates of the stars.**

Parameter space is 4D:
- Field center (RA and Dec)
- Field scale
- Field orientation (which way is N and E?)

Fields typically less than $10^{-6}$ of the sky:
- Not even the scale or the orientation is known.
- Cannot brute-force look over the whole sky in 4D.
Diamonds in the sky

2) Make a quad (“diamond”) with 4 bright stars in the image

Now, search the whole sky for the same quad.
Hashing the quads

How to match *something* against a huge number of entries?

Not by brute-force (comparing one by one):
  • Too many comparisons
  • Too slow for those that happen to be at the end.
Hashing the quads

How to match *something* against a huge number of entries?

Not by brute-force (comparing one by one):
- Too many comparisons
- Too slow for those that happen to be at the end.

Use a map (dictionary) algorithm:

- You do not lookup *helicopter* in a dictionary by starting at the first word on “A”, going word by word.
- You go directly to the page with *hel*.
- Only a few words are compared with *helicopter*.
- Searches are quick and take (nearly) constant time.

An algorithm turns *something* into a code (*hash*)
- Hashes can be indexed then searched easily.
Hashing the quads

- Most separated stars \((A, B)\) define coordinate system.

- \(C\) and \(D\) are measured in this coordinate system.

- The 4-element vector \((X_C, Y_C, X_D, Y_D)\) is the hash code.
Hashing the quads

Hash properties:

- Invariant under scaling and rotation.
- Varies smoothly with the position of the stars → (small changes in the coordinates make small changes in the code).
- Uniform distribution.
Indexing the sky

Now that we have the word (hash), we need the dictionary (index, hashtable).

Take catalogs (USNO-B1 + 2MASS), calculate hashes over the whole sky, and store in an index.
Picking the diamonds

Cannot expect to match quad exactly: positions have errors.

Look over the neighborhood of a quad, finding all similar quads (quick and easy).

How to decide which similar quad is the real match?

Each possible match is a hypothesis that can predict something testable: where else in the image other stars are expected to be?
Picking the diamonds

Use a Bayesian decision between two models:

• Foreground model (\(F\)): uniform probability of finding a star in the image + blob of probability around each catalog star (the match is the right solution).

• Background model (\(B\)): uniform probability only (the match is wrong).

Accept hypothesis if the Bayes factor

\[
K = \frac{p(D|F)}{p(D|B)}
\]

is larger than a threshold \((10^9)\).
Picking the diamonds

The threshold for $K$ is determined from the utility functions

And the prior probability of a random pick being right, set to $p(F)/p(B)=10^{-6}$.

The first acceptable match is used as solution. If none is accepted, another quad in the image is tried.
Tests – SDSS sources and images

- Index built to match SDSS scale (for efficiency).
- r band Images classified as excellent in SDSS.

| Phase       | Images Recognized | Unrecognized | Percent Recognized |
|-------------|-------------------|--------------|-------------------|
| USNO-B: 0.1 s | 172,882           | 9,339        | 94.87             |
| USNO-B: 1 s  | 181,826           | 395          | 99.78             |
| USNO-B: 10 s | 182,158           | 63           | 99.97             |
| USNO-B: 15 s | 182,160           | 61           | 99.97             |
| 2MASS       | 182,211           | 10           | 99.99             |
| Original images | 182,221         | 0            | 100.00            |
Tests – SDSS sources and images

All SDSS “excellent” images identified unequivocally from few stars tried.
Tests – SDSS sources and images

Still works well with worse r-band images:

| CPU time (per image) | Percentage of Images Recognized |
|----------------------|--------------------------------|
|                      | Excellent | Good   | Acceptable | Bad    |
| 0.1 s                | 94.87     | 94.85  | 94.57      | 84.11  |
| 1 s                  | 99.78     | 99.74  | 99.64      | 96.58  |
| 10 s                 | 99.97     | 99.94  | 99.94      | 99.11  |
| 60 s                 | 99.97     | 99.94  | 99.95      | 99.18  |
## Tests – SDSS and GALEX

### SDSS bands

| CPU time (per image) | Percentage of Images Recognized | \( u \) | \( g \) | \( r \) | \( i \) | \( z \) |
|----------------------|---------------------------------|--------|--------|--------|--------|--------|
| 0.1 s                |                                 | 87.80  | 93.88  | 94.87  | 93.59  | 94.36  |
| 1 s                  |                                 | 98.58  | 99.73  | 99.78  | 99.73  | 99.75  |
| 10 s                 |                                 | 99.82  | 99.96  | 99.97  | 99.96  | 99.96  |
| 60 s                 |                                 | 99.84  | 99.96  | 99.97  | 99.96  | 99.96  |

### GALEX (NUV)

| CPU time (per image) | Percentage of Images Recognized |
|----------------------|---------------------------------|
| 1 s                  | 74.46                           |
| 10 s                 | 93.56                           |
| 100 s                | 98.95                           |
| 1000 s               | 99.74                           |
Tests – False positives

Results must be reliable:
- Thresholds set to avoid false positives, even if at the cost of more false negatives.
Tests – False positives

False positives only in images with serious problems:

Search image

Catalog match
Tests – False positives

False positives in images with serious problems:

Search image

Catalog match
Using Astrometry.net – local install

[user@computer]$ solve-field --use-sextractor image0025.fits
Reading input file 1 of 1: "image0025.fits"
Computing image percentiles...
Extracting sources...
> WARNING: default.sex not found, using internal defaults
----- SExtractor 2.8.6 started on 2011-10-11 at 11:45:23 with 12 threads
Measuring from: "Unnamed" / 3072 x 2048 / 0 bits FLOATING POINT data
(M+D) Background: 615.207 RMS: 16.4187 / Threshold: 24.628
Objects: detected 2963 / sextracted 2477
> All done (in 1 s)
Sorting brightness using MAG_AUTO and BACKGROUND columns failed; falling back to MAG_AUTO.
Solving...
Reading file "/image0025.axy"
Field 1 did not solve (index index-219.fits, field objects 1-10).
(...) Field 1 did not solve (index index-204-11.fits, field objects 21-30).
  log-odds ratio 155.889 (5.03287e+67), 30 match, 1 conflict, 56 distractors, 50 index.
  RA,Dec = (203.459, -65.9432), pixel scale 0.462407 arcsec/pix.
  Hit/miss: ++++++++-+-++-+-++-+---++---+--+++++-++---++-+c-+-------------------++
  +-------------+-----+(best)-------------
Field 1: solved with index index-204-10.fits.
Field 1 solved: writing to file ./image0025.solved to indicate this.
Field: image0025.fits
Field center: (RA,Dec) = (203.5, -65.94) deg.
Field center: (RA H:M:S, Dec D:M:S) = (13:33:50.006, -65:56:34.647).
Field size: 23.6449 x 15.7673 arcminutes
Creating new FITS file "/image0025.new"
Creating index object overlay plot...
Creating annotation plot...
Your field contains:
  NGC 5189
Using Astrometry.net – local install
Using Astrometry.net – local install
Using Astrometry.net – besides local install

Web service (http://nova.astrometry.net/)

Anonymous use makes images public. Privacy and more options through an OpenID (Google, Yahoo, etc.)
Using Astrometry.net – besides local install

Flickr: images uploaded to astrometry group
http://www.flickr.com/groups/astrometry

M81 and M82
The galaxy pair of M81 and M82 in this quick jab at the sky. 3x10m luminosity, 3x10min Mog each for a total 1h of exposure before the clouds rolled in. This image is starting to show what the gear
Summary

• Often it is necessary to obtain the celestial coordinates from images.

• Problem cannot be solved by brute-force.

• Astrometry.net solves the problem reliably and efficiently:
  → Candidate solutions quickly found from hashes of quads.
  → Solution selected by Bayesian decision using the rest of the field.
  → Identifies correctly every good quality image.
  → Runs automatically and quickly.
  → False positives are rare.
  → Local install, webservice and Flickr interfaces.

http://www.astrometry.net
AJ 139, 1782L (2010) - http://dx.doi.org/10.1088/0004-6256/139/5/1782

http://www.ppenteado.net/ast/pp_astrometry_201110.pdf