Photometric Surveys: Catalog generation with HEALPix and VO

Jiří Nádvorník

Astronomical Institute Academy of Sciences, Czech Republic

June 15, 2015

What do we do

- Light curve generation
- Special requirements
  - Incremental - flexible for any data addition
  - No fixed grid for image positions

VO usage

- Publishing:
    - GAVO DaCHS
    - SIA - Whole images
    - SCS - Photometric points
    - 'SSA' - Light curves
- Processing:
    - HEALPix
- Clients:
   - SPLAT-VO, Aladin, Topcat

Our data

- Data origins:
  - Remote control of 1.54 Danish Telescope in Chile
  - Different working groups, different targets, different topics
    - minor planets, variable stars, SMC
  - No fixed grid (typical light curve survey)
  - Differential astrometry and photometry **not usable**
  - 'Garbage' data recycling
- 100 thousand images, UBVRI filters
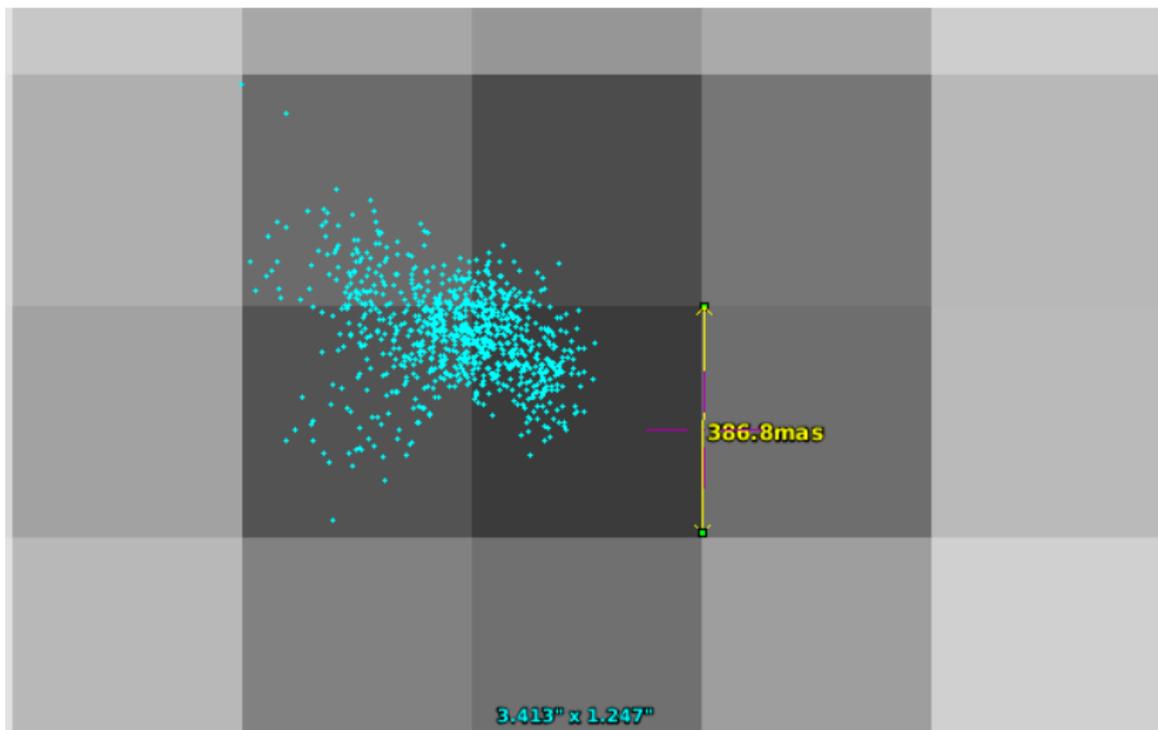- 400 million measurements (astrometry + photometry)
- 6 million objects

Data distribution
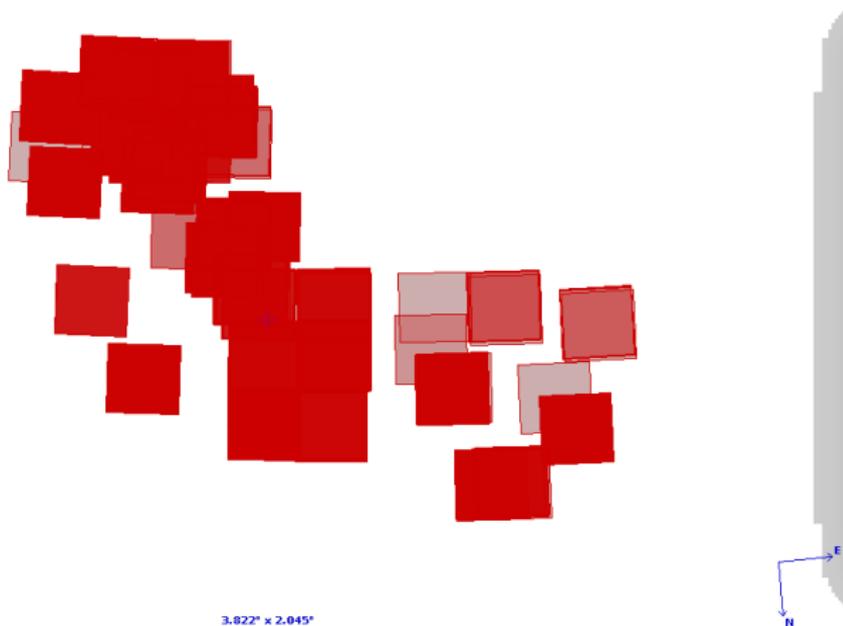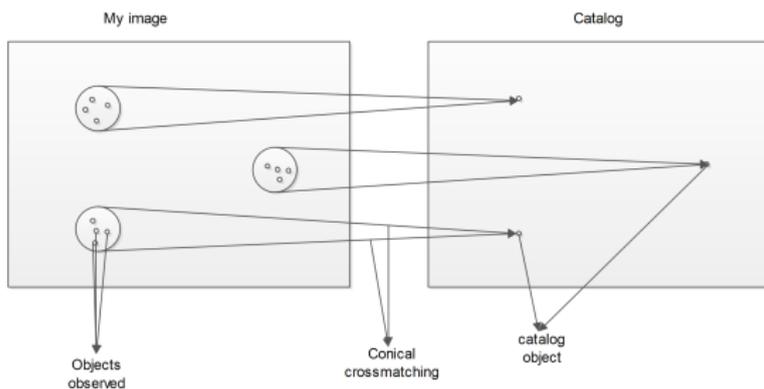
NGC330 13x13 arcmin

Cluster

On-line catalog

- Catalog cross-match with fixed radius
- Cons:
  - Catalog inaccuracies, duplicities
  - Can't produce light curves for objects not in catalog
  - Cannot produce transients
  - Best result with PPMXL: cca 70 % of data used (with errors)

Our own catalog

- Usage of 100 % data
- Independent of catalog faults
- Can identify transients
- Catalog generation process:
    1. Compute astrometry + photometry for each individual image (Munipack, UCAC4)
    2. Cluster these points - produce catalog of objects
    3. Generate light curves for whole survey (on the fly using SSA + datalink)

Implementation of catalog generation

- C++ application with HEALPix library
- Integration with GAVO DaCHS package for VO publication
- Clustering problem
  - NP-Hard in general (quantum computer needed)
  - Best K-means fails for thousands of clusters (vs. 6 million)
  - We know cluster sizes $\Rightarrow$ We can define overlaps
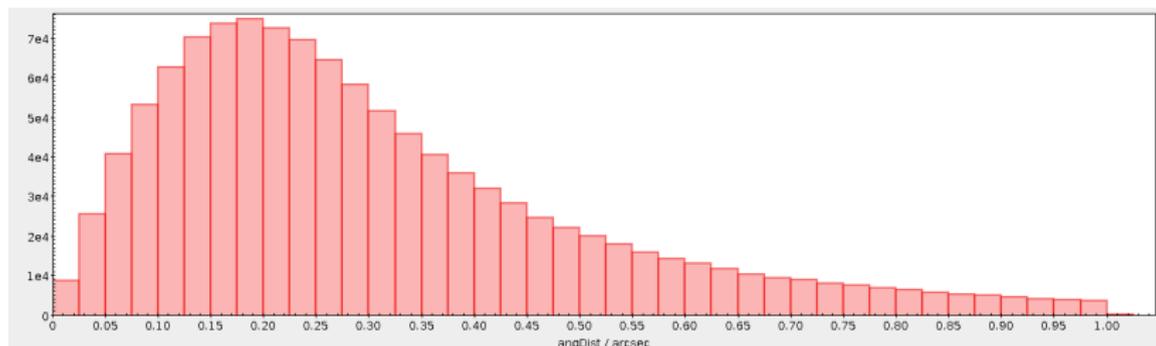 - HEALPix for parallelization

HEALPix usage

- We can specify overlaps
- Pixels cover the same area
- C++ library
- Fast
- Great support
- Great Aladin debugging

HEALPix overlaps
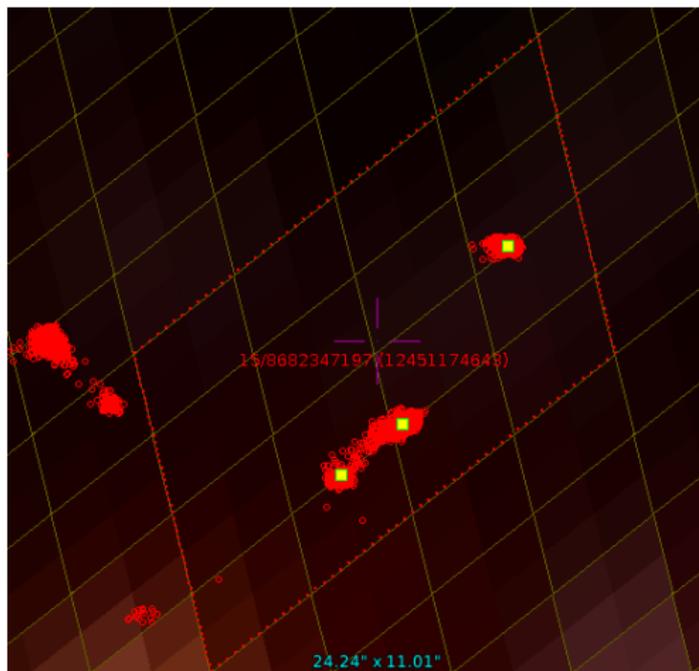
Results

- Cca 2 h per whole dataset for linear algorithm (12 cores, 32 GB RAM)
- K-means by order slower, but scales linearly
- Very good scalability
- Exceptional quality with linear algorithm
- Acceptable memory complexity cca 8 GB per 100 mil. observations

Demo

Achievements

- We can produce incremental light curves for any photometric survey
- Independent of catalog coverage
- HEALPix parallelization:
    - Clustering problem reduced from 6 mil. to less than 10 clusters
    - Can be reused on any kind of spherical coordinates (astronomy, geodesy)
- Light curves published with 'standard' VO protocols (SSA vs. TimeSeries?)

Acknowledgments

Petr Škoda
Markus Demleitner
Martin Reinecke
Filip Hroch
Supported by grant GACR 13-08195S

Links

- The C++ application
  https://gitlab.com/nadvornik-ji/AstroClustering
- Pipeline for astrometry and photometry
  http://munipack.physics.muni.cz/
- Publishing software http://soft.g-vo.org/
- HEALPix library for C++ application
  http://healpix.sourceforge.net/

Questions