Cross-matching Engine for Incremental Photometric Sky Survey

Student: Jiří Nádvorník
Supervisor: Petr Škoda

FACULTY OF INFORMATION TECHNOLOGY
CZECH TECHNICAL UNIVERSITY IN PRAGUE

Introduction

The goal is to produce **light curves** from **any photometrical survey**, no one has done it yet. A light curve is a graph displaying brightness of one physical object over a period of time.

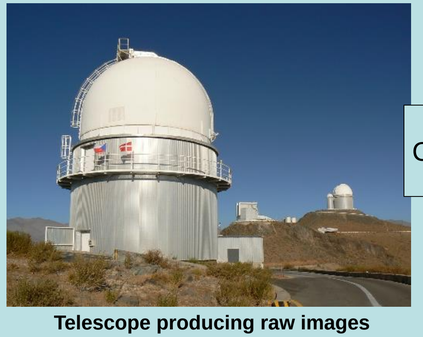
Standard solution:
- Relies on a predefined grid of images, we cannot
- Cannot identify overlapping objects, we can
- Cannot be parallelized

Most complex problem is **identifying** physical objects (**clustering** spherical coordinates).

Difficulties to overcome
- 400 million observations
- **Parallel** Clustering of **Big Data**
- Needs to be fast
- Needs to run on a **modest machine**

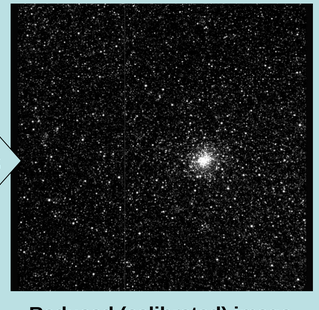
Data gathering

Remote control of Danish robotic DK154 telescope in Chile



Telescope producing raw images

Calibrating images



Reduced (calibrated) image

Collecting images

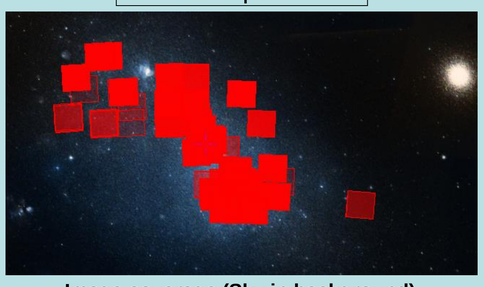


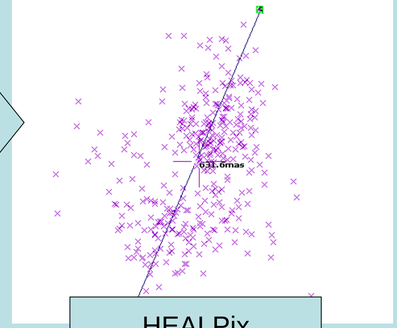
Image coverage (Sky in background)

Extracting observations with Munipack

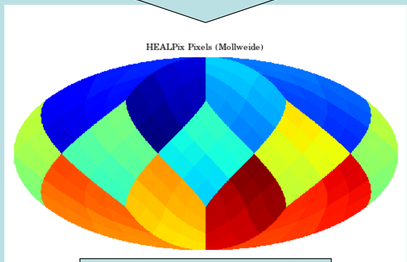
Observations			
Object	Right asc. [deg]	Declination [deg]	Brightness [mag]
Not known yet	13.15437936	-72.8014664181	19.1522
Not known yet	13.1543684677	-72.8014613066	19.2179
Not known yet	13.1542871262	-72.8014367811	19.3138
Not known yet	13.1543485507	-72.8014548044	19.1899
…	…	…	…

Identifying cluster sizes

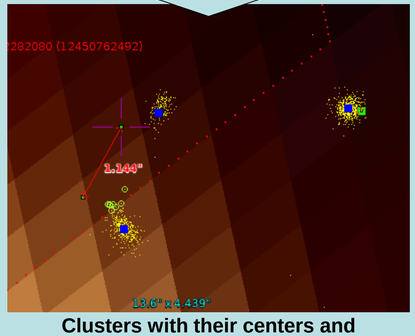
Observations of one object (1 cluster)



HEALPix tesselation



Parallelized K-means clustering



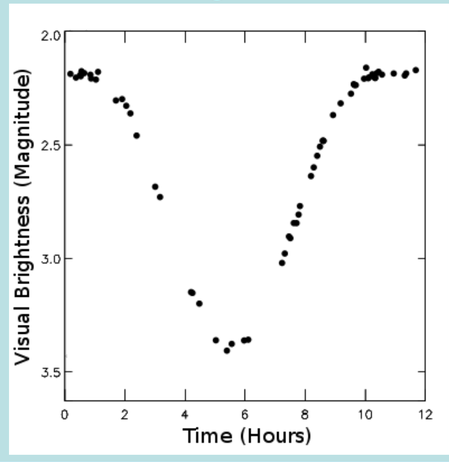
Clusters with their centers and underlying HEALPix pixels

Publication

Results

- Clustering **400 mil.** observations **under 2h** (10 cores, 32GB RAM)
- Consuming ~8GB RAM per 100 mil. observations
- Not dependent on number of clusters
- **High precision** compared to other strategies used nowadays

Final light curve



Opportunities

- Can process **any kind** of photometrical survey
- **Universal parallelization** scheme can be used for processing any kind of spherical data (**astronomy**, **geodesy**)
- May be used for CRTS project on Caltech to identify overlapping objects