Robust Registration of Astronomy Catalogs

Johns Hopkins Whiting School of Engineering

Fan Tian, Tamás Budavári, Amitabh Basu

Johns Hopkins, Department of Applied Mathematics and Statistics

Abstract

Due to a small number of reference sources, the astrometric calibration of images with a small field of view is often inferior to the internal accuracy. An important experiment with such challenges is the Hubble Space Telescope (HST). A possible solution is to cross-calibrate overlapping fields instead of just relying on standard stars. Following Budavári and Lubow (2012), we use infinitesimal 3D rotations for fine-tuning the calibration but re-formalize the objective to be robust to large number of false candidates in the initial set of associations. Using Bayesian statistics, we accommodate bad data by explicitly modeling the quality which yields a formalism essentially identical to M-estimation in robust statistics. Our preliminary results show great potentials for these methods on simulated catalogs where the ground truth is known.

Simulation

MOCK UNIVERSE As shown in fig. 1, we demonstrate the problem on simulations to images taken by HST Advanced Camera for Surveys (ACS) on Wide Field Channel.

CATALOGS By applying a small random perturbation to sources in the mock universe simulating astrometric uncertainty, and taking 3D rotations, we obtain two transformed catalogs, as shown in fig. 2. A selection interval has also been applied to source properties when generating catalogs.

Simulation parameters:

- Image size: 202" × 202"
- No. of sources: 1500
- Source directions: unit vectors in radians.
- Source properties: random Gaussian(0,1)
- Astrometry uncertainty: 0.05"

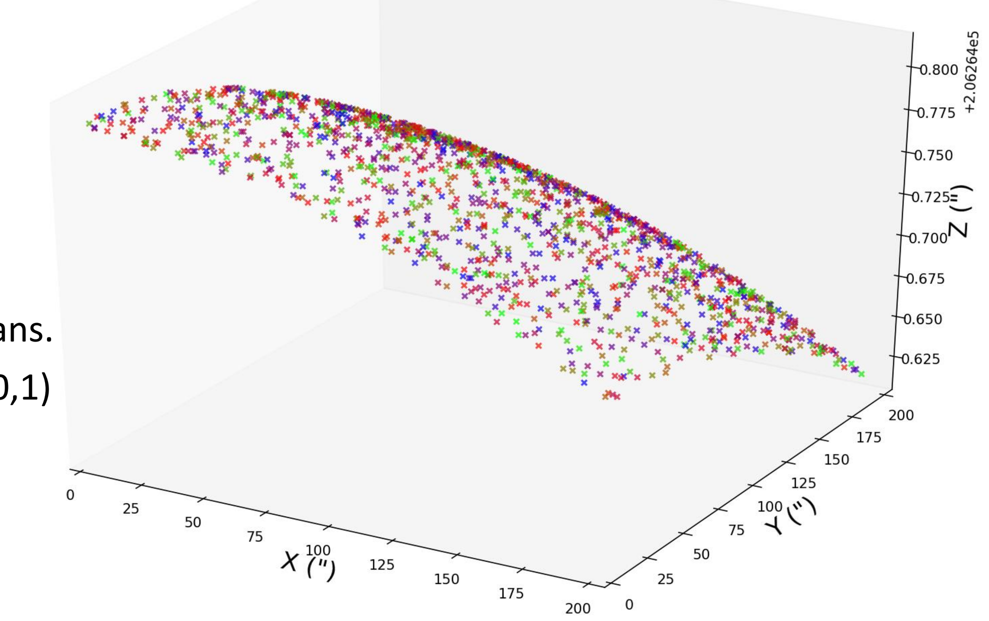

Figure 1: 3D mock universe. Colored by source properties.

PARTITION MODEL Following the partition model in Budavári & Loredo (2015) (also see Budavári & Basu, 2015), sources from two catalogs within a certain distance threshold are clustered together under the hypothesis that they potentially correspond to the same underlying object. For two catalogs scenario, there are two cases:

- Only one true matching pair
- No true matching exists in the cluster

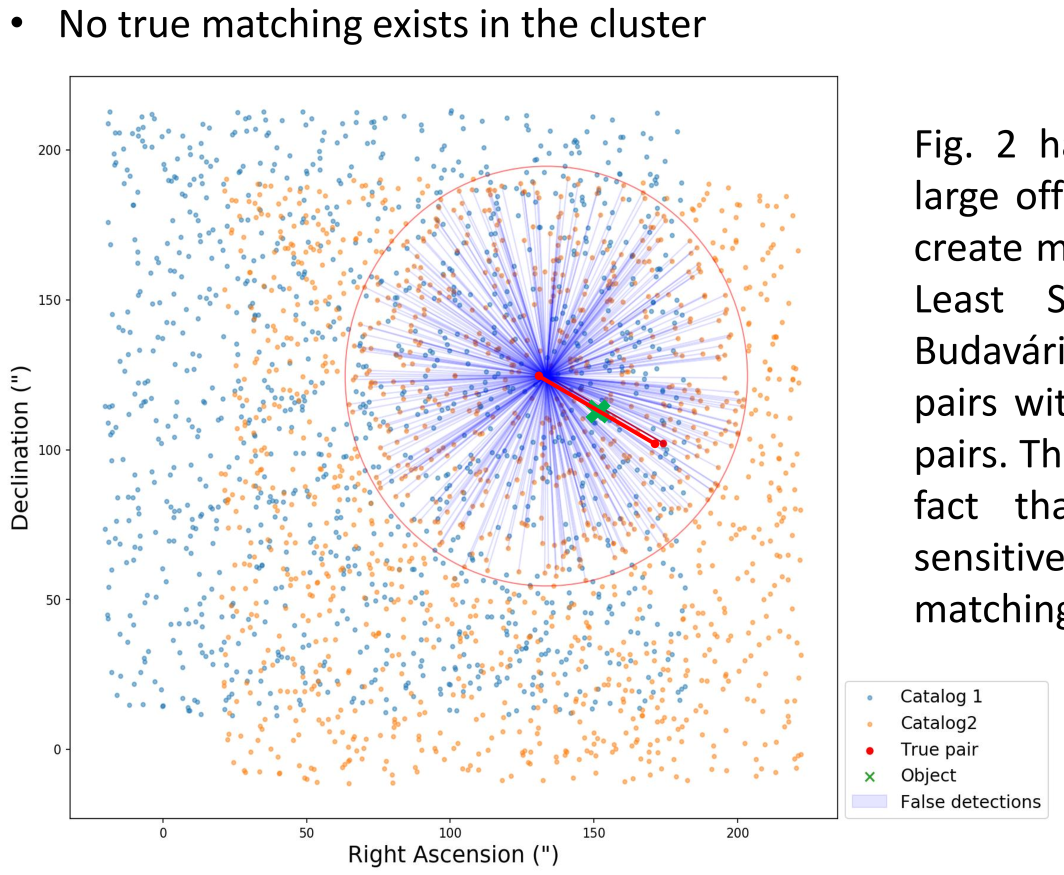

Fig. 2 has shown that, for images with large offset, a larger search radius would create many bad matching pairs. Previous Least Squares method described in Budavári & Lubow (2012) considers all pairs within the search radius being true pairs. This is therefore problematic for the fact that Least Square estimation is sensitive to outliers, i.e. the bad matchings.

Figure 2: Illustration of a true pair and all false detections within a search radius of 70". Maximum offset ~ 50".

Robust Estimation

BAYESIAN FORMALISATION Given a dataset D of sources $\boldsymbol{r}$ and calibrators $\boldsymbol{c}$, we derive the 3D infinitesimal transformation vector $\boldsymbol{\omega}$ from the posterior pdf

$$p(\boldsymbol{\omega}|D) \propto p(\boldsymbol{\omega}) \int d\gamma\, p(\gamma) \prod_q \left[\gamma \ell_q^{G}(\boldsymbol{\omega}) + (1-\gamma)\ell_q^{B}(\boldsymbol{\omega})\right]$$

The likelihood functions for a 'good' or a 'bad' pair candidate are given by Fisher distribution of $\ell_q^{G}(\boldsymbol{\omega}) = F(\boldsymbol{c}_q; \boldsymbol{r}'_q(\boldsymbol{\omega}), \kappa_q)$ and Uniform distribution of $\ell_q^{B}(\boldsymbol{\omega}) = \frac{1}{4\pi}$ respectively.

OPTIMIZATION The optimization for $\boldsymbol{\omega}$ is the maximization of the likelihood function

$$\max_{\boldsymbol{\omega}} \prod_q \left[\frac{2\gamma_*}{\sigma_q^2} \exp\left\{ -\frac{\left[\boldsymbol{c}_q - (\boldsymbol{r}_q + \boldsymbol{\omega}\times\boldsymbol{r}_q)\right]^2}{2\sigma_q^2} \right\} + (1-\gamma_*) \right]$$

Let $\Delta_q = \boldsymbol{c}_q - \boldsymbol{r}_q$, take $-\log$ of the likelihood function, we obtain the robust $\rho(\boldsymbol{\omega};\sigma)$ function.

At minimum $\tilde{\boldsymbol{\omega}}$, the vanishing gradient yields the solution

$$A\tilde{\boldsymbol{\omega}} = b$$

for:

$$A = \sum_q \frac{w_q}{\sigma_q^2}(I - \boldsymbol{r}_q \otimes \boldsymbol{r}_q);$$

$$b = \sum_q \frac{w_q}{\sigma_q^2}(\boldsymbol{r}_q \times \boldsymbol{c}_q)$$

$$w_q = W\left(\sqrt{\frac{(\Delta_q - \boldsymbol{\omega}\times\boldsymbol{r}_q)^2}{\sigma_q^2}}\right)$$

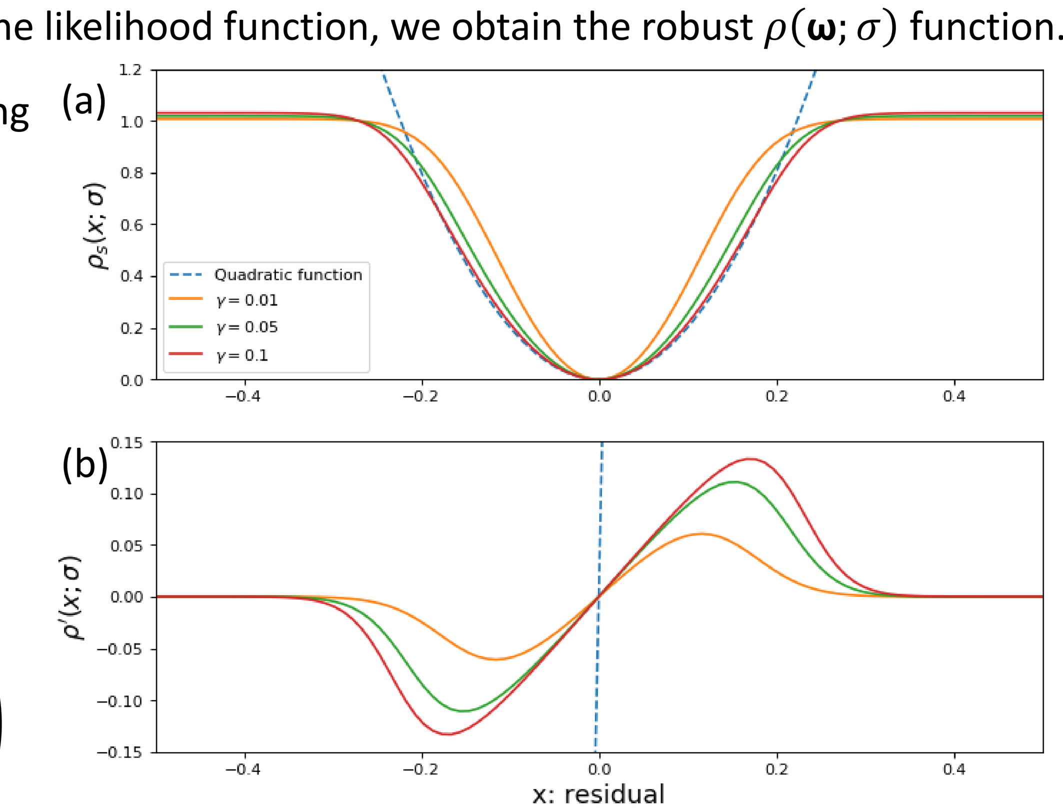

Figure 3: Plot for different γ parameters: (a) scaled ρ function to [0,1]; (b) influence function; with comparison to a quadratic function.

Implementation

The robust registration algorithm is implemented in Python for the following pipeline:

Input source lists → Partition dataset, for search radius R → Obtain estimator $\tilde{\boldsymbol{\omega}}$ with iterative optimization method → Correct with $\boldsymbol{r}' = \boldsymbol{r} + \tilde{\boldsymbol{\omega}} \times \boldsymbol{r}$

Initialize $\{w_q\}_0 \equiv \mathbf{1}$; $\sigma_0 \approx \frac{1}{3}R$; stopping tolerance $\boldsymbol{\tau}_{\text{stop}}$

1. Set $t = 0$
2. **while** $\frac{\|\nabla F(\omega_t)\|}{\|\nabla F(\omega_0)\|} > \tau_{\text{stop}}$ **do**
3. Compute $\tilde{\boldsymbol{\omega}}_t$ by solving $A\boldsymbol{\omega} = b$
4. Update on weights by $\{w_q\}_t = \mathrm{W}(\tilde{\boldsymbol{\omega}}_t)$
5. Update on $\sigma_{t+1} = \alpha * \sigma_t$ for $\alpha < 1$
6. **if** $\sigma_f < \sigma$, **then**
7. $\sigma_f = \sigma$
8. **end if**
9. **end while**
10. **return** approximate solution $\tilde{\boldsymbol{\omega}}_t$

METHOD COMPARISON To draw a valid comparison of the proposed method, we tested both the Least Squares and Robust M-estimation methods as follows:

(1) Two catalogs with a small offset, with increasing search radius

(2) A set of simulations with increasing offset, with a fixed large search radius.

As the ground truth is known, we are able to directly evaluate the accuracy after correction. Here we take the average separation between true pairs as the measurement of image offset. A successful correction would align two catalogs to the calibrator direction resulting in a small offset approximately σ after correction.

Results and Discussion

IMPROVED ASTROMETRY Fig. 4 shows correction results for example catalogs in Fig2.

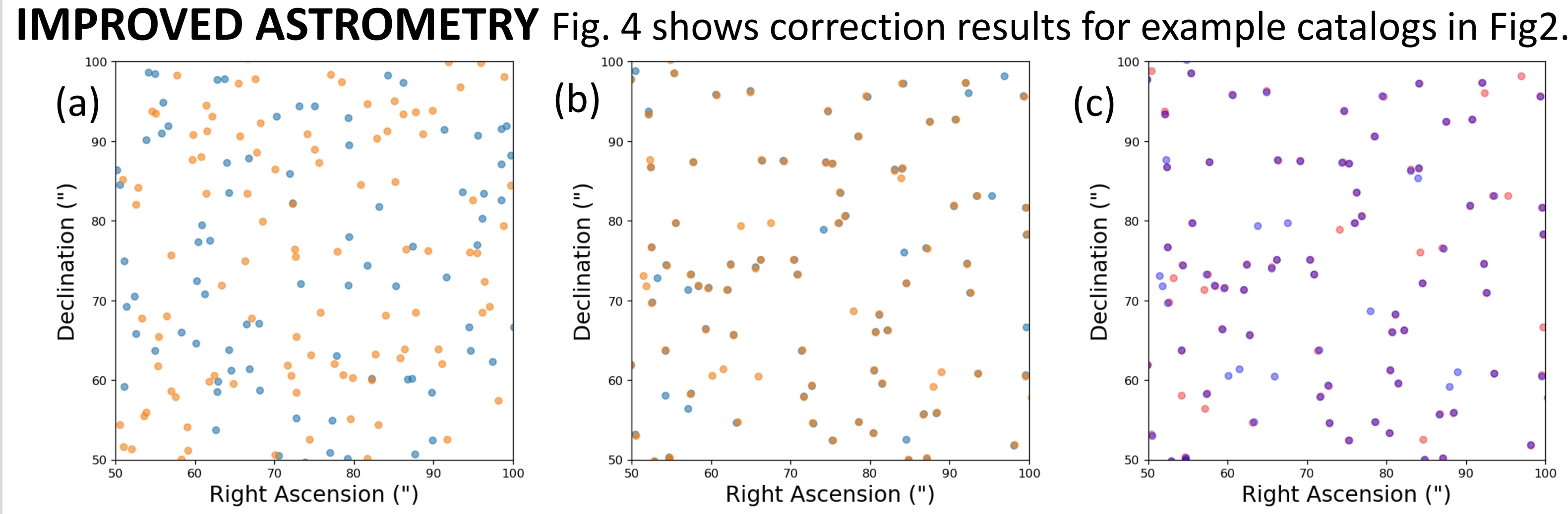

Figure 4: (a) Shifted catalogs; (b) Corrected catalogs with robust estimation $\tilde{\boldsymbol{\omega}}$; (c) Ground truth direction

COMPARISON RESULTS Fig. 5 draws comparison for Least Squares estimation results with Robust M-estimation.

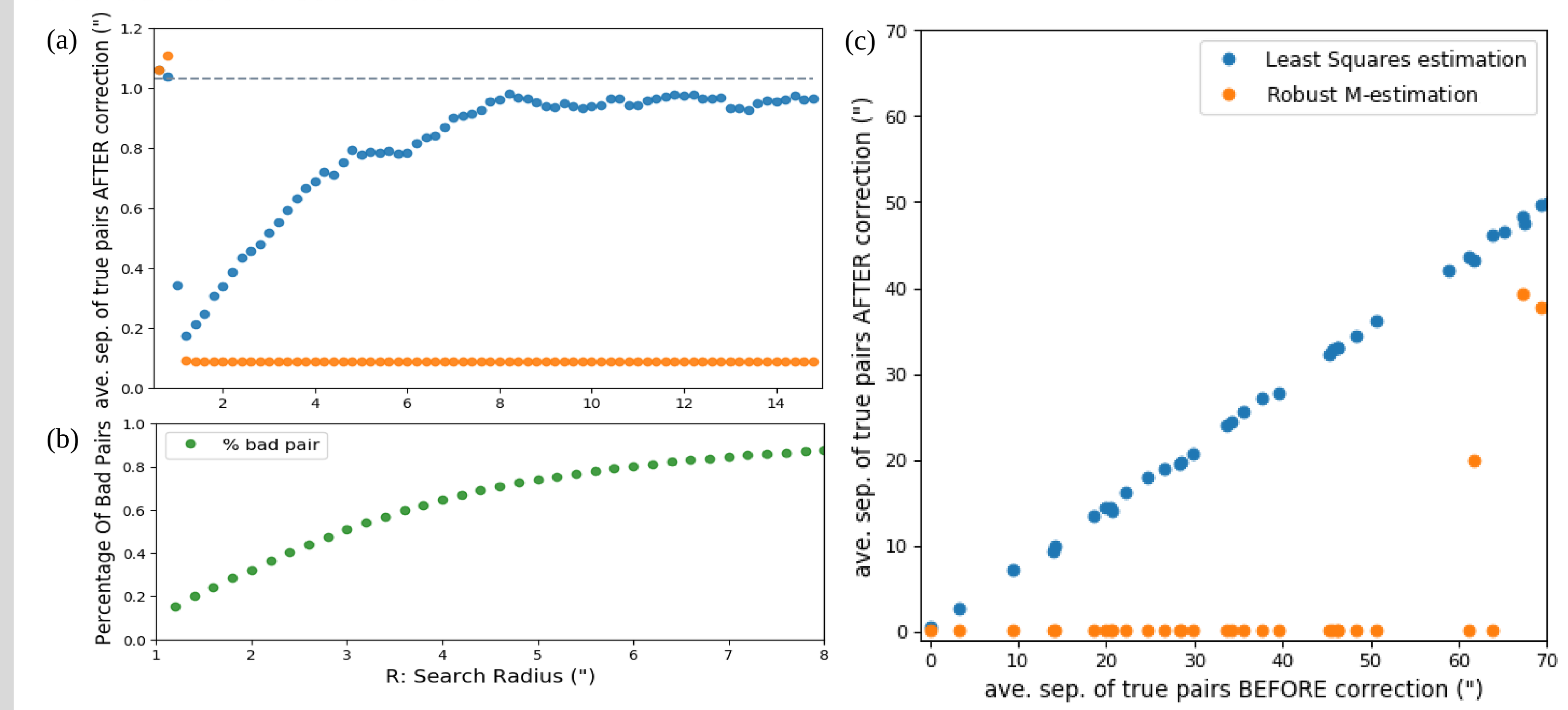

Figure 5: (a) (b) Least squares/Robust-M comparison for small offset with increasing search radius; (c) Least squares/Robust-M comparison for increasing offsets with large search radius.

LIMITATIONS AND FUTURE WORK One limitation is that, current simulation tests have been successful for up to 1 arcminute offset. But Hubble Source Catalog (HSC) has found cases with offsets up to 100 arcseconds. We are working on addressing these scenarios too. Moreover, while testing our implementation on different values of the uncertainty parameter σ, we have encountered certain numerical issues. The recovery is more likely to fail for small σ values due to the exponential term in the objective. We have succeeded on this by artificially assign large σ during iterations and converge to the desired value at later steps. We are currently working on automating the algorithm with this implementation.

We have tested the algorithm on a set of HSC and Gaia DR1 & DR2 images. The results have shown a successful correction on both HSC-Gaia DR1 and HS-Gaia DR2 regardless the very different geometry of those images. In future, we will extend the use to HSC-Gaia DR2 with large offsets .

References

[1] Budavári, T. & Basu, A 2015. Probabilistic Cross-Identification in Crowded Fields as an Assignment Problem. *The Astronomical Journal*, 152, 4, 86-90.

[2] Budavári, T. & Loredo, T. 2015, Probabilistic Record Linkage in Astronomy: Directional Cross-Identification and Beyond. *The Annual Review of Statistics and Its Application*, 2, 113-39.

[3] Budavári, T. & Lubow, S. H. 2012. Catalog Matching with Astrometric Correction And Its Application to the Hubble Legacy Archive. *The Astrophysical Journal*, 761, 2, 188-198.

Contact

Fan Tian

Email: ftian4@jhu.edu

Johns Hopkins University, Baltimore, MD, 21218