Transient Discovery with PTFIDE: Image Differencing & Extraction

Frank Masci & the iPTF Collaboration

iPTF Summer School, August 2015

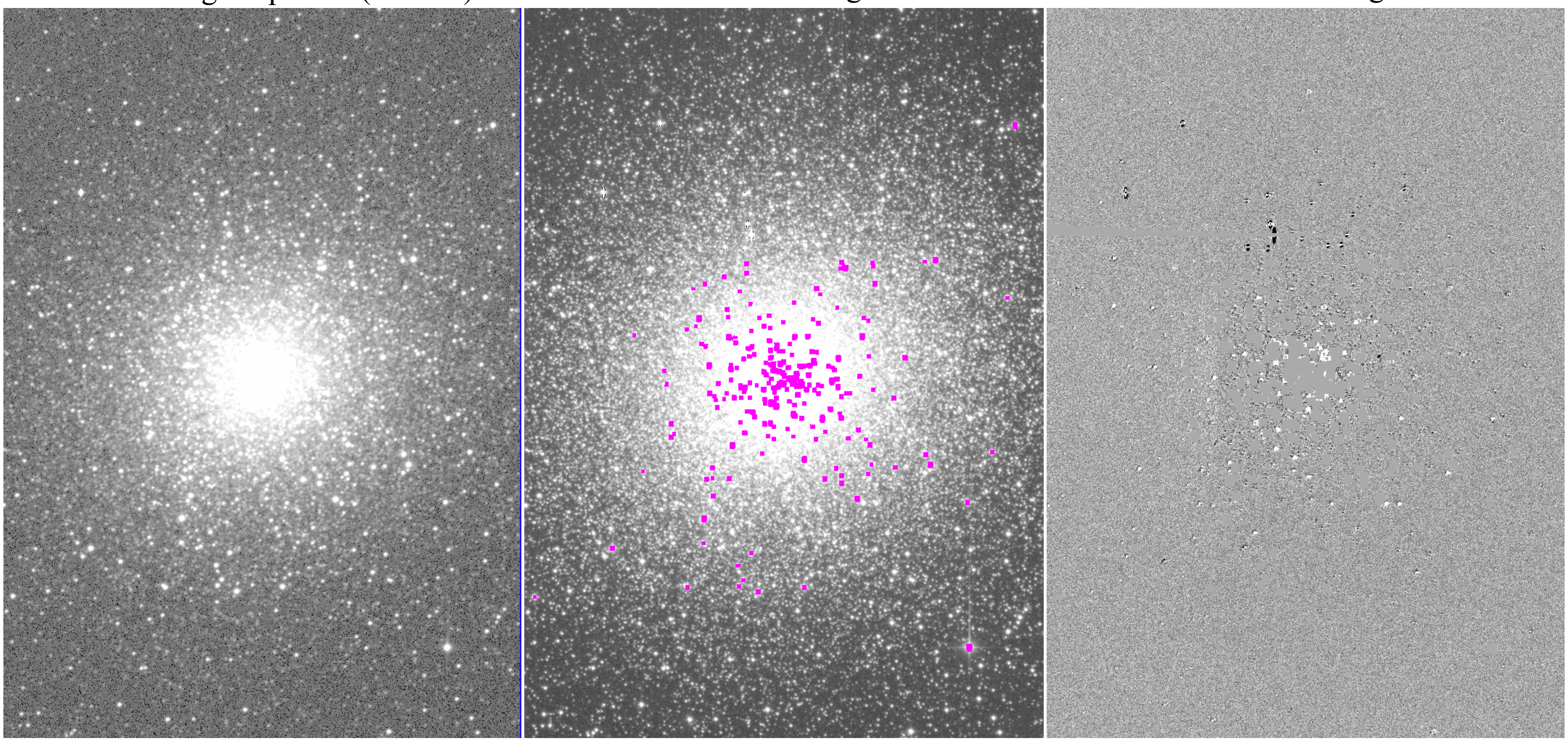
Definitions & Introduction

- **PTFIDE:** PTF Image Differencing and Extraction software for iPTF, ZTF (and the future…)
- **IDE:** IPAC Discovery Engine (IPAC = Infrared Processing & Analysis Center)
- **Image differencing**: discover transients by suppressing everything that's static in space and time
- **Transients:** any object that has varied in flux, or "suddenly" appeared or disappeared, or has moved (e.g., asteroid) in a new image exposure relative to some historical image (benchmark image)
- New exposure image: **"science image"**
- Historical (benchmark) image: **"reference image"**
  - a stack (co-add) of several or more high-quality historical science exposures
  - has higher signal-to-noise ratio than a single exposure image

E.g., M13 globular cluster

- **Enormous benefit:** image-differencing suppresses regions with high-source confusion, improving ability to discover flux variables and transients: the needles in the hay!

Science image exposure (R-filter) – Reference image = Difference image

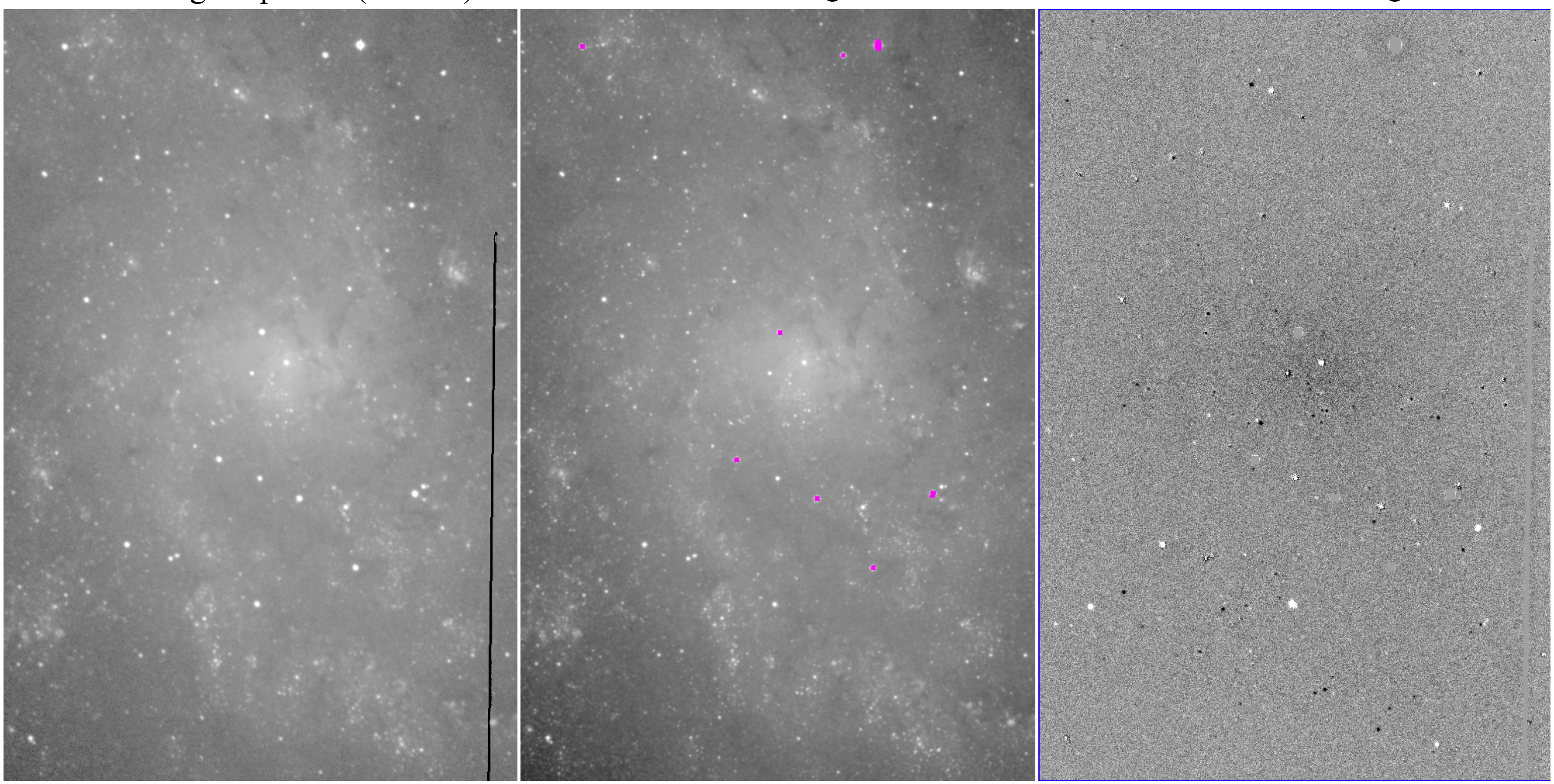


- Bad / saturated pixel regions: colored magenta (zeroed in difference)

E.g., M33 galaxy core

- **Another benefit:** image-differencing suppresses regions with complex backgrounds and emission: enhancing discovery potential, but also photometric accuracy of transient / variable candidates

Science image exposure (R-filter) – Reference image = Difference image



?

PTFIDE processing flow

- Quality-filtered single CCD-image exposure (*sci*)
- Reference (*ref*) image with source catalog
- Processing parameters

ref/sci source matching for throughput (gain) matching of *sci* to *ref* (need < 1% accuracy)

Relative astrometric refinement of *sci* to *ref*: orthogonal shifts only (need <~ 0.2″ accuracy)

Reproject, resample & "warp" reference image onto *sci* image frame; update masks

Match smoothly varying spatially-dep backgrounds

Make point-source cutouts/ coadds from *sci* & *ref* for matching-kernel derivation

Derive PSF-matching convolution kernels to match seeing/resolutions (spatially dependent)

Apply PSF-matching kernels and compute two image differences: *sci* – *ref* & *ref* – *sci* with masks and uncerts

Quality assurance metrics for difference images; good/bad decision before proceeding

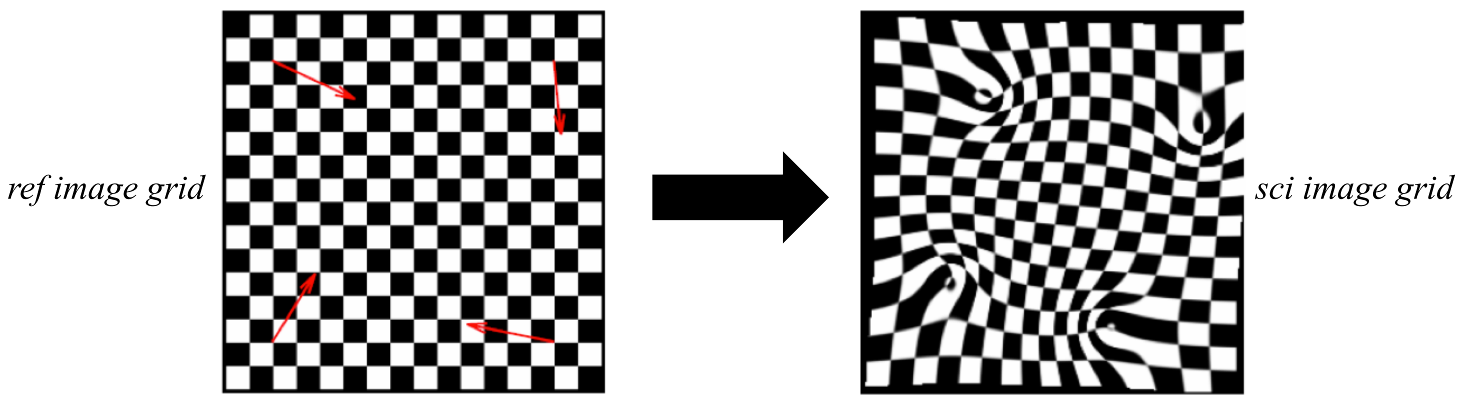
Estimate spatially varying PSF, extract transient candidates with PSF-fitting & aperture photom;
→ catalog + metadata for both difference images

Loose filtering of transients using PSF-fitting metrics

Compute source metrics and "features" to support machine-learned vetting

Reference image to science frame reprojection

- Archival reference image is “warped” (and resampled) onto science image grid using science image distortion polynomial coefficients
- Camera (field-of-view) distortion is calibrated upstream as part of astrometric calibration



- Astrometric / distortion calibration of input science image is crucial!
- If wrong (even slightly), astrometry of reprojected reference image will also be wrong and residuals will result in difference image (more later)

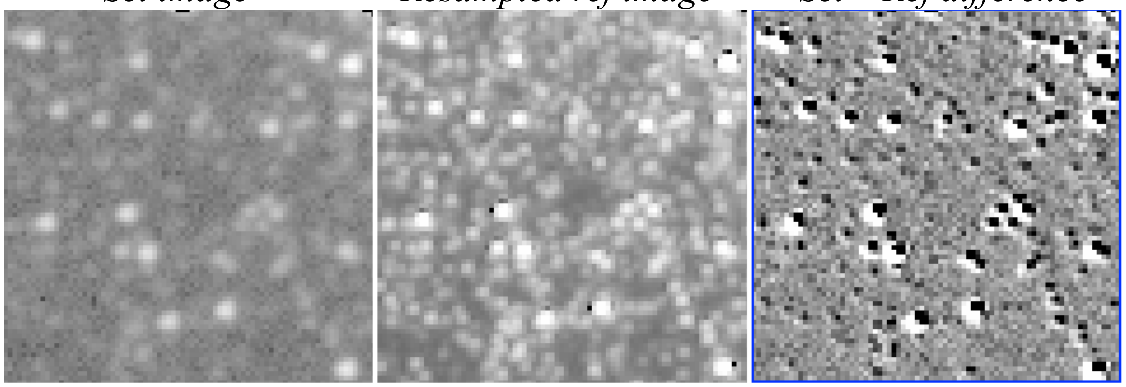
Example difference after *ref* to *sci* reprojection

PTFIDE also refines the astrometry of input science image (relative to the reference image)

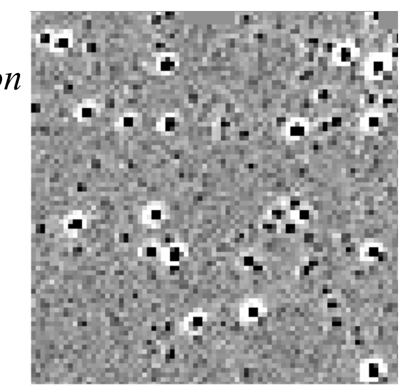
- only applies constant global Δx, Δy corrections to *sci* image pointing before *ref* reprojection
- computed using point-source matching between science and reference image

Sci image *Resampled ref image* *Sci – Ref difference*

zoom on M13
cluster field



with Δx, Δy correction

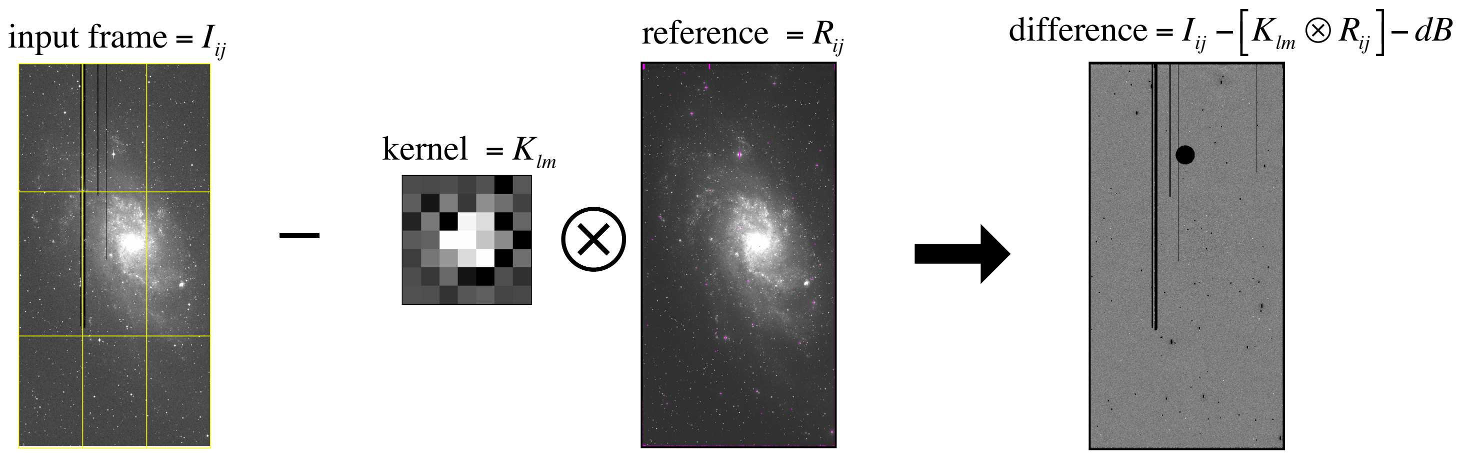


Generic PSF-matching model

- **Goal:** match the seeing profiles in the (resampled) *ref* and (gain-matched) *sci* images
- We assume the science image I can be modeled from a higher S/N, better "seeing" reference image R, a PSF-matching convolution kernel K, differential background dB, and noise term:

$$I_{ij} = \left[K(u,v) \otimes R_{ij} \right] + dB + \varepsilon_{ij}$$

- **Unknowns:** PSF-matching kernel $K(u,v)$ to convolve with better seeing image, and dB



(single chip with **M33**)

- Since seeing is a slow-varying function of position, solve for PSF-matching kernel over a 3 x 3 grid

Derivation of PSF-matching kernel

- PSF-matching entails finding an optimum convolution kernel K by minimizing some cost function, e.g., chi-square:

$$\chi^2 = \sum_{i,j}\left[\frac{I_{ij} - M_{ij}}{\sigma_{ij}}\right]^2 \equiv (I - M)^T \Omega_{\mathrm{cov}}^{-1} (I - M)$$

where M is the "model" image: $$M_{ij} = \left[K(u,v) \otimes R_{ij}\right] + dB$$

- We discretize the kernel $K(u,v)$ into a 9 x 9 pixel image (a 2D array of delta functions) and then estimate the 81 pixel values therein (coefficients K_{lm}):

$$K(u,v) = K_{lm}\delta(u-l)\delta(v-m)$$

- Model image can then be written:

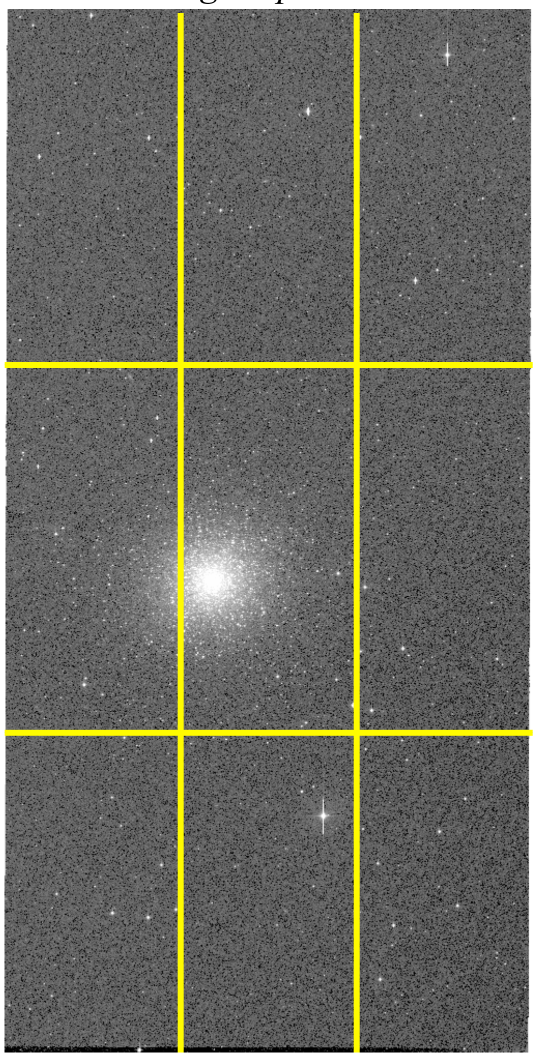
$$M_{ij} = dB + \sum_l \sum_m K_{lm} R_{(i+l)(j+m)}$$

- The K_{lm} can be solved using standard linear-least squares via $\partial\chi^2 / \partial K_{lm} = 0$ and inverting the matrix system

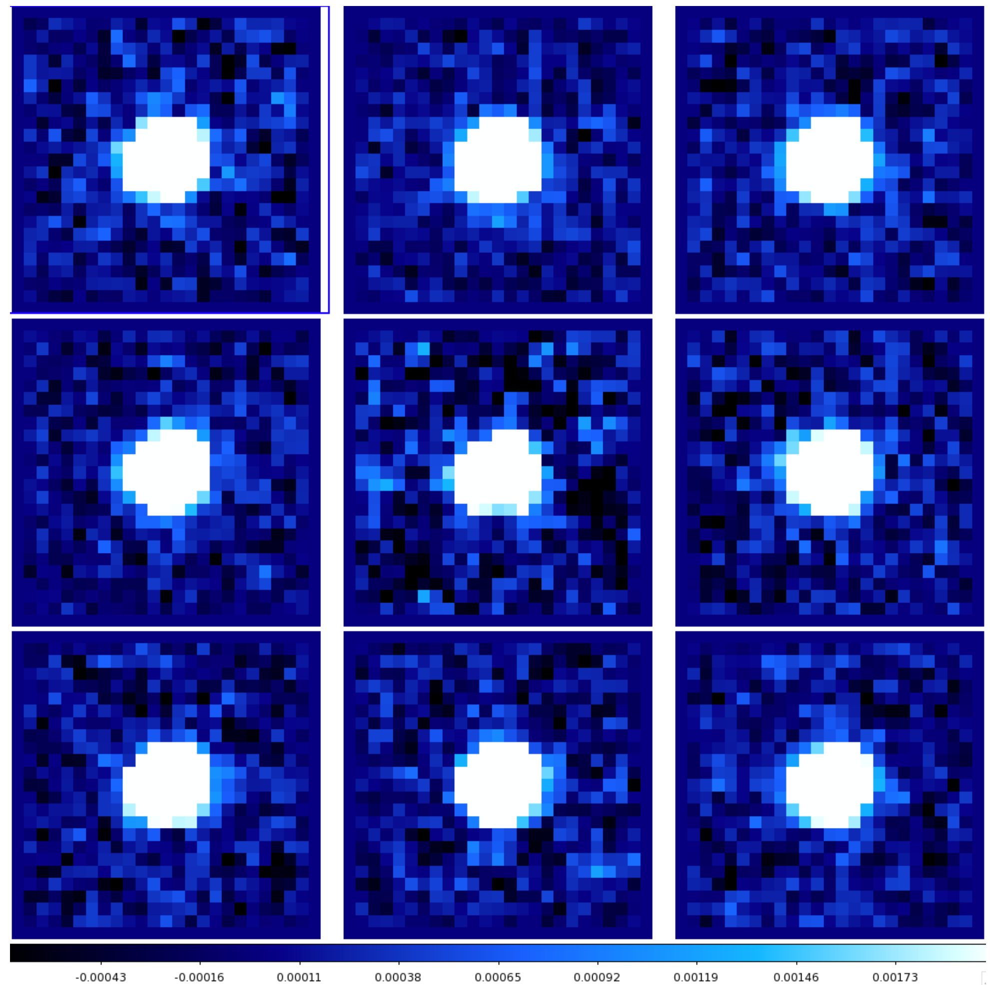
- The above delta-function-basis for $K(u,v)$ is more flexible: can take on more general shapes

Input PSF-images from *sci-image* for deriving PSF-matching kernel

***science** image exposure with M13*

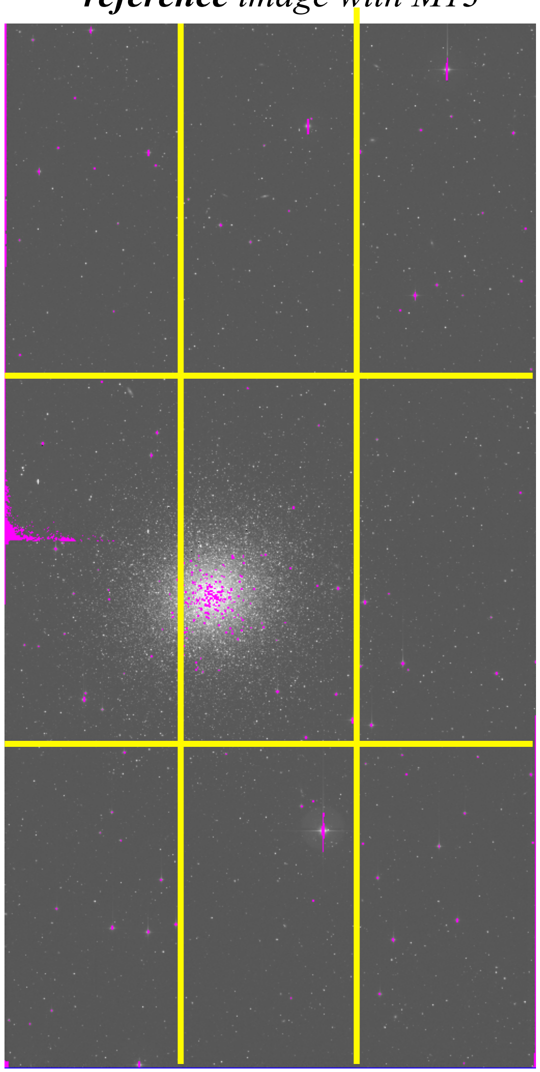


*PSF (co-add) products over **sci-image** partitions*

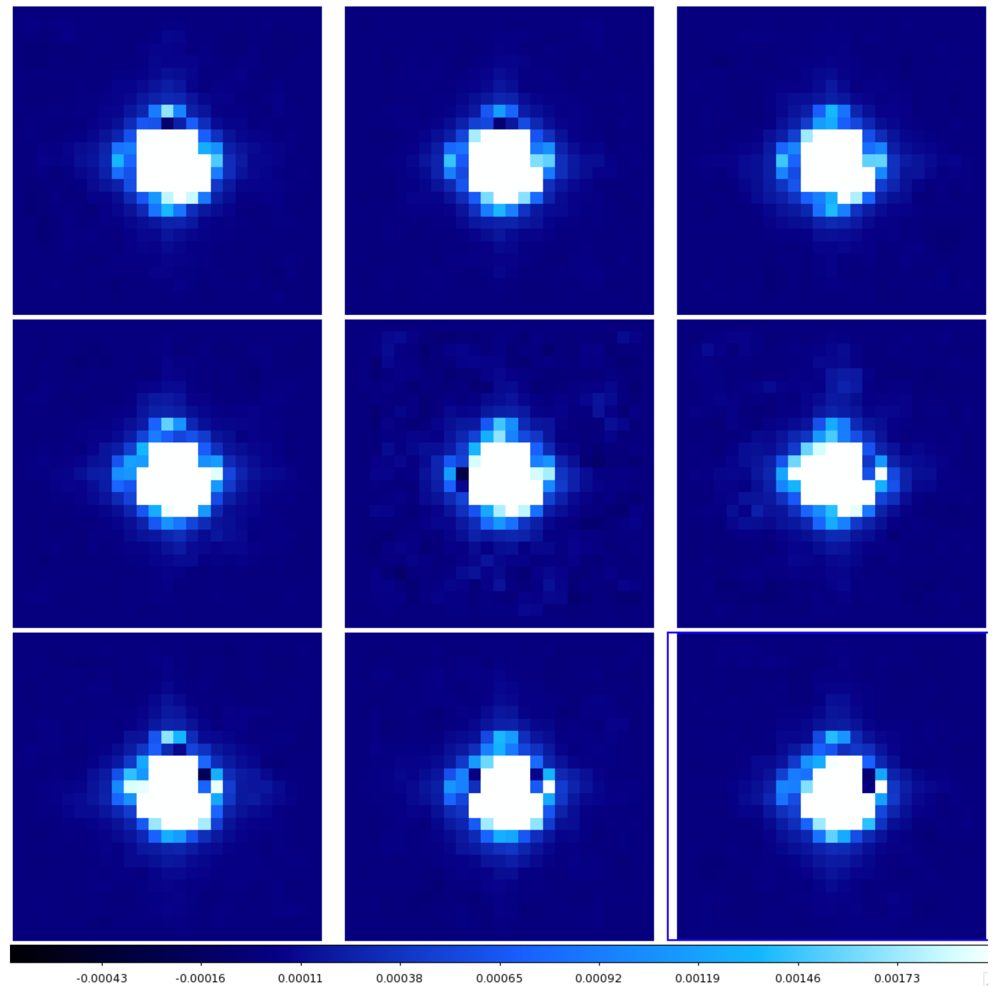


Input PSF-images from *ref-image* for deriving PSF-matching kernel

***reference** image with M13*

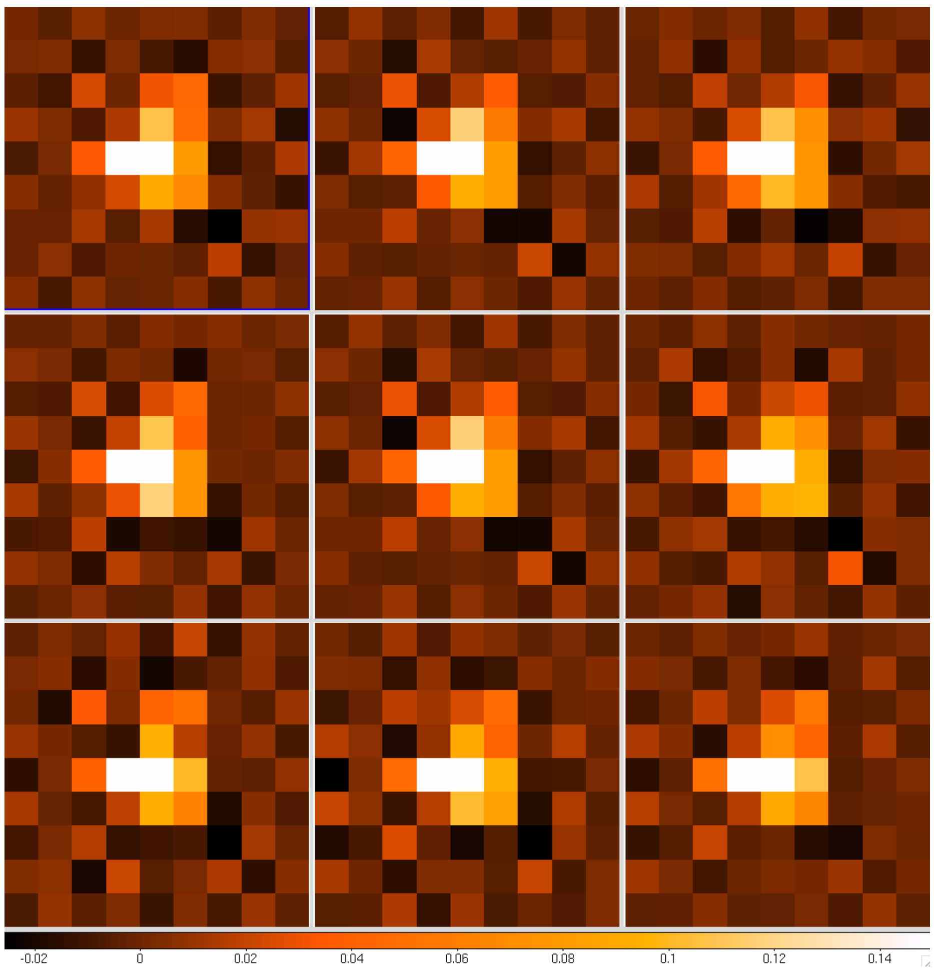


*PSF (co-add) products over **ref-image** partitions*



Final solutions for PSF-matching kernels

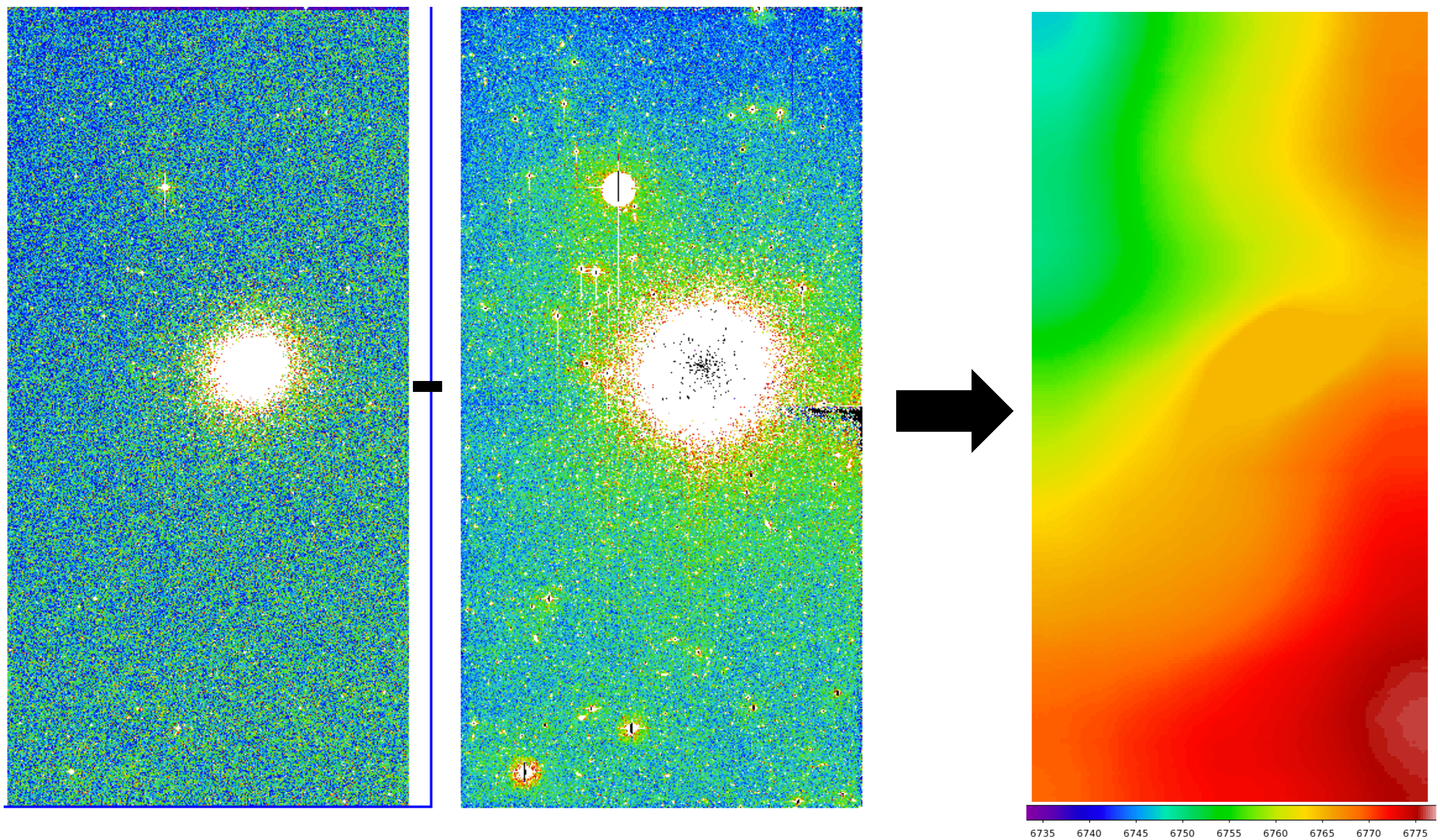
Convolution kernels for the 9 image partitions to match the *sci* and *ref* image PSFs for the M13 test case



Pre-conditioning step
(prior to PSF-matching)

- Compute low-pass filtered, smoothly-varying differential background and correct science image to match reference image background: $sci_{new} = sci_{old} - \langle sci_{old} - ref_{resampled} \rangle_{filt}$
- Helps improve photometric accuracy on difference images later

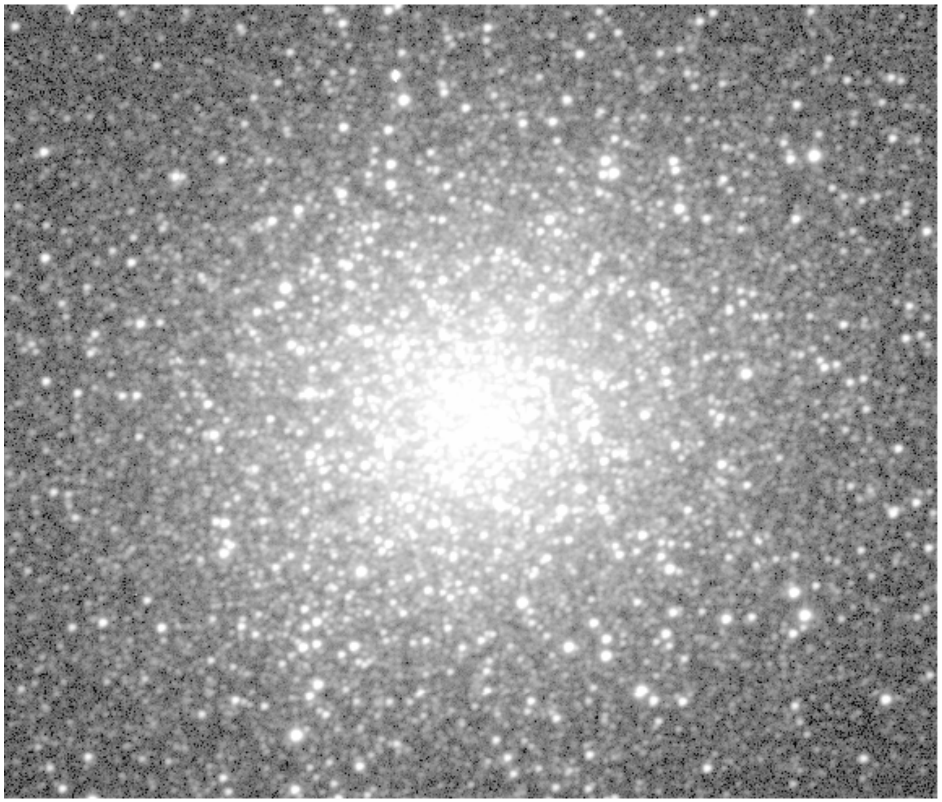
sci image *ref image* *differential background*



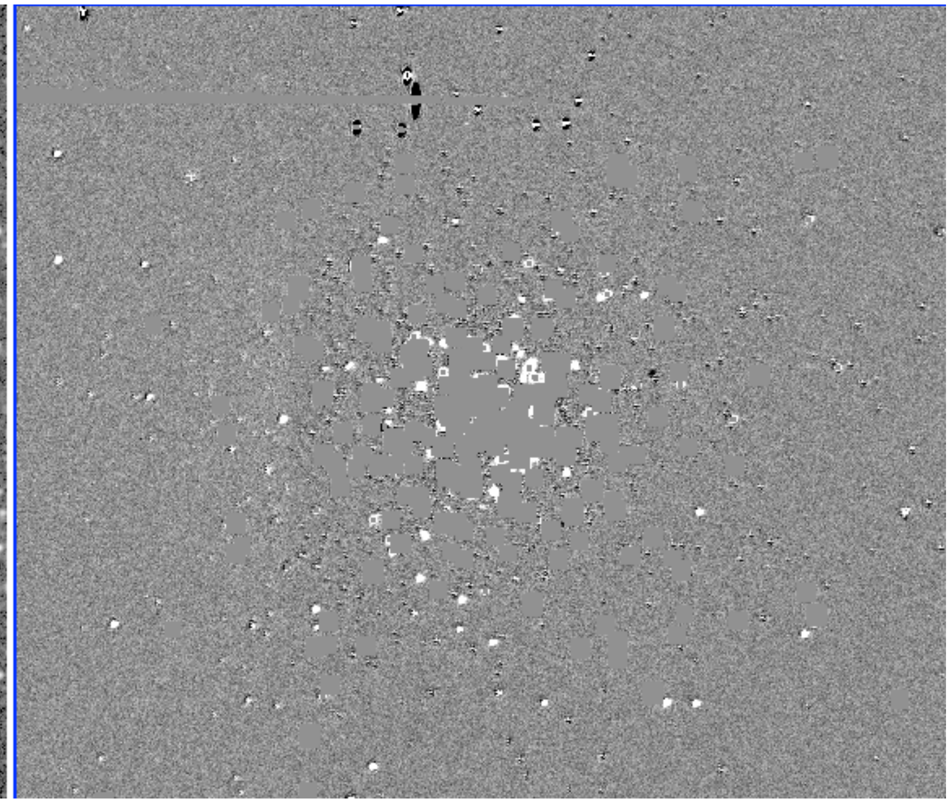
Final difference image:
zoom on M13 globular cluster

bad/saturated pixels in difference replaced by zero in difference

science image exposure (~ 9’ x 9’ *zoom*)



sci – *K* (x) *ref difference image*

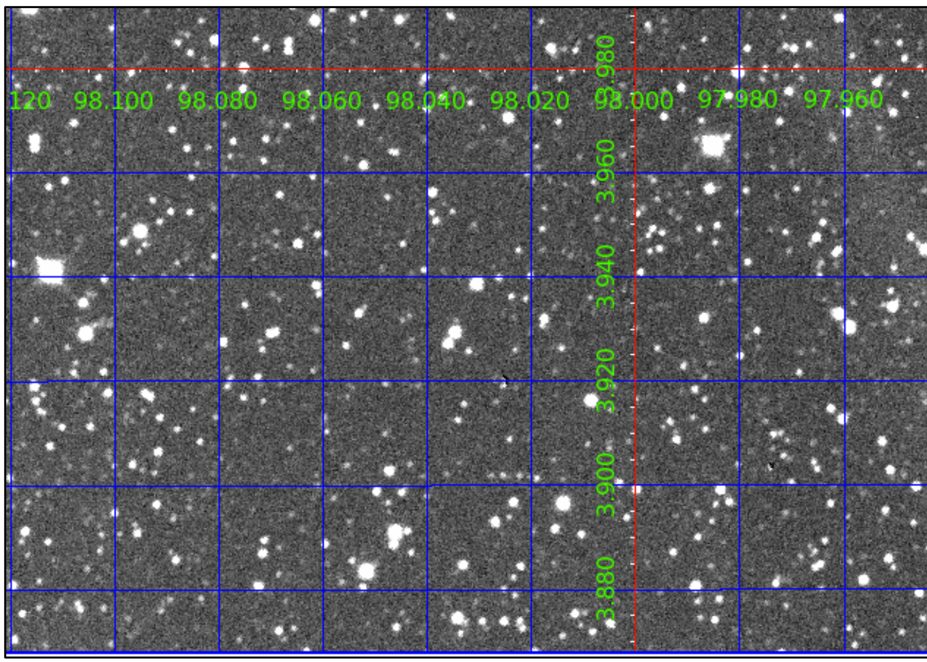


“Good” difference in Galactic Plane

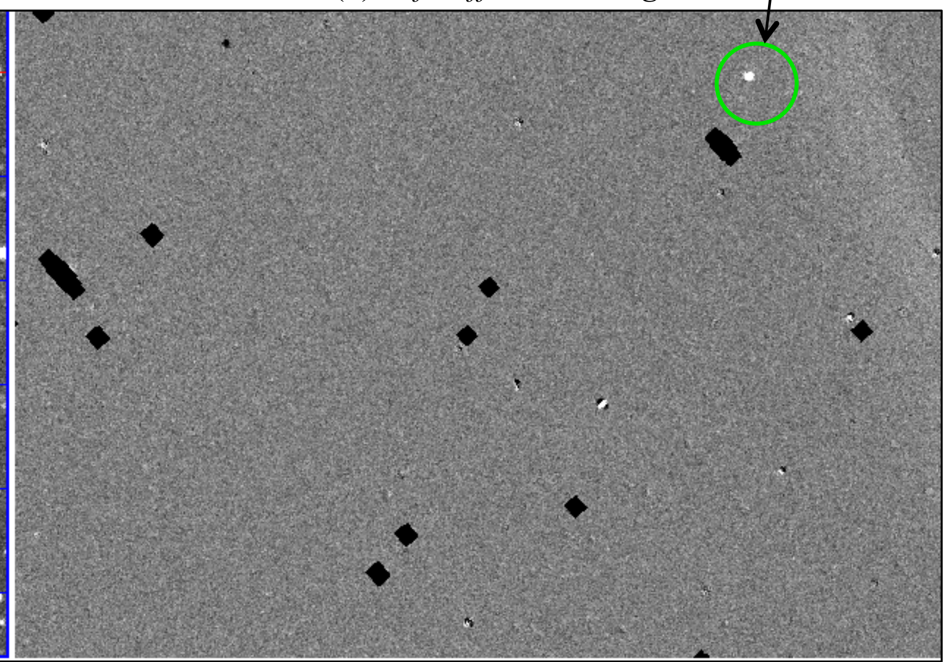
When upstream astrometric/distortion calibration is near perfect, it works!

known variable

science image exposure (~ 10’ x 7’ *zoom*)



Sci – K (x) *Ref difference image*

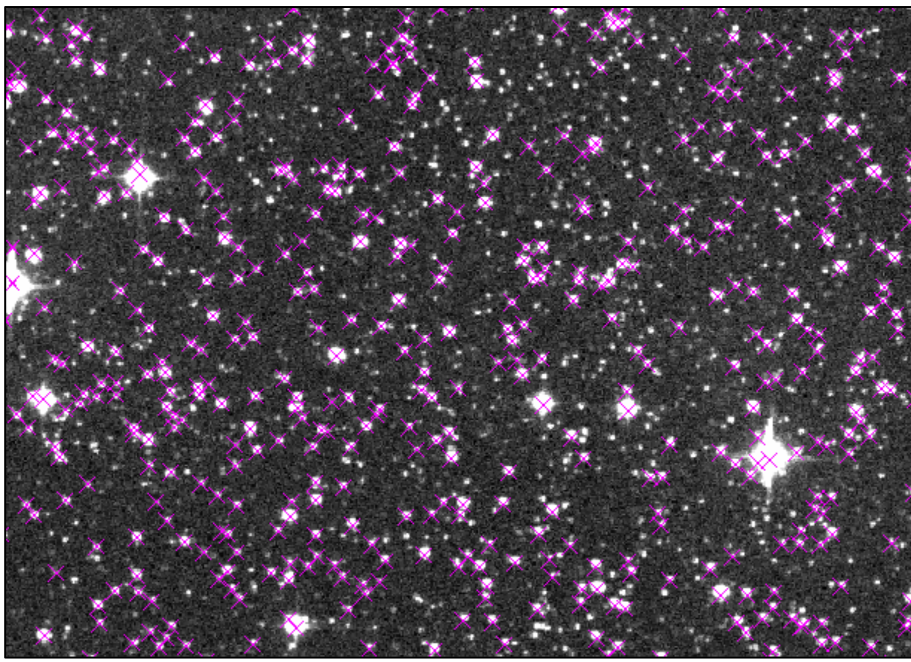


coordinate grid is galactic

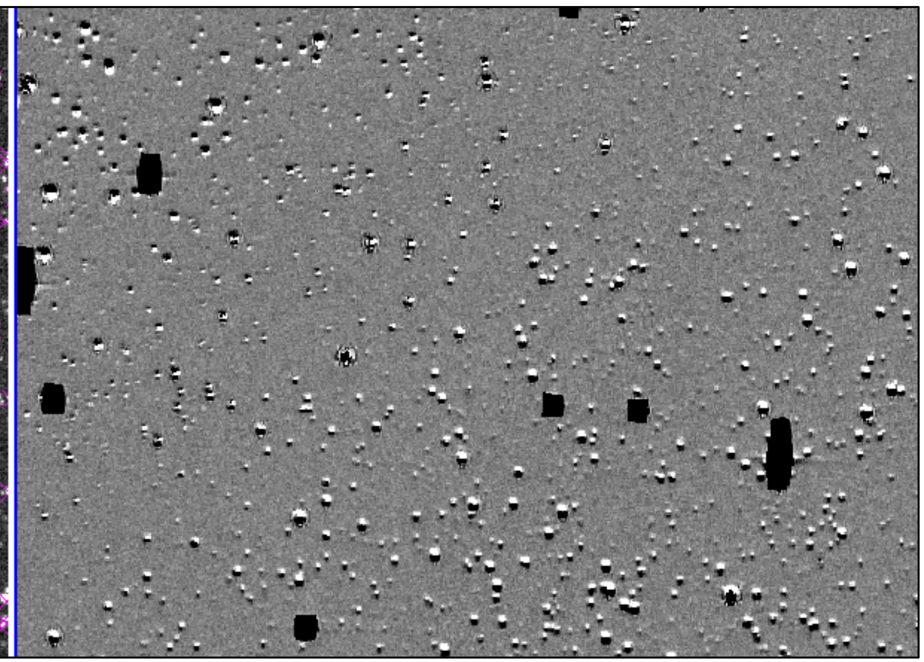
When things go wrong:
e.g. “bad” difference in Galactic Plane

When upstream astrometric / distortion calibration is “slighty” wrong (even 1-pixel from image edge-to-edge!)

science image exposure (~ 12’ x 8’ *zoom*)



Sci – K (x) *Ref difference image*



magenta crosses: 2MASS positions

Crowded-field conundrum(s)

- Crowded fields are a challenge (e.g., galactic plane)!

- Instrumental calibration is more difficult:
  - astrometry and photometric-gain matching all require source matching of some sort

- Source-matching is ambiguous and messy in crowded fields
  - naïve nearest neighbor matching using some radial tolerance is not robust
  - use of aperture photometry to calibrate relative gain is not optimal

- Current success rate for good (usable) difference images in galactic plane: ~ 50%
  - bad subtractions strain the candidate extraction and ML vetting steps downstream
  - no transients are extracted/stored from *really* bad subtractions: impacts survey completeness

Candidate extraction and photometry

- Candidate transients are detected on difference images and their fluxes measured using both PSF-fitting and aperture photometry

- PSF-fitting provides:
  - better photometric accuracy to faint fluxes; provides de-blending ability
  - diagnostics to distinguish point sources from artifacts (false-positives) in diff. images
  - maximizes reliability of candidates since most transients are point sources

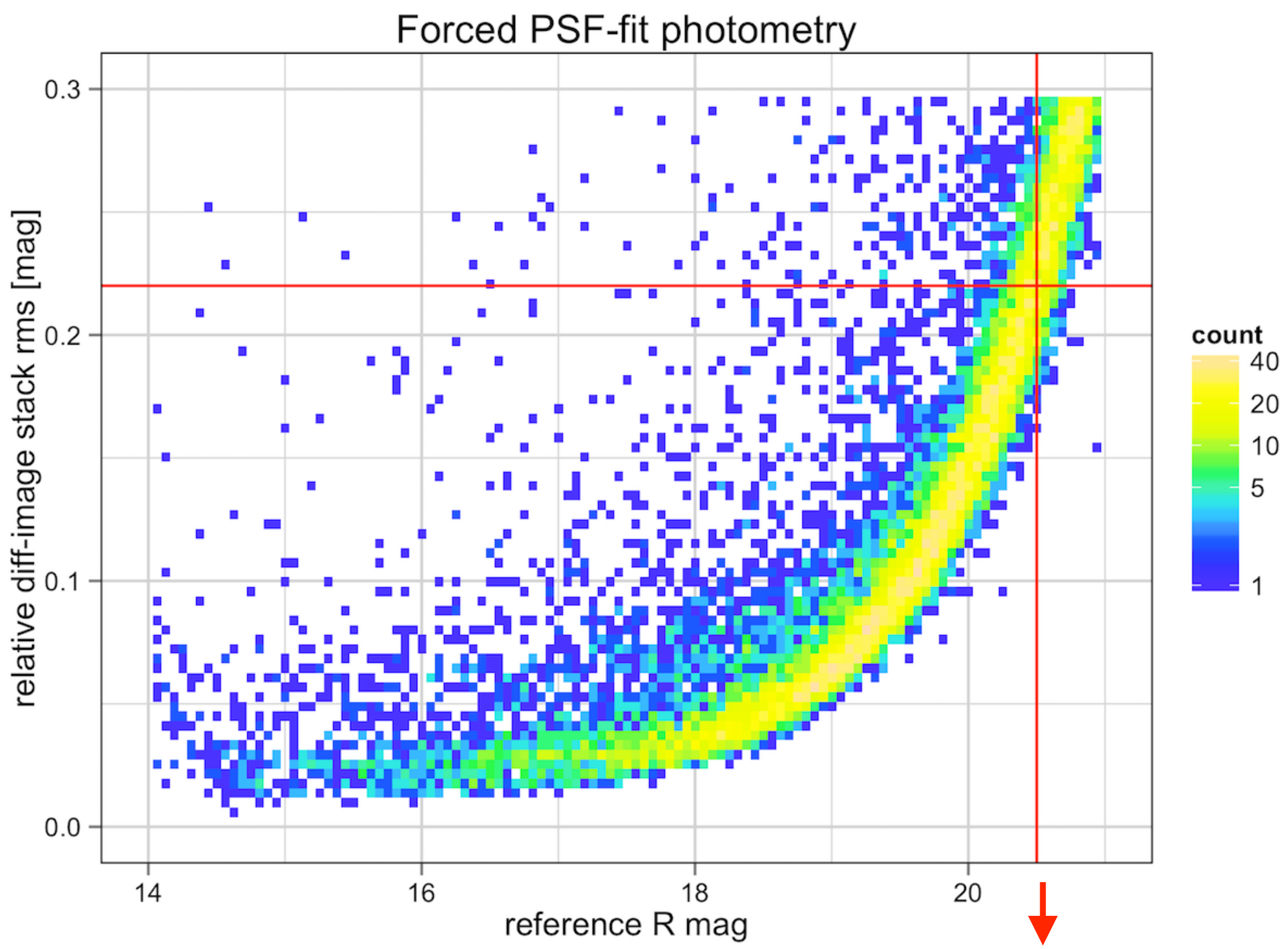
- Difference-image photometry provides “AC photometry” or “relative photometry”.
- Absolute photometry for lightcurve generation is sometimes referred to as “DC photometry”, e.g., if have a variable star with a time-average reference baseline flux f_{ref} , then:

$$mag_{DC} = ZPMAG - 2.5\log_{10}\left[f_{AC} + f_{ref}\right]$$

- Where fluxes f_{AC} and f_{ref} pertain to the <u>same</u> photometric zeropoint (*MAGZP*)

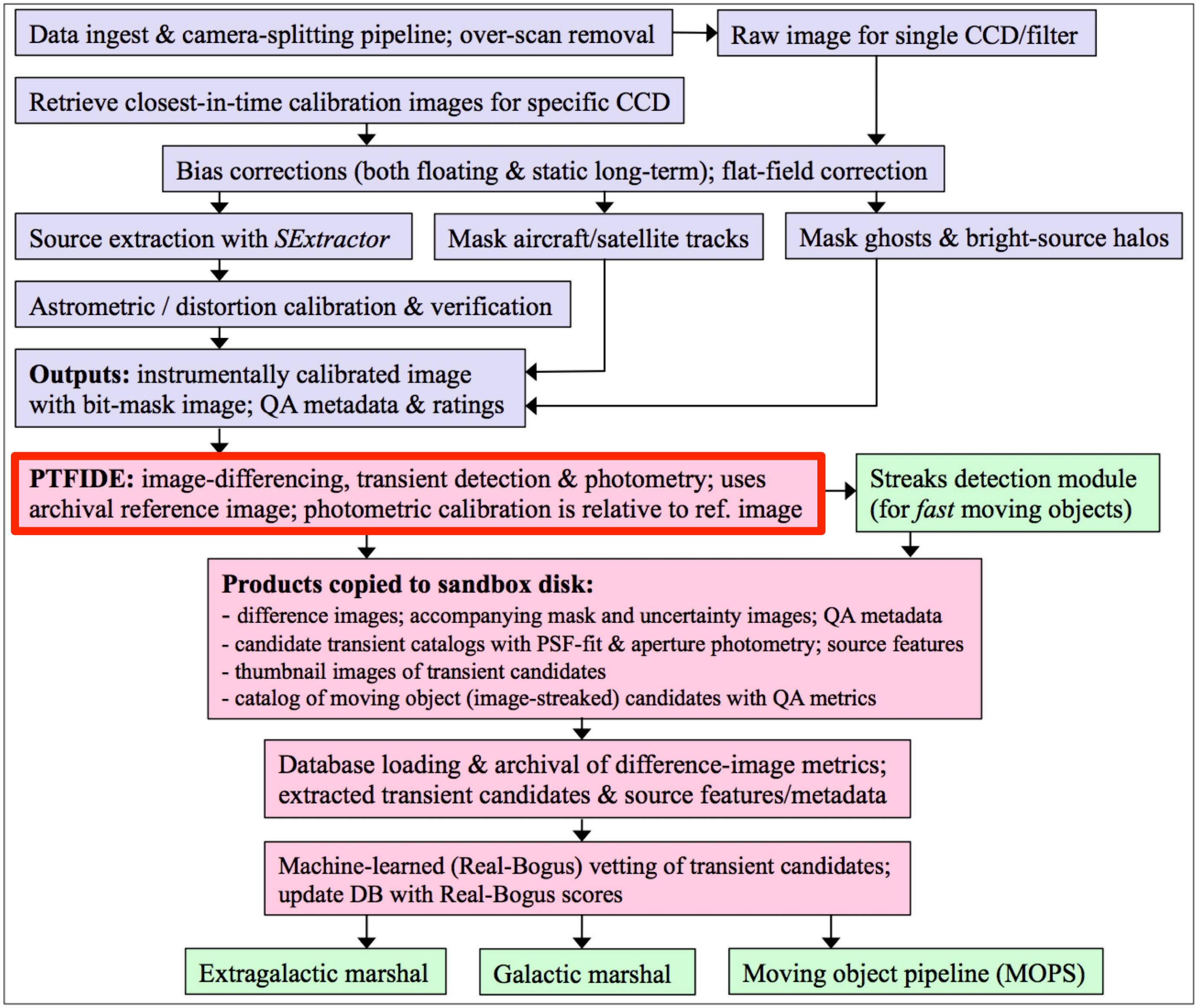
AC + DC = science or noise?

Photometric sensitivity in PTFIDE difference images from PSF-fitting



5σ limit ~ 20.5 mag

The bigger picture:
“real-time” pipeline at IPAC/Caltech



Status of operations at IPAC/Caltech

- Real-time pipeline is currently running PTFIDE, extracting transient candidates
- Populating a database with candidates, source-features, and image-subtraction metadata
- Machine-learned vetting is in place (JPL - Caltech effort)
- Dissemination and follow-up being performed by the various science marshals in near real-time
- Ongoing tuning and refinements in response to on-sky performance and science analyses
- A moving-object discovery pipeline (MOPS) is also in place: uses outputs from PTFIDE
- Offline tools also in place to enable archival research, e.g., a “forced-photometry” service

Summary / Some lessons learned

- PTFIDE with machine-learned vetting is now running in near real-time at IPAC/Caltech to support discovery and archival research
- Algorithms and software are generic enough to use on future projects, e.g., ZTF
- Refinement and optimization continues: from telescope to vetted candidates to lightcurves
- Galactic-plane challenge continues
  - Note: we are not starved for candidates to follow up!
  - Disseminating what's worthy of follow-up is a separate challenge
- Fix problems at the source (e.g., instrumental and sociological) rather than patch downstream
- More eyes on intermediate data products, not only final vetted candidates. Feedback is important!
- We are still learning!

Further reading

- More detailed presentation on PTFIDE:
  http://web.ipac.caltech.edu/staff/fmasci/home/miscscience/masci_lsst_ztf_Nov2014.pdf
- **Old** white paper on PTFIDE:
  http://web.ipac.caltech.edu/staff/fmasci/home/miscscience/ptfide-v4.0.pdf
- Photometry on PTFIDE products:
  http://web.ipac.caltech.edu/staff/fmasci/home/miscscience/forcedphot.pdf
- In preparation:

The IPAC Image Subtraction and Discovery Pipeline for the Palomar and Zwicky Transient Facilities

Frank J. Masci[1], Russ R. Laher[2], Umaa D. Rebbapragada[3], Gary B. Doran[3], Brian D. Bue[3], Mansi Kasliwal[4], Eric Bellm[4], Eran O. Ofek[5], Jason Surace[2], David L. Shupe[1], Carl J. Grillmair[2], Ed Jackson[2], Adam A. Miller[3,4,8], Adam Waszczak[6], Tom Barlow[4], Lin Yan[1], Yi Cao[4], S. Bradley Cenko[7], George Helou[1], Thomas A. Prince[4], and Shrinivas R. Kulkarni[4]
fmasci@caltech.edu

August 15, 2015

ABSTRACT

We describe a new, near real-time transient-source discovery pipeline for the *intermediate* Palomar Transient Factory (iPTF) and its successor, the Zwicky Transient Facility (ZTF), with an emphasis on algorithms for PSF-matching, image subtraction, detection, photometry, quality-assurance, and machine-learned vetting of extracted candidates. We review the metrics and

Back up slides

Machine-learned candidate classification: real or bogus?

- **Definitions:**

  RB = Real or Bogus candidate, above some probability threshold

  FPR = False Positive Rate: fraction of bogus transients incorrectly classified as real (max tolerable = 1%)

  FNR = False Negative Rate: fraction of real transients incorrectly classified as bogus

Performance analysis as of ~ May 2015 (Umaa Rebbapragada & Gary Doran, JPL):

- **Measure FNR at 1% FPR**
- **NERSC:**
  - **RB2: 7.7% (Brink et al. 2012)**
  - **RB4: 17.0%, RB2 had 42% on same set**
  - **RB5: 5% (LANL)**
- **PTFIDE:**
  - **3%**

Research described in this presentation was carried out at the Jet Propulsion Laboratory under a Research and Technology Development Grant, under contract with the National Aeronautics and Space Administration.