STREAMING ALGORITHMS FOR OPTIMAL COMBINATION OF IMAGES AND CATALOGS

3/9/2015

Tamás Budavári / Johns Hopkins University

Processing Observations

- Take images on the mountain & process at home
  - IRAF, IDL, Python, PyRAF, ...
- Sloan Digital Sky Survey did single pass by design
  - The PHOTO pipeline
- The time-domain is different!

Goals

- From hundreds of visits, we need
  - High signal-to-noise ratio
  - Sharp images
  - Deep catalogs

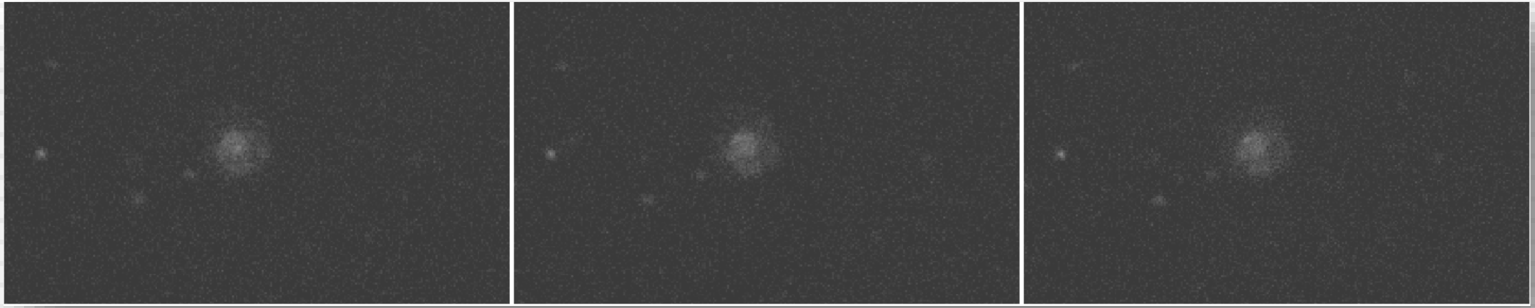
Traditionally

- Batches
  - Collect images of next release & combine them
  - Extract deep catalog & time series
- Long wait time & inefficient processing
  - Could we do it incrementally?

Outline

- Blind deconvolution of multi-epoch imaging
- On-the-fly catalogs for time-domain surveys

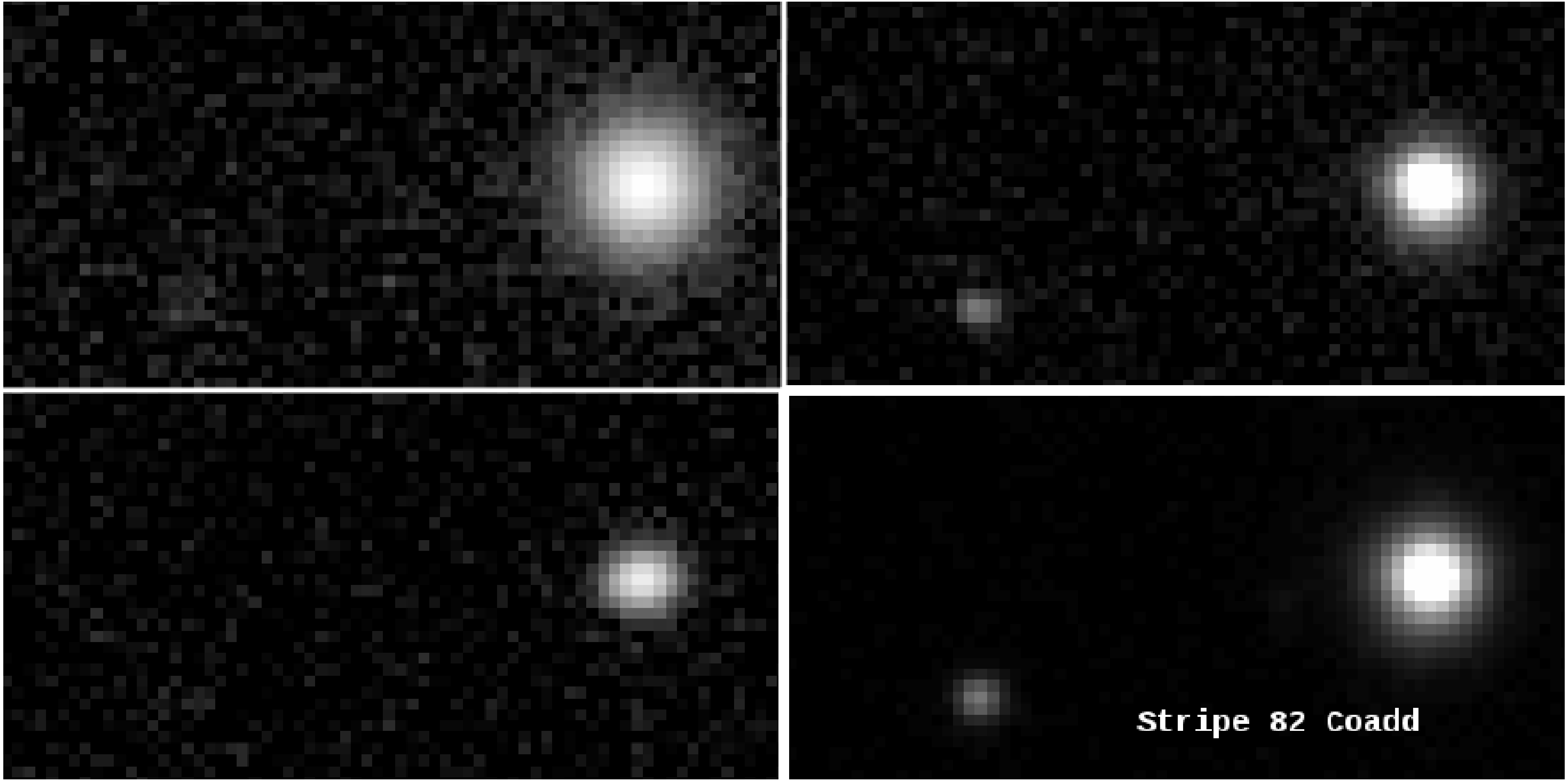
Combining Images



Current Methods

- Brute-force summing of images is incorrect
- Lucky imaging uses only the best images
- Convolve to worst acceptable PSF & coadd

Throwing away a lot of information!


Stripe 82 Coadd

Simple Model for Exposures

- Background image convolved with unknown point-spread function
- Plus the noise

- Solve for x?

$$y_t = f_t * x$$

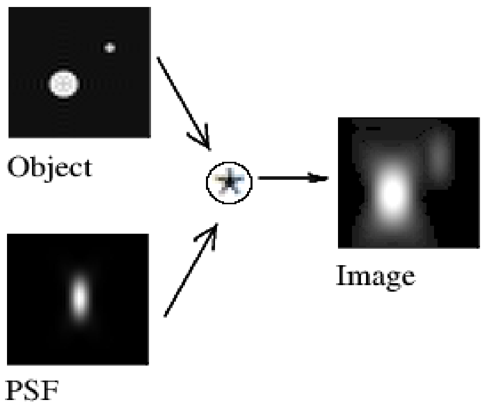


Image Deconvolution

- Correcting Hubble's optics & R-L deconvolution
  - See *White (1994), Starck+ (1994), Lauer (1994,2002), …*
- Now it's different with hundreds of exposures
  - With different PSFs

Multiepoch Deconvolution

- We solve for background image & all the PSFs
  - Breaks the degeneracy of the single-image case
  - Iterative incremental approach
    - Solve for PSF of each incoming image
    - Update model image
    - Repeat until convergence

Computational Optics

- Elegant optimization
  - Richardson-Lucy updates for Poisson likelihood func
  - Gaussian yields similar updates

$$x_{t+1} = x_t \odot u_{t+1}$$

The devil is in the details!

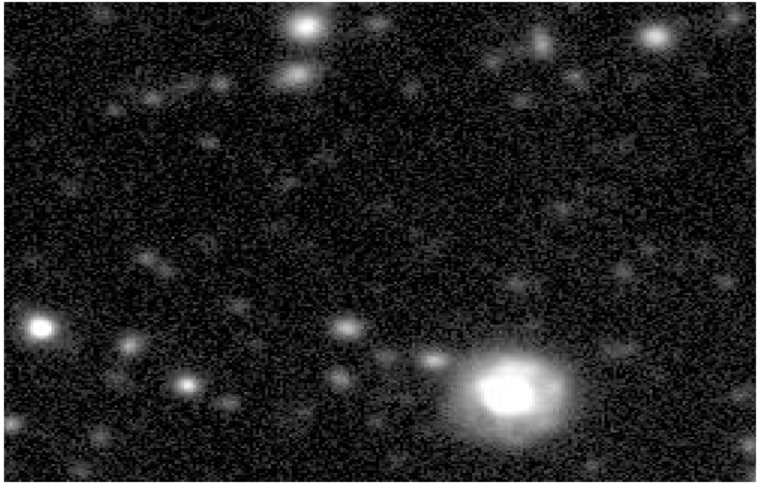
Priors & Likelihoods

- Stars are point sources
  - Regularization
- Modified likelihoods
  - Masking saturated & bad regions
  - Damped variants & robust stats
- Controlling convergence
  - Update clipping, …

Coadded & Reconstructed

- Coadding
  - Brings out faint sources
  - But blurs the images
- We deconvolve
  - For high-res details

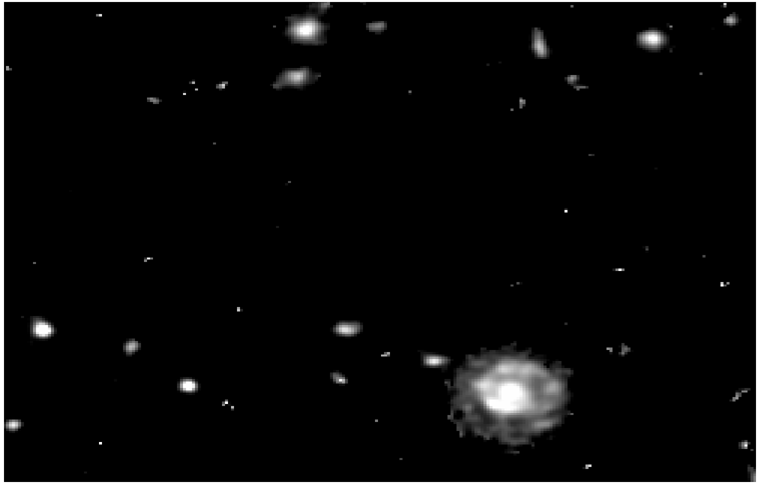
Coadded Image



Coadded & Reconstructed

- Coadding
  - Brings out faint sources
  - But blurs the images
- We deconvolve
  - For high-res details

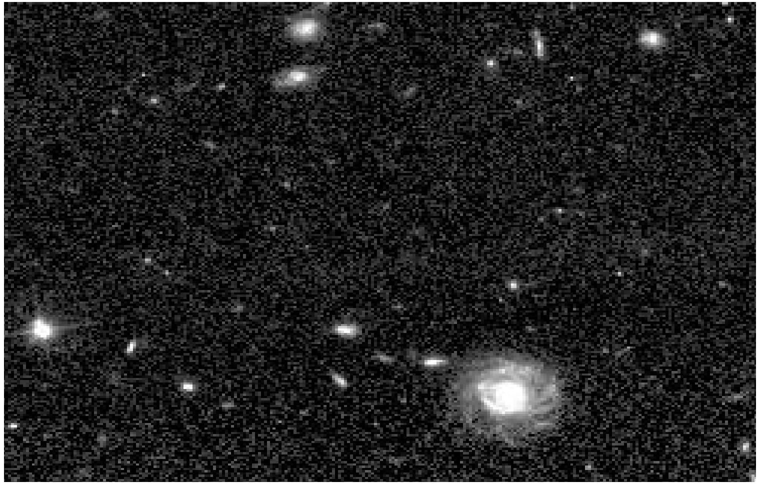
Deconvolved Image



Coadded & Reconstructed

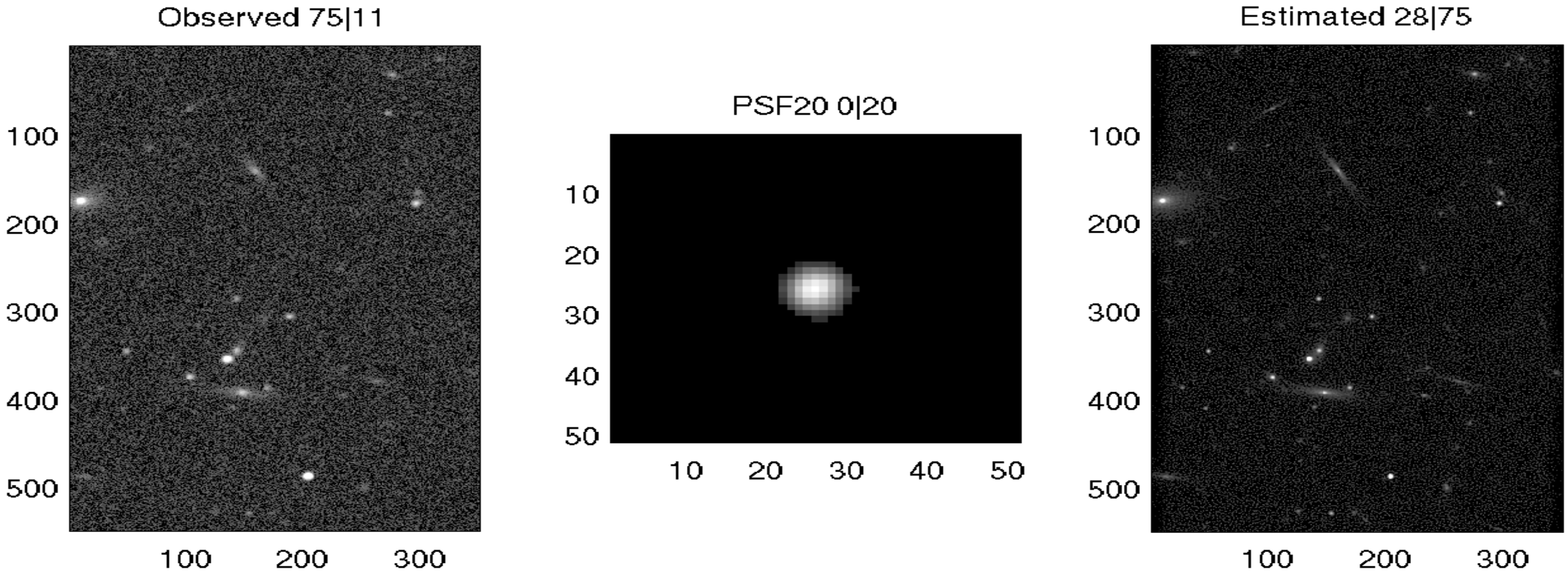
- Coadding
  - Brings out faint sources
  - But blurs the images
- We deconvolve
  - For high-res details

Hubble Image



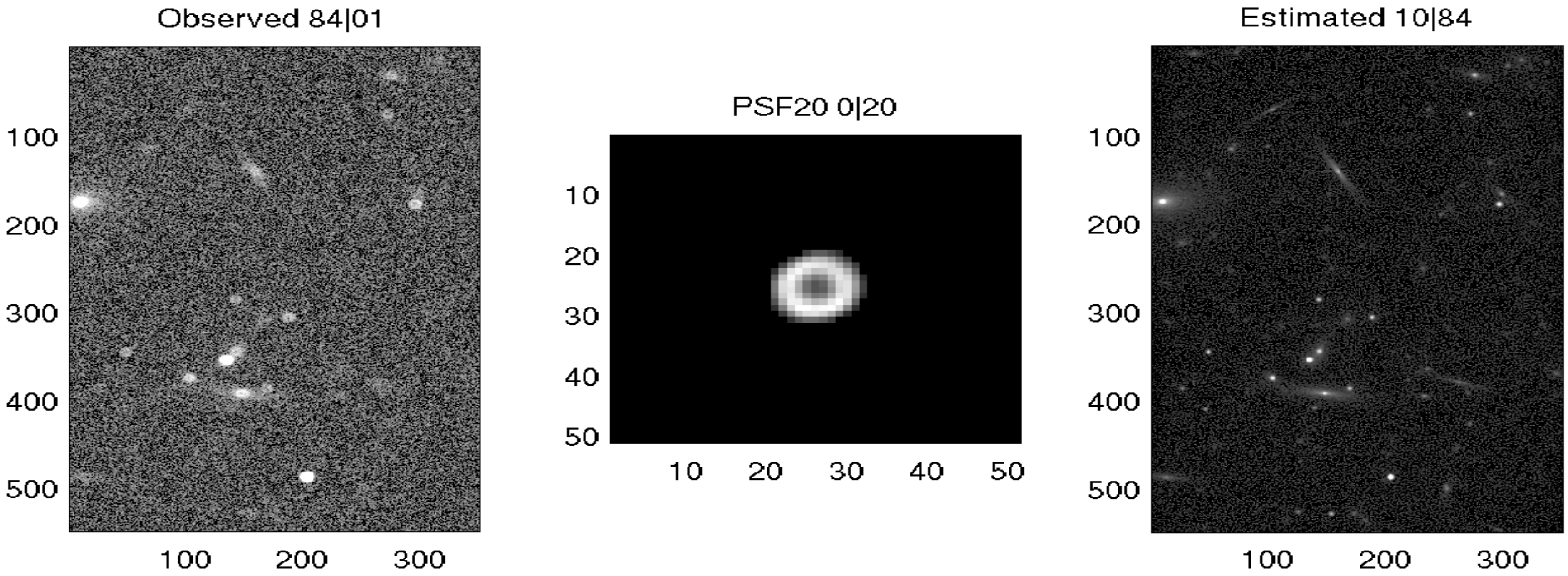
LSST ImSim

Work with Andy Connolly



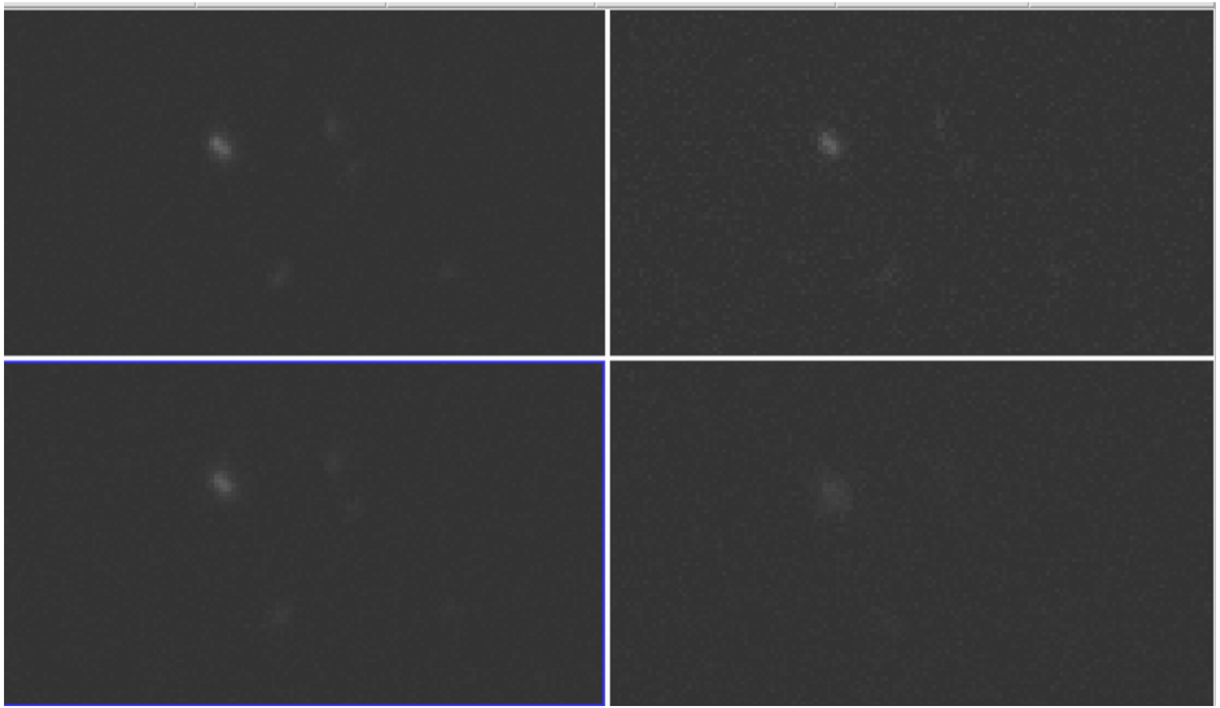
LSST ImSim

Work with Andy Connolly

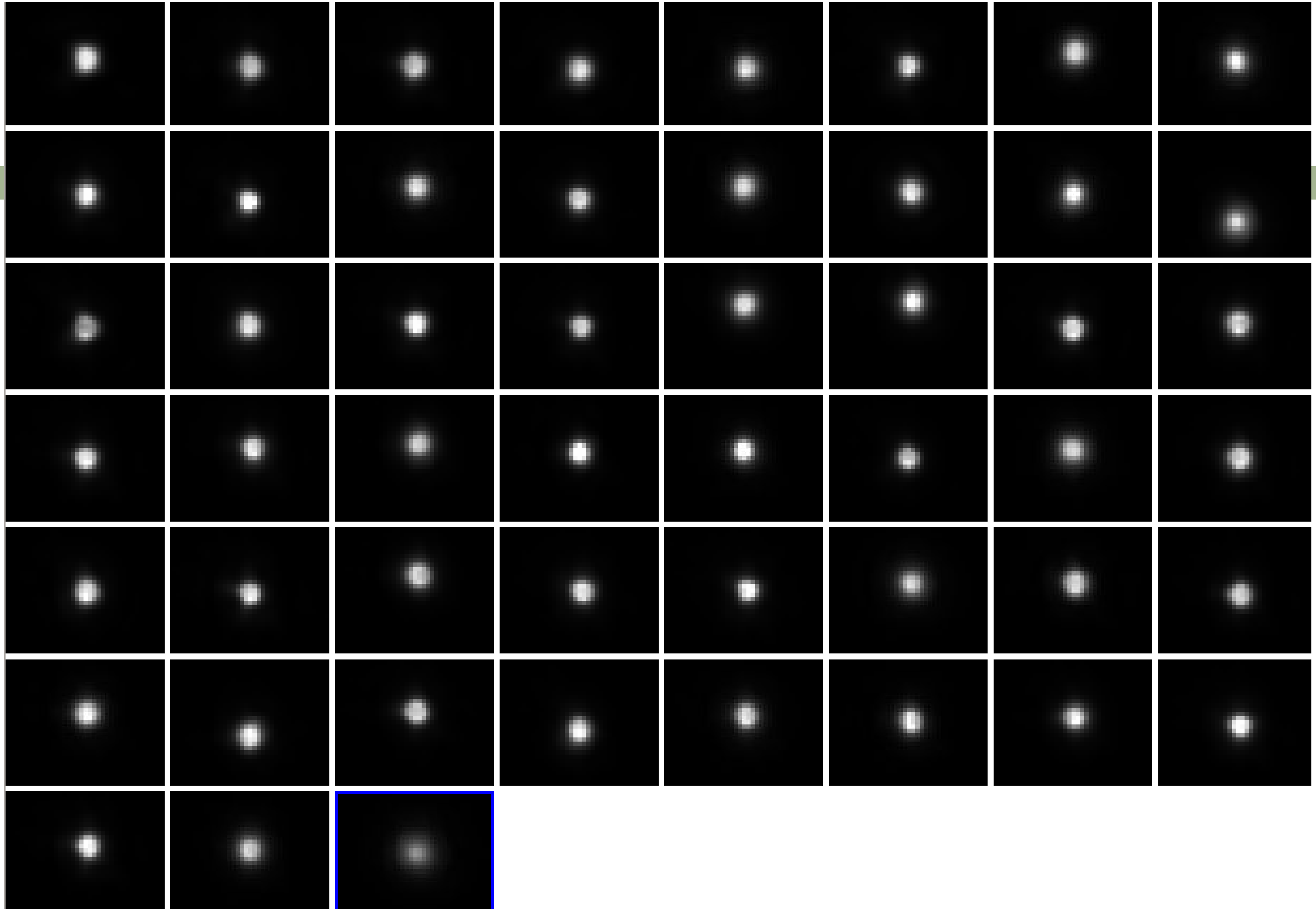


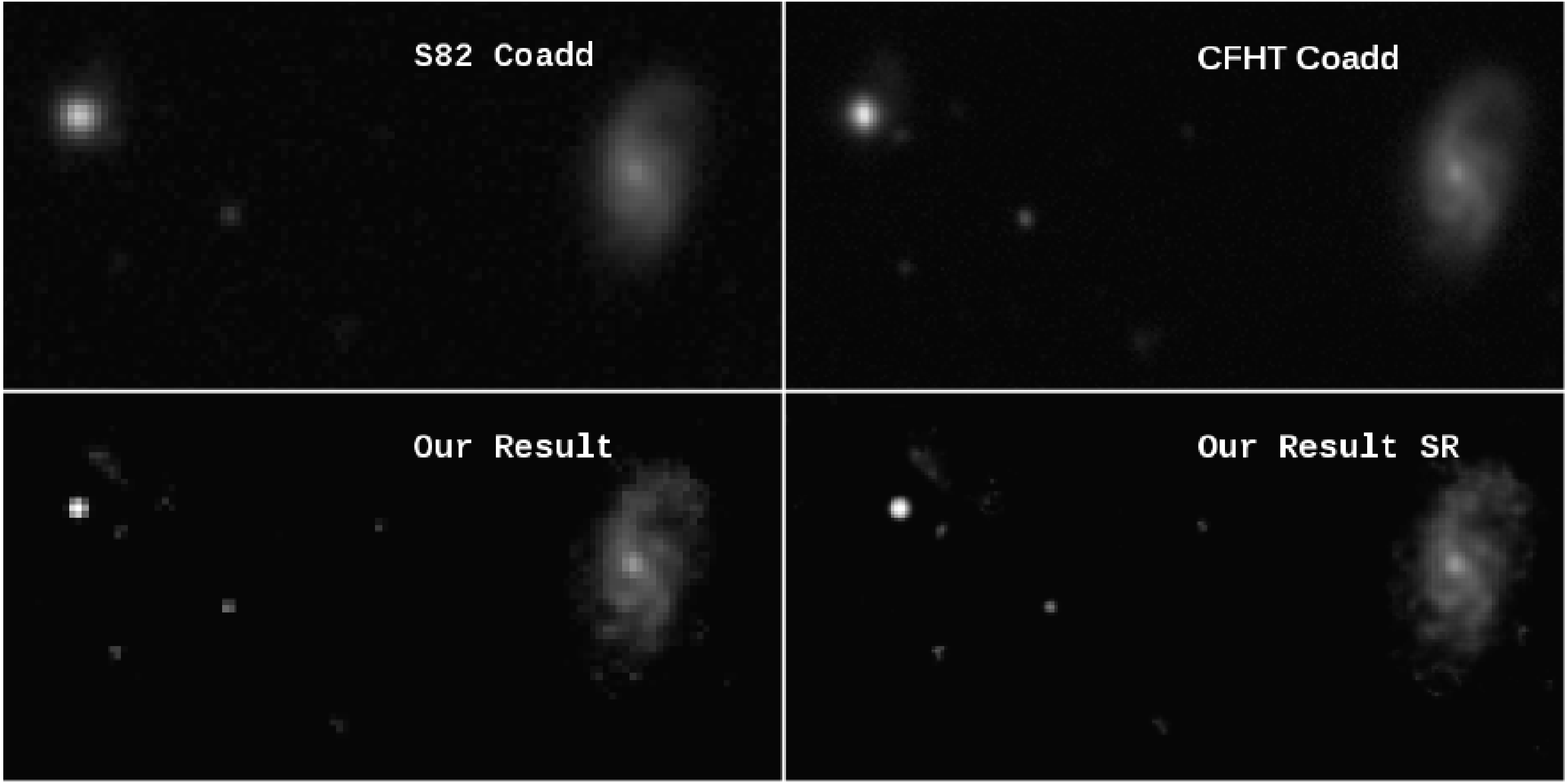
SDSS Stripe 82

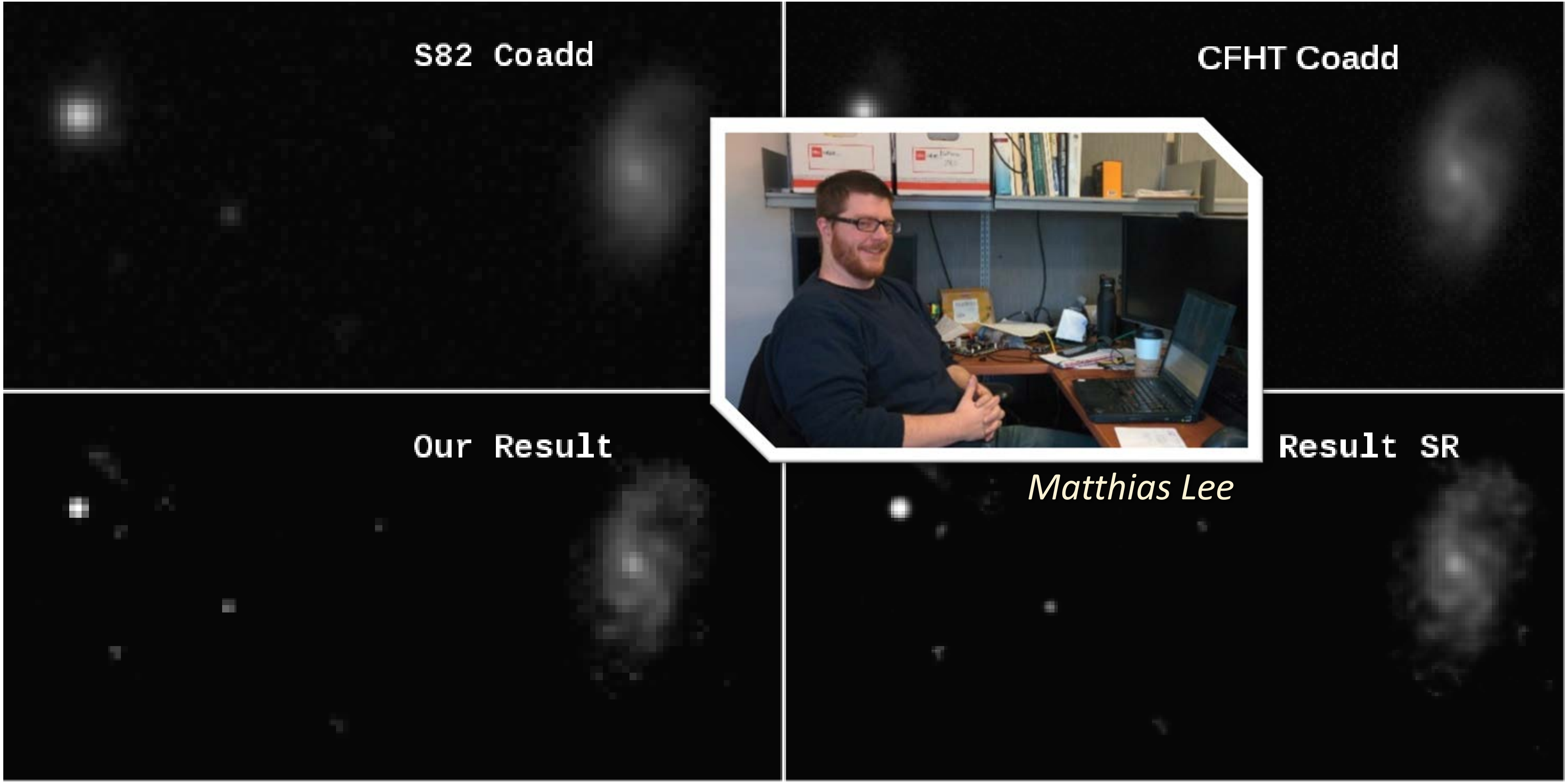
- Single-epoch samples



PSFs




S82 Coadd
CFHT Coadd
Our Result
Our Result SR



Matthias Lee

Beyond Pretty Pictures

- Statistical comparison to ground truth
  - To deeper exposures
  - To LSST ImSim

Tools and Future Work

- Fast GPU implementation in Python + CUDA
- Include sky background solution
- Source extraction from results
- Need pipeline code
- Get more science out!

[Funded by NSF AAG]

Alternatively on Catalogs?

- Coadd images & perform forced photometry
  - Ideal but registration tricky, processing slow, …
- Extract source lists from each image
  - Combine source lists into a catalog

Need to dig in the noise to go deep!

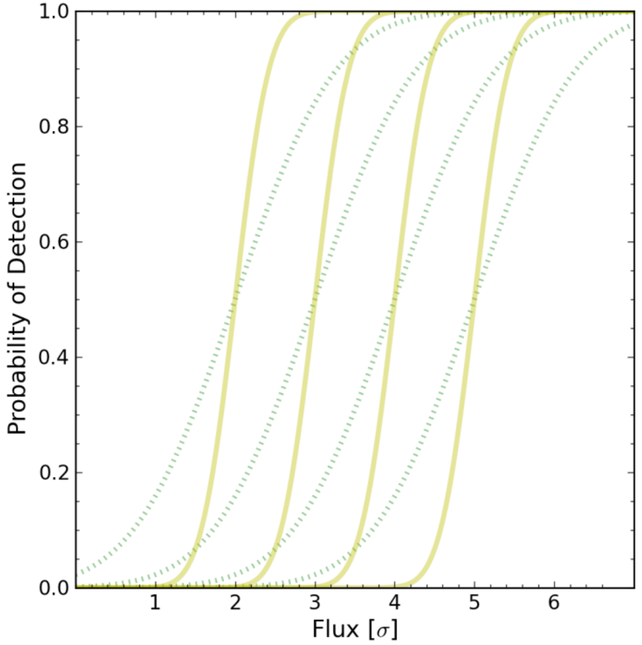
Detection Probability

- Measured flux is true + normal error $f_i = f + \epsilon_i$
- Probability of detection

$$P_f \equiv P(f_i > f_D | f) = \frac{1}{2}\,\mathrm{erfc}\left(\frac{f_D - f}{\sigma\sqrt{2}}\right)$$

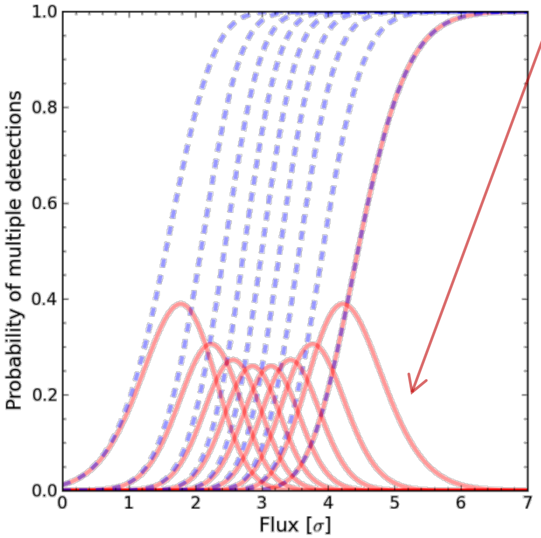
Detection Probability

- Measured flux is true + normal error $f_i = f + \epsilon_i$
- Probability of detection
  - As a function of the true flux
    - Thresholds at *2-, 3-, 4- & 5*σ
  - Sharper for 9-way stacks



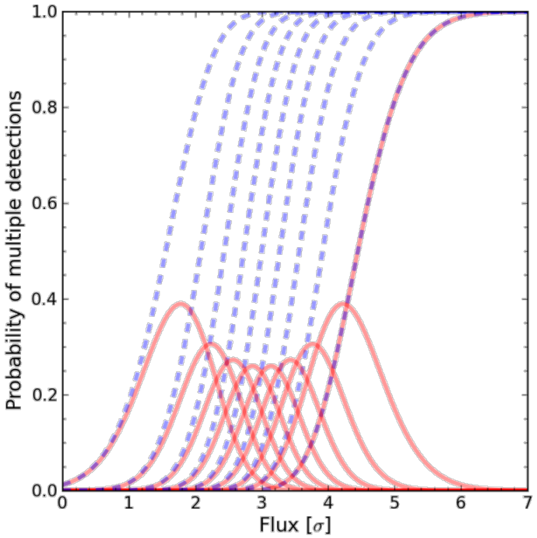
Detection Probability

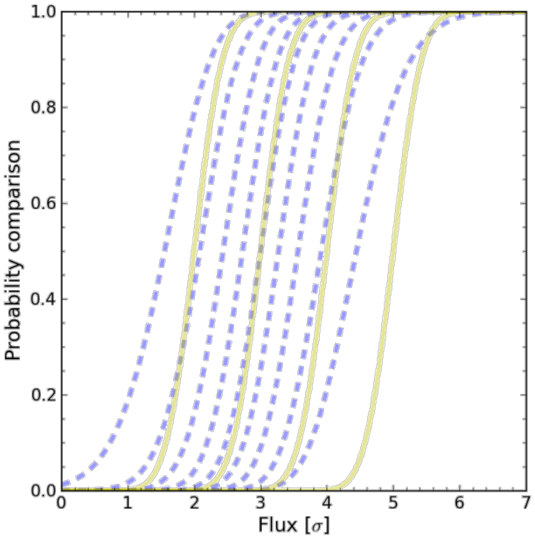
- Multiple exposures
  - Binomial $$P(n|k,f) = \binom{k}{n} P_f^n \, (1-P_f)^{k-n}$$



Detection Probability

- Multiple exposures
  - Binomial $$P(n|k,f) = \binom{k}{n} P_f^n \left(1-P_f\right)^{k-n}$$





What is a Real Source?

- Is it “real” or just “noise” ?
  - Bayesian hypothesis testing

$$B = \frac{L_{\mathrm{real}}}{L_{\mathrm{noise}}}$$

We weed out the noise as we go!

Summary

- Processing on streams of data
  - Linear scaling and monitoring of results
- Multiepoch deconvolution works
  - GPU speed makes this practical
- On-the-fly catalog aggregation

Incremental approaches that scale