

# LSST Data Management Working Group Report (Static/Shape Pipeline)

Attending: Andy Connolly (chair), Charles Alcock, Chuck Claver, Inwoo Han, Chris Jermaine, Tod Lauer, Tony Tyson, Dave Wittman

## Section 1. Research and Development Prior to Construction

The working group discussed a number of substantial R&D tasks that need to be accomplished in the near-term in order that the LSST proceed to an implementation phase. While the tasks outlined below do not represent a complete set of algorithmic challenges we believe they are critical path items for the development of the LSST Static/Shape Pipeline.

1. Algorithm Development
  - a. Coaddition of images accounting for differential chromatic refraction.
  - b. Optimal weighting of data under variable seeing and transparency conditions.
  - c. Calculation of the geometric distortion and PSF across a frame and its correction on a field or CCD basis.
  - d. Optimal image shape extraction under conditions with a variable PSF.
  - e. Optimal deblending of sources.
  - f. Optimal photometry of faint galaxies (i.e. sources without an a priori model for their image profiles).
  - g. Image coaddition and source detection in color space.
2. Calibration Procedures
  - a. Requirements and techniques for photometric calibration of multicolor photometry across the full sky.
  - b. Absolute calibration of the LSST photometric system.
  - c. Astrometric accuracy requirements for the shape pipeline and calibration of astrometry across an LSST field.
3. Development Resources
  - a. LSST imaging simulator.
4. Science Resources
  - a. Metadata for science analyses including Montecarlo realizations and structures for describing the selection function across the sky.
  - b. Deliverables required by individual science cases and their time frames.

## Section 2. Issues in developing the R&D components for the LSST

A more detailed discussion of the requirements for these individual R&D tasks was undertaken by the working group to determine the scope and nature of the algorithms and how current state-of-the-art analyses and telescopes addressed these issues.

1. Coaddition of images taken over a wide range in conditions (transparency, airmass, filters)
  - a. The primary concern about image coaddition comes from the differential chromatic refraction (DCR) across a single filter or between filters for multicolor data. In catalog space (where the relative colors of a source are known) this can be corrected for in a simple fashion (i.e. it is a smooth function of zenith distance). In image space and for sources below the detection threshold of a single exposure (where we have no a priori color information) the correction is more complicated. Based on the LSST scheduling model it is likely that image offsets in excess of 0.2 arcseconds will be present within the data and that they may introduce biases into any shape measures. To date most lensing surveys ignore or reduce this problem by concentrating on low airmass fields. Algorithms for correcting an image are, therefore, lacking. Approaches discussed by the working group, for tackling this problem, focused on an iterative approach (initially a mean correction based on the color of the sky then measuring the colors of sources and repeating). Challenges such as correcting on a pixel-by-pixel level and the need for color information for all sources that we wish to correct were also discussed. The magnitude and impact of this effect needs to be addressed in the near future as it has an impact on the need for an atmospheric dispersion corrector on the LSST. Current observational data sets have the necessary range of airmass and colors to test these effects.
  - b. With the wide range of PSF and imaging conditions the development of optimal coaddition techniques that can account for the variable conditions is essential. Approaches such as the convolution techniques proposed by Kaiser or maximum likelihood image reconstruction techniques were discussed and the need for detailed analyses that characterize how well these techniques work was identified. Any implemented technique must be able to account for image rotation present within an Alt-Az system and be general enough to provide coaddition algorithms for both the static and variable pipelines. Some discussion focused on developing new data structures that would facilitate the coaddition of the data.
  - c. The initial two algorithm developments discussed above rely on a detailed understanding of the geometric distortion and PSF across an LSST field. A case study is required to understand how much of this information can be derived from wavefront sensing and the metrology of the telescope. The trade off between requiring stars distributed across a frame (with the associated problem of the stars being predominantly red and therefore having a different DCR than blue galaxies) and the difficulty in understanding the optical state of the telescope needs to be addressed.
  - d. A general agreement arose for a simple image stacking timeline where the stacks are accomplished over 1 month, 6month, 1 yr, 2 yr and 4 yr etc baselines. Some concerns were raised about dealing with moving sources for the longer baseline coadditions and the need for an iterative approach to this problem was discussed.

2. As systematics will likely drive the quality of the science that can be achieved with the LSST the calibration of the system (photometrically and astrometrically) was seen as a major R&D effort.
  - a. Techniques such as those developed by the 2MASS and SDSS groups for calibrating the internal photometric system of the telescope by tying fields together through cross scans and overlaps was considered an appropriate approach for the LSST. Work on how well this strategy would work given the observing strategy of the LSST was seen as a near term goal of the R&D development. How these self calibration techniques deal with variable transparency across a field needs to be quantified. While calibration of the internal system was seen as the priority (as the LSST will define a photometric system that others will adopt) it is clear that we must be able to relate the number of photons detected to absolute flux density. Techniques for tying the LSST to an absolute system and the fundamental uncertainties on such measures (including how we determine the response of the system on a chip by chip basis and for the lifetime of the survey) need to be addressed.
  - b. A well defined basis for the required accuracy for the LSST photometric and astrometric system needs to be defined by the science drivers for the project.
  - c. The impact of these requirements and the number of passbands needs to be modeled for the individual science cases (e.g. the photometric redshifts for the weak lensing studies) to understand how they impact on the survey.
3. Metadata and extracting science from the survey
  - a. A requirement for a telescope facility database was discussed. There was general agreement that a database that would store the state of the telescope (actuators, temperatures etc) at any given time was required. This facility database would record changes to the system (engineering and software) and be linked to both the QA components and to the science databases. The need for science results to be correlated with the state of the telescope was seen as important for understanding the presence of systematics in the system
  - b. In order to extract science from the database we need to know the observing conditions (depth of a particular region of the sky, the seeing for a given observation, whether a bright star meant that an object could not be detected, and Galactic reddening as a function of position). All of these components are required to understand the selection function of the data. Data structures for representing these selection function in a compact manner (and that can be efficiently queried) need to be developed. Approaches varying from Montecarlo simulations where fake observations are injected into the data stream to analytic measures of the variance on the sky need to be incorporated into the science databases and pipelines
4. Colors and image coaddition of multifrequency data has been discussed in the literature but not at the level required by the LSST pipeline. Image detection in multiple passbands has the potential to identify sources with a broad range of spectral energy distributions. How these images are coadded (accounting for

- DCR) and whether source detection in color space provides new science drivers need to be researched.
- a. Current studies of imaging data show that there is a strong covariance between the colors of sources. Covariances of the colors (and all measured parameters) must be propagated through to the science database.
  - b. New methods for defining fluxes in a multifrequency environment (e.g. principal components) need to be considered for storing and representing the data.
5. Data products for the LSST
- a. A definition of the data products produced by the static pipeline is required. The need for a “best” sky for the full survey was raised whereby the survey area is covered by data with the optimal seeing conditions. To accomplish this the quality of data in the database, as a function of position of the sky, needs to be fed back into the scheduler.
  - b. The interface between the variability pipeline and static pipeline needs to be defined. The static pipeline will likely measure all of the parameters required for the variability studies that can be accomplished at the catalog level. The need for multiple pipelines or branches and how these are interfaced (at the reduction and database level) was not well defined
6. Compression
- a. Compression as a mechanism for reducing the I/O rate and storage requirements needs to be investigated. Lossy compression has a checkered record in astrophysics (with the concern that it introduces systematic effects into the data). For some science requirements it can, however, reduce the network requirements for transferring data. Identifying these science applications and their ability to tolerate compression is a necessary task.
7. Definition of survey components
- a. It was considered important that the commissioning period of the LSST be defined early on and the tasks that are required to be accomplished during this period be specified. These components can include the definition of the initial calibration, an initial multicolor survey and defining the initial image templates for the coadded and variability pipelines
8. Large objects and interpolation between chips
- a. The definition of what the sky looks like prior to observations was deemed a necessary task. The presence of bright stars or galaxies that cross a chip boundary needs to be taken into account when analyzing a frame. Scattered light from sources near to the telescope beam must be determined and accounted for. Interpolation across interchip boundaries must be demonstrated to work.
9. Data processing requirements
- a. No initial concerns were expressed about the processing requirements for the static pipeline.

### **Section 3. Unresolved issues: Key issues that need resolving in the R&D phase**

1. Differential Chromatic refraction in image coaddition.
2. Optimal stacking algorithms.
3. Photometric and astrometric calibration techniques.
4. Optimal shape extraction.
5. The ability to use metrology as a first order correction for the PSF and geometric distortions across the LSST field.
6. Measuring photometry for faint sources (where no template exists).
7. Deblending of photometric sources.
8. The development of an LSST simulator.

### **Section 4. Team Makeup and Relations**

A number of teams in the US and Europe are currently working on many of the challenges that the LSSST will face. These include Erben, Kaiser, Lupton, Mellier, Monet, and Wittman. Much of the observational data required to develop the algorithms discussed above exists but needs to be put into a common environment for the development teams. Potential data sources include DLS, CFH Legacy, ESSENCE, and SMACHO. These need to be supplemented by an LSSST simulation facility that provides the data at the cadence of the LSST (i.e. previous surveys do not have short exposures). It is clear that we require a collaborative approach with these extant surveys and the new generation of programs that are getting underway (e.g. PanSTARRS, DES, WYIN-ODI). The intellectual resources required to accomplish the R&D path are limited and should be shared amongst the groups. Common test data and development environments, simulators and visualization tools are essential.

### **Section 5. Resources required for near term completion**

To accomplish these goals a requirement is naturally funding. The group felt that the difficulty is not just, however, in raising the funds for these algorithm developments but more in getting the people with the necessary skills onboard and working at the 100% level. Precursor data appear to be either in place or accessible, simulated data needs to be developed (though not initially at the level of a full end-to-end simulator). Access to data from the new generation of telescopes with wavefront sensors to determine how well the geometry and psf can be calibrated was considered important.

### **Section 6. Milestones and R&D Roadmap**

The milestones and task lists required to accomplish the R&D development goals are clearly not well defined. To reach a stage where the construction can begin on a 2-3 year timeframe requires that over the near term we

1. Breakdown of the R&D task from a high level design to individual components
2. Define expertise and developer groups
3. Define science requirements (astrometry, photometry)
4. Compile test data sets (real, simulated) for testing the imaging coaddition, wavefront sensing PSF estimation and photometric and astrometric calibration
5. Start the algorithm development and testing for DCR corrections, PSF estimation, system calibration
6. Prototype implementations on precursor data 12/05
7. Interface development with the data model by 6/06