



LARGE SYNOPTIC SURVEY TELESCOPE

Large Synoptic Survey Telescope (LSST)  
Data Management

# Crowded Field Photometry in LSST Data Release Production

Jim Bosch, Robert Lupton, Colin Slater

DMTN-129

Latest Revision: 2019-11-14

## Abstract

A proposal for how to integrate traditional detect/fit/subtract crowded stellar field processing algorithms with the DRP pipelines.



## Change Record

Version	Date	Description	Owner name
1	2019-10-21	Unreleased.	Jim Bosch

*Document source location:* <https://github.com/lsst-dm/dmtn-129>



# Contents

<b>1 Proposal Overview</b>	<b>1</b>
<b>2 Pipeline Changes</b>	<b>2</b>
2.1 Single-Epoch Processing . . . . .	2
2.2 Joint Calibration . . . . .	3
2.3 Coaddition . . . . .	3
2.4 Coadd Processing . . . . .	4
2.5 Forced Photometry . . . . .	6
<b>3 Compute Resources</b>	<b>7</b>
<b>A Background and Context</b>	<b>8</b>
<b>B References</b>	<b>9</b>
<b>C Acronyms</b>	<b>9</b>

# Crowded Field Photometry in LSST Data Release Production

## 1 Proposal Overview

This document is a proposal to add traditional crowded field processing (iterative detection and subtraction of stars, simultaneous fitting of all or more stars in each frame, and an assumption that there are no galaxies) as a bifurcation of the Data Release Production deep processing pipelines, utilizing the same single-epoch processing and possibly the same coadds, but generating (in general, though not always) distinct Objects and ForcedSources.

This approach differs from the baseline plan described in Appendix A in that it does not rely at all on the applicability or extensibility of a (still-incomplete) high-latitude deblender to stellar fields, though it does depend on many other existing or planned DRP pipeline steps being improved in their ability to handle crowded fields. Overall it is no more ambitious algorithmically than existing state-of-the-art codes, and could in fact leverage some of these codes, *in part*.

We believe this plan is much more viable than “just running” an existing stellar field code on LSST data. That would involve:

- integrating the code with our data access, configuration, and process control middleware so it can be run at scale;
- repeating many pipeline steps that the existing/planned DRP pipelines will also perform;
- configuring, tuning, and debugging such a code to perform well (scientifically and computationally) on LSST data.

The last step must be done regardless, of course, but we do not believe starting from an existing code (in its entirety) will make it substantially easier. And our judgement is that the computational and (especially) development costs of running an existing external code largely as-is are likely to be higher than the costs associated with integrating just the core of an existing code (or writing one from scratch) as described below.

## 2 Pipeline Changes

### 2.1 Single-Epoch Processing

DRP's single-epoch processing steps would be shared between the crowded-field and high-latitude pipeline forks, and hence would need to be enhanced in some ways to support the former. The purpose of these steps is primarily about characterizing the images for later steps, not producing measurements directly, and this will be particularly true for stellar field processing; this proposal does not attempt to produce science-grade Source measurements in crowded fields (just Object and ForcedSource measurements).

These pipeline steps are already required to be robust in crowded fields to support image subtraction, which also depends on them. In the baseline, this robustness could come at the expense of accuracy - because image subtraction does not depend sensitively on background subtraction or PSF estimation, we could probably get away with using simpler models that are easier to fit in crowded fields, though it is also true that doing better on those characterizations may make image subtraction easier. Perhaps more importantly, it may in fact be necessary to do quite well on all of these in order to generate good templates for image subtraction in crowded fields, especially when DCR correction is involved. Still, it is probably true that direct crowded-field processing downstream of these pipeline steps would require them to work better in at least some ways than the baseline effectively requires.

There may be relatively little qualitatively new algorithm development here, because our goal is not to detect and measure all sources in single-epoch images; instead we simply need enough robust measurements to characterize the image. Running the iterative detection/-fitting/subtraction code described in Coadd Processing on single epoch images as well may be necessary, but that should be driven by empirical testing rather than assumed; it is possible that some aspects of image characterization could get easier in crowded fields (especially only moderately crowded fields) because the number of effectively isolated unsaturated stars may actually go up.

After initial robustness improvements to these steps, accuracy improvements would need to be driven by the needs of downstream processing (e.g. residuals from poor subtraction in the detection/fitting/subtraction loop might indicate a need for better PSF models).

## 2.2 Joint Calibration

Joint calibration consists of steps that perform joint fits to single-epoch catalogs from multiple images to derive consistent photometric and astrometric calibration.

As it does not deal directly with images and hence blending, it should not require substantial modification for crowded fields, especially if we rely on it only for calibration purposes. In that case, even if the single-epoch input catalogs have considerably more sources, as long as a subset of those sources can be matched without confusion, the fitting algorithms that form the core of these steps do not need special handling for crowded fields at all.

Complications arise if we decide to use these steps for astrometric measurements (i.e. proper motion and parallax) as well. This would actually be the standard approach to astrometric measurement - LSST's baseline plan to instead fitting models to the postage stamps of individual stars is both more ambitious and less proven - and hence it is an attractive fallback option if the baseline proves unworkable. In this case, we want to perform joint fitting on all stars, not just the subset needed for calibration, and hence processing crowded fields requires the fitting algorithms to scale to much larger numbers of sources. Perhaps more importantly, it would require full-depth, science-grade Source measurements in crowded fields - something we explicitly hoped to avoid.

A good balance might be to not attempt joint-fitting astrometric measurements in crowded fields, or to impose a higher magnitude limit on the sources included, even if we do use this approach to perform astrometric measurements to full (single-epoch) depth in the rest of the survey. We hope be able to obtain proper motions in crowded fields from coadd processing instead, and *may* be able to obtain parallaxes from information derived during forced photometry.

## 2.3 Coaddition

Most of the qualitatively new handling of crowded fields we propose here happens in coadd processing, but we expect this to run on the same suite of coadds used for baseline processing. Specifically, we expect to need a suite of either direct or Kaiser-style coadds in all bands and different non-overlapping epoch ranges (we are envisioning yearly coadds, though the exact timescale we use is TBD). We may also use a slightly different suite of (smoothed) coadds

for deep detection, but not fitting (instead of smoothing direct coadds); optimal detection coadds would be produced as a byproduct of Kaiser coaddition, but can also be produced independently.

Crowded-field processing's does highlight and clarify an existing problem with our baseline plans for coaddition: direct coadds have discontinuous PSFs and noise properties wherever the set of input images changes. Direct coaddition does not attempt to control where these edges land, instead assuming that the fraction of objects that lie on an edge is small (so they can be flagged and thrown away). That assumption is demonstrably false when there are more than a handful of epochs (as will always be the case for LSST), though in that case the severity of any given discontinuity is also reduced. Our baseline plan is to use some combination of multi-epoch fitting and per-object coadds for all final deep measurements, with direct coadds with discontinuities used only for preliminary measurements, but this approach has never acknowledged the existence of blends and includes no way to deal with them.

An alternative is to build coadds – either direct or Kaiser – in small rectangular cells in which we do not include epochs that have an edge within that cell. This ensures that discontinuities only appear (but always appear) at the edges of the cells. This makes deblending and other algorithms that involve convolving some model with the PSF possible (by using different PSFs to evaluate convolved models in different cells), including the kind of simultaneous fitting we will need to do in crowded fields. While it is becoming increasingly likely that we will need to take this approach even in the baseline, but processing crowded stellar fields would make it a certainty.

## 2.4 Coadd Processing

The baseline plan for detection on coadds is to detect on the suite of coadds described in the previous section, and then merge both above-threshold regions and the peaks within them heuristically into a single consistent catalog. This is broadly consistent with the approach taken by traditional crowded field codes, but with the caveat that in crowded field codes this is done iteratively, with simultaneous fitting and subtraction of detected stars at each iteration.

Our proposal is to start this iteration after the initial merge across bands and epoch ranges; the initial above threshold regions (“parent footprints”) would thus be the same as what our baseline high-latitude pipelines would start from. From here, we would simultaneously fit point-source models for all post-merge peak to the full suite of coadds, with flux (probably

assumed constant; maybe marginalized over), position, and proper motion all free parameters. These models (or perhaps just the brightest and/or most securely fit) would then be subtracted from those coadds, and new peaks detected in the residuals (again on the same suite of coadds). This procedure would then iterate (independently over each parent footprint) until no new significant peaks are detected in the residuals. Something like this iteration is the core of all existing crowded-field photometry codes, but each of these has its own heuristics for merging peaks, identifying new peaks, and robust simultaneous fitting. While it is unclear whether we gain from using external code directly for this component – combining PSF model evaluation (which we already have in hand) and an off-the-shelf sparse solver is relatively straightforward – we do expect to at least learn a lot from existing codes in terms of what heuristics are useful.

The relationship between the Objects detected via this approach and the Objects detected via the baseline high-latitude pipeline is something we would like to leave for future research. The worst-case scenario is that these sets of Objects must be considered disjoint – even though some will definitely be derived from the same initial peaks, the presence or absence of neighboring Objects during deblending can completely change the definition of the primary Object. It is more likely that we will be able to identify some – but not all – Objects as the same in both pipelines. One intriguing possibility is that we could actually feed the high-latitude pipelines the complete set of Objects identified by the iterative crowded-field processing; while crowded-field processing would in general “shred” galaxies by treating them as multiple, distinct point sources, it may be easier for a high-latitude pipeline to construct robust blend families that include galaxies by putting these back together than by doing its own iterative detection. Ideally, that would let us feed information on galaxy shredding back to the crowded-field pipelines, at least in the form of flags, but this is beyond the established state of the art in crowded field processing, and thus not something we can guarantee.

While the simultaneous fitting described above produces photometric measurements for every star in the Object catalog, the proposal here is to use the output parameters of these fits only for astrometric measurements. Because the suite of coadds we fit to includes different ranges of epochs (at least in later data releases), we should be sensitive to proper motion as well as absolute position. Parallax is probably impossible to constrain directly from these coadds. Flux parameters fit here are considered preliminary measurements that will be superseded by photometry measurements.



## 2.5 Forced Photometry

Just as in the baseline high-latitude processing, we will use the catalog derived from coadd processing to perform forced photometry on single-epoch images, assuming point-source models for all sources (extended sources with variability are both rare and nearly impossible to model).

There are a few important differences between crowded-field forced photometry and baseline forced photometry, however:

1. Baseline forced photometry will be performed on both single-epoch difference images and direct images, unless we can conclusively demonstrate that one of these approaches outperforms the other. Crowded-field forced photometry will be performed only on direct images.
2. Crowded-field forced photometry will involve a simultaneous fit (with positions fixed and only amplitudes varying) of all objects in each epoch. The degree to which objects will be fit simultaneously in the baseline forced photometry is still unclear; on difference images, most objects can probably be measured independently, but we have not yet identified any approach that can be expected to yield good deblended forced photometry on direct images in the presence of galaxies.
3. We explicitly expect the fluxes produced by crowded-field forced photometry to be aggregated into the steady-state fluxes reported in the Object table, as not relying on coadd-based fluxes gives us more flexibility in how coadds are built and how flux parameters and handled in coadd processing. While it may be useful to aggregate some (direct image only) forced photometry results in the baseline case, baseline coadd processing involves many different measures of photometry, and most of these are intentionally quite different from the point-source measurements used in forced photometry.
4. The input catalogs are in general different – crowded-field forced photometry will be based on the positions derived from the crowded field side of the coadd-processing pipelines, while baseline forced photometry will of course use the positions derived from the the baseline side of those pipelines. While those positions may be the same for some parent footprints (isolated objects, at least), in general the number of children as well as their positions will differ.

Overall, the picture for how we would perform crowded-field forced photometry is substantially clearer than the picture for how we would perform forced photometry in the baseline case, reflecting the fact that a pipeline that acknowledges the presence of galaxies is substantially more complicated.

While it is critical for positions to be held strictly fixed during forced photometry in order to yield the best fluxes, we will also consider either performing separate fitting in which positions can vary per-epoch or evaluating the derivatives of the likelihood with respect to position (at the coadd-derived positions), as this information might provide constraints we could combine to yield improved proper motion estimates and/or parallax measurements. To our knowledge, this has never been demonstrated and hence we cannot assume it will be a workable approach.

### 3 Compute Resources

As processing would bifurcate in coadd processing into distinct baseline high-latitude and crowded-field pipelines, this proposal necessarily involves more compute resources than the baseline alone if both branches are run over the full survey. The additional compute cost will likely be a small fraction of the baseline compute cost, however, and is quite possibly in the noise of our estimates of the baseline compute cost. As a result, the additional compute cost involved in running the crowded-field branch over the full survey could probably be completely recouped by not running the baseline processing in the most crowded parts of the survey.

The same is also true for database storage costs – the set of columns produced by the crowded field pipeline branch will be a tiny fraction of the set of columns produced by the baseline pipeline branch. And while the crowded field branch may indeed produce more rows than the baseline branch (by design!), the area of sky where the difference is greater than even a factor of two will be quite small, so we expect the smaller number of columns to be much more important overall. Taking advantage of that smaller number of columns of course requires the Object table be split into separate tables for crowded-field measurements and baseline measurements (or some more complex normalization).

## A Background and Context

The DM baseline plan for direct-image processing crowded stellar fields has long been to make “best-effort” improvements to detection, deblending, and measurement algorithms designed for relatively high galactic latitude processing (the LSST 18000 deg<sup>2</sup> footprint). This might mean incorporating some aspects of traditional crowded-field processing (cf. DAOPHOT ALL-STAR), such as iterative detection and subtraction, but changes focused on improving performance in crowded stellar fields would be regarded as lower priority than making the processing perform well in uncrowded fields. Our primary plan – never a detailed one – centered around extending an SDSS-style deblending algorithm by increasing the probability that a peak would be treated as a point source in crowded regions.

The lack of a clear mandate/requirements in the “best effort” language has run up against the difficulty of building even a high-latitude pipeline that can handle the amount of (mostly galaxy) blending at LSST depths, making it increasingly unlikely that the baseline would leave DM with any time to focus on crowded stellar field performance during construction. Moreover, while we still regard the extensions we envisioned adding to our pipelines as sound ideas, they would need to build on a solid deblending pipeline that simply does not exist yet, making it impossible to accurately gauge what our performance might be like on stellar fields until extremely late in construction. This is of course true of high-latitude field processing as well, though there we can point to requirements that say we must improve. Perhaps more importantly, it is clear that high-latitude deblending at LSST depths is well beyond the current state of the art, and hence science collaboration members unsatisfied with our processing will probably not be able to point to external code or algorithms that would do substantially better. This is not true for crowded stellar fields, where algorithms have long existed that can process extremely crowded stellar fields much better than either our current algorithms or those we would be likely to have by the end of construction (absent any focus on crowded fields). These codes assume that the presence of background galaxies can be ignored, however, which much will be less true (again, due to depth) for LSST than for any previous major crowded-field survey, so it is also not guaranteed that traditional crowded-field processing would work better on LSST data, especially for regions of intermediate crowding and lower extinction.

LSST does already have a clear mandate to do image differencing in crowded stellar fields, which includes generating alerts in crowded fields. This means that algorithms that feed image differencing are already required to perform robustly in crowded fields. In many cases

(such as PSF and background estimation), image differencing has little or no dependence on the quality of those algorithms, essentially requiring only that they produce not-unreasonable outputs (which may still be nontrivial in extremely dense fields). In others (such as astrometric registration), image differencing may depend critically on how well these steps perform.

## **B References**

### **References**

## **C Acronyms**