The Pan-STARRS PS1 Image Processing Pipeline

The Pan-STARRS PS1 Image Processing Pipeline

Eugene Magnier, Nick Kaiser, Ken Chambers

Institute for Astronomy, University of Hawaii, Woodlawn Dr., Honolulu, HI 96820

ABSTRACT

The Pan-STARRS PS1 Image Processing Pipeline (IPP) performs the image processing and data analysis tasks needed to enable the scientific use of the images obtained by the Pan-STARRS PS1 prototype telescope. The primary goals of the IPP are to process the science images from the Pan-STARRS telescopes and make the results available to other systems within Pan-STARRS. It is also responsible for combining all of the science images in a given filter into a single representation of the non-variable component of the night sky, defined as the "Static Sky". To achieve these goals, the IPP also performs other analysis functions to generate the calibrations needed in the science image processing, and to occasionally use the derived data to generate improved astrometric and photometric reference catalogs. It also provides the infrastructure needed to store the incoming data and the resulting data products.

1. BACKGROUND

The Pan-STARRS Telescope #1 (PS1) on Haleakala is nearing completion, and is expected to start survey operations in early 2007. It is designed to act as a test-bed for the commissioning, testing, and calibration of the Pan-STARRS hardware and software in anticipation of the full Pan-STARRS array. PS1 uses a 1.8m primary mirror to image a roughly 7 square degree field on a 1.4 Gigapixel CCD camera. With 0.26 arcsecond pixels, the images from PS1 will be well-matched to the sub-arcsecond seeing of Haleakala. The large entendue of PS1 will make it the most efficient survey telescope for the near future. The PS1 Survey Mission is expected to last three years, and will produce in the vicinity of 1.8PB of raw image data. The PS1 Survey Mission will consist of multiple survey components, including a very large area survey covering three quarters of the full sky (the 3π survey) and targeted fields observed frequently and repeatedly for both extensive temporal coverage and depth. A major driver for the observing strategies is to efficiently detect solar system objects, including potentially hazardous asteroids. The telescope will use 6 filters, including a wide filter (*w*) for surveys specific to the solar-system, as well as Sloan-like *griz* filters and a *y* filter at the reddest end of the camera sensitivity. This last filter will exploit the high quantum efficiency of the detectors at 1 micron.

The Pan-STARRS PS1 Image Processing Pipeline (IPP) is designed to perform essentially all of the image analysis tasks required to extract the science results from the survey data. Like other large-scale surveys (e.g., the Sloan Digital Sky Survey or the 2 Micron All-Sky Survey), end users will have access to derived data products, not the raw image data stream. The IPP will perform the individual image calibrations, image combinations, and object measurements needed to characterize the astronomical sources detected in the images. During clear, dark nights, the PS1 telescope will produce images at sustained rate of one 3GB image every 45 seconds, for periods as long as 10 hours. The IPP needs to perform the data processing with a high enough throughput to keep up with the raw image data. The resulting data products, and the extensive supporting metadata stream, will be made available to the other components of the project, including the Published Science Products System (PSPS), which will make the data products available to users via a database and sophisticated query mechanisms. Other science clients which will perform additional interpretation of the science data products will also receive subsets of the full IPP output data stream.

2. IPP OVERVIEW

2.1. Scope

The IPP receives data from two Pan-STARRS subsystems: the Camera, from which it receives the large volume of image data, and OTIS (Observatory, Telescope and Infrastructure Subsystem), from which it receives metadata describing the images and the environmental conditions. The location of the IPP computing hardware will evolve over the course of the first year of the PS1 project. Initially, during the start of the commissioning phase, a subset of hardware will be located at the summit facility on Haleakala. The Maui High-Performance Computing Center (MHPCC) is in the process of relocating its computer room. When the high-bandwidth network connection is established from the PS1 site to MHPCC and the new facility is available, the IPP hardware will be relocated to the MHPCC computer room in Kihei.

The users of the IPP output are all systems internal to the Pan-STARRS project. They consist of: 1) The Preferred Science Clients, which receive specified data products on short timescales. 2) The Moving Object Processing System (MOPS), which is one of the Preferred Science Clients, but has the distinction of being a component funded by Pan-STARRS. It will receive the detections of all transient objects. 3) The Published Science Products Subsystem (PSPS), which will receive all data products of interest to the community external to the Pan-STARRS data processing systems, and will act as the long-term archive and publishing clearinghouse.

2.2. Analysis Tasks

The IPP performs several types of data analysis in a regular fashion. The most obvious of these is the science image analysis, from which the measurements of individual astronomical objects are actually derived. In preparation for this critical function, the IPP must also analyze the calibration images needed by the science image analysis. Downstream from the science image analysis, the IPP must perform data calibration on the collection of object detections, yielding improved calibrations and improved reference catalogs for astrometry and photometry.

The IPP science image analysis is separated into two major stages: the analysis of individual images (historically called "Phase 2") and the analysis of groups of images taken of the same portion of the sky ("Phase 4"). This division is illustrated in Fig. 1. The individual images are analyzed independently, as they arrive from the telescope. At this stage, the standard image detrending steps (bias, flat, etc) are performed, as discussed below. Objects are detected in these images, and used to perform astrometric and photometric calibrations. Stars and non-stellar objects are distinguished, but only limited effort is spent at this stage on characterizing the extended sources. Phase 2 ends with calibrated individual images and tables of objects from those images. In the second major stage, the images which correspond to the same portions of the sky are combined. Several image combinations may be performed. Sets of individual science images may be combined into a single, high-quality image which has been cleaned of cosmetic defects. Comparison between this image, or the individual images, and an archival reference image of the same location (the "Static Sky" image) may be performed. Image difference techniques are used to detect the variable, transient, and moving sources. Finally, the new images may be combined with the Static Sky image in order to improve the signal-to-noise. All of the images described above will have object detection and classification performed on them. The objects detected in the summed image stacks are called "Phase 4 Σ", while the image difference detections are called "Phase 4 Δ" detections.

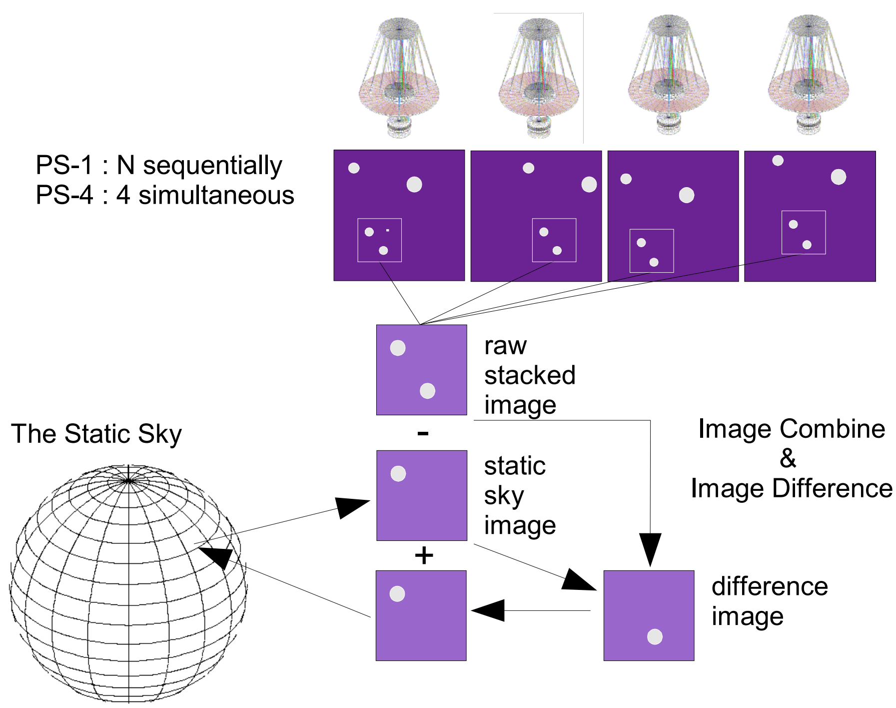

Fig. 1. IPP Image Analysis Stages

Additional science image analysis is performed on the Static Sky images. It is in this stage that detailed analysis of the shapes of extended objects is performed. Since the Static Sky is the most sensitive to the faint galaxies, and since the galaxies do not change with time, this is the logical place to perform this type of analysis. In the Static Sky analysis, the data from all five filters will be analyzed at the same time to improve the signal-to-noise for fainter sources. Common parameters such as the location of the galaxy may be fitted to a single value in all filter images. An important outstanding issue is how to account for the variations in seeing between the multiple images which contribute to the Static Sky images. The measured galaxy shapes require a good estimate of the effective PSF for each object. The baseline IPP code will initially estimate the impact of the effective PSF on the parameters of the observed galaxy shapes. However, a possible extension of the software will allow for the PSF to be convolved with the model *before* fitting to the observed flux distribution. If this analysis may be made sufficiently efficient, or enough computing resources are available, it may be possible to perform this type of PSF-convolved fitting across all input images simultaneously without sacrificing information in the image warping and stacking process.

The IPP is also responsible for generating a high-quality astrometric and photometric reference catalog from the collection of measurements of the astronomical sources. As the images are analyzed, the information about each object is supplied to the IPP object database software called DVO (the Desktop Virtual Observatory). Several programs interact with this database to iteratively improve the calibration of the individual images and to update the astrometric and photometric reference catalog. This analysis can be viewed as a very large least-squares problem, in which the astrometric or photometric parameters of the images are solved for to minimize the residuals for individual objects, while the positions and magnitudes of individual objects are adjusted to minimize the residuals for individual images. This analysis will likely be limited to the objects which have been observed with sufficient signal-to-noise ratios to have a strong detection, but the resulting image parameters can then be used to characterize all objects in the images. As a natural consequence of this analysis, the objects which have significant residuals even after the iterations have run their course can be identified as photometric variables or objects which detectable proper-motion and/or parallax. Images which were obtained until less-than-ideal conditions will also be detected, and potentially excluded from this analysis.

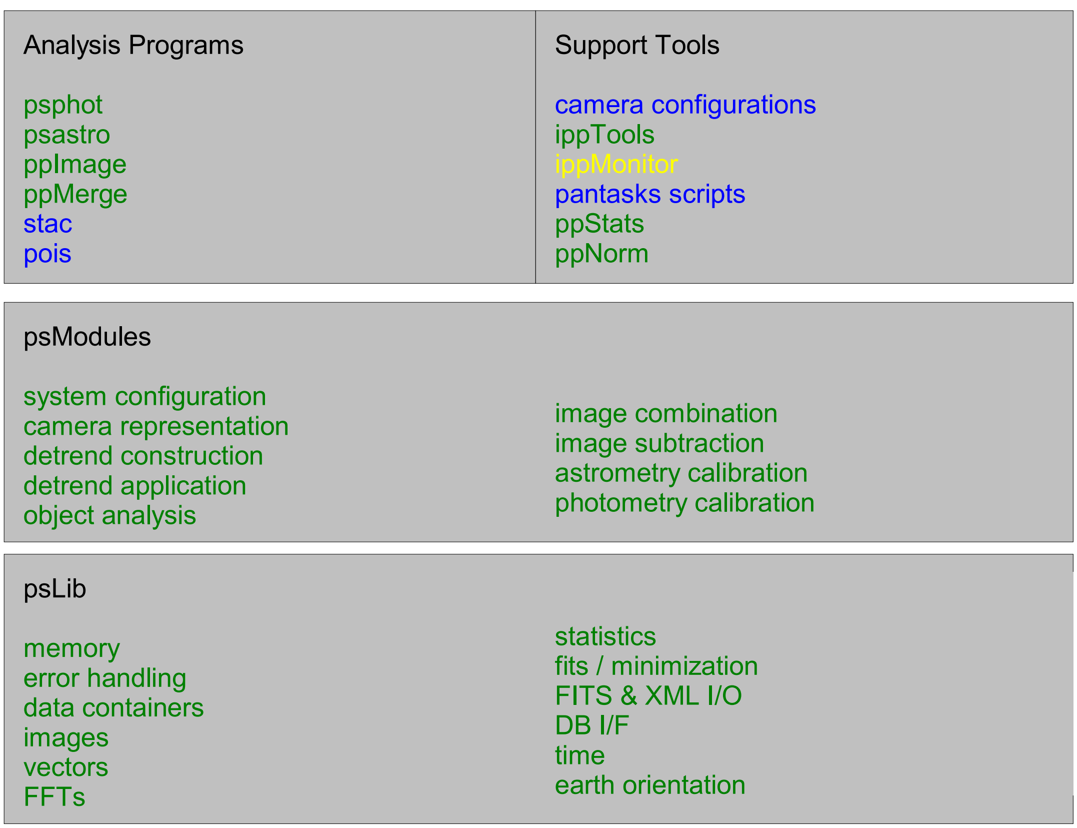

Fig. 2. IPP Architecture

2.4. IPP Software and Hardware Organization

The programs used to perform the image and object analysis stages are designed as stand-alone UNIX programs. These programs may be run manually, if desired. For the bulk of IPP operations, the programs must be automatically launched for specific images when needed. All automatic analysis tasks are tracked within the IPP Metadata Database, which also records the Q/A statistics produced by the analysis steps. The operation of the IPP is managed by a program called "PanTasks" which schedules specific analysis tasks based on the current state of the Metadata Database.

The IPP image analysis tasks are well suited to parallel processing. Not only are individual images processed in Phase 2 independently, but most of the computational effort for each of the chips of the mosaic camera may be processed without reference to the other chips. The analysis of different patches of the sky ('sky cells') in Phase 4 may be performed independently as well. PanTasks uses the associated program, "pcontrol" to distribute analysis tasks across a computer cluster. The IPP can take advantage of the natural parallelization units (chips and sky cells) to make interesting optimizations for the parallel processing. For example, each chip may be assigned to a specific computer, allowing all of the Phase 2 operations on that chip, as well as all calibration image analysis for that chip, to be performed on the same computer used to store the data. This type of optimization minimizes the load on the network switch since most data access is to a local disk array.

In order to facilitate testing and development, and to encourage flexibility, the IPP is built in a layered fashion. The lowest level functions define basic data types, statistical analysis, and I/O operations. These are written in C and collected together into a library called "psLib". The next layer consists of functional elements which are more specifically related to astronomical data analysis, such as detecting sources in an image or performing a bias subtraction. These are built from the psLib elements, and are also collected into a library called "psModules". Making use of these two libraries are the top-level analysis programs which are run on the cluster computers. There are also a collection of support functions which interface with the Metadata Database and provide isolation of the database design from the analysis programs. Fig. 2. illustrates the relationship between these three layers of the IPP Software Architecture.

The major analysis programs defined by the IPP are:

- **ppImage:** responsible for all basic single-image analysis, including detrending and invoking the photometry and astrometry analysis.
- **ppMerge**: responsible for all detrend image stacking operations, including combining raw fringe images
- **psPhot**: PSF modeling, object detection and classification. The functionality of this program is also available as a library module for use by ppImage.
- **psAstro**: Astrometric calibration, including retrieval of reference sources from available databases, and simultaneous mosaic solutions. The functionality of this program is also available as a library module for use by ppImage.
- **Stac**: image warping and combinations. This program is a key element of the Phase 4 analysis.
- **POISub**: image difference analysis, including PSF-kernel matching using the Alard-Lupton method as an option.

2.5. IPP Milestones

Over the next 6 months, the IPP will pass several major software milestones. At these times, major portions of the IPP functionality will be completed. We list below the planned milestones and the corresponding IPP functionality.

IPP Release 1 (Sep 22, 2006) : Phase 1-3, Detrend Creation, Infrastructure, ISP Support

Release Contents: Analysis programs for Phases 1-3 and the Detrend Creation analysis, PanTasks (the controller/scheduler), PanTasks scripts for Phase 1-3 and Detrend Creation, the Metadata Database tools (ippTools) needed to track the process flow, Nebulous, and DVO.

Release Capabilities: Detrend image creation (bias, dark, flat, fringe), image detrending (excluding convolved guide kernels), single-image positive object detection and classification, astrometric calibration, object detection databasing, basic zero-point analysis. Real-time ISP transparency measurements.

IPP Stage 2 (Oct 30, 2006) : Phase 4, static sky pixel definitions

Release Contents: Phase 4 analysis programs, Phase 4 PanTasks scripts, Static Sky pixel layout, interface to Magic, psphot and ppImage upgrades.

Release Capabilities: Image warping, Image differencing, Image stacking, object modeling for difference images (loading PSF from an external source, fitting positive and negative sources), analysis of the OTA guide kernel, detrend images convolved with the guiding kernel.

IPP Stage 3 (Dec 30, 2006) : Static Sky, AP Catalog tools, output interfaces

Release Contents: Updates to psphot; DVO crawler tools (uniphot, relphot, relastro); external interfaces.

Release Capabilities: Static sky version of the photometry analysis, with more-complete characterization of galaxy parameters and the simultaneous multi-filter photometry; tools to perform the astrometric and photometric calibration of the all-sky survey data; the interfaces to provide data to the MOPS, PSPS, and other science clients as defined.

IPP Stage 4 (Mar 30, 2007) : Code Freeze for Operational Readiness Review and the Survey Start

Release Contents: Updates to programs identified in the commissioning.

Release Capabilities: Improvements determined on the basis of the commissioning with GPC-1 (possibly new PSPhot PSF models, temperature dependent selects of input detrend data as needed, etc).

3. SOFTWARE DEMONSTRATIONS

3.1. PSPhot

In the following section, we illustrate the current status of one of the major IPP components, PSPhot. PSPhot models the image PSF using any of a number of possible analytical models. The parameters of the models are allowed to vary by position in the image to account for variations due to the optics and seeing. PSPhot uses the PSF model to distinguish stellar and non-stellar objects, assigning a probability for each object that the shape is consistent with a PSF. PSPhot measures photometry of objects using the PSF model for stellar sources, and any of several

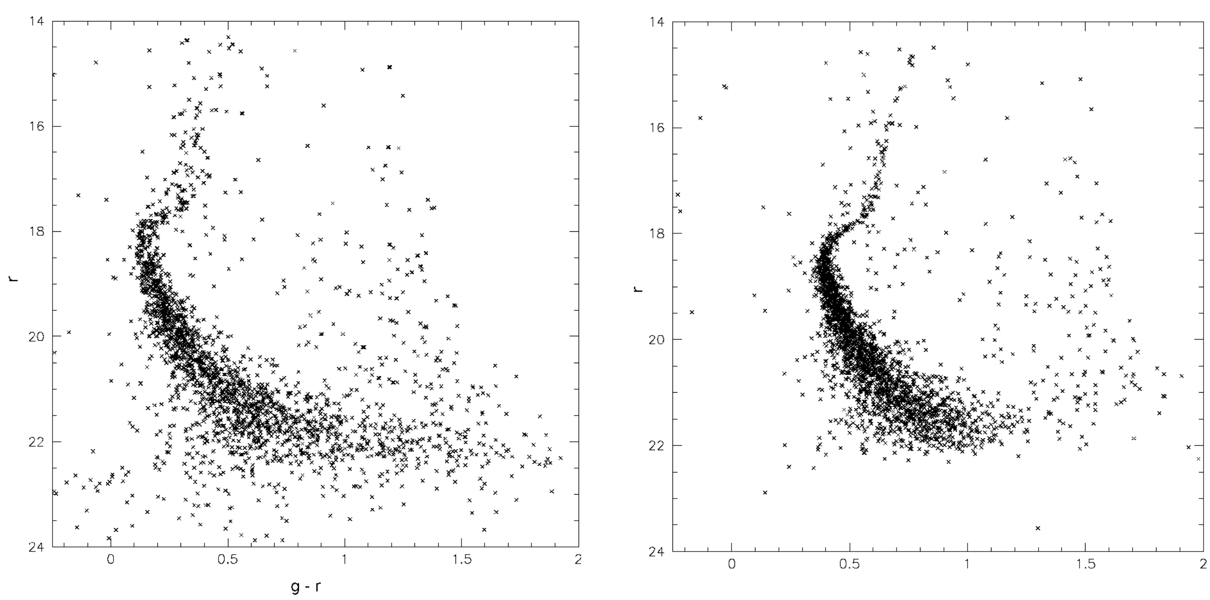

Fig. 3. PSPhot and SDSS Photo Comparison.

Left: M13 photometry with Photo. Right: M13 photometry with PSPhot possible alternate models for extended sources. It also measures an aperture correction for each image, and uses the curve of growth of the PSF model to scale all photometry to a common-sized aperture.

PSPhot is designed to perform well in crowded fields. It has the ability to detect potentially blended sources and perform a simultaneous fit to a collection of nearby sources. It also uses a two stage detection process, in which a quick, linear subtraction of the PSF model, with position fixed by the object centroids, is used to make the detailed fitting process more robust.

Recent tests by the team led by R.H. Lupton have show the efficacy of these approaches. These researchers have used SDSS images of the Galactic plane region in the vicinity of M13 to compare the behavior of PSPhot with the Photo routine used by SDSS (Lupton REF) and the well-used program DAOphot (Stetson REF). Photo was designed to perform well at high-galactic latitude and to measure galaxies well, since SDSS is primarily an extragalactic survey concentrated on the northern Galactic polar cap. However, some observations called the "Transition Strips" have been obtained across the Galactic plane, in order to connect the northern and southern Galactic cap regions. The SDSS group led by R. Lupton intend to use PSPhot for these high-density regions. We illustrate the ability of PSPhot to perform crowded field photometry in regions where Photo performs poorly using preliminary results kindly provided by R.H. Lupton.

Fig. 3. shows a pair of color-magnitude diagrams of the region including the cluster M13. The CMD on the left was measured with the SDSS routine Photo, while the one on the right used PSPhot. The giant branch is clearly more ragged and ill-defined in the Photo version of the analysis. Comparisons with DAOphot measurements of the same images show an excellent correspondence, with a slight improvement in the DAOphot photometry over the PSPhot photometry. Lupton and his team are exploring the PSPhot parameters and code details to improve even further the analysis of such crowded fields.

3.3. ISP Image Analysis

An important component of the Pan-STARRS PS1 Telescope is the atmospheric transparency monitor called the Imaging Sky Probe. This device uses a 120 mm lens to image a roughly 3 degree field of view, comparable to that

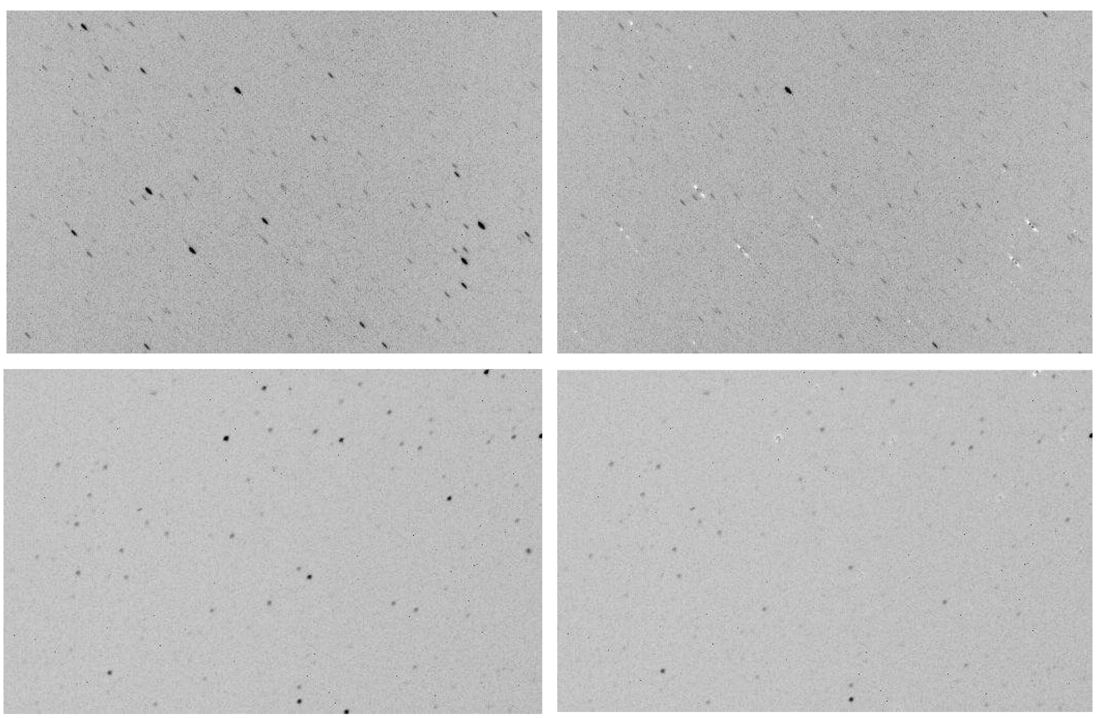

Fig. 4. PSPhot PSF modelling of a *g*-band ISP image. Left column: input images. Right column: PSF-subtracted images. Top row: data from the upper-right hand corner of the camera. Bottom row: data from the lower-left hand corner of the camera. Note the substantial image PSF variations, which PSPhot models well, yielding only minimal residuals. The remaining objects were below the subtraction theshold for this analysis.

of the Gigapixel camera of PS1. The field is imaged onto a small (2k square) commercial detector. The goal of the system is to use frequent observations of relatively bright stars to measure the transparency of the atmosphere. The photometry is initially tied to the Tycho and 2MASS photometry of the stars, but is later improved by determining an internal magnitude system for the stars in the appropriate magnitude range. PSPhot will be used to perform the photometric measurement of objects in the images. One interesting issue with the ISP images is that there is substantial image shape variations within individual images in the *g*-band. PHPhot must successfully model the PSF variations to perform an accurate measurement. Fig. 4. illustrates the ability of PSPhot to perform such an analysis. In this figure, we present an example of two regions of a single ISP *g*-band image, showing the clear variations in the PSF from one corner of the chip to the other. We also show the image after the PSF model has been measured and subtracted by PSPhot. The small residuals show that the PSF variations are being captured in the PSF model. Remaining objects in the subtraction are mostly below the subtraction threshold; there are also some objects for which the PSF model was a poor fit and which were skipped.

4. SUMMARY

The Pan-STARRS PS1 Telescope is nearing operational readiness. The Pan-STARRS PS1 Image Processing Pipeline (IPP) is also nearing operational readiness, and will be completed before the survey operations start. The IPP is designed to be flexible, robust, able to achieve the stringing scientific accuracy requirements for the system, and capable of keeping up with the very high data rate. We have discussed elements of the software infrastructure and the hardware architecture. More complete documentation is available on the Pan-STARRS web site at http://pan-starrs.ifa.hawaii.edu. Testing of portions of the software are underway, and we have illustrated specific examples of the behavior of one of the components, PSPhot.

5. ACKNOWLEDGEMENTS

The design and construction of the Panoramic Survey Telescope and Rapid Response System by the University of Hawaii Institute for Astronomy is funded by the United States Air Force Research Laboratory (AFRL, Albuquerque, NM) through grant number F29601-02-1-0268.

6. REFERENCES

1. Lupton Photo REF

2. Stetson DAOphot REF