Astronomical Data Analysis Software and Systems XXIII
ASP Conference Series, Vol. 485
N. Manset and P. Forshay, eds.

Development of a Database System for Data Obtained by Hyper Suprime-Cam on the Subaru Telescope

Y. Yamada,[1] T. Takata,[1] H. Furusawa,[1] Y. Okura,[1] M. Koike,[1] H. Yamanoi,[1] N. Yasuda,[2] S. Bickerton,[2] N. Katayama,[3] S. Mineo,[2] R. Lupton,[3] J. Bosch,[2] C. Loomis,[3] H. Miyatake,[3] P. Price,[2] K. Smith,[3] D. Lang[2,3]

[1]*National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan*

[2]*Kavli IPMU, Kashiwa 277-8583, Japan*

[3]*Princeton University, Princeton, NJ 08544, USA*

Abstract. We are developing a database system for the Hyper Suprime-Cam (HSC) data on the Subaru Telescope in preparation for the HSC Survey. Since HSC has a huge field of view (1.5 degree diameter), it will produce a huge amount of data. Here, we make a brief report on the prototype of our database.

1. Introduction

Hyper Suprime-Cam (HSC) is an optical and near-infrared (0.4-1.1 μm) wide-field camera for the prime focus of the Subaru Telescope, located on the summit of Mauna Kea in Hawaii. The combination of its huge field of view (1.5 degree diameter) and the large mirror (8.2 m) are unprecedented, and make it very efficient at wide-field surveys of the faint astronomical sky. HSC has 104 science CCDs, over an order of magnitude increase compared to Suprime-Cam (the current wide-field imager on Subaru, with a 34'$\times$27' field of view), with a corresponding increase in the raw data rate. Our aim is to store this data efficiently and supply them in a useful way for the astronomical community.

We are currently developing a database system to manage images obtained by HSC and catalogs of detected source on processed images. These catalogs include not only detected sources in stacked images, but also those in each frame (CCD), for the purpose of detecting variable objects, weak lensing measurements, etc. Moreover, we have been granted time as part of a Strategic Survey Program to survey $\sim$ 1400 square degrees with multiple colors over the next 5 years. We expect the number of rows in the object table to exceed 5×10^9. Here, we describe the HSC database, including the structure of tables, the data flow, the table partitioning plan for dealing with such large data, and efficient functions for querying the database.

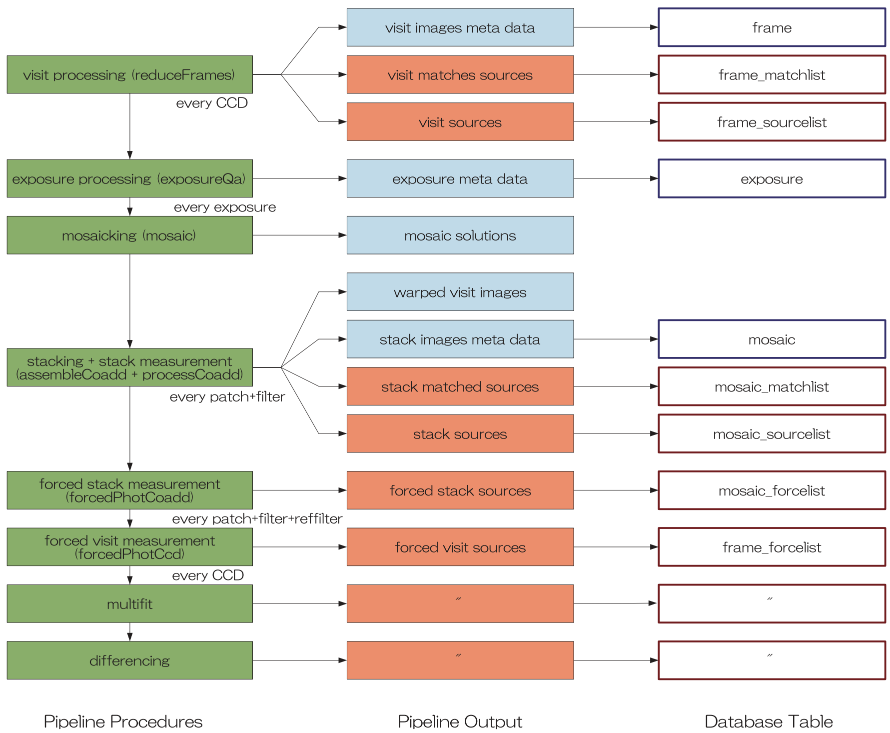

Figure 1. Flow chart for the HSC data pipeline showing the output and database tables.

2. Data Flow

The HSC images are analyzed by a pipeline developed as a partnership between the National Astronomical Observatory of Japan (NAOJ), Kavli IPMU and Princeton University. The whole process are roughly divided into 3 parts:

1. The pipeline produces reduced images and catalogs from the raw frame data;
2. A mosaicking and stacking process generates "skytile" images and catalogs;
3. For the objects detected on each skytile image, the pipeline performs forced photometry in the stacked skytile images in all bands and the reduced (unwarped) images.

3. Data Volume

In Table 1 we show the expected data volume for our catalogs. We will produce catalogs of objects detected on individual frames, whereas most previous surveys produce catalogs only for objects of stacked images. We expect the final catalog will contain in excess of 5×10^9 objects.

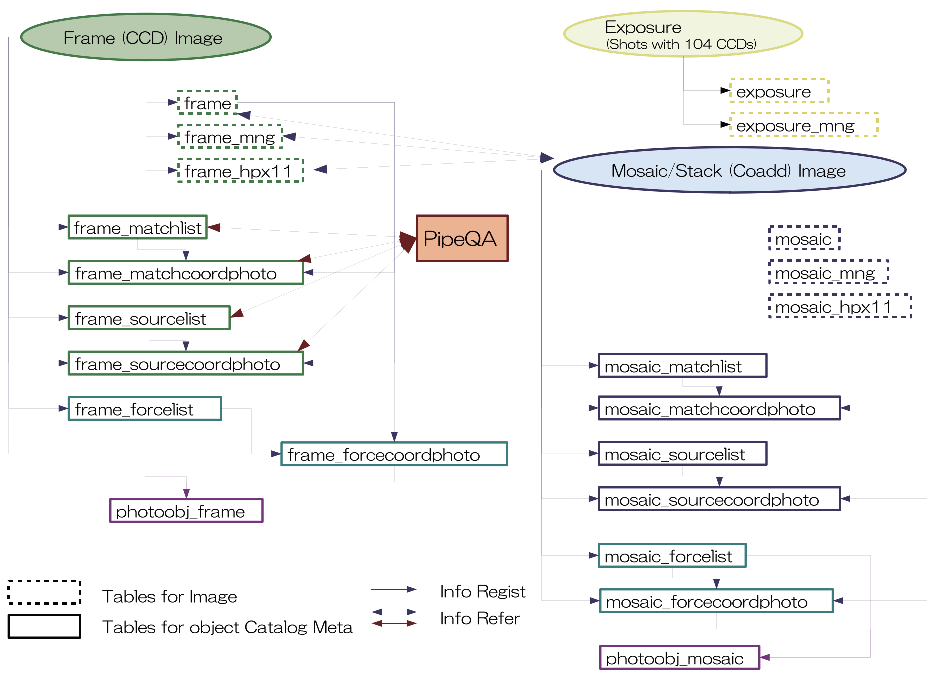

Figure 2. Relation of our database tables. Boxes of broken and solid line indicate tables for management and object catalogs, respectively.

4. Database

Here we outline the design of our database.

4.1. Tables

We show in Figure 2 the relation between our database tables. Tables named `*list` are for the metadata of pipeline output fits file catalogs, while tables named `*coordphoto` are for the calculated sky coordinates and magnitudes. The sky coordinates are calculated from the pixel coordinates in the `*list` tables and WCS information from the pipeline. The magnitudes are calculated from the fluxes in `*list` tables and the magnitude zero point in the management table. Since the catalog tables are so large, we have plan to divide catalog tables into groups of skytiles (known as "tracts").

Table 1. Data Volume for HSC Survey

Surveys	Area	Num. of CCD Image	Amount (GB)	CCD Sources	Num. of Stack Images	Amount (GB)	Stack Sources
Wide	2000 sq. deg	2350400	202134	5184000k	2000×5filters	10166	800000k
Deep	4 pointings	247936	21300	540000k	18×8filters	1453	15000k
Ultra-Deep	2 pointings	305344	26300	360000k	2×11filters	223	3000k
Total	-	2903680	249734	6084000k	10166	46676	818000k

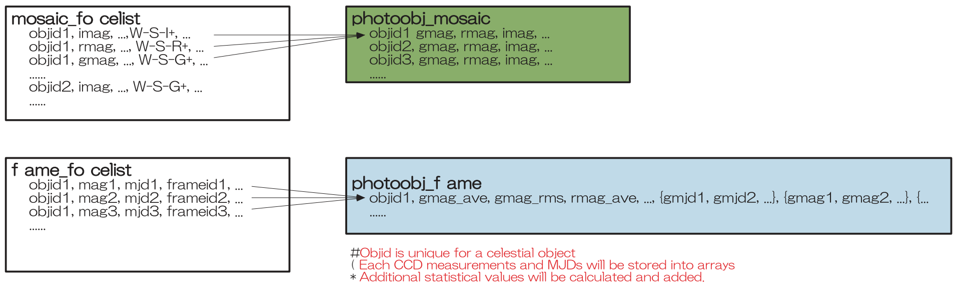

Figure 3. Production of the `photoobj` table. The `objid` is a unique identifier for each celestial object.

4.2. Photoobj Table

Tables `Photoobj_frame` and `photoobj_mosaic` gather multiple band forced photometry measurements into a single record for each object (Figure 3). These tables are similar to SDSS system. For `photoobj_frame`, the frame measurements and MJDs for one object are stored in arrays. Additional statistical values (mean and standard deviation) will also be calculated and included.

4.3. Stored Functions

Stored functions in the database are important for constructing efficient queries. Currently we have installed functions for calculation of statistical values (quantile, median, etc.), for the HealPix index system (Górski et al. 2005) and for spatial searching with a cone (radial) or rectangle. These functions are implemented in the C and C++ languages.

For the spatial searches, the catalog tables (such as `frame_sourcelist`, etc.) have columns `cx, cy, cz`, containing Cartesian coordinates. First, a rough selection is made of objects within $2R$ ($+-R$ for `cx, cy, cz` of the search center) and the distances between the search center and these objects are calculated. Finally, objects with distances smaller than the specified searching radius are selected.

4.4. Summary

We have described the structures of our database for the Hyper Suprime-Cam pipeline. We still face many challenges, not only in operating the data management/distribution but also in catalog searches due to the large amount of data to be produced by HSC. Future work will involve using better database indices, partions, stored functions and table design. Our goal is to present an efficient system to the scientist users.

Acknowledgments. This work is supported by the Subaru Telescope project. The Hyper Suprime-Cam project is also supported by Grant-in-Aid for Scientific Research of Japan Society for the Promotion of Science (JSPS).

References

Górski, K. M., et al. 2005, ApJ, 622, 759. `astro-ph/0409513`