Pipeline Processing of Infrared-Array-Camera Images from the Space Infrared Telescope Facility (SIRTF)

Russ Laher[1] and Mehrdad Moshir[2]

SIRTF Science Center, MS 100-22, California Institute of Technology, Pasadena, CA 91125

To be presented at the *Workshop on Computational Optics and Imaging for Space Applications*, NASA/Goddard Space Flight Center, May 10-12, 2000.

Abstract. The Space Infrared Telescope Facility (SIRTF), the fourth and final element of NASA's Great Observatory program, is scheduled for launch into an Earth-trailing solar orbit in December 2001. Its Infrared Array Camera (IRAC) will provide 5.12×5.12 arcminute images of the celestial sky in four infrared bands centered at 3.6, 4.5, 5.8 and 8.0 μm simultaneously. Two InSb and two Si:As focal-plane-array (FPA) systems, each with 256×256 pixels, are used for raw image acquisition in the two shortest and two longest wavelength bands, respectively. The pixels are read out in four multiplexed channels, opening up the possibility of four separate bias drifts in the image data. Prior to distribution of the data to the relevant observers, the raw images will undergo several stages of automated processing at Caltech's SIRTF Science Center (SSC) to remove instrument artifacts and transform them into basic calibrated data (BCD) products. The image-data processing will be done for each band independently, and will include the following steps: 1) transformation of InSb data into the positive "sense"; 2) conversion of the integer image data to real numbers; 3) truncation correction; 4) detection and correction of wrapped-around negative data; 5) barrel-shift and Fowler-sampling number normalization; 6) electronic bandwidth correction; 7) latent-image detection; 8) dark-current subtraction; 9) dark-current channel-offset normalization; 10) linearity correction; 11) non-uniformity correction; 12) cosmic-ray/radiation-hit detection; 13) engineering-to-astronomical units conversion; and 14) quality-assurance characterization. The "science-data processing thread" will require several calibration products generated by at least five calibration threads of the pipeline. There are separate calibration threads for estimating the dark current, detector linearity, and image non-uniformity (field-flatness). There is also an ancillary thread that uses a Kalman filter for noise-mitigated estimates of the image non-uniformity measurements in time. A fifth calibration thread will provide gain and read-noise estimates for the images on a pixel-by-pixel basis. The scale factor that is required for the final step of converting data numbers (DN) into absolute flux densities will be determined by non-automated analysis. A "calibration server" will determine the latest and/or most-suitable calibration products to use in pipeline-reduction of a given data set. The pipeline design calls for modular software elements written in the UNIX-style of command-line inputs and outputs, with namelist capability for parameters that change infrequently. Higher-level scripts written in either Perl or C-shell will chain the relevant software elements into the various processing threads. Other scripts running under an executive (OPUS) will manage the automated pipeline processing with little operator intervention. All raw and processed images will be stored in FITS (Flexible Image Transport System) format, and will be archived at the SSC. Both raw and processed images and intermediate data products, as well as the calibration data used in the processing, will be made available to the appropriate SIRTF observers. Following a data-validation period, all SIRTF data will be made publicly available.

[1] SIRTF IRAC Data-Processing Cognizant Engineer (laher@ipac.caltech.edu)
[2] SIRTF Downlink Data-Processing Task Lead (mmm@ipac.caltech.edu)

1.0 INTRODUCTION

As the scheduled launch date of December 2001 for the Space Infrared Telescope Facility[1] (SIRTF) fast approaches, the level of activity associated with SIRTF data-processing software design and development is ramping up at the SIRTF Science Center (SSC). In particular, several key software elements for processing data from SIRTF's onboard Infrared Array Camera (IRAC) have been implemented and unit-tested in the past year. Furthermore, the architectures of the various processing chains, or threads, needed to estimate the calibration data and reduce raw images, or Data-Collection Events (DCEs), into Basic Calibrated Data (BCD) have matured. High-level thread scripts have been written to direct the data flow through the software elements of the threads. Integrated software testing using both instrument test data and simulated data is now under way.

The SIRTF data flows along a circuitous route before finally reaching the SSC. Starting at the spacecraft, the data will be transmitted via telemetry to a ground receiving station where it will be sent to Jet Propulsion Laboratory (JPL) for processing by the Flight Operations System (FOS). The data will be separated into FITS (Flexible Image Transport System) images, housekeeping, pointing, and spacecraft-engineering data. The FITS image format is preferred by astronomers, who are the ultimate consumers of SIRTF data. The data will then be shipped to the SSC for processing into a calibrated form, that is, the BCD, which will be readily useable by the astronomy science community. The BCDs will be archived at the SSC, and will be eventually made available to the public.

This paper describes the architectures of the SSC's IRAC-data processing threads and the algorithms of its associated software elements that have been or are planned to be implemented in support of this aspect of the SIRTF mission.

2.0 IRAC DATA-PROCESSING PIPELINE

The pipeline threads described in this section process one or more raw FITS images of the same infrared (IR) band at a time. The production thread takes a single raw image and generates a calibrated image. A major requirement is that the processing of data for a given IR band must be independent of the other bands, even though the fields of view of the 3.6 and 5.8 μm bands, and the 4.5 and 8.0 μm bands are directly overlapping. The calibration threads typically process several images of a given type for good statistical results. The processing is done for the data associated with each pixel in the image.

In its most basic form, a FITS image of IRAC data consists of an ASCII-text header and a binary image-data portion. The header contains useful ancillary information about the image, such as where the telescope was pointing when the image was acquired and to which IR band the image data correspond. The binary portion consists of a single plane of 256×256 pixels of 16-bit unsigned integer data. Both FITS header information and the image data itself are used in the pipeline processing.

2.1 Production Thread

The production thread is used for routine processing of science data. Other data taken for deriving calibration constants are processed by the calibration threads described in the next section. The calibration threads must be executed at least once prior to production thread operation, in order to generate the needed calibration quantities. Figure 1 depicts the data flow through the production thread. Generally, each box in the processing chain of Figure 1 represents a stand-alone software program, written in either C or FORTRAN, which is suitable for testing separately as a unit.

The first or TRANHEAD step translates the numerous FOS-encoded keywords in the FITS-file header, such as “A0123D00”, into human readable keywords, such as “INSTRUME”. It also averages groups of similar keywords to reduce the size of the header where possible, and derives quantities needed for the processing, such as the date/time of image acquisition, image exposure time, etc., which are then written to the output FITS header.

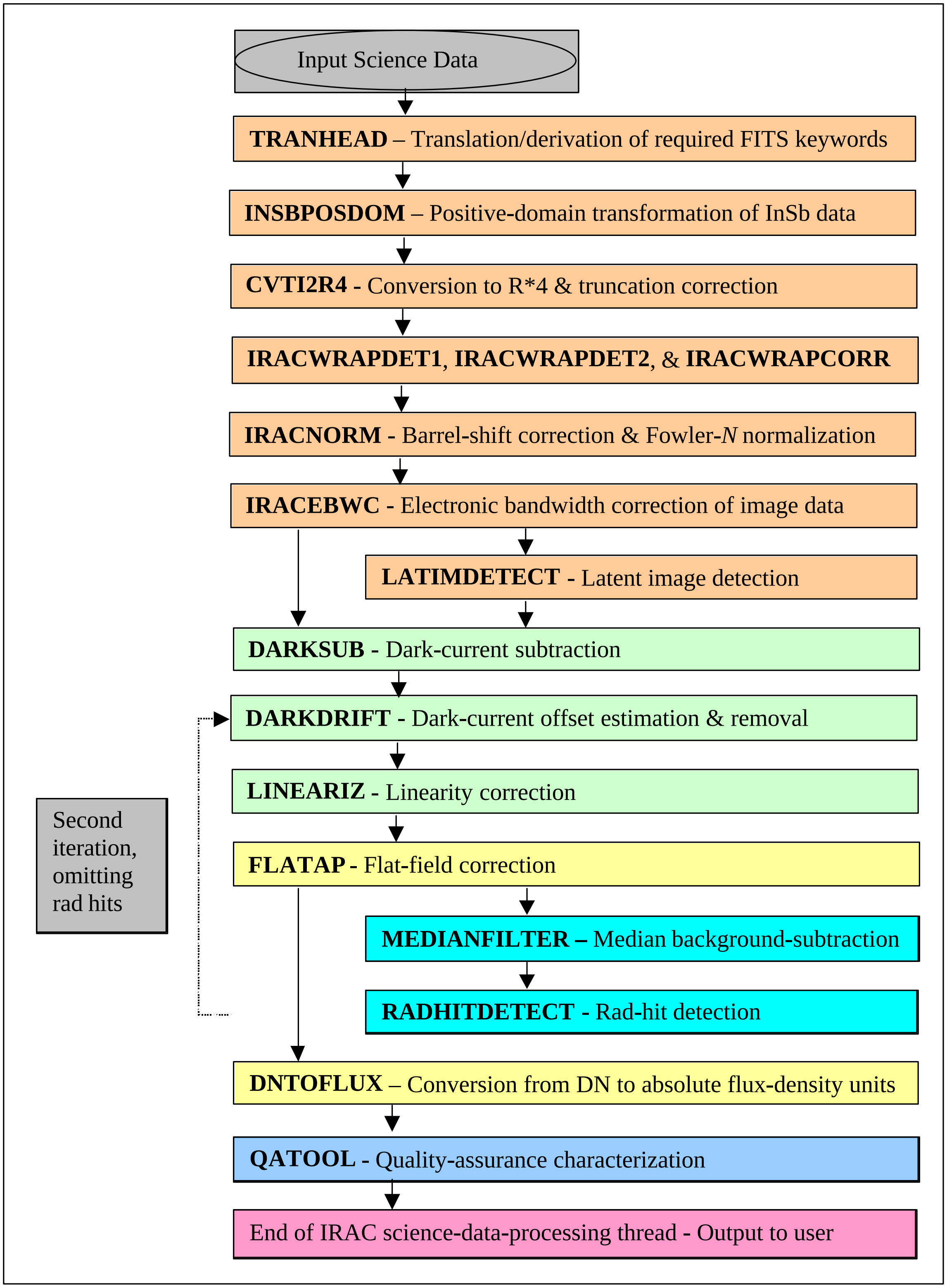

Figure 1. IRAC Science-Data-Processing Thread

The INSBPOSDOM step is executed only for band-1 (3.6 μm) and band-2 (4.5 μm) IRAC data. This is due to the conditioning of the data by readout electronics for these bands, which cause the data to decrease with increasing irradiance on the associated InSb focal plane arrays (FPAs). The purpose of this step is to transform the data so that it increases with increasing irradiance like the Si:As data from the two longer wavelength bands.

The CVTI2R4 step converts the 16-bit unsigned integer data into 32-bit floating data. It also optionally removes the bias in the data caused by truncation of the data's precision, which amounts to adding 0.5 DN (data numbers) to the data if it was right-bit-shifted on the spacecraft. This condition is determined by examination of the appropriate FITS keyword by the pipeline script, which then loads either 0.0 or 0.5 into CVTI2R4's command line.

The IRACWRAPDET1, IRACWRAPDET2, and IRACWRAPCORR steps transform data values in the high end of the unsigned 16-bit range, which has been reserved for the negative data values expected due to noise or bias drift, into small negative values. Generally, values greater than M, nominally 50000, will be "wrapped around" in the 2-s complement sense by this software. The four multiplexed readout channels for each IR band are handled with separate M-value settings. IRACWRAPDET2 is currently just a stub, but it will eventually employ a more sophisticated algorithm than the simple thresholding used by IRACWRAPDET1, in order to distinguish radiation hits from wrapped-around values.

The IRACNORM step normalizes the data. It divides the data by the number of Fowler samples that were accumulated onboard the spacecraft during the data integration period. It also multiplies the data by 2^N, where N is the number of bits the data were right-shifted prior to transmission to the ground.

The IRACEBWC step performs electronic bandwidth correction, which is basically a time-domain deconvolution to remove the relatively slower response of the readout electronics from the underlying signal. This correction has been characterized separately on the ground for each of the four IR bands.

The LATIMDETECT step thresholds the data to determine the locations of bright pixels that may cause ghosting, or latent images, in subsequent data frames. A mask image of bright pixel locations is produced to facilitate image analysis and mitigation of this effect. This software element's companion reporting process, which combines LATIMDETECT results for relevant prior data frames into a latent-image map for the current frame, will be implemented within the next couple of months.

The DARKSUB step subtracts from the data the dark-current, which is determined via the DARKCAL calibration thread (see next section).

The DARKDRIFT step corrects the data for drifts in readout-channel biases since the last dark-current calibration, which can occur independently in each of the four multiplexed readout channels. DARKDRIFT has flexible operating modes, which allow the drift corrections either to be interpolated from "local" dark-current images (taken near the time of science-data acquisition), or estimated from the science data itself. The software also has in place a stub for temperature compensation of the drift corrections, which allows for variations in temperature between times of acquisition of the dark-current and science data. If deemed necessary, the IRAC instrument team will provide the correct temperature-compensation transfer function for this software.

The LINEARIZ step converts the data from observed values into linear values. That is, it linearizes the data over its entire dynamic range according to the linear portion of the response of each pixel, which tends to go nonlinear at high DN values. The linearity correction coefficients used by LINEARIZ are determined via the LINCAL calibration thread (see next section).

The FLATAP step performs the nonuniformity correction (also known as field flattening) to the data. This step requires a "flat" from prior execution of the FLATFIELD calibration thread (see subsection 2.2.3). The flattening is done by dividing the input image by the flat on a pixel-by-pixel basis. If the flat was determined from data acquired by flooding the FPA with the transmission-calibrator (TC) lamp (with the shutter closed), then an additional correction to make the TC illumination uniform is required. This is made by comparing a flat determined from sky data (“sky flat”) and a flat determined from TC data (“TC flat”) taken at nearly the same time.

The MEDIANFILTER step produces a local-median-filtered image using a nominal, 5×5-pixel, sliding input window, where the local median value is computed at a given pixel by centering the window on that pixel and using data values in the window. The local-median image is subtracted from the original image to produce a background-subtracted image. Note that this is an ancillary product used only in the next step (RADHITDETECT); the final output of this thread is a BCD that includes the celestial background.

The RADHITDETECT step detects radiation (cosmic-ray) hits in the background-subtracted image by application of an artificial neural network (ANN). The ANN is a standard hidden-layer type with sigmoidal nonlinearities, a 5×5-pixel input grid, 20 hidden nodes, and one output node[2]. The ANN output is a value in the range [0, 1], and corresponds to the posterior probability of rad-hit presence. The ANN weights and biases are computed via off-line backpropagation training on ten workstations in parallel, in order to speed-up the training to three days per IR band. Data acquired during in-orbit check-out (IOC) will be used for ANN training and testing. Simpler filters, which have adequate but lower performance, will be used in place of RADHITDETECT until the ANN weights and biases are available. Testing at the SSC using simulation and cyclotron data has shown that this method of rad-hit detection has a false-detection probability of 4×10^{-4} at a threshold setting corresponding to a probability of true detection of 0.80. This performance is five times better than that of the conventional local-median spike filter. The products generated by RADHITDETECT include an image of rad-hit-presence probabilities, and a table of locations and strengths of rad-hit presence probabilities that exceed the rad-hit detection threshold.

The DNTOFLUX step converts the data values in DN units to some absolute units that have yet to be determined, such as Janskys (=10^{-26} W/m^2/sec). The conversion factor will be determined from analysis of suitable IOC data by an astronomer specializing in photometric calibration. Optionally, this software generates an additional BCD image with Not-a-Number (NaN) markers at the locations of bad pixels identified prior to and during the processing.

The QATOOL step performs various statistical calculations on the output BCD image for the purpose of providing a quantitative measurement of the data quality. Among the many statistics computed are the median, trimmed mean and standard deviation, and datascale[3], which is a very robust estimator of distribution width that is relatively insensitive to outliers. The statistics are computed over the entire image, separately over the image quadrants, separately over the four readout channels, and also over a user-definable box in the image.

Although a final decision has yet to be made, there is the possibility that a second iteration through some of the software elements in the production thread thread will be done with the detected rad-hits excluded, in order to improve the final BCD product.

2.2 Calibration Threads

The calibration processes described in this section will produce an output mask image to indicate the problematic pixels of various types that may be encountered, in addition to their regular calibration products.

2.2.1 Dark-Current Estimation

The dark-current-estimation thread includes pre-processing by the orange blocks in Figure 1, followed by the DARKCAL process, which robustly averages a number of dark images (acquired with the shutter closed and stimulator lamps off) to produce an estimate of the dark current for a given IR band and integration time. Included in this processing is an outlier-rejection scheme, and uncertainty estimation. Prior to the averaging at each pixel across input images, the bias is computed for each readout channel separately and subtracted off. This yields an estimate of the dark-current on relatively long time scales, which is used in subsequent processing by the DARKSUB step of the pipeline.

2.2.2 Linearity Estimation

The linearity-estimation thread includes pre-processing by the orange blocks in Figure 1, followed by the LINCAL process, which analyzes a number of images taken for different integration times with the stimulator lamp(s) on and the shutter closed. The integration times must produce data that span its dynamic range to an extent suitable for determining the linear response of each pixel, and characterizing the nonlinear response range and saturation threshold level. After rejecting outliers, the remaining data are processed via a least-squares fit of linear values versus observed values. The fit coefficients are used in subsequent processing by the LINEARIZ step of the pipeline.

2.2.3 Nonuniformity Estimation

The thread for nonuniformity estimation, or field flattening, includes pre-processing by the orange and green blocks in Figure 1. It then executes the FLATFIELD process, which analyzes a set of equal-integration-time images of a sky region with a uniform background (with telescope dithering between image acquisition for point-source rejection in the median calculations). Alternatively, it can also analyze a set of TC images taken with the shutter closed, but in this case, it needs to apply a correction for nonunifomities in the TC illumination. Outlier rejection is included in this processing. At each pixel, a normalization factor, also known as a flat-field correction factor, is computed for equalizing the response of the associated detector to all others in the FPA. The flat-field correction factors for all detectors in the FPA are assembled into a FITS image, known as a “flat”. As an option, pixel-by-pixel statistics of the stacked images are computed and outputted for use by the gain/read-noise estimation thread (see subsection 2.2.5). A flat produced by this thread is used in subsequent processing by the FLATAP step of the pipeline.

2.2.4 Nonuniformity Tracking

This thread deals with the possibility of noisy flat-field estimates by implementing a separate Kalman filter[4] at each image pixel to make improved state estimates of flat-field correction factors propagating in time. This form of filtering assumes independent, Gaussian-distributed process and measurement noise. The inputs to the FLATTRACK process used in this thread include the noise-mitigated flat-field state estimates at the previous time and the current flat-field estimates at the current time from the FLATFIELD process. FLATTRACK outputs a noise-mitigated flat-field state estimate at the current time. This thread is executed separately for flats associated with each different integration time of interest. A filtered flat produced by this thread can be used in place of a non-filtered flat in production processing.

2.2.5 Gain/Read-Noise Estimation

The GAINREAD process is executed in this thread after pre-processing by several runs of the FLATFIELD thread for data sets with different integration times, where the range of integration times result in good coverage of the data’s dynamic range at every image pixel. The GAINREAD process employs the conventional signal-versus-variance technique[5] to estimate the gain and read noise at each image pixel. This involves least-squares fitting a line to the input data at each pixel. The gain, in units of electrons per DN, corresponds to the inverse of the line’s slope, and the read noise, in units of electrons, is derived from the line’s y-intercept.

3.0 CALIBRATION SERVER

A "calibration server" will be implemented for the purpose of providing the appropriate calibration data to the science-data processing thread. For example, production processing of a raw image will require, among other things, subtraction of a dark current image taken with the *same* integration time. Thus, the calibration server must be capable of searching a "library" of dark-current images for the correct one to apply.

The calibration server's "brain" will be programmed with sets of rules specific to the various types of observations possible, many of which will dedicate a certain amount of observing time for acquisition of new raw calibration data, which are processed by the calibration threads and used to update the calibration data base. The development of this vital component of the SIRTF software subsystem is currently in an early stage; however, significant progress in this area is expected to occur over the next six months.

4.0 DISCUSSION AND CONCLUSIONS

The data-processing pipeline, with its various threads for generating BCDs and calibration products, has a modular design. Each software element in the pipeline is a stand-alone program with well-defined inputs and outputs. Software control is provided by parameter passing via command-line and/or namelist. This design facilitates upgrading the methods employed in the software elements as the development progresses. It also allows the software elements to be easily reconfigured as the processing requirements evolve and the interplay among the processes becomes better understood. This design philosophy has served the development effort very well so far.

The software elements have been developed to be as general and flexible as possible. For example, data memory is allocated dynamically by reading the image size from the FITS header. This allows many of the IRAC software elements to be used in pipelines for the Infrared Spectrograph (IRS) and the Multiband Imaging Photometer for SIRTF (MIPS), which are the other science instruments onboard SIRTF. These instruments produce images of different sizes than the IRAC, so it is important that the software be flexible enough to handle this without having to re-compile. It is noteworthy that the pipeline for the MIPS 24-μm data is in many ways similar to that for the IRAC data; some differences are that the MIPS 24-μm images are smaller (128×128 pixels), and require an additional correction for the "droop" effect.

For initial non-automated pipeline-thread testing, Perl-wrapper scripts have been written to chain the software elements together such that the output of one process becomes the input to the next process. Either these scripts, or translated C-shell versions of these scripts will be implemented for the final testing under an automated-processing executive.

Pipeline-thread testing is now under way. In late March, several test images from the IRAC instrument were processed by FOS and made available for testing at the SSC. Exercises were conducted with the dark-current, linearity, and flat-field calibration threads, as well as the science-data processing thread. Here are some preliminary results of the testing. The flat-field results using a small 8.0-μm image set indicate a preliminary image flattening to within 1.2% (1-σ). The dark-current results for the 3.6-μm data show a pseudo-mux-bleed artifact in the data, which is not present in the 4.5-μm data, even though the same FPA technology is utilized. For removal of this effect from the 3.6-μm data, another software element (not shown in Figure 1) will be required.

The success realized thus far has proven the robustness of the pipeline design. Between now and launch the additional software that is required will be developed and software refinements will be made. Much more testing with simulated and instrument data is planned.

5.0 ACKNOWLEDGEMENTS

The support of the National Aeronautics and Space Administration (NASA) is gratefully acknowledged.

6.0 NOMENCLATURE

ANN	Artificial Neural Network
BCD	Basic Calibrated Data
DCE	Data Collection Event
FPA	Focal Plane Array
FOS	Flight Operations Software
IOC	In-Orbit Check-Out
IR	Infrared
IRS	Infrared Spectrograph
IRAC	Infrared Array Camera
JPL	Jet Propulsion Laboratory
MIPS	Multiband Imaging Photometer for SIRTF
NaN	Not a Number
NASA	National Aeronautics and Space Administration
SIRTF	Space Infrared Telescope Facility
SSC	SIRTF Science Center
TC	Transmission Calibrator

7.0 REFERENCES

1. Fanson, J., G. Fazio, J. Houck, T. Kelly, G. Rieke, D. Tenerelli, and M. Whitten, The space infrared telescope facility (SIRTF), Proc. SPIE, Vol. 3356, pp. 478-91, August 1998. Further and more up-to-date information on the SIRTF mission can be found at http://sirtf.jpl.nasa.gov/sirtf.
2. Pao, Y.-H., *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley, Reading, Mass., 1989.
3. Rousseeuw, P. J. and C. Croux, Alternatives to the median absolute deviation, Journal of the American Statistical Association, Vol. 88, pp. 1273-83, 1993.
4. Bar-Shalom, Y. and T. E. Fortmann, *Tracking and Data Association*, Academic Press, Boston, 1988.
5. McLean, I., *Electronic Imaging in Astronomy: Detectors and Instrumentation*, ISBN 0-471-96972-9, Wiley and Sons, 1997.