

ACE: Astronomical Cataloging Environment

Francisco G. Valdes

IRAF Group, NOAO¹, PO Box 26732, Tucson, AZ 85726

Abstract. A new IRAF package under development for detecting, evaluating and cataloging objects in astronomical images is summarized.

Astronomers are producing ever more and deeper images of the sky. While an image contains all the information about that view of the sky within some passband and at some resolution, for many astronomical investigations what is needed is a catalog of the sources in the image. Reducing the image to the significant information about the sources – position, brightness, classification, etc. – makes working with very large areas of the sky more tractable.

The Astronomical Cataloging Environment (ACE) is a new IRAF package being developed for cataloging sources in images and manipulating the catalogs. The reasons for developing ACE, given a number of other similar programs, are to implement new or improved features and algorithms and to provide integration with IRAF. While ACE will include a variety of tools for cataloging and analysis, this paper outlines the features and algorithms of the basic detection and cataloging tool, excluding object classification.

The output of the detection task consists of a combination of object regions, a catalog, and log information. The object regions are recorded in a pixel mask with values combining the object number and various flags. The pixel mask uses a run-length encoding that is compact and efficient to use in the algorithms. These masks can be used for a variety of purposes and algorithms outside of ACE. The catalog may be in a variety of standard formats (e.g., binary table or text file). The contents of the catalog are specified by the user in a catalog definition file. This can select any of the measured values and define expressions based upon them.

All algorithms provide for application of user specified pixel masks and exposure maps. Maps include a constant value, a low resolution image to interpolate or a full image. The pixel masks select pixels to be ignored. An exposure map is used to adjust estimates of the standard deviation of a pixel by the square root of the map value. This is typically used when multiple, dithered exposures are combined by *averaging* the overlapping pixels where the number of overlapping pixels at each point in the final image may vary. When all images have the same exposure time the exposure map would be the number of images that contributed to each pixel.

¹National Optical Astronomy Observatories, operated by the Association of Universities for Research in Astronomy, Inc. (AURA) under cooperative agreement with the National Science Foundation.

1. Sky Determination

Detection of sources in images begins with estimating the probability distribution of the background sky signal *at every image pixel*. In practice only the mode μ and the standard deviation σ are estimated, and a normal distribution is assumed for estimating the significance of deviations from the mode. Sky information may be specified by the user, separately for both the mode and σ in the form of maps, or determined by ACE from the image data. When ACE measures the sky from a science image, it uses one or two passes through the data and then updates the estimates when detecting the objects. The first pass fits 1-D functions to a subsample of lines using sigma clipping iteration to eliminate objects. Unrejected pixels are used to fit 2-D surfaces for μ and σ , and coefficients are recorded in the image header.

A second pass divides the image into large, equal-sized blocks and estimates μ and σ in each block using histograms of the deviations $\delta = (I - \mu)/\sigma$, where I is the pixel value and μ and σ are from the first pass sky. The deviation histogram is truncated at some number (e.g., 2.5) to reject object pixels. When the truncated histogram contains too few pixels, the block and its neighbors are rejected to eliminate effects from large galaxies and bright stars. The μ and σ of the deviation histograms are evaluated and scaled back to data values using the means $\langle\mu\rangle$ and $\langle\sigma\rangle$ from the computation of δ . Blocks are averaged into larger blocks, and any missing blocks are interpolated. The block values are stored as reduced-resolution sky and sky sigma maps that are interpolated to get sky values for every pixel in the science image.

2. Object Detection

Object detection consists of finding pixels with deviations δ outside specified limits. Object pixels have positive deviation, and "negative" objects (i.e., negative pixel deviations) may either be detected or ignored. Deviations within the limits are used to update the sky by accumulating and evaluating histograms as before.

Adjacent object pixels (user-selected either along the four pixel edges or also along the diagonals) are assigned to the same object region. A candidate object is detected when no further pixels are added to an object region.

The depth of detection is set by pixel noise that may be reduced by convolving the image data at the expense of reducing resolution. Matching the PSF size produces optimal convolution, but any function will improve faint object detection. ACE permits use of any convolution function and recognizes symmetries in the kernel to optimize computation. When image data are convolved, the values of σ used in computing δ are reduced by the root sum of the kernel weights.

Typical detection thresholds δ are 3–5, which ensures that some pixels outside the thresholds will not be part of real objects. To avoid too many spurious detections ACE provides two detection-stage criteria: that the number of pixels in a candidate object must exceed a certain value, and that at least one of the mean and maximum of δ must exceed a specified statistical significance.

3. Splitting Merged Objects

Regions of connected pixels that are just above the background sky will sometimes include more than one astronomical object due to the overlap of their light. While all such overlapping objects can't be separated, many can be by setting a higher threshold. Conceptually this means finding saddles in the light distribution between the objects.

ACE finds such thresholds by taking the detected regions and repeatedly raising the threshold by a small amount to see whether or not the pixels separate into two or more subregions. When a separation occurs the original region is marked as a "multiple" object and the subregions are added to the list of detected objects. The threshold is then raised further, with the split regions included as separate objects, to look for further splitting. This continues until there are fewer pixels in a region than the criteria for defining a new object.

The amount of time taken to split merged objects is related to the number of thresholds used. To limit the number of thresholds while still considering separations at all thresholds from near sky to the peaks of the brightest objects requires using non-linearly spaced thresholds. Near sky background, where most objects merge and faint objects quickly disappear, a linear spacing is used with steps that are a fraction of the sky standard deviation. When the threshold reaches a certain value the spacings become non-linear.

Rather than read through the image for each threshold, possibly convolving the data, a special in-memory splitting image of the positive deviations is created during the detection stage. Pixels that are not in objects, i.e., those below the positive detection threshold, are set to zero in the splitting image. The other pixels are converted into bin numbers corresponding to the chosen thresholds to produce a discrete number of short integer values. The splitting image is compressed using run-length encoding of the non-zero values. This format, which excludes pixels not in objects, can be efficiently searched for objects by running through the run-length segments. As the splitting goes through the thresholds, lower thresholds are removed from the run-length segments of the splitting image.

4. Growing the Object Regions

Threshold detection always misses the wings of the light distribution in the detected objects. Also the method of increasing thresholds shrinks the regions identified with the split objects. To improve this situation, once the object regions have been identified there is an option to "grow" the regions, i.e., the borders of the object regions are extended by adding pixels that touch the borders. Objects are permitted to grow on the other side, even after they've touched a neighbor on one side. When adding a pixel in one pass that touches more than one object region, there is a contention for that pixel. This is resolved by assigning the pixel to the object with the greater flux defined as the sum of sky-subtracted pixel values. This operation can be done efficiently using the run-length mask encoding of the object regions.

5. Cataloging

The preceding steps all deal with defining the regions occupied by the objects in the image. The next step is to compute catalog parameters and statistical errors over the object regions, efficiently accessed through the object region mask. The variances of object pixels are represented by $\sigma_{object}^2 = (\sigma^2 + (I - \mu)/g)/t$ where g is a gain factor scaling the pixel values to a Poisson variance and t is the relative exposure from the exposure map. All the terms are functions of position as specified by maps. If no exposure map is given, a default of unity is used; if no gain map is given, the Poisson term is omitted. The μ and σ may be values initially specified or sky values derived earlier, or new values may be specified. By not giving a gain map and using an externally derived σ , any noise behavior for the objects can be defined.

Rather than giving a list all possible quantities, a categorized summary follows. Position information includes the sky-subtracted intensity centroid over the object region, the position of the maximum pixel and the limits of the object region along the image axes. The positions may be measured in image pixel coordinates and in celestial coordinates. Celestial coordinates require a coordinate system in the image header.

Photometric quantities include fluxes within the object region and within circular apertures of specified radii, a Kron flux and an apportioned flux for split objects. The apportioned flux takes the original flux of a merged object and divides it among the split objects in the same ratio as the fluxes of the split objects measured at the higher isophote where the objects separated. The fluxes may be expressed in magnitudes if desired. The circular aperture fluxes include a simple partial pixel correction at the boundaries.

Shape information includes the second moments of the light distribution from which an ellipticity and position angle are derived. Miscellaneous information includes catalog entry identifications, the number of pixels in the object and flags of various kinds.

6. Multi-image and Difference-image Cataloging

For these two common types of astronomical imaging programs, ACE uses the celestial coordinate systems to align the images. ACE provides the ability to use one of two or more image (or a combination thereof) to define object regions and then determine the sky and evaluate objects in individual images using the common regions.

ACE also allows specifying two images (one of which could be a stack of all exposures) that are registered and differenced internally to detect sources in the difference. Distinguishing real differences from artificial ones (e.g., caused by point spread function variations and registration errors) can be difficult. Such effects will produce significant positive and negative differences centered on bright and narrow sources such as bright stars. Such false flux differences are minimized by requiring that a flux ratio threshold be exceeded in a comparison between the two images.