A Systematic and Automated Technique for Detection of Star Clusters Applied to New GLIMPSE Data

Emily Mercer, Dan Clemens, and the GLIMPSE Team

The Galactic Legacy Infrared Mid-Plane Survey Extraordinaire (GLIMPSE) Team is currently producing a large database of infrared point sources in the inner Galaxy. The database, which includes a high-reliability catalog and an archive of point sources, is being generated from mid-infrared images taken by the Infrared Array Camera (IRAC) aboard the *Spitzer Space Telescope*. We have developed an algorithm that uses the GLIMPSE point source catalog and archive to systematically search for and detect star clusters. The software operates by locating and characterizing significant over-densities in the sky distribution of point sources. The routine relies on advanced statistical tools for detecting the clusters and measuring several of their properties, including location, angular size, and member density. Our cluster-finder has been tailored to also search for and detect clusters in the 2MASS point source catalog. Using both the GLIMPSE and 2MASS catalogs, the software is able to detect globular, open, and young embedded clusters. Consequently, we are now producing a complete and up-to-date list of clusters located in the inner Galaxy. This new list of clusters, specifically the young embedded clusters, will allow us to identify regions of ongoing star formation in the Milky Way.

Cluster-Finding Method

- **Bin** point source catalog into image
- **Filter** image
- **Detect** over densities in image
- **Run** EM algorithm over a subimage
- **Select** best mixture model of clusters and background
- **Output** properties of found clusters

About GLIMPSE

GLIMPSE is a fully sampled, confusion limited, 4-band near- to mid-infrared survey of the inner two-thirds of the Galactic disk with a spatial resolution of ~2". IRAC is being used to image 220 square degrees at wavelengths centered on 3.6, 4.5, 5.8, and 8.0 microns in the Galactic longitude range of 10 deg to 65 deg on both sides of the Galactic center and in Galactic latitude +/- 1 deg.

The team is providing the following products: a high reliability GLIMPSE Point Source Catalog (GPSC), a GLIMPSE Point Source Archive (GPSA; ~5 sigma), and a Mosaiced Image Atlas of the entire surveyed area at all four IRAC bands, all of which will be made available via the IPAC data archive. We expect to catalog about 50 million sources detected with S/N >= 5.

New Clusters Found in the GLIMPSE Region

l14.13-0.65

l4.16-0.07

l332.84-0.59

l338.38+0.11

Spatial Distribution of Clusters in GLIMPSE Field

25	Newly found
130	Dutra & Bica
109	Open
4	Embedded
1	Globular
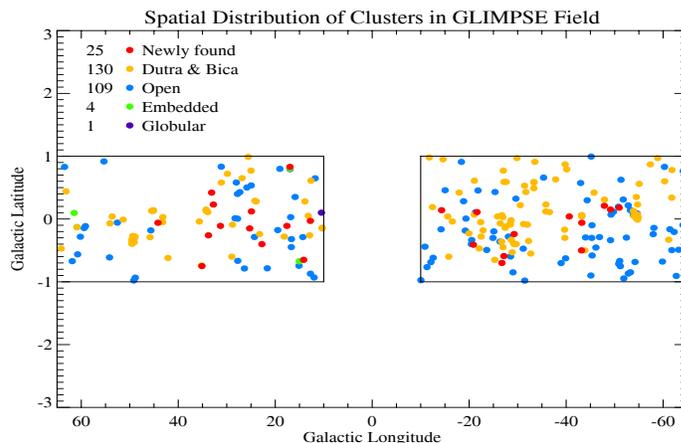
The Detection Algorithm

The automated cluster-finding routine utilizes the Expectation-Maximization (EM) algorithm (Dempster, Laird, Rubin, 1977, JRSS-B). The EM algorithm is a statistical tool we apply to over-dense regions of the GLIMPSE point source distributions. It implements a mixture of Gaussians and iteratively computes the center and size of each potential cluster, until these values converge. The E-step computes the likelihood of each point source belonging to the either a cluster or background component. The M-step updates the values of the center and size for each cluster, based on the probabilities calculated in the E-step. Finally, the best mixture model of potential clusters and the background is chosen using the Bayseian Information Criterion (BIC). The BIC is a model selection tool that penalizes models with additional parameters.