The Astrophysical Journal, 635:560–569, 2005 December 10

NEW STAR CLUSTERS DISCOVERED IN THE GLIMPSE SURVEY

E. P. Mercer,[1] D. P. Clemens,[1] M. R. Meade,[2] B. L. Babler,[2] R. Indebetouw,[3] B. A. Whitney,[4] C. Watson,[2] M. G. Wolfire,[5] M. J. Wolff,[4] T. M. Bania,[1] R. A. Benjamin,[6] M. Cohen,[7] J. M. Dickey,[8] J. M. Jackson,[1] H. A. Kobulnicky,[9] J. S. Mathis,[2] J. R. Stauffer,[10] S. R. Stolovy,[10] B. Uzpen,[9] and E. B. Churchwell[2]

Received 2005 June 10; accepted 2005 August 16

ABSTRACT

A systematic and automated search of the extensive GLIMPSE mid-infrared survey data of the inner Galaxy was carried out to uncover new star clusters. This search has yielded 59 new clusters. Using our automated search algorithm, these clusters were identified as significant localized overdensities in the GLIMPSE point-source catalog (GLMC) and archive (GLMA). Subsequent visual inspection of the GLIMPSE image mosaics confirmed the existence of these clusters plus an additional 33 heavily embedded clusters missed by our detection algorithm, for a total of 92 newly discovered clusters. These previously uncataloged clusters range in type from heavily embedded to fully exposed clusters. More than half of the clusters have memberships exceeding 35 stars, and nearly all the clusters have diameters of 3′ or less. The Galactic latitude distribution of the clusters reveals that the majority are concentrated toward the Galactic midplane. There is an asymmetry in the number of clusters located above and below the midplane, with more clusters detected below the midplane. We also observe an asymmetry in the number of clusters detected in the northern and southern halves of the Galaxy, with more than twice as many clusters detected in the south.

Subject headings: infrared: stars — open clusters and associations: general

1. INTRODUCTION

Discoveries of new star clusters in the disk of the Milky Way have significantly increased in recent years. The Two Micron All Sky Survey (2MASS; Skrutskie et al. 1997) was used to discover the majority of these new clusters (e.g., Dutra et al. 2003; Bica et al. 2003b), which were previously undetectable in optical surveys due to dust extinction at low Galactic latitudes. Until recently, the method of detecting clusters at optical and infrared wavelengths has been limited to visual inspection, yielding some 2000 optical and near-IR clusters. Recently, there have been attempts to automate the detection of star clusters by locating peaks in the stellar surface density, and these methods have been applied to the 2MASS and Deep Near Infrared Survey of the Southern Sky (DENIS) point-source catalogs (Ivanov et al. 2002; Reylé & Robin 2002). To date, these automated efforts have yielded 10 new clusters.

Developing an accurate, unbiased census of star formation in the inner Galaxy is important for understanding the structure of the inner Galaxy. Most stars form in clusters, which are deeply embedded within molecular clouds. Since embedded clusters are the most likely sites of recent star formation, they can be used to study the current Galactic distribution of star formation. Because of high dust extinction toward the Galactic plane, embedded clusters are often only detectable at near- to mid-IR wavelengths. Thus, the star clusters that can be detected in infrared sky surveys represent a particularly useful tool for elucidating the structure and nature of the star-forming inner Galaxy.

The Galactic Legacy Infrared Mid-Plane Survey Extraordinaire (GLIMPSE; Benjamin et al. 2003; Churchwell et al. 2004) project has mapped most of the first and fourth quadrants of the inner Milky Way ($|l| = 10^\circ$–65° and $|b| \leq 1^\circ$) at mid-IR wavelengths of 3.6, 4.5, 5.8, and 8.0 μm using the Infrared Array Camera (IRAC; Fazio et al. 2004) on the *Spitzer Space Telescope* (Werner et al. 2004). The GLIMPSE team has produced a database of point sources extracted from the mid-IR images, which include a high-reliability point-source catalog (GLMC, with about 3×10^7 objects) and an archive of point sources (GLMA, $\sim 5 \times 10^7$ objects). The GLIMPSE team is also producing image mosaics of the inner Galaxy. With the GLIMPSE mosaics and database of point sources, it is now possible to obtain a more accurate census of star-forming regions in the inner Galaxy and a more complete view of our Galaxy's structure.

In this paper, we describe the components and application of a parametric algorithm to detect star clusters in the inner Galaxy systematically and report the results of using this algorithm on the GLIMPSE database of point sources. In § 2, we describe the point-source data sets used, the GLMC and GLMA. A detailed description of the cluster search algorithm is given in § 3. The results of applying the algorithm on the GLMC and GLMA and a parallel visual inspection of the GLIMPSE image mosaics are given in § 4, and a discussion of the results follows in § 5.

2. GLIMPSE DATA

We used the GLIMPSE high-level data products derived from the IRAC data acquired toward the GLIMPSE survey area, which spans some 220 deg^2 of the inner Galaxy. Basic calibration of the GLIMPSE IRAC frames was performed by the *Spitzer* Science Center pipeline (*Spitzer* Science Center 2004). The data were further processed through the GLIMPSE pipeline (Benjamin et al.

[1] Institute for Astrophysical Research, Boston University, 725 Commonwealth Avenue, Boston, MA 02215.
[2] Astronomy Department, University of Wisconsin, 475 North Charter Street, Madison, WI 53706.
[3] Department of Astronomy, University of Virginia, P.O. Box 3818, Charlottesville, VA 22903.
[4] Space Science Institute, Boulder, CO 80301.
[5] Department of Astronomy, University of Maryland, College Park, MD 20742.
[6] Physics Department, University of Wisconsin–Whitewater, Whitewater, WI 53190.
[7] Radio Astronomy Laboratory, University of California at Berkeley, 623 Campbell Hall, Berkeley, CA 94720.
[8] Department of Astronomy, University of Minnesota, Minneapolis, MN 55455.
[9] Department of Physics and Astronomy, University of Wyoming, Laramie, WY 82071.
[10] *Spitzer* Science Center, California Institute of Technology, Pasadena, CA 91125.

2003). Point sources were extracted from each IRAC frame using a modified version of DAOPHOT (Stetson 1987). The modified version incorporates a local background determination when searching for stars, since many of the GLIMPSE frames have strong background variations with position. Multiple detections of each point source were then band-merged to create a single list of sources.

The GLIMPSE point-source catalog (GLMC) consists only of high-reliability sources. To meet the reliability requirement of $\geq$99.5%, each source must be detected twice in any one of the four IRAC bands (3.6, 4.5, 5.8, or 8.0 μm) and at least once in an adjacent band. The GLIMPSE point-source archive (GLMA) consists of all sources with signal-to-noise ratios $\geq$5 above the local background in at least one band. Note that the GLMA has a less stringent point-source extraction criterion and a weaker reliability criterion and therefore includes more sources. Information provided in the GLMC and GLMA includes source fluxes (with errors), positions (with errors), local stellar density, local sky brightness, and source quality flags. More details of the GLMC and GLMA source selections are provided in the GLIMPSE Science Data Products Document (Meade et al. 2005) and the GLIMPSE Quality Assurance Document (E. B. Churchwell et al. 2005, in preparation). The catalog and archive contain 30 million and 47 million sources, respectively.

3. AUTOMATIC DETECTION ALGORITHM

We developed an algorithm to search for star clusters in the GLIMPSE point-source catalog and archive in a systematic and unbiased manner. The software locates and characterizes significant local overdensities in the sky distribution of point sources. It uses statistical models for detecting clusters and measuring several of their properties, including location, angular size, and number of members. In general, a statistical model is defined as a set of probability distribution functions that describe the data. The statistical models we use to locate and characterize clusters are explained in the following subsections. Briefly, the algorithm operates in three steps: (1) it computes initial guesses of potential cluster locations by detecting regions of locally high stellar density, (2) it estimates the locations and sizes of potential clusters by fitting two-dimensional Gaussians, and (3) it removes false cluster detections by checking the statistical significance of each cluster.

3.1. *Detecting Clusters Automatically*

The method of detecting clusters begins with finding local density peaks in the two-dimensional (l, b) distribution of GLIMPSE point sources. To find peaks, we divide square degree regions (e.g., $1^\circ \times 1^\circ$) of the GLIMPSE point-source catalog or archive into smaller spatial bins to create a "raw image." The square degree regions are limited to 1° spans in l and b because of the strong background variations with position seen in the GLIMPSE image mosaics. The bins themselves are square and nonoverlapping; their size is a free parameter (e.g., $0\overset{\circ}{.}01 \times 0\overset{\circ}{.}01$).

Assigned to each bin is the count corresponding to the number of point sources that fall in the (l, b) range spanned by the bin. Star counts are expected to follow a Poisson distribution, since the total number of stars is very large ($N_{\rm stars} \sim 10^5$ deg^{-2}) and the small bins will yield average counts that are small ($\sim$10). When the average number of stars per bin is approximately constant over the entire region, the star counts follow a statistical model considered to be "homogeneous" (Trumpler & Weaver 1953). A homogeneous Poisson model is characterized by a single Poisson distribution that describes all bins. In this case, the bins with counts that are greater than some significance threshold (e.g., $\geq 4\,\sigma$) above the mean could be selected as locations of potential clusters. Under the assumption of such a constant background, our initial search detected few clusters. We found $\sim$15 previously cataloged clusters and $\sim$15 uncataloged clusters. It is possible, however, to refine the statistical model and so improve the algorithm's sensitivity to finding new clusters.

In the GLIMPSE data, we find large-scale trends of a decreasing areal density of stars with increasing angular distance from the Galactic plane and from the Galactic center (Benjamin et al. 2005). On a smaller angular scale ($\sim 1^\circ \times 1^\circ$), the data do not follow these trends. Rather, the areal density of stars fluctuates because of regions of high extinction (e.g., dark clouds) and diffuse emission. To detect clusters in $1^\circ \times 1^\circ$ regions, where the GLIMPSE background is highly variable, we adapted the homogeneous Poisson model to account for the changes in the local stellar density. Because the areal density is nonuniform, the background distribution of stars is better described by a nonhomogeneous Poisson model. A nonhomogeneous Poisson model can be defined as the multiple Poisson distributions that best describe the counts seen in the bins. Under this approach, the constant background condition exists only over relatively small regions ($\sim 5' \times 5'$, $N_{\rm stars} \sim 8000$), for which a single Poisson distribution describes each such region.

Due to the Poisson nature of this adaptive model, regions with higher stellar density will produce higher noise counts, and regions with lower stellar density will have lower noise counts. Therefore, a global significance threshold based on large-area averaged star counts cannot be used to detect density peaks in regions with very different stellar densities. A robust way of detecting peaks in the stellar density when the average bin count fluctuates with location is to equalize the noise level of the higher density regions with the noise level of the lower density regions. Since different Poisson distributions describe different regions in the raw image, it is useful to transform the raw image so these different distributions have the same noise scales. Different regions can then be compared in an unbiased way. We refer to this process as "standardization" and define it to be the subtraction of the (local) mean and division by the (local) standard deviation.

To standardize a star count raw image, the local means and local standard deviations must first be determined. Since the variance of a Poisson distribution is equal to its mean, we need only estimate the local means. This is done by smoothing each raw image using a median filter with a circular kernel (e.g., angular radius $\sim 0\overset{\circ}{.}05$). The smoothed image will reflect changes in the background on size scales of arcminutes. Using the median-filtered image, we standardize the raw images by subtracting the median-filtered images and dividing by the square root of the median-filtered images.

Standardizing each raw image is necessary to allow comparison of the significance of overdense bins drawn from different local star count distributions that are characterized by different noise count values. This processing effectively centers and normalizes the bin values so they have means of zero and variances of unity. The standard deviation of the standardized image is then approximately constant over the entire region. Moreover, the distribution of pixel values will now follow a Gaussian, provided the original average number of stars per bin was large enough (e.g., $\geq$5). After the standardization processing, the locations of cluster candidates can now be selected using a global significance threshold (e.g., $\geq 4\,\sigma_0$, where σ_0 is the noise level of the full standardized image). Standardizing the raw image using the local means enables a more even and unbiased approach for detecting stellar overdensities in the GLMC and GLMA. Compared to the homogeneous

background approach, the number of detected, previously unknown clusters more than tripled when the nonhomogeneous background model and standardization process were applied.

Magnitude and color selection criteria were also implemented in the search for local overdensities. The range of magnitudes in each IRAC wave band was divided into three categories, "bright," "mid," and "faint," by separating the corresponding magnitude histogram into thirds. Typical magnitude ranges of the 3.6 μm band are 3.5–13.0, 13.0–14.0, and 14.0–15.5 for the bright, mid, and faint cuts, respectively. Similarly, colors were divided into three categories: "red," "green," and "blue." From analysis runs employing these magnitude and color selection criteria, we found that only the bright magnitude cut of the shortest IRAC wave band (3.6 μm) produced samples of GLMC and GLMA sources with cluster candidates that could be identified as overdensities by the algorithm. Hence, we detected clusters via two sets of criteria: using no color/magnitude cuts and using the bright 3.6 μm cut.

3.2. *Characterizing Clusters Automatically*

An adequate first-order model for representing a stellar cluster in star count data is a two-dimensional angular Gaussian distribution. The angular components of the Gaussian are the cluster center, $\boldsymbol{\mu}$, and size, $\boldsymbol{\Sigma}$. As described in § 3.1, the background of noncluster point sources is best approximated by a nonhomogeneous Poisson model. Following Fraley & Raftery (2002), we assume that the local distribution of stars in a region of the GLIMPSE catalog and archive can be represented by a mixture of a local Poisson background component and a number, k, of two-dimensional Gaussian cluster components.

The probability density of a star being found near a direction $\boldsymbol{x} = (x, y)$ in a solid angle A is given by (Fraley & Raftery 2002, eq. [11])

$$f(\boldsymbol{x}) = \tau_0 \frac{b(\boldsymbol{x})}{\mathrm{A}} + \sum_{i=1}^{k} \left\{ \frac{\tau_i}{2\pi\sqrt{|\boldsymbol{\Sigma}_i|}} \exp\left[-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_i) \right] \right\}, \quad (1)$$

where the first term represents the local background stellar density and the second term corresponds to the sum of the k Gaussian cluster components. We define $b(\boldsymbol{x}) = \mathrm{background}(\boldsymbol{x})/\mathrm{mean}(\mathrm{background})$, where $\mathrm{background}(\boldsymbol{x})$ is the intensity of the median-filtered image at $\boldsymbol{x}$. When the background is uniform $b(\boldsymbol{x}) = 1$, otherwise $b(\boldsymbol{x})$ is proportional to the local stellar density background, estimated via median filtering. The parameters τ_i are the relative proportions of the mixture components ($\tau_i \geq 0$; $\sum_{i=0}^{k} \tau_i = 1$) and can be interpreted as follows. Higher values of τ_0 give a greater likelihood that a randomly selected star belongs to the background rather than to a cluster. Given that N stars are contained within region A, approximately $\tau_0 N$ stars belong to the background and $\tau_i N$ stars belong to the ith cluster.

The two-dimensional Gaussian components of equation (1) allow for noncircular clusters. As such, $\boldsymbol{\mu} = (\mu_x, \mu_y)$ and $\boldsymbol{\Sigma}$ is the covariance matrix,

$$\begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix}.$$

When the matrix elements are such that $\sigma_{xy} = \sigma_{yx} = 0$ and $\sigma_{xx} \neq \sigma_{yy}$, the cluster is elliptical and aligned along the x- or y-axis. If, however, $\sigma_{xy} = \sigma_{yx} = 0$ and $\sigma_{xx} = \sigma_{yy} = \sigma^2$, then the Gaussian component reduces to the more familiar expression, $1/(2\pi\sigma^2) \exp[-(x^2 + y^2)/\sigma^2]$, for $\boldsymbol{\mu} = (0, 0)$. In equation (1), the notation $|\boldsymbol{\Sigma}_i|$ refers to the determinant of the covariance matrix, $(\boldsymbol{x} - \boldsymbol{\mu}_i)^T$ refers to the transpose of the vector, and $\boldsymbol{\Sigma}_i^{-1}$ refers to the inverse of the covariance matrix.

Once initial values for the locations and sizes of candidate clusters were found from searching the standardized raw images (see § 3.1), they were further refined using the expectation-maximization (EM) algorithm (Dempster et al. 1977). The EM algorithm sequentially improves the fit of the model, equation (1), to the data by maximizing the likelihood of the model's fit,

$$L(X|\tau_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \prod_{j=1}^{N} f(X_j), \quad (2)$$

where X_j is the jth point source in a given solid angle, A, of the GLMC (or GLMA) and N is the total number of point sources in A. The value of L represents the quality of the model's fit for a given set of parameters, where higher values of L indicate better fits.

The EM algorithm consists of two steps. In the expectation, or E step, each point source is assigned $k + 1$ values, which are the probabilities that the source belongs to the background or to one of the k potential clusters (calculated in eq. [1]). In the M step, the model parameters (τ_i, $\boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}_i$) that maximize the likelihood computed in the E step are determined. Each step of the EM algorithm yields updated parameters that improve the likelihood L given by equation (2). Alternating between the two, the E and M steps are repeated until the updated values of the location and size estimates of each potential cluster meet a convergence criterion. For a detailed description of the EM algorithm and its applications, see McLachlan & Krishnan (1997).

3.3. *Selecting Clusters Automatically*

Some Gaussian components of our background + cluster mixture model could represent "false" clusters. A false cluster is generally due to an anomalous overdense bin that has been modeled as a Gaussian cluster component. Although a false cluster could represent a valid statistical overdensity in the population of stars (spanning several bins), a true clustering of stars is not present. In principle, a mixture model with more allowed components will give a better fit to the data than a model with fewer components. These false cluster components might not, however, significantly improve the overall fit and can usually be removed without negative effects.

Once all candidate clusters have been characterized by the EM process, we attempt to remove false clusters. This is done quantitatively using the Bayesian information criterion (BIC; Fraley & Raftery 2002). The BIC is a model selection tool that assigns an information quantity to models with different numbers of parameters. The BIC is defined as

$$\mathrm{BIC} = 2\ln(L) - m\ln(N), \quad (3)$$

where L is the likelihood, m is the total number of parameters in equation (1), and N is the total number of point sources in region A. If a Gaussian component represents a false cluster, then removing it from the model will not decrease the likelihood significantly. That is, removing a false cluster reduces m but does not significantly reduce L. Thus, the BIC value will increase if a false cluster is removed. Conversely, if a real cluster is removed, L significantly decreases, which thereby lowers the BIC value.

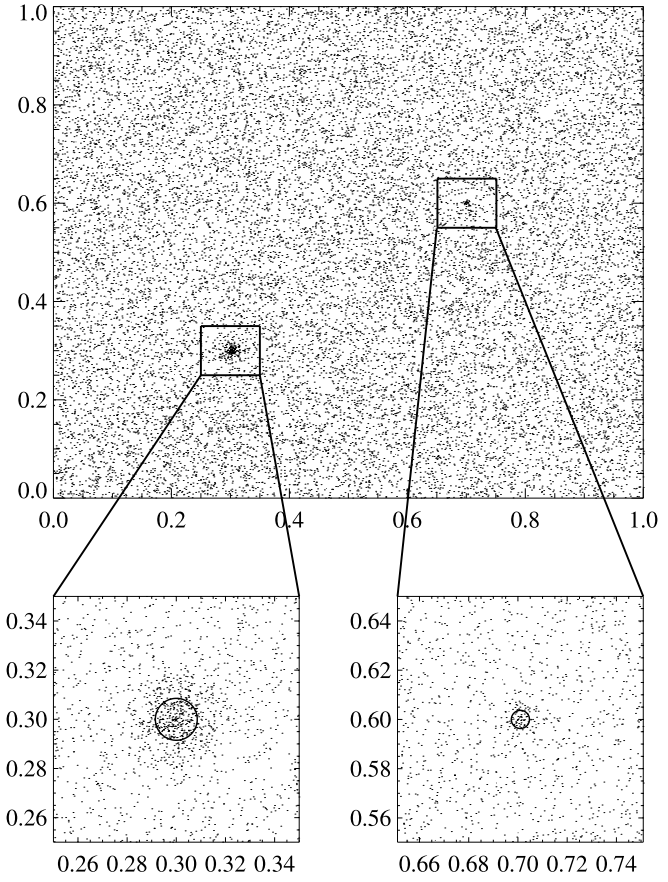

FIG. 1.—Simulated point-source catalog of 100,000 sources (reduced to 20,000 here for ease of display) with two Gaussian clusters. Circles represent the sizes of the clusters as found by the detection algorithm.

We test for false clusters by iteratively excluding the least probable of the Gaussian components and reestimating the mixture model, equation (1). The least probable component is defined as that with the lowest proportion τ_i. If the reduced model has a higher BIC value, we declare that the excluded component lacked significant information and remove it from our list of potential clusters. This reduction process is iterated until all components of the mixture model are deemed significant, according to the BIC.

3.4. *Simulating Clusters*

We tested our cluster-finding algorithm using a simulated point-source catalog. The simulated catalog contained 100,000 sources uniformly distributed over a $1^\circ \times 1^\circ$ area, similar to the source densities in regions of the GLMA. Gaussian distributions of point sources were inserted in the catalog to represent clusters. The algorithm described in § 3.1 successfully detected all the artificial clusters that were added to the simulated catalog. One example of our simulations is shown in Figure 1. The simulation contains two cluster components superimposed on the uniform catalog. The larger cluster has 500 members centered at (0.°300, 0.°300) where

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$

with $\sigma = 0.'500$, and the smaller cluster has 100 members centered at (0.°700, 0.°600) with $\sigma = 0.'250$. The cluster-finding algorithm detected both of these clusters and returned parameters close to the initial conditions. The larger cluster was found to contain 419 members with $\boldsymbol{\mu} = (0.^\circ300, 0.^\circ300)$ and $\sigma = 0.'511$. The smaller cluster was found to contain 85 members with $\boldsymbol{\mu} = (0.^\circ701, 0.^\circ600)$ and $\sigma = 0.'221$.

4. RESULTS

We applied our search algorithm to the inner Galactic plane mid-IR GLIMPSE point-source data and found 91 cluster candidates that satisfied our search criteria (see § 3). Of those clusters, 32 were previously cataloged. The remaining 59 are cluster candidates with high likelihood. Most of the new cluster candidates were detected in the GLMA rather than in the GLMC. The GLMA is a more complete data set, since the source selection criteria for it are less stringent than those for the GLMC. Therefore, clusters are easier to identify in the GLMA because their overdensity contrasts are higher.

The algorithm was also applied to the 2MASS point-source catalog data that span the GLIMPSE survey area to compare with our results. Only six of the 59 new GLIMPSE clusters were detected. Since only six of the previously uncataloged clusters were found in the 2MASS point-source catalog, this demonstrates the enhanced utility of our automated search technique coupled with the extraordinary sensitivity of the GLIMPSE data. Of the 32 previously cataloged clusters our algorithm also detected, more than half were also detected in our search of the 2MASS catalog.

4.1. *Visual Inspection*

Subsequent visual inspection of the GLIMPSE image mosaics allowed us to confirm the presence of a cluster or stellar association at or near each of the 59 locations identified by our automated search. From visual inspection of the 2MASS images, we found that only 24 of these 59 new clusters show apparent clusterings of stars in the near-IR. Our examination of the GLIMPSE images revealed that the new clusters have the morphologies of either young embedded clusters or exposed open clusters, depending on their association with extended emission. Of the 59 new clusters, 29 are associated with extended emission in all four IRAC bands. An example of one of the embedded clusters with extended emission detected in the GLMA is shown in Figure 2. For comparison, optical and near-IR images of the cluster, as well as a plot of the locations of GLMC and GLMA point sources, are also shown.

We also performed a visual inspection of the GLIMPSE image mosaics to search for heavily embedded clusters missed by our automated detection algorithm. Our search yielded an additional 33 such clusters that do not appear in any cluster catalog. We estimated visually the center coordinates and size of each of these clusters. The number of members per cluster was then calculated by counting the number of GLMA sources located within the circular area defined by the cluster radius and center coordinates.

4.2. *Cluster Catalog*

The list of all 92 new clusters detected in the GLIMPSE survey data is given in Table 1. For each of the 59 clusters automatically detected, our algorithm returned an *estimate* of the center coordinates, angular size, and number of members, $N_{\rm stars}$. Recall that $N_{\rm stars}$ is assigned based on the relative proportion (τ_i) described in § 3.2. All clusters detected by the automated routine are labeled A in column (8) of Table 1. Most of these clusters were detected using the bright magnitude cut of the 3.6 μm band and are identified as such in column (9) of Table 1. Those clusters associated with mid-IR extended emission, those detected in the automated search of the 2MASS point-source catalog, and

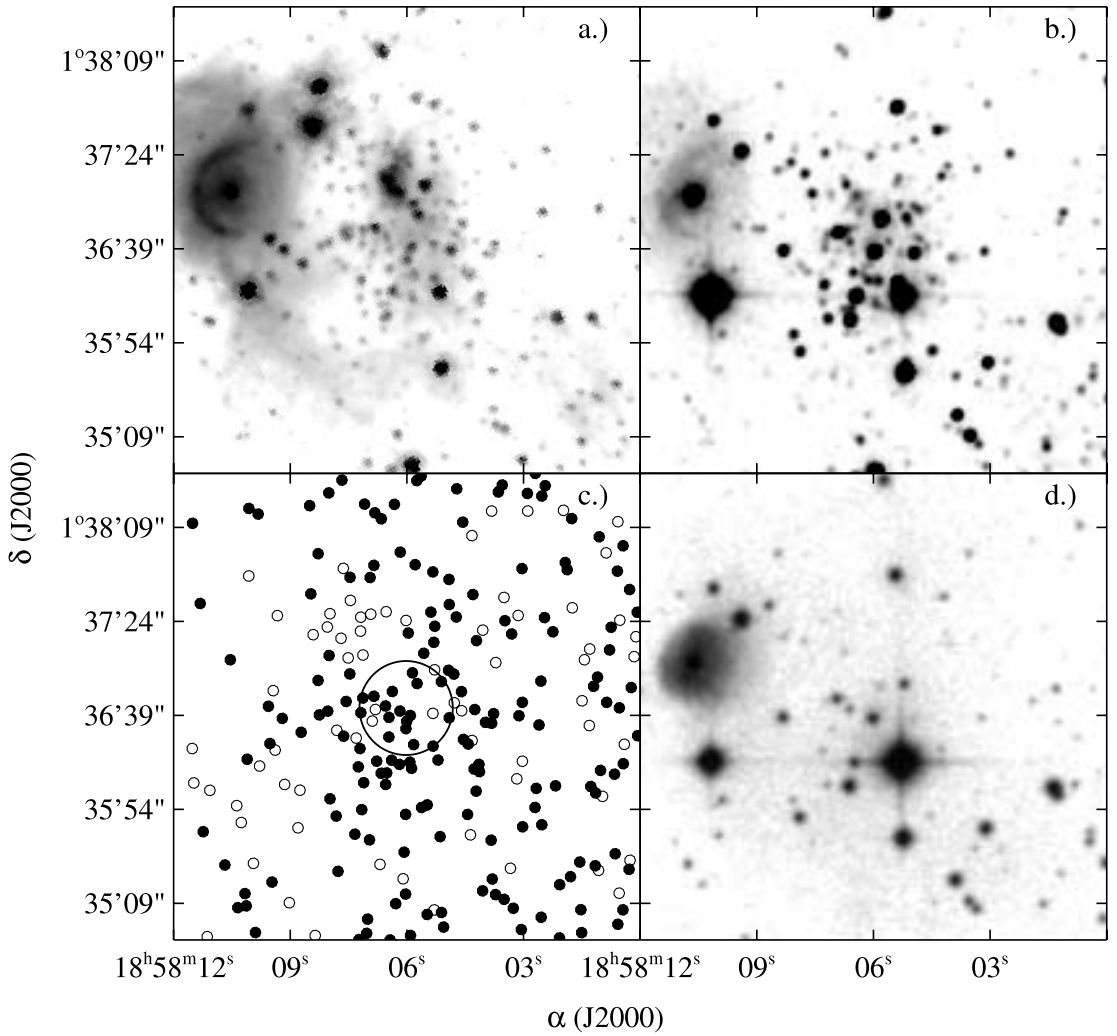

FIG. 2.—Comparison of cluster 14 (*a*) in the mid-IR (GLIMPSE 3.6 μm); (*b*) in the near-IR (2MASS *H* band); (*c*) GLMC point sources (*filled circles*) and GLMA point sources (*open and filled circles*); (*d*) in the optical (Digitized Sky Survey [DSS] *B* band). The large circle in (*c*) represents the size of the cluster as found by the detection algorithm.

those that appear in the 2MASS images have also been identified in column (9). The 33 clusters detected by visual inspection of the GLIMPSE mosaics are listed in Table 1 with *estimates* of their center coordinates, angular size, and number of members. These visually detected clusters are labeled V in column (8) of Table 1.

4.3. *Cluster Distributions*

The sky distribution of the 92 newly detected clusters (plus GLIMPSE GC-01; Kobulnicky et al. 2005) is shown in Figure 3, alongside the distributions of 2MASS clusters (Bica et al. 2003b; Dutra et al. 2003) and optical open clusters (Dias et al. 2003). Within the GLIMPSE survey area, we detected more than 2 times as many new GLIMPSE clusters in the southern half ($295^\circ < l < 350^\circ$) of the Galaxy as in the northern half ($10^\circ < l < 65^\circ$). Similar north-south asymmetries of clusters are also observed in the distributions of near-IR and optical clusters. There are approximately 40% more 2MASS clusters and ~75% more optical clusters in the south.

We also observe an asymmetry in the number of clusters located above and below the Galactic midplane. We detected ~40% fewer GLIMPSE clusters above the Galactic midplane than below it. This midplane asymmetry is also observed in the distributions of 2MASS and open clusters, although it is less pronounced (~25% fewer).

The Galactic longitude distributions of GLIMPSE, 2MASS, and optical clusters are shown in Figure 4. In the north, the trend in the number of GLIMPSE clusters decreases with increasing longitude. However, in the south, there are two peaks in the number of clusters, near $l = 330^\circ$ and $l = 313^\circ$. The 2MASS distribution also shows a peak in the number of southern clusters near $l = 330^\circ$ and a peak in the number of northern clusters near $l = 47^\circ$. These enhancements of GLIMPSE and 2MASS clusters could trace the location of spiral arm tangencies. The number of optical clusters appears uniform with increasing longitude in the north and mostly uniform in the south, with a small enhancement near $l = 310^\circ$.

The Galactic latitude distributions of GLIMPSE, 2MASS, and optical clusters are shown in Figure 5. The majority of GLIMPSE clusters are concentrated within $0^\circ.5$ latitude of the midplane. There is a population of 2MASS clusters slightly offset from the midplane. The distribution of optical clusters appears mostly uniform, both above and below the midplane. A closer examination of these distributions shows that the standard deviations of the latitudes of GLIMPSE and 2MASS clusters (limited to those located in the GLIMPSE survey area) are similar. That is, the spread

TABLE 1

NEW STAR CLUSTERS DISCOVERED IN THE GLIMPSE SURVEY DATA

ID (1)	l (deg) (2)	b (deg) (3)	α (J2000.0) (4)	δ (J2000.0) (5)	R (arcmin) (6)	N_{stars} (7)	Detection Method[a] (8)	Comments (9)
1	12.753	−0.146	18 13 55	−17 56 55	0.8	47	A	b,c
2	13.994	−0.124	18 16 20	−16 50 51	0.7	36	V	c,d
3	14.129	−0.644	18 18 30	−16 58 31	0.6	87	A	d,e,f
4	15.176	−0.638	18 20 32	−16 02 59	0.7	77	A	b,c,f
5	17.594	−0.110	18 23 19	−13 40 02	0.6	114	A	d,e,f
6	18.911	−0.054	18 25 38	−12 28 38	0.3	22	A	b,f
7	18.956	−0.056	18 25 44	−12 26 18	0.6	63	A	b,f,g
8	20.640	−0.022	18 28 49	−10 55 55	0.5	11	A	d,e,f
9	22.752	−0.399	18 34 08	−09 14 02	0.3	48	A	d,e,f
10	23.221	−0.335	18 34 47	−08 47 17	0.8	86	A	b,g
11	29.957	−0.193	18 46 41	−02 44 07	0.5	25	A	b,f
12	30.790	−0.078	18 47 48	−01 56 30	0.8	27	A	b,c,f
13	33.771	−0.261	18 53 53	+00 37 39	0.5	51	A	d,e,f
14	35.125	−0.747	18 58 05	+01 36 38	0.3	45	A	c,d,e,f,h
15	38.926	−0.354	19 03 40	+05 10 14	0.8	61	V	c,d
16	38.948	−0.530	19 04 20	+05 06 33	0.3	18	A	b,c,f
17	42.258	−0.209	19 09 19	+08 11 45	0.3	21	A	b,e,f
18	42.432	−0.264	19 09 50	+08 19 30	1.1	86	V	c,d
19	43.213	+0.961	19 06 53	+09 34 57	0.4	44	A	b,e,f
20	44.177	−0.072	19 12 25	+09 57 40	0.5	78	A	d,e,f,h
21	48.603	+0.028	19 20 30	+13 55 24	1.8	257	V	c,d
22	52.234	+0.740	19 25 00	+17 27 38	1.0	78	V	c,d
23	53.772	+0.164	19 30 13	+18 32 15	0.4	45	A	b,e,f,g
24	56.249	−0.158	19 36 30	+20 32 58	0.7	39	V	c,d
25	62.932	+0.086	19 50 06	+26 28 00	0.6	53	A	c,d,g
26	296.296	−0.489	11 53 15	−62 36 31	0.4	47	A	d,e,g
27	296.719	+0.360	11 58 21	−61 52 11	0.4	66	A	b,e,g
28	297.393	−0.625	12 02 24	−62 58 06	0.8	59	V	c,d
29	297.513	−0.769	12 03 12	−63 07 56	1.3	144	V	c,d
30	298.756	−0.408	12 14 32	−62 58 49	0.3	39	A	d,e,f,g,h,
31	298.863	−0.434	12 15 27	−63 01 16	1.7	189	V	c,d
32	300.132	−0.086	12 26 52	−62 49 27	0.3	28	A	c,d,e,g
33	300.506	−0.176	12 30 05	−62 56 50	0.8	34	V	c,d
34	302.433	−0.105	12 47 03	−62 58 21	1.3	159	V	c,d
35	304.846	+0.080	13 08 09	−62 43 46	0.3	47	A	b,e,g,h
36	305.384	−0.237	13 13 04	−63 00 21	0.8	43	V	c,d
37	305.835	−0.061	13 16 52	−62 47 22	1.2	122	V	c,d
38	309.084	+0.171	13 44 20	−62 04 02	0.9	122	A	d,g
39	309.271	−0.609	13 47 19	−62 47 27	0.9	31	A	d,f
40	309.890	+0.400	13 50 36	−61 40 12	0.9	40	V	c,d
41	310.990	+0.436	13 59 28	−61 22 00	0.3	33	A	b,c,g
42	311.190	+0.641	14 00 37	−61 06 59	1.4	38	A	b,c,f
43	311.207	+0.770	14 00 28	−60 59 15	0.9	29	A	b,c,f
44	311.483	+0.360	14 03 36	−61 18 29	0.8	30	V	c,d
45	311.488	−0.456	14 05 34	−62 05 24	0.8	72	V	c,d
46	311.901	+0.085	14 07 36	−61 27 12	0.9	50	V	c,d
47	311.936	+0.205	14 07 35	−61 19 42	0.9	65	V	c,d
48	311.979	+0.224	14 07 53	−61 17 51	0.3	26	A	b,e,f
49	312.975	−0.434	14 17 31	−61 36 57	0.2	17	A	b,c,f
50	313.279	−0.335	14 19 39	−61 25 19	0.2	21	A	c,d,e,f
51	313.788	+0.709	14 20 42	−60 16 04	0.6	38	V	c,d
52	314.226	+0.343	14 25 03	−60 27 35	0.6	41	V	c,d
53	315.981	−0.284	14 40 08	−60 22 20	0.4	39	A	d,e,f,g
54	316.772	−0.060	14 45 10	−59 50 24	0.7	65	A	b,c,f,g
55	316.844	−0.028	14 45 34	−59 46 50	0.1	11	A	b,c,f
56	317.422	+0.082	14 49 19	−59 25 54	0.5	29	A	b,c,f
57	319.164	−0.362	15 03 00	−59 01 23	0.8	96	A	b,c,g
58	319.402	−0.013	15 03 19	−58 36 09	1.1	106	V	c,d
59	320.245	+0.436	15 07 13	−57 47 52	1.0	113	V	c,d
60	321.124	−0.520	15 16 36	−58 10 07	0.6	101	A	b,c,g
61	326.142	−0.371	15 45 59	−55 10 28	0.7	61	V	c,d
62	326.675	+0.528	15 45 05	−54 08 14	1.8	155	A	c,d,f,h
63	326.905	−0.299	15 49 50	−54 38 42	0.8	24	A	b,c,f

TABLE 1—*Continued*

ID (1)	l (deg) (2)	b (deg) (3)	α (J2000.0) (4)	δ (J2000.0) (5)	R (arcmin) (6)	N_{stars} (7)	Detection Method[a] (8)	Comments (9)
64	327.554	−0.828	15 55 37	−54 38 38	0.2	26	A	b,e,f
65	328.192	−0.570	15 57 50	−54 02 09	1.2	58	V	c,d
66	328.196	−0.477	15 57 27	−53 57 44	1.0	90	V	c,d
67	328.312	−0.589	15 58 33	−53 58 21	1.0	89	V	c,d
68	328.809	+0.878	15 54 46	−52 31 47	0.5	78	A	d,e,f
69	329.656	−0.481	16 04 53	−53 00 30	0.9	171	V	c,d
70	329.697	+0.584	16 00 27	−52 10 49	0.4	54	A	b,e,f
71	331.278	−0.411	16 12 25	−51 51 43	0.0	33	A	b,c,f
72	331.342	−0.363	16 12 30	−51 46 59	0.6	52	A	b,c,f,h
73	331.571	−0.009	16 12 01	−51 22 06	0.3	22	A	b,e,f
74	332.162	−0.455	16 16 45	−51 17 04	0.5	14	V	c,d
75	332.364	+0.294	16 14 22	−50 36 13	0.4	45	A	b,f
76	332.652	−0.620	16 19 43	−51 03 37	1.0	55	V	c,d
77	332.782	+0.021	16 17 27	−50 30 39	0.5	51	A	b,e,f
78	333.089	−0.498	16 21 08	−50 39 57	0.7	55	A	c,d,f
79	334.522	+0.810	16 21 42	−48 43 40	0.7	37	A	c,d,f
80	336.776	+0.085	16 34 12	−47 36 16	0.4	33	A	b,c,g
81	338.384	+0.111	16 40 24	−46 23 38	0.6	65	A	b,c,f
82	341.302	−0.293	16 52 56	−44 26 03	0.3	25	A	c,d,f
83	342.166	+0.292	16 53 28	−43 23 42	0.4	44	A	b,e,f
84	344.291	−0.640	17 04 38	−42 18 13	0.2	22	A	b,f
85	344.301	−0.568	17 04 22	−42 15 07	0.2	16	A	b,e,f
86	344.770	−0.282	17 04 40	−41 42 21	0.6	41	A	b,f
87	345.357	−0.130	17 05 55	−41 08 47	0.8	113	A	b,c,g
88	345.491	+0.307	17 04 31	−40 46 31	0.9	50	V	c,d
89	345.494	+0.351	17 04 20	−40 44 47	1.0	49	V	c,d
90	345.719	+0.155	17 05 53	−40 41 09	0.7	55	V	c,d
91	348.208	+0.429	17 12 26	−38 31 27	1.1	110	V	c,d
92	349.825	−0.578	17 21 22	−37 47 19	1.1	76	A	c,d,f

NOTE.—Units of right ascension are hours, minutes, and seconds, and units of declination are degrees, arcminutes, and arcseconds.
[a] Clusters detected automatically are labeled A; clusters detected visually are labeled V.
[b] Association of stars.
[c] Associated with mid-IR extended emission.
[d] Cluster of stars.
[e] Detected using no magnitude cut.
[f] Detected in 2MASS catalog with our algorithm.
[g] Appears in 2MASS images.
[h] Detected using "bright" 3.6 μm magnitude cut.

in latitudes of the GLIMPSE and 2MASS clusters is approximately the same. The standard deviation of optical cluster latitudes is larger than that of both GLIMPSE and 2MASS, likely because these clusters are located nearby.

The latitude distribution of all the clusters was fit with an exponential. The distribution and fit are shown in Figure 6. The exponential that best describes the data is $N_{\rm clusters}(|b|) = N_0 e^{-|b|/b_0}$, where $N_0 = 65 \pm 5$ and the angular scale height is $b_0 = 0\overset{\circ}{.}66 \pm 0\overset{\circ}{.}07$. We removed the GLIMPSE clusters from the distribution and refit the exponential to test for an increase in the angular scale height. A larger scale height when GLIMPSE clusters are absent from the distribution would suggest that GLIMPSE clusters probe lower Galactic latitudes. We found no significant increase in the scale height for the 2MASS and optical only clusters sample.

We examined the distribution of cluster membership numbers and cluster diameters of the 59 automatically detected clusters. The 33 visually identified clusters were not considered in our analysis, as the estimates of their size and membership were not determined in a mathematically rigorous manner. Figure 7 shows the membership number distribution, and Figure 8 shows the size distribution. We find that more than half of the new GLIMPSE clusters have 35 stars or more. Our membership numbers are consistent with previous derivations for the minimum stellar density of a cluster. Lada & Lada (2003) define a stellar cluster as a group of 35 or more stars, based on evaporation timescales. The distribution of cluster diameters shows that most of the GLIMPSE clusters are compact, with diameters less than 2′. For cluster sizes of ∼1′ (the peak of the size distribution), if those correspond to 1 pc, the median distances are ∼3 kpc.

5. DISCUSSION

Using the 2MASS images, Dutra et al. (2003) and Bica et al. (2003a, 2003b) cataloged 346 infrared clusters, cluster candidates, or stellar associations. Of these, 130 are located in the GLIMPSE survey area, but only 15% are returned by our automated search of the 2MASS point-source catalog. Since our algorithm did not detect all previously known infrared clusters, it implies that our method of detecting clusters might not be as efficient as visual inspection. It would also imply that there are more GLIMPSE clusters remaining to be found. Because the GLIMPSE survey is confusion limited and because of the varying backgrounds, any

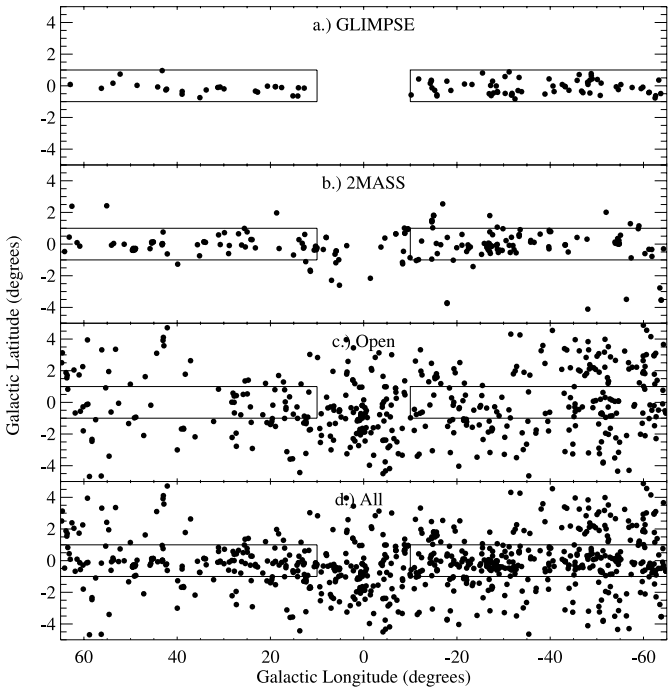

FIG. 3.—Galactic distributions of (*a*) the 92 new GLIMPSE clusters and GLIMPSE GC-01; (*b*) previously discovered 2MASS clusters; (*c*) optical, open clusters; (*d*) all clusters. Thin rectangles denote the GLIMPSE survey region.

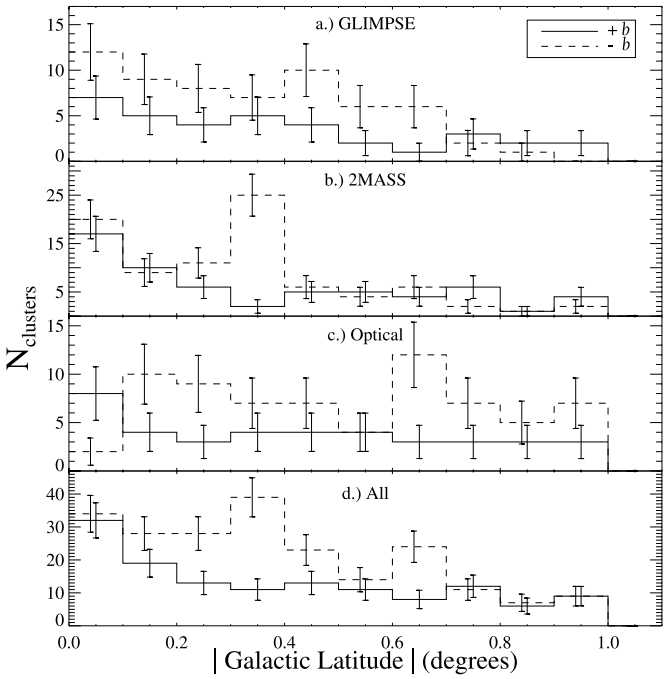

FIG. 5.—Latitude distributions of (*a*) GLIMPSE clusters and GLIMPSE GC-01; (*b*) 2MASS clusters; (*c*) optical clusters; (*d*) all clusters. The solid lines show the distributions of clusters above the midplane, and the dashed lines show the distributions of clusters below the midplane. Error bars were determined using the standard deviation of the binomial distribution.

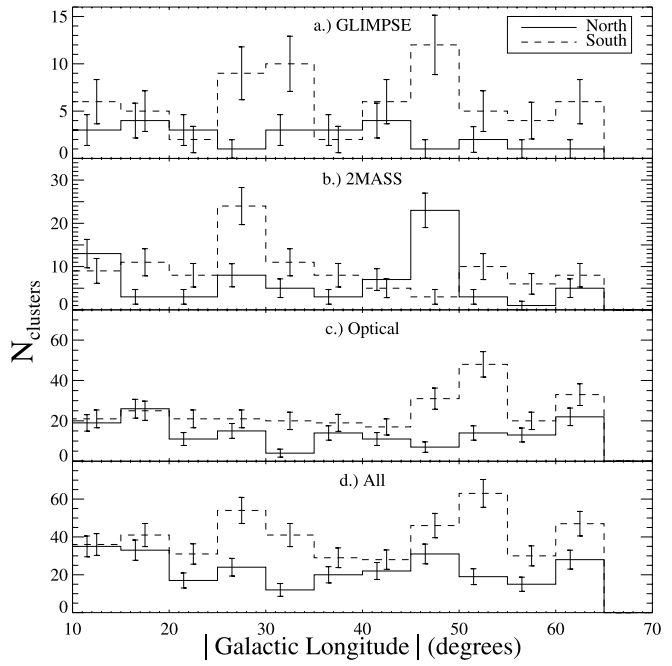

FIG. 4.—Longitude distributions of (*a*) GLIMPSE clusters and GLIMPSE GC-01; (*b*) 2MASS clusters; (*c*) optical clusters; (*d*) all clusters. The solid lines show the distributions of clusters in the northern Galaxy, and the dashed lines show the distributions of clusters in the southern Galaxy. Error bars were determined using the standard deviation of the binomial distribution.

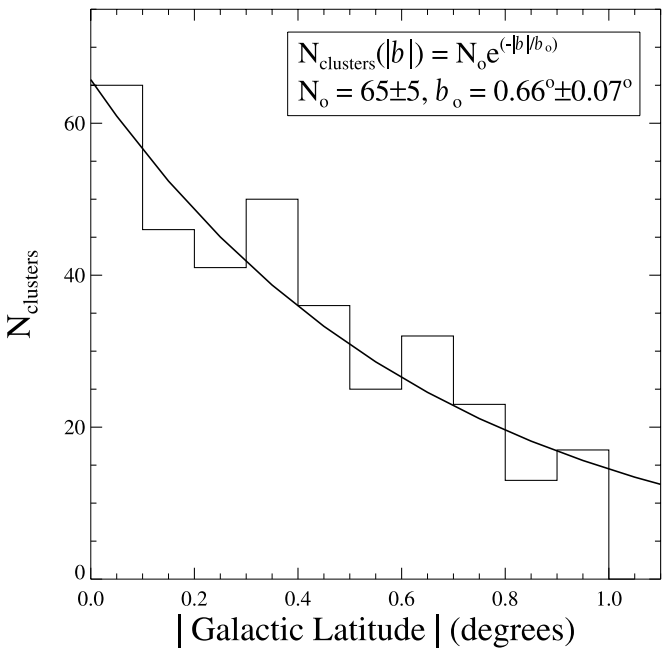

FIG. 6.—Latitude distribution of all clusters (GLIMPSE, 2MASS, and optical) within 1° of the Galactic midplane. An exponential curve was fit to the distribution to obtain the angular scale height.

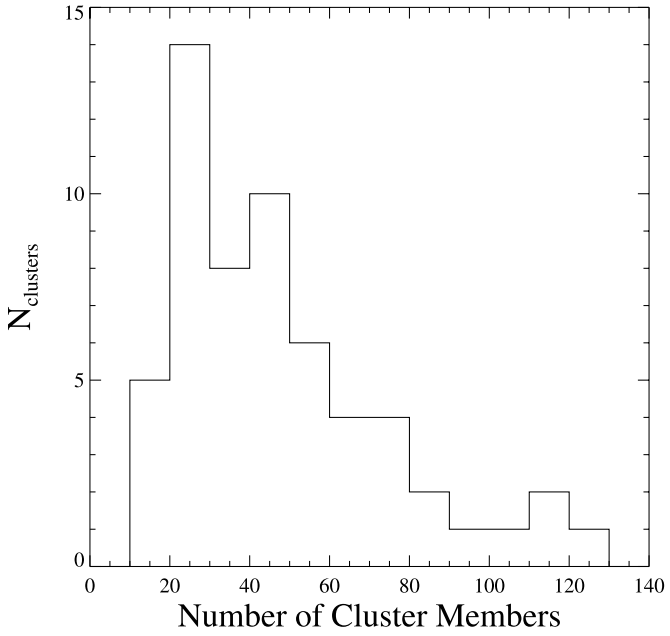

FIG. 7.—Distribution of cluster memberships for the 59 GLIMPSE clusters automatically detected.

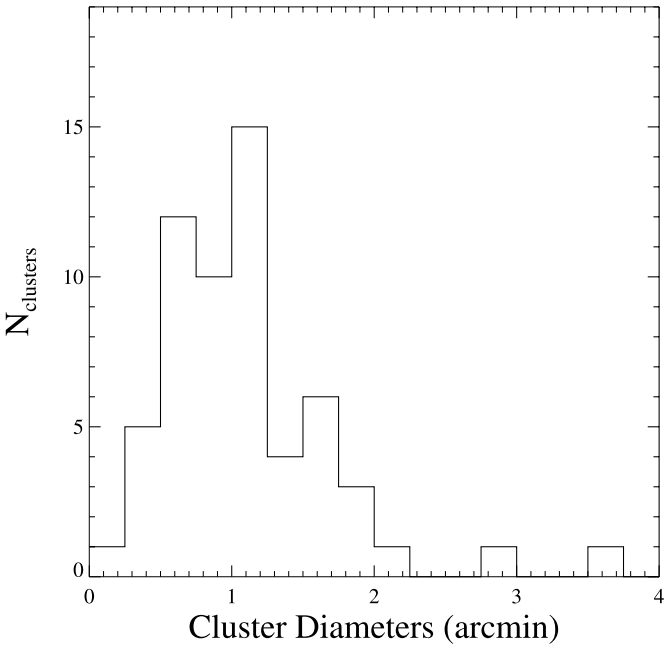

FIG. 8.—Distribution of cluster diameters for the 59 GLIMPSE clusters automatically detected.

detection method is unlikely to be perfect. We have shown, however, that our automated technique works better than simple overdensity automated searches.

In addition to the embedded and open clusters detected in the GLIMPSE data, our algorithm also detected the only two known globular clusters located in the GLIMPSE survey area. One of the globular clusters, 2MASS GC-01, was previously discovered serendipitously in the 2MASS images (Hurt et al. 2000). The other globular cluster, GLIMPSE-C01, was recently discovered by our GLIMPSE team (Kobulnicky et al. 2005) and also found serendipitously.

Heavily embedded clusters seen in the GLIMPSE image mosaics are typically not discovered by our search algorithm. Rather, these clusters are identifiable only by visual inspection. Because these clusters are deeply embedded in gas and dust, they often do not contain enough detectable cluster members to meet our detection significance threshold. The low cluster membership can be attributed to missed detections by the GLIMPSE point-source extraction routine, to extended emission hiding stars, and/or to the clusters being inherently sparsely populated or too dense to resolve individual members. There are, however, only about 30 of these uncataloged, deeply embedded clusters (see Table 1) in the GLIMPSE survey region.

One conclusion could be that mid-IR wavelengths alone are not sufficient for detecting star formation regions, particularly those that contain stellar clusters. The near-IR might be thought to be more appropriate, since it is less contaminated by extended emission. The standard deviations of the latitude distributions of the GLIMPSE and 2MASS clusters are similar, however, suggesting that both wave bands probe latitudes near the Galactic midplane equally well.

The discovery of these new clusters detected in the GLIMPSE data raises several questions with regard to Galactic star formation: How well do we understand the current rate of star formation in our Galaxy? Based on the standard rate, what is the expected number of young embedded clusters? Are we finding a lower than expected number of these young clusters? Are the slight overdensities in cluster numbers at some longitudes revealing spiral arms or other effects? We intend to investigate these issues using the extensive GLIMPSE survey data to better understand the current state of Galactic star formation.

6. SUMMARY

Mid-IR data from the GLIMPSE survey of the inner Milky Way were systematically searched for star clusters. Our automated search routine discovered 59 previously uncataloged infrared star clusters or stellar associations in the GLIMPSE point-source catalog and archive. Of these new clusters, 35 appear embedded or associated with extended emission, while the remainder appear as mid-IR–bright open clusters. We have also visually identified 33 heavily embedded clusters in the GLIMPSE image mosaics that were missed by our detection algorithm.

The automatic detection method used star count map renormalization and statistical models to locate and characterize the clusters. We modeled the GLIMPSE point-source catalog and point-source archive as a mixture of nonhomogeneous Poisson backgrounds with k two-dimensional Gaussian cluster components. The expectation-maximization algorithm was used to improve the cluster locations and to estimate their sizes and number of members. False cluster detections were removed using the Bayesian information criterion technique.

The latitude and longitude distributions of the clusters reveal two distinct asymmetries. More clusters are found below the Galactic midplane, and more clusters are detected in the southern half of the Galaxy. Most of the clusters have memberships greater than 35 stars. This is consistent with the minimum number needed for a cluster to remain intact. Finally, the clusters are also found to be compact in angular size, with diameters less than $3'$.

Support for this work, part of the *Spitzer Space Telescope* Legacy Science Program, was provided by NASA through contracts 1225025 (Boston University), 1224653 (University of Wisconsin),

1224988 (Space Science Institute), 1224681 (University of Maryland), 1256801 (University of Wisconsin–Whitewater), 1242593 (University of California, Berkeley), 1253153 (University of Minnesota), and 11253604 (University of Wyoming) issued by the Jet Propulsion Laboratory, California Institute of Technology under NASA contract 1407.

REFERENCES

Benjamin, R. A., et al. 2003, PASP, 115, 953
———. 2005, ApJ, 630, L149
Bica, E., Dutra, C. M., & Barbuy, B. 2003a, A&A, 397, 177
Bica, E., Dutra, C. M., Soares, J., & Barbuy, B. 2003b, A&A, 404, 223
Churchwell, E. B., et al. 2004, ApJS, 154, 322
Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977, J. R. Stat. Soc., 39, 1
Dias, W. S., Alessi, B. S., Moitinho, A., & Lépine, J. R. D. 2003, in EAS Pub. Ser. 10, Galactic & Stellar Dynamics, ed. C. Boily et al. (Les Ulis: EDP Sciences), 195
Dutra, C. M., Bica, E., Soares, J., & Barbuy, B. 2003, A&A, 400, 533
Fazio, G., et al. 2004, ApJS, 154, 10
Fraley, C., & Raftery, A. 2002, J. Am. Stat. Assoc., 97, 458
Hurt, R. L., Jarrett, T. H., Kirkpatrick, J. D., Cutri, R. M., Schneider, S. E., Skrutski, M., & van Driel, W. 2000, AJ, 120, 1876
Ivanov, V. D., Broissova, J., Pessev, P., Ivanov, G. R., & Kurtev, R. 2002, A&A, 394, L1
Kobulnicky, H. A., et al. 2005, AJ, 129, 239
Lada, C. J., & Lada, E. A. 2003, ARA&A, 41, 57
McLachlan, G. J., & Krishnan, T. 1997, The EM Algorithm and Extensions (New York: Wiley)
Meade, M., et al. 2005, GLIMPSE Data Product Document (Madison: Univ. Wisconsin), http://data.spitzer.caltech.edu/popular/glimpse/20050415_enhanced_v1/Documents/glimpse_dataprod_v1.5.pdf
Reylé, C., & Robin, A. C. 2002, A&A, 384, 403
Skrutskie, M. F., et al. 1997, in The Impact of Large Scale Near-IR Sky Surveys, ed. F. Garzon et al. (Dordrecht: Kluwer), 25
Spitzer Science Center. 2004, *Spitzer* Observer's Manual (Pasadena: SSC), http://ssc.spitzer.caltech.edu/documents/som/
Stetson, P. B. 1987, PASP, 99, 191
Trumpler, R. J., & Weaver, H. F. 1953, Statistical Astronomy (New York: Dover)
Werner, M., et al. 2004, ApJS, 154, 1