

A MULTI-DIMENSIONAL REDUCTION METHOD OF LARGE-SCALE SURVEY DATABASE

Y. LEE¹, Y. S. KIM^{1,2}, H. W. KANG¹, J. H. JUNG¹, C. H. LEE¹, I. S. YIM¹,
B. G. KIM¹, H. G. KIM¹, AND K. T. KIM²

¹Korea Astronomy and Space Science Institute, Daejeon 305-348, Korea

²Chungnam National University, Daejeon 305-764, Korea

E-mail: yulee@kasi.re.kr

(Received November 13, 2012; Accepted January 16, 2013)

ABSTRACT

We present a multi-dimensional reduction method of the surveyed cube database obtained using a single-dish radio telescope in Taeduk Radio Astronomy Observatory (TRAO). The multibeam receiver system installed at the 14 m telescope in TRAO was not optimized at the initial stage, though it became more stabilized in the following season. We conducted a Galactic Plane survey using the multibeam receiver system. We show that the noise level of the first part of the survey was higher than expected, and a special reduction process seemed to be definitely required. Along with a brief review of classical methods, a multi-dimensional method of reduction is introduced; It is found that the ‘background’ task within IRAF (Image Reduction and Analysis Facility) can be applied to all three directions of the cube database. Various statistics of reduction results is tested using several IRAF tasks. The rms value of raw survey data is 0.241 K, and after primitive baseline subtraction and elimination of bad channel sections, the rms value turned out to be 0.210 K. After the one-dimensional reduction using ‘background’ task, the rms value is estimated to be 0.176 K. The average rms of the final reduced image is 0.137 K. Thus, the image quality is found to be improved about 43% using the new reduction method.

Key words: survey; molecular cloud; multi-beam receiver; reduction

1. INTRODUCTION

Molecular clouds, which are mostly but irregularly distributed on the Galactic Plane, are the birthplaces of all stars and planets. The study of these molecular clouds starts with a survey covering a substantially large region of the Galactic Plane; The formation and distribution of various kinds of molecular clouds, including giant molecular clouds (GMCs) and dense cores, and their physical properties had been investigated with the database homogeneously constructed. Survey of these molecular clouds had been conducted in many different ways for the last three decades (Dame et al., 1987, 2001; Sanders et al., 1986; Heyer et al., 1998; Lee et al., 2001; Jackson et al., 2006). These surveys were conducted with single-beam (or pixel) receivers initially, and later with multi-beam array receiver systems to use the telescope time more efficiently.

The present receiver system of Taeduk Radio

Astronomy Observatory (TRAO) is the fifteen beam array imported from Five College Radio Astronomy Observatory (FCRAO). However, as this receiver with Schottky-diode mixers is an old type and has been getting aged, six of them were out of order at the moment and only 9 pixels are operating in 2012 season. More explanation of the receiver system was introduced in Lee et al. (2010). The database obtained from these surveys are usually in large scale. A consistent and homogeneous reduction process would be required for the large database. This reduction process can be tested in various methods. A fast reduction method for a larger scale database obtained from single-dish radio telescopes was introduced in previous paper (Lee, 2001).

Recently, we started a survey toward the second quadrant Galactic Plane (TOGS: TRAO Outer Galaxy Survey; Lee et al., 2010, 2013) in ¹³CO(1-0) using TRAO’s multi-beam receiver system. The first target region of TOGS

is the exact Galactic Plane, the partial region of ^{12}CO Outer Galaxy Survey (OGS; Heyer et al., 1998); The longitude range is from 102.5° to 141.5° , and the latitude range is from -1° to $+1^\circ$. Its sampling grid is 50 arcsec and the velocity resolution is 0.63 km s^{-1} . However, at the first year operation, our new survey data obtained using this receiver system were found to be noisier and more spurious than expected, and thus, more specific and refined reduction process seemed to be required. It was found that the performance of TRAO receiver system had been downgraded and not optimized at initial stage, thus the quality of the data was fairly low, though its performance has been upgraded recently.

We dealt with database of the first year observation: the longitude range of the survey is from 108° to 112.3° , and the latitude range is from -1° to $+1^\circ$. Classical and previous reduction methods of spectra obtained from single-dish radio telescope are reviewed in Section 2. We present a new fast method of data reduction in Section 3. In Section 4, the results and their statistics are discussed and summarized.

2. REVIEW OF CLASSICAL REDUCTION METHODS

As the volumes of survey data were substantially enlarged, faster methods were required to take off heavy load of data reduction work. While an extended explanation of a faster reduction method had been described in detail by Lee (2001), in this section we briefly review classical and previous reduction methods for three-dimensional (cube) spectral survey data obtained from a single-dish radio telescope. Polynomial baseline subtraction, bad channel elimination, and subtraction of ripple patterns (or sine waves) are the most common and basic reduction processes of the spectra. The classical reduction methods of spectroscopic data obtained from a large-scale survey were mostly based on manual mode, reducing one-by-one, though one may run a limited batch job for a stack of spectra. SPA (FCRAO and TRAO), CLASS (IRAM-Grenoble), Karma (Australia), DRS (Nagoya University) are the reduction packages (set of tasks with similar theme) provided by individual observatories equipped with single-dish radio telescopes. Though each one has its own merits and defects, CLASS seems to be the most widely used package as it provides various analysis tasks (programs) as well as reduction tools. Before reducing spectra one by one, one may transform a stack of raw spectra onto a three-dimensional grid, constructing a

cube data; velocity and galactic coordinate (v, l, b), or velocity and equatorial coordinate (v, α, δ). Each section of these cube data can be displayed on a two-dimensional image displayer, such as saomage DS9 (provided by Harvard-Smithsonian Center for Astrophysics) or ximtool, when using IRAF.

By transforming stacks of spectra into image and displaying it before any careful reduction process, one can recognize the patterns of the baselines, bad pixels, and the quality of data visually all at once. This process can be done within several astronomical reduction packages. In this paper we will focus on IRAF, which is widely being used in most of astronomical communities. Transformation of spectra to image before conducting individual reduction had previously been suggested by FCRAO software group, when ‘fcrao’ reduction package was developed (Heyer & Carpenter, 1991, in private communication). Within IRAF the obtained spectra can be transformed to IRAF data format or FITS format. For the spectroscopic survey data, it would be three-dimensional cube data and it can be handled as a regular database within IRAF. Image displaying and recognizing the patterns are the major merits of this method. Another merit is displaying the composite spectra of any image section up to a few hundred spectra at once. However, ‘fcrao’ reduction package within IRAF does not include statistics and ripple pattern subtraction routines, thus the ‘fcrao’ reduction routines are not enough to efficiently handle a large amount of spectra; ‘Baseline’ routine subtracts a polynomial baseline from a spectrum. The coefficients in the polynomial are determined from a least square fit to the observational data in the specified intervals. After the fitted baseline has been subtracted, the rms noise level is calculated for the data in the specified intervals. Though there are higher order polynomials, linear, second and third order fittings are usually used. To subtract ripple pattern(s) we can use ‘SINE’ command within SPA. It subtracts a sine wave, for which the period, phase, and amplitude are obtained from a least squares fit to the observed spectrum in the specified intervals.

It was reported by Lee (2001) that subtraction of baselines and ripple patterns could be done very efficiently using ‘background’ task within IRAF. After the fitted baselines and ripple patterns are subtracted, the rms noise level is estimated for specified velocity intervals, and for any subsection of the images.

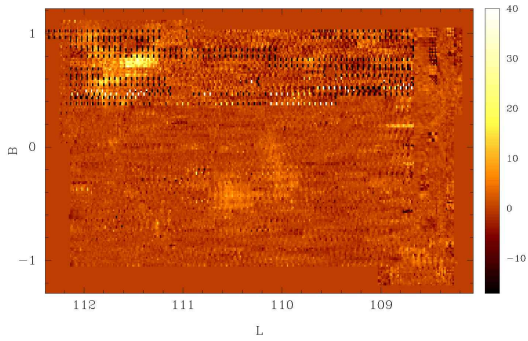


Fig. 1. Raw integrated intensity map of the TOGS Region I. The upper part of the image is noisier than the other part, as the initial stage of receiver system was not stable, but later on it was stabilized. The x axis represents galactic longitude, and y axis is galactic latitude. Unit of the color scale is K km s^{-1} .

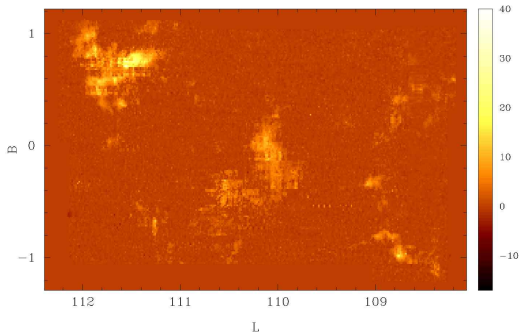


Fig. 2. The integrated intensity map of the TOGS Region I after initial reduction process along the velocity axis only using ‘background’ task. Unit of the color scale is K km s^{-1} .

3. A MULTI-DIMENSIONAL REDUCTION METHOD

The most important reduction process of the survey data is how to subtract the baseline efficiently and homogeneously. As mentioned above, Lee et al. (2001) showed that it could be a very good choice to use an IRAF task, ‘background’ in package of ‘twod.longslit’ for baseline removing process, instead of using regular methods provided in several observatories’ reduction packages.

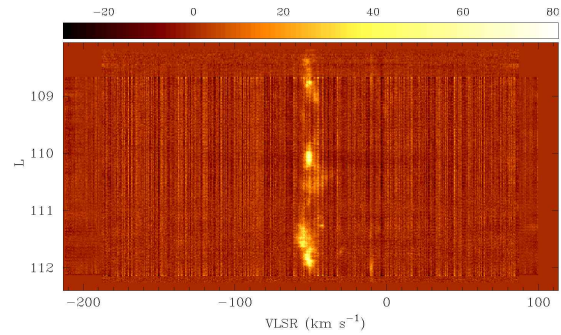


Fig. 3. Spatial-velocity integrated intensity map of the TOGS Region I after one dimensional reduction process using ‘background’ task for velocity direction was applied; The x axis represents velocity and y axis represents galactic longitude. Unit of the color scale bar is K arcmin .

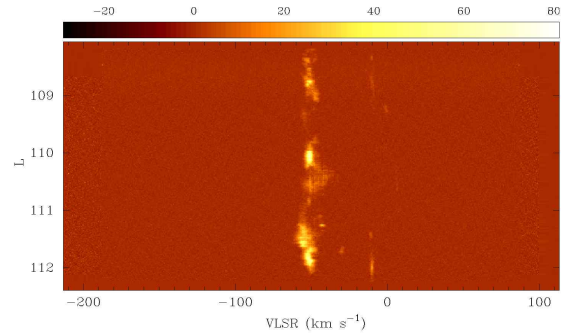


Fig. 4. Spatial-velocity integrated intensity map of the TOGS Region I after three dimensional reduction method was applied. Unit of the color scale bar is K arcmin .

The ‘background’ task provides a very efficient way of subtracting baselines, fitting each row (line) or column of an image interactively or in a batch job mode with varying functions specified by the sample parameters. Before applying the functions to a spectrum, we can set up low and high rejection limits in units of the residual sigma of each spectrum. If ‘low_reject’ and/or ‘high_reject’ value are larger than zero, the sigma of the residuals between the fitted points and the fitted function is computed, and

those points whose residuals are less than $\text{low_reject} \times \sigma$ and greater than $\text{high_reject} \times \sigma$ are excluded from the fit. Points within a ‘grow’ pixels of a rejected pixels are also excluded from the fits. ‘grow’ is the number of pixels within the distance of the rejected pixels to avoid the emission related feature. The function is then refit without the rejected points. Other detailed explanation was described in Lee et al. (2001).

In Figure 1, the integrated intensity map of the Region I of TOGS is represented. As the initial performance of the TRAO receiver system was not in optimum status, first part of the data was much noisier than expected (Figure 1). After an initial reduction (for one-direction, velocity) of data using ‘background’ task, which is the usually conducted, it apparently seemed to be well reduced as the integrated intensity (over velocity direction) map was clean much more upgraded (Figure 2). Its statistics showed that the noise distribution of negative pixel values was almost natural gaussian (see Section 4). Some bad channel elimination and baseline subtraction along the velocity direction was applied using the method introduced by Lee (2001). However, when integrated over the longitude direction and latitude direction instead of velocity direction, it is found that the spatial-velocity integrated intensity map (L-V map) show very noisy map (Figure 3). It is very obvious that there might be certain characteristic pattern of channels, which prompted into view when integrated. Another spatial-velocity integrated intensity map (B-L map) gives the similar result, which we did not presented in this paper. Though the cause of this pattern is still unknown and remained to be discussed, it is likely to be combination of various unstable performance of the aged receiver system and the correlator’s weird responses. In fact, out of 512 channels of the correlator, the first 40 and last 40 channels often show much more noisier values than the rest of channels. When integrated over the well-performing channels except these bad channels, the results show the same pattern as Figure 3. No matter what would be the reason, it does imply that the more refined reduction process is definitely required. We looked into the images for all three directions in detail using IRAF image displayer DS9, and checked their statistics for various velocity sections and several spatial sections.

As we mentioned that after reduction for one direction (velocity axis) using the ‘background’ task, the spatial velocity integrated intensity map (L-V map) seemed to be still noisy (Figure 3). We tested several ways to improve

the image quality, lowering its noise level, and found that applying this method to other two directions endows us the best result. When using ‘background’ task, one may conduct fitting procedure interactively and individually, or in a batch job mode with a function to the columns or lines specified by the sample parameter. This function may be a Legendre polynomial, Chebyshev polynomial, linear spline, or cubic spline of a given order or number of spline pieces. This function is then subtracted from the entire line or column to create an output line or column. Before applying the functions to a spectrum, we can set up low and high rejection limits in units of the residual sigma of each spectrum, which will be excluded from the fit. Points within a distance of ‘grow’ pixels of a rejected pixel are also excluded from the fit.

The most important parameters were found to be high rejection and the growing radius as the real emission features can explicitly be excluded by introducing the proper high rejection parameter and rejection growing radius. The searching process of these two parameters were very tricky, thus we tested several ways to find the best combination. After several trial and error, we found that the proper values of the parameters which endow us the best reduction results; Fitting function is spline3, order of fitting function is 10, high rejection sigma is 3, and the rejection growing radius is 3. After subtracting the baselines for all three directions (for three dimensions), the L-V map is presented in Figure 4.

4. REDUCTION STATISTICS AND DISCUSSION

We present the rms sensitivity images (Figures 5, 6 and 7) of raw data, after one dimensional reduction and three dimensional reduction, respectively. These rms images, which represent noise level, were produced using the ‘newimage’ task within IRAF. More reddish and darker color represents lower rms, and the brighter color is noisier and higher rms. In fact, total black color in the background means zero rms value, where no data were obtained. The upper part was observed when the TRAO receiver system was installed two years ago and its performance was not optimized. The rest part (from middle to bottom) except the righthand part was observed after it became more stabilized in 2011. The righthand part was observed year at the following season when it was more stabilized in 2012. It is noticeable rms value of the image was substantially upgraded when comparing the Figure 5 (raw image), Figure 6 (after one-direction reduction) and

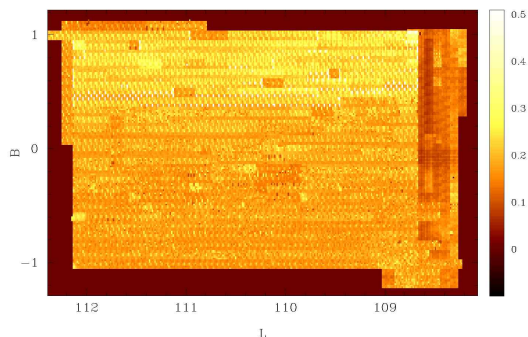


Fig. 5. The rms sensitivity image of raw data. Unit of the color scale bar is K.

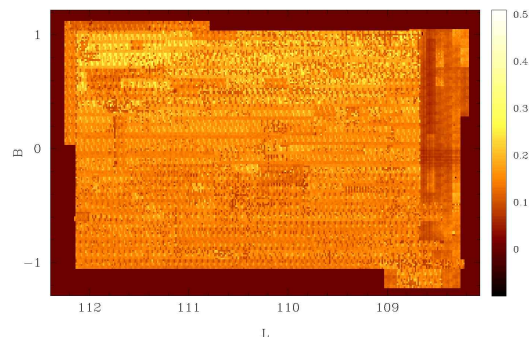


Fig. 7. The rms image after final reduction. Unit of the color scale bar is K.

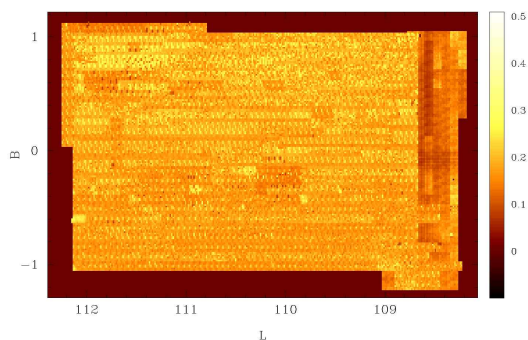


Fig. 6. The rms sensitivity image after one direction reduction. Unit of the color scale bar is K.

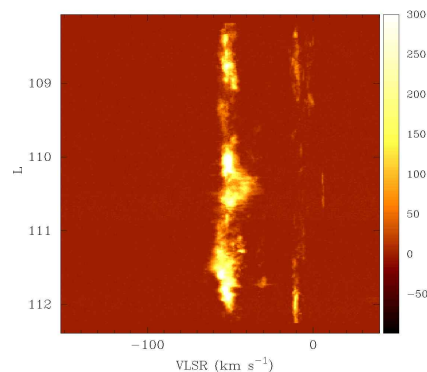


Fig. 8. ^{12}CO spatial-velocity integrated intensity map of the Region I. Unit of the color scale bar is K arcmin.

Figure 7 (after final reduction). In fact, the rms value of Figure 5 is 0.241 ± 1.598 K, and it is 0.137 ± 0.039 K for Figure 7, which were estimated using ‘imstatistics’ task within IRAF. The large error of the raw image is caused by the many bad channels involved, especially at the upper part of the image. After primitive baseline subtraction and elimination of these bad channels, the rms value turned out to be 0.210 ± 0.120 K. After one dimensional reduction using ‘background’ task, the rms value is estimated to be 0.176 ± 0.035 K (Figure 6). One may noticed that the rms values of the final rms image (Figure 7) are different; It is 0.154 ± 0.047 K for the upper part, 0.139 ± 0.029 K for the middle part, and 0.100 ± 0.028 K for the righthand part, and again, the average rms of the whole image is 0.137 ± 0.039 K. Thus, the image quality was improved about 43% in

terms of rms value after final reduction. The rms image of TOGS can be compared with that of other surveys, such as Galactic Ring Survey (GRS; Figure 7 of Jackson et al., 2006). GRS is the most recent large survey database toward the Galactic Ring. The difference between high rms and low rms values are much larger than our result as their observing mode was two-fold, initially position-switching mode and the later On-The-Fly (OTF) mode. Their rms map showed that the OTF mode gave much better result.

More important result is that the L-V integrated intensity map (Figure 5) is substantially improved. All the stripe-like features were remarkably eliminated. To compare our result to the ^{12}CO L-V integrated intensity map, we obtained it from OGS (Heyer et al., 1998). The L-V map of OGS for the same region of TOGS is presented

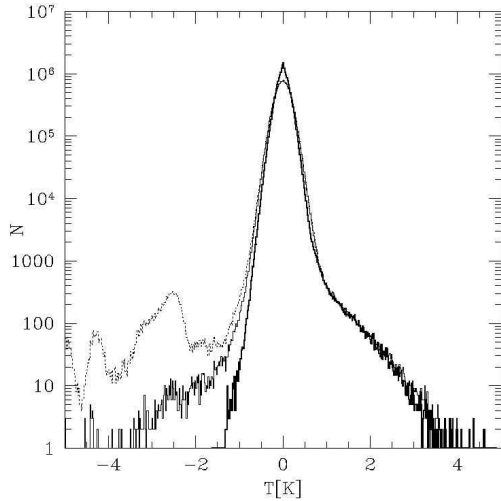


Fig. 9. Distribution of all voxels as a function of temperature. The dotted line shows raw data, the thin solid line for the data after excluding the bad channels of the first and last part of correlator channels, and the thick solid line represents the distribution of final reduced data.

in Figure 8. As the velocity coverage of OGS was much smaller than TOGS, its range of x axis is narrower. The emission features of both maps are remarkably similar to each other. Most of ^{12}CO emission feature is clearly noticeable in ^{13}CO L-V map, even the weak part of them. Though it is meaningless to mention the rms error bar ranges, the rms value of Figure 3 is estimated about 4.43 K in pixel unit, and that of Figure 4 (after final reduction) is 0.82 K in pixel unit.

Another way to evaluate the reduced image statistically is checking the temperature distribution histogram. A histogram of the number of l , b , v voxels (pixels in 3d) versus temperature for the ^{13}CO survey data of the Region I is shown in Figure 9. The temperatures shown cover from -5 K to $+5$ K, though a few voxels extend to 5.75 K, and the bin size of the histogram is 0.1 K. The raw data are presented with the dotted line. As mentioned previously, the first and last part of the correlator behaved weirdly, and thus many bad channels, showing especially negative values, and the rms sensitivity could not be estimated properly. The distribution of these negative features can be notified in lefthand side. After excluding these bad-data channels, the thin solid line represents its

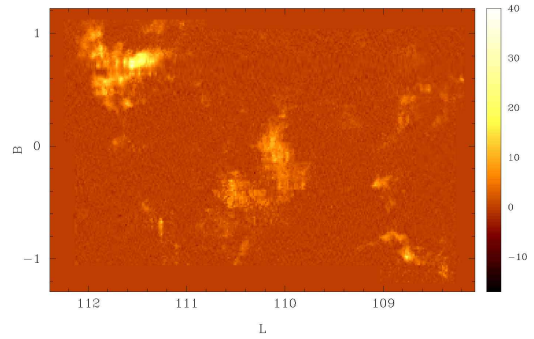


Fig. 10. Integrated intensity map of the TOGS Region I after final reduction. Unit of the color scale is K km s^{-1} .

distribution after initial reduction in one direction. After final reduction process, thick solid line represents the temperature distribution; The negative part is almost perfect gaussian which is the remark of well reduced data (Figure 9). The final integrated intensity map of the surveyed region was presented in Figure 10 after all the reduction process. This map also shows a complete emission features when comparing to ^{12}CO integrated intensity map (not shown here), similarly to Figures 4 and 8 of L-V integrated intensity maps. Considering these maps, the temperature distribution (Figure 9) and rms sensitivity improvement, our method of reduction seems to be working properly and efficiently, and most importantly it would be one of the fastest way to reduce noisy database.

REFERENCES

- Dame, T. M., Ungerechts, H., Cohen, R. S., de Geus, E. J., Grenier, I. A., May, J., Murphy, D. C., Nyman, L.-A., & Thaddeus, P., 1987, A Composite CO Survey of the Entire Milky Way, *ApJ*, 322, 706
- Dame, T. M., Hartmann, D., & Thaddeus, P., 2001, The Milky Way in Molecular Clouds: A New Complete CO Survey, *ApJ*, 547, 792
- Heyer, M. H., Brunt, C., Snell, R. L., Howe, J. E., Schloerb, F. P., & Caroentem, J. M., 1998, The Five College Radio Astronomy Observatory CO Survey of the Outer Galaxy, *ApJS*, 115, 241
- Jackson, J. M., Rathborne, J. M., Shah, R. Y., Simon, R., Bania, T. M., Clemens, D. P., Chambers, E. T., Johnson, A. M., Dormody, M., Lavoie, R., & Heyer, M. H., 2006, The Boston University-Five College Radio

- Astronomy Observatory Galactic Ring Survey, ApJS, 163, 145
- Lee, Y., 2001, A Fast Reduction Method of Survey Data in Radio Astronomy, JKAS, 34, 1
- Lee, Y., Stark, A. A., Kim, H. G., & Moon, D. -S., 2001, The Bell Laboratories ¹³CO Survey: Longitude-Velocity Maps, ApJS, 136, 137
- Lee, Y., Jung, J. H., Kang, H. W., Lee, C. H., Kim, H. G., Yim, I. S., & Kim, B. G., 2010, Observational Test Study of TRAO Outer Galaxy Survey Comparing to FCRAO Outer Galaxy Survey, PKAS, 25, 23
- Lee, Y., Jung, J. H., Kim, Y. S., Kang, H. W., Lee, C. H., Kim, H. G., Yim, I. S., Kim, B. G., & Kim, K. T., 2013, in preparation
- Sanders, D. B., Clemens, D. P., Scoville, N. Z., & Solomon, P. M., 1986, Massachusetts-Stony Brook Galactic Plane CO Survey. I - (b, V) Maps of the First Galactic Quadrant, ApJS, 60, 1