Searching for Milk Way Substructures using a matched filter

Queiroz A.,Santigo B., Luque E., Balbinot E.
UFRGS

July 11, 2014

Abstract

This work has the main goal of detect new stellar objects in our Galaxy. To achieve this goal we use the matched filter technique combined with a simulator of simple stellar populations. The first use of this method was made on the photometric data from the Sloan Digital Sky Survey III that includes the entire region covered by Baryon Oscillation Spectroscopy Survey on the Galactic South plus the previous imaging regions from the Sloan Digital Sky Survey. The initial step is to validate the method on the known satellites from the Sloan Digital Sky Survey. We successfully detected both dwarfs galaxies and star clusters, among the stellar clusters detected there are Koposov I and II, Whiting I, Segue 3, Balbinot 1 and Palomar 13. Among the dwarf galaxies there are Canes Venatici I and II, Hercules, Leo IV, Leo V, Leo T, Coma Berenices, Segue I and II and Wilman I. We also detected the tails of the Sagittarius dwarf, showing that the method is ready to be applied to new data sets, such as the Dark Energy Survey.

1 Introduction

The search of stellar substructures in our Galaxy can help us understand several of its aspects, such as the structure of the stellar halo, the initial episodes of mass assembly and star formation in the Galaxy, the matter content and gravitational potential in the dark halo, and even the nature of the dark matter particles themselves. An increased census of faint Milk Way (MW) satellites will also allow the extension of scaling relations towards very small masses, radii, luminosities and surface brightnesses. Before the advent of large and homogeneous surveys, such as the Sloan Digital Sky Survey (SDSS; http://www.sdss.org/) and The Two Micron All Sky Survey (2MASS), the MW was known to host only a handful of satellites galaxies, including the two Magellanic Clouds and a few brighter dwarfs. The situation has dramatically changed in the past 15 years, when about 15 new faint satellites were discovered, mostly using SDSS, most of them considerably less luminous than a typical globular cluster. The recently initiated Dark Energy Survey (DES, http://www.darkenergysurvey.org/) is also expected to reveal extra MW satellites. In fact, Tollerud et al (2008, ApJ, 688, 277) have predicted from 19 to 37 new objects from DES, assuming that the detection characteristics of this survey are similar to those from SDSS. Rossetto et al (2011, ApJ, 141, 185), on the other hand, argue that DES will probe the main sequence and its turn-off in a range of distances not reached by SDSS, therefore increasing the detectability of satellites out to 120kpc.

In this report we present a method to detect resolved stellar systems by means of a matched filter (MF). This work is an extension of the Sparse code, developed by Balbinot et al 2011 (MNRAS, 416, 393) to detect a tidal tail around the low galactic latitude globular cluster NGC 2298. These authors, and others before them, have used the matched filter to detect stars belonging to a previously known population, such as a star cluster , Grillmair C. J. 2012 (ASP,458,219), Davenport et al 2009 (AAS,41,321), To search for unknown structure, it is necessary to build a grid of filters for different stellar populations, spanning a wide range of ages, distances, and metallicities. This is done here by using simulated simple stellar populations (SSPs). It is also necessary to search for concentrations (or peaks) of filtered stars on the sky on a large number of maps. This is done with the aid of the Sextractor code (Bertin & Arnouts 1996). In Section 2, we explain the method in more detail, discussing the implementation of the matched filter, the grid of SSP simulations, and the pipeline based on Sextractor runs. The data used are presented in Section 3. Section 4 describes the main results. Finally, in Section 5 we present our conclusions and perspectives.

2 Method

2.1 Matched Filter

The Matched Filter (MF) is a technique that has several applications for signal processing. In the context of Astronomy, it has been used mainly to detect low density SSPs on imaging data. We here use it to find new clusters and dwarf galaxies following the work of Balbinot et al 2011 (MNRAS, 416, 393). We start by modeling the number of stars as a function of position on the sky (α,δ) and of colour (c) and magnitude (m).

$$N(\alpha, \delta, c, m) = N_{cl}(\alpha, \delta, c, m) + N_{bg}(\alpha, \delta, c, m) \tag{1}$$

The first term on the right hand side corresponds to the contribution by the object we want to discover, whereas the second term makes up the background. Both quantities may be split into a normalization term and a probability distribution function (PDF). By definition, an SSP is described by a single CMD distribution, but the number density of its stars may vary as a function of position. We thus write:

$$n_{cl}(\alpha, \delta, c, m) = \xi_{cl}(\alpha, \delta) f_{cl}(\alpha, \delta, c, m) \tag{2}$$

where ξ_{cl} and f_{cl} are the SSP number normalization and PDF on the CMD plane, respectively. In the case of the background stars belonging to the Galactic field population, we know that both the number and the CMDs vary as a function of position across the sky. So, we write

$$n_{bg}(\alpha, \delta, c, m) = \xi_{bg}(\alpha, \delta) f_{bg}(\alpha, \delta, c, m) \tag{3}$$

Note that the functions f_{cl} and f_{bg} are statistics functions that describe the probability that a randomly picked star belonging to the SSP or to the background is found in a given CMD position. Their integral over the CMD plane, therefore, normalizes to unity. With the definitions above, eq (1) then becomes

$$N(\alpha, \delta, c, m) = \xi_{cl}(\alpha, \delta) f_{cl}(c, m) + \xi_{bg}(\alpha, \delta) f_{bg}(\alpha, \delta, c, m) \tag{4}$$

We construct the f_{cl} and f_{bg} PDFs by means of Hess diagrams in bins of 0.01 in colour and 0.1 in magnitude. For the sake of short notation, we refer to a CMD bin using the index j. The sky is also divided into bins that we choose to have an area of 1x1 $arcmin^2$. These bins are labeled by i, for short. Then we can transform the equations of n_{cl} and n_{bg} in discrete functions.

$$n(i, j) = \xi_{cl}(i) f_{cl}(j) + \xi_{bg}(i) f_{bg}(i, j) \tag{5}$$

The left hand side is expected to be the observed number of stars in a given position in spatial and CMD space. If the actual number of stars observed in a catalog is N(i,j), we can then build the variance between this number and that expected from the model for each spatial bin:

$$S^2(i) = \sum_j \frac{[N(i,j) - \xi_{cl}(i) f_{cl}(j) - \xi_{bg}(i) f_{bg}(i,j)]^2}{\xi_{bg}(i) f_{bg}(i,j)} \tag{6}$$

The term in the denominator expresses the expected Poisson fluctuation in the star counts, which, for simplicity, we to be dominated by the background. Minimizing this equation and solving for $\xi_{cl}(i)$

we have the number of observed stars that, according to the model given by eq (4), are more similar and consistent to the model SSP.

$$\xi_{cl}(i) = \frac{\sum_j N(i,j) f_{cl}(j)/f_{bg}(i,j)}{\sum_j f_{cl}^2(j)/f_{bg}(i,j)} - \frac{\xi_{bg}(i)}{\sum_j f_{cl}^2(j)/f_{bg}(i,j)} \tag{7}$$

In the end, as output of the code we get a stellar density map provided by the $\xi_{cl}(i)$ array. In practice, we model the background PDF, $f_{bg}(i,j)$, using samples taken directly from our target stellar catalog. We do that under the assumption that the contamination by any yet to be detected SSP, is insignificant. As for the SSP PDF, we make use of simulated samples, as described in the next subsection.

2.2 Simulations

Since we do not know a priori what stellar population we will find, we create a grid of SSP simulations with the code GenCMD. GenCMD uses PADOVA isochronous Bertelli, G et al 2007 (ASP,374,41) for different assumed distances and randomly selects stellar masses following a predefined initial mass function (IMF). Currently, we are adopting a Kroupa et al 2001 (MNRAS,322,231) IMF for that purpose. Given each stellar mass, we interpolate among the isochrone entries to draw absolute magnitudes in the desired filters. These are converted into "measured" apparent magnitudes using the assumed model distance and a reddening map. For this latter we use Schlegel et al 1998(APJ,500,525).
Positions on the sky are simulated assuming a Hubble modified density profile. We simulate several SSPs at various ages, metallicities and distances. Four grids of parameters were defined. One of them covers a more restricted parameter range, specially in distance. It also has a narrower binning. It is described by Table 1. This initial grid was applied to the BOSS footprint, in the search for new star clusters. None was found but previously known clusters were recovered (see Section 4). In this initial application, it became clear that there was no need for a very narrow binning in age and metallicity, since the f_cl were largely degenerate. A smaller grid also helps circumvent limitations in storage and processing. As a result, we made a second grid to be applied to other datasets. It is more extended in distances in order to allow detection of faint dwarfs in the outer reaches of the galactic stellar halo. But it is much more coarsely binned in all parameters. It is described in Table 2.

Parameters	Lower limit	Upper limit	steps
Log (Age(yrs))	9.0	10.1	0.1
Metallicity, Z	0.001	0.019	0.002
Distance (kpc)	10	40	2

Table 1: *Parameter grid used to simulate SSPs for the search of star clusters in BOSS. It includes a total of 1485 models. Lets indicated it by Grid 1*

Parameters	Lower limit	Upper limit	steps
Log (Age(yrs))	9.0	10.2	0.3
Metallicity, Z	0.0002	0.001	0.007
Distance (kpc)	10	200	10

Table 2: *Parameters used to simulate SSPs for the search in SDSS and other surveys. It includes a total of 228 models. Lets indicate it by Grid 4*

We use each one of these models in the application of the MF. As mentioned, Grid 1 was used in the BOSS data with the goal of finding mostly clusters of stars, while Grid 4 was used in the other SDSS catalogs around known dwarf galaxies and clusters. Beyond Grid 4 we used more two grids that differ only by the metalicity from Grid 4, Grid 3 goes to Metallicity from 0.0001 to 0.0101 in steps of 0.005Z and Grid 2 goes from 0.001 to 0.019 in steps of 0.006Z.

2.3 Objects Detection

It is a bit difficult to analyze every density map, that goes out of sparse. For example on the Grid 1, there would be 1485 maps to be studied. Its not very practical to do this kind of analyzes manually, so as a solution for this problem we use the code SEXtractor, that search for peaks of density in each one of these maps. Any object found by SEXtractor in each map, was putted on a list and then as a result we had a very huge list with repeated objects, because one object probably could have been detected by more than one SSP model.
Not all of the objects founded by SEXtractor were really real that are of course some bad signal and spurious things in the catalogs, to try eliminate these things we made a cut based on the signal of the object and the area covered by it. As we search for clusters and galaxies we are searching for compact things so we do not want a flux per area to be too small, we made a cut on the final list of objects to take only the detections with density greater then 15 pixies. That cut was based on the detection of the globular cluster Whiting 1.
Then finally we organized the objects by importance (more probable to be a structure), and the criteria for importance that we used was the number of repetitions of the object, in other words, the number k of models that detected the same object by the MF in all N models simulated:

Systematically run Sparse on N simulated models $\Rightarrow$ N density maps filtered by MF $\Rightarrow$ in k maps the object have been detected.

And then we had our final list of detected objects in order of importance. On table 3 in section results we show a table with the most important detections we made with the MF on the area of BOSS and SDSS.

To validate if the objects founded are really real, we should always check on its colour magnitude diagrams (CMDs) to see if it is consistent with a stellar population. So another tool that makes part of our detections is building CMDs of the final object list. This is done systematically with a code that uses the equatorial positions and the dimensions of the object. We pick the major axis, parameter of SExtractor output and then multiplied it by 5 to get a radius of a circle, centered on the RA and DEC of the detection, where we selected the area of the CMD. We compared this CMD with a CMD of field stars, sampling a region in a ring with the initial radio being one hundred times the major axis. We have mask the areas doest contain data around this rings for both areas to cover an effective area.
We can see the CMD of Canes Venatici I in figure 1, where certainly if the object was not known it would be confirmed that it is a stellar population. We can clearly see its over density compared with the background. There are cases in which the CMD does not reveal much about the stellar population detected as it was the case for Hercules Dwarf Galaxy we can see its CMD on figure 2. After seen examples like this one we knew that we could not discriminate a candidate only by its CMD, its important to make a thorough analysis of the candidates looking for imaging data, seen the density map and trying to improve the decontamination by the background on these finals CMDs. For the sake of this case and others we have applied a decontamination algorithm on the CMDs by subtracting the background sample for the area of the object, we can see the results for this application on the Hercules CMD in figure 5 ?!. We did examined the density maps for these objects, as is shown in figure 6 for an accurate analysis. For last we adjust a radial density profile based on a King profile

from where we can take the more aproximate size of the object. These Density profiles are shown in figures 6 and 7 for Canes Venatici I and Hercules Dwarfs.

3 Data

We applied the MF on the the area covered by BOSS that is a survey of the project Sloan Digital Sky Survey III (SDSS-III, http://www.sdss3.org/index.php) and in small regions around some known objects on SDSS. BOSS has imaging 2,000 deg^2 of the sky in the south hemisphere and SDSS has imaging more than 8,000 deg^2 of the sky. Both surveys were observed on the filters ugriz.

4 Results

After the steps: Simulations $\Rightarrow$ MF application $\Rightarrow$ Sextractor we got a list ordered by importance of objects that could or not could be a substructure of our Galaxy. For the moment we did not confirmed the detection of a new stellar structure, but we detected efficiently structures that are already known, which validates the use of the method described. On table 4 we can check the detections and the parameters of Sextractor's output from these known objects.

Ultra Faint Dwarfs	Area (sq.deg)	N^{o} Models	Rank	N^{o} objects	CMD	Conv FWHM -SE (pixels)	Density (n^{o} of stars per pixels^2)	Grid of models	Major Axis (pixels)
Canes Venatici I	6x9	106	1	50	yes	5.0	14.256	2	2.527
Canes Venatici II	9X8	16	1	363	yes	1.5	14.534	2	0.621
Coma	8X8	28	1	33	yes	5.0	15.065	2	1.45
Hercules	9X9	10	1	1351	maybe	2.5	42.504	3	1.983
Leo IV	9X9	30	3	3867	yes	3.0	10.7	4	1.052
Leo V	9X9	11	3	1464	yes	4.0	6.789	4	1.11
Leo T	9X9	84	1	668	yes	2.5	9.368	4	0.968
Segue 1	9X9	16	3	344	yes	4.0	11.133	4	1.799
Segue 2	9X9	10	2	510	yes	4.0	7.391	4	1.024
Ursa Major I	9X9	17	3	1889	yes	3.0	6.308	4	2.355
Willman 1	9X9	81	1	306	yes	5.0	6.468	4	2.022
Star clusters									
Balbinot 1	9X9	510	1	1108	yes	No conv	15.8780	1	1.768
Koposov 1	9X8	62	1	4438	maybe	No conv	13.443	2	0.289
Koposov 2	9X9	13	1	272	yes	1.5	26.683	4	0.534
Pal 13	9X9	1192	1	110	yes	No conv	16.8750	1	2.042
Segue3	9X9	1075	1	5242	maybe	No conv	16.7990	1	1.908
Whiting 1	40X40	1154	1	1477	yes	No conv	1031.2740	1	0.8040

Table 3: *Know objects founded by the SparSEx, and its respectives parameters: The area of the catalog used as sample for the detection with the method, number of models that have detected the object, Rank it is classified by the greater quantity of number of models, N objects the total number of detected objects in that catalog, CMD if it does seem with a stellar population, Conv(FWH-SE) Sextractor parameter used as the FWH of a Gaussian convolution to smooth the density map, the density of the detected object, The grid of simulated SSPs used on the detection of the object and finely the major axis of the object.*

The list of detected objects from SExtractor's output comes with the parameters of positions and major axis. In the beginning we were using these parameters to built a CMD of the detections, but as we can see on figure 1 the central position of the object, (Koposov II), was a bit displaced. Plotting the distribution of positions around the central position given by SExtractor we can find the real center, in figure 1 it is indicated by a yellow circle. Around this new center we have chosen the points to built the CMDs of the detections. The area chosen for the built of the CMD was a circle around the center of the object with a radius with the size of the core radius taken from the king profile of the objects shown on figures 3, 6 and 9. then we compared the objects CMD with a background sample, in a area of a ring around the object. this ring is 20 times the core radius away from the object and has to covers the same area of the object's CMD.

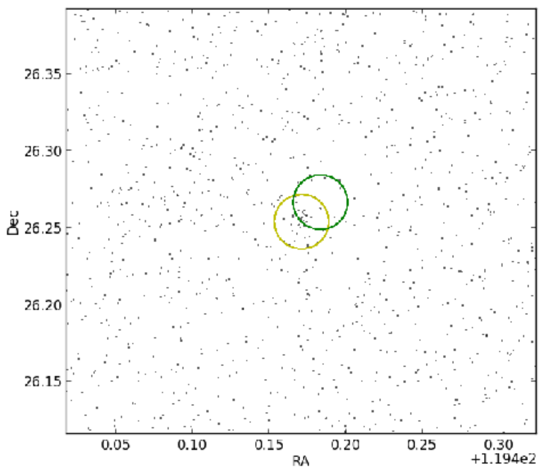

Figure 1: *Distribution of equatorial positions for an area around the detection of Koposov II, the green circle indicates th central position of the cluster given by SExtractor. We have adjust the center for the yellow circle by eye.*

In the sequence we show the results for two dwarf galaxies (Canes Venatici I e Hercules) and for the cluster Koposov II, first we show on figures 2, 5 and 8 the density maps of the resulting application of the MF with a model that had contrasted the object with the best signal. We can see in these figures the over densities detected by SExtractor. Then we have compute the density profiles for these substructures shown in figures 3, 6 and 9 from where we have got the core radius for the objects. We then built the CMDs in an area of the core radius size, where we can see if these CMDs have characteristics of a SSP. On the CMD of Canes Venatici I, 4 we can see the its Giant red branch and its horizontal branch, and the turn off of stars. Nevertheless for Hercules and Koposov II, shown in figures 7 and 11, it demands a bit more of creativity to see something. But these CMDs match with the ones founded by its discoverer, Belokurov et al 2007(AJ,654,906) and Koposov et al 2007(AJ,669,337).

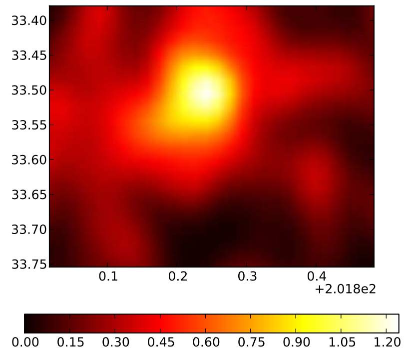

Figure 2: *Density map of a detectation for Canes Venatici I, resulting for the aplicantion of MF with a model with 10 kpc of distance, 9.9 log(years) and 0.001z, Smoothed with a gaussian of FWH 5.0 pixels. We can see its density over the backgorund.*

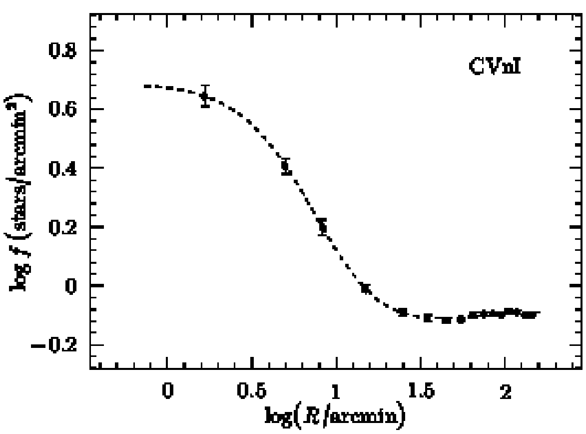

Figure 3: *Radial density profile of Canes Venatici I, the dashed lines represent the best fit of the king profile. From that we take the core radius that for the case of CVnI is* $r_c = 4'.86 \pm 0'.07$

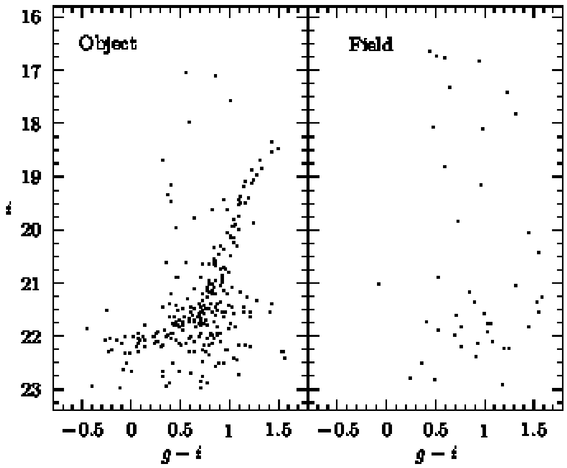

Figure 4: *Left panel (objeto): Colour magnitude diagram of Canes Venatici I. We can see clearly its Giant red branch and its horizontal branch, that represent an over density compared with its bakcround, right panel (campo). The CMD was built over a area with the size of the core radius discoverd with the density profile of 3*

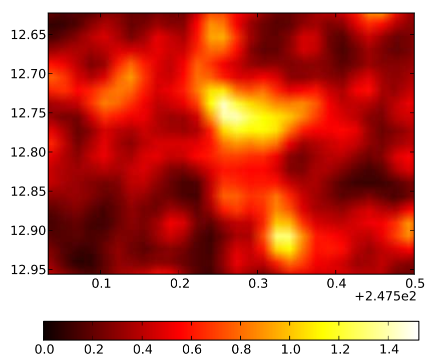

Figure 5: *Density map of a detectation for Hercules, resulting for the aplicantion of MF with a model with 10 kpc of distance, 10.20 log(years) and 0.001z, Smoothed with a gaussian of FWH 2.5 pixels. We can see its overdensity with the backgorund though it is not so clear as in Canes venatici I.*

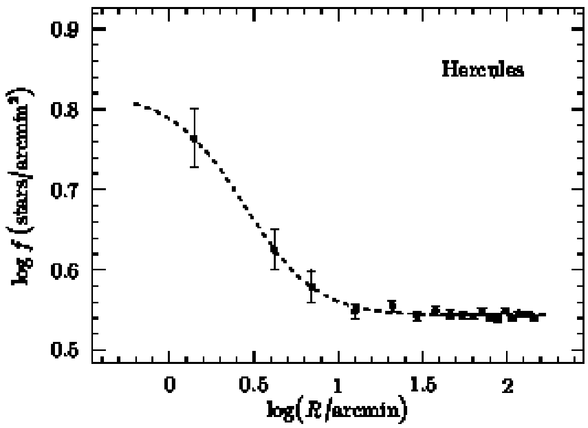

Figure 6: *Radial density profile of the dwarf galaxy Hercules, the dashed lines represent the best fit of the king profile. From that we take the core radius* $r_c = 2'.41 \pm 0'.16$

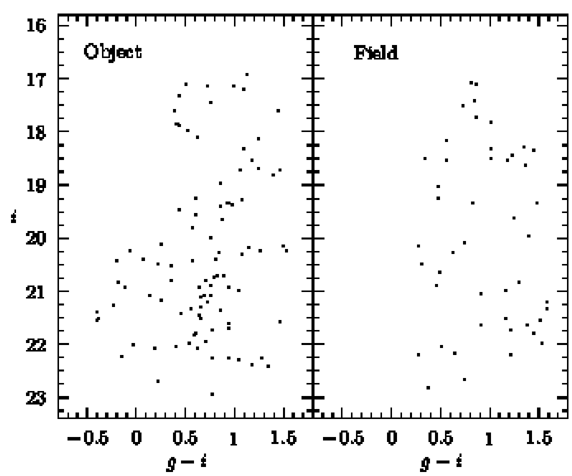

Figure 7: *Left panel (objeto): Colour magnitude diagram of the dwarf Galaxy Hercules, it is not very easy to see the caracteristics of SSP on this CMD Belokurov et al 2007(AJ,654,906), the CMD do not differ much in density with the bakcround,right panel (campo). In this case a good decontamination in crucial*

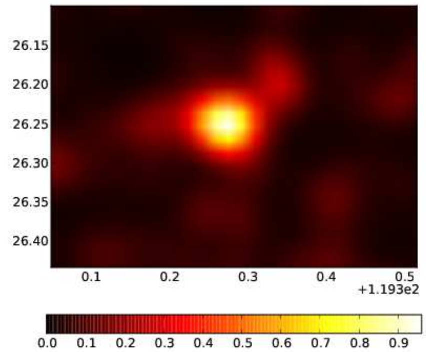

Figure 8: *Density map of the detectation for the stellar cluster Koposov II, resulting for the aplicantion of MF with a model with 170 kpc of distance, 9.30 log(years) and 0.002z. We can see its overdensity with the backgorund.*

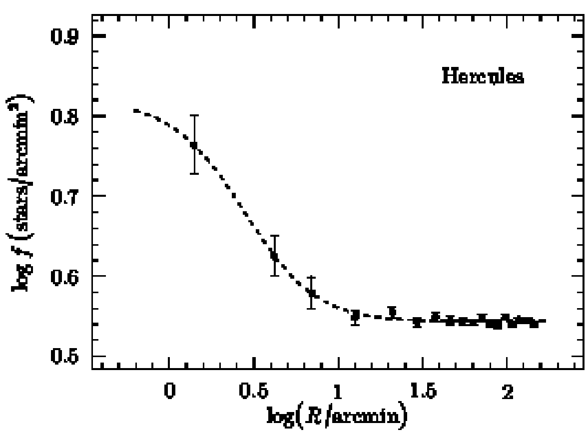

Figure 9: *Radial density profile of the stellar cluster Koposov II, the dashed lines represent the best fit of the king profile. From that we take the core radius* $r_c = 0'.53 \pm 0'.05$

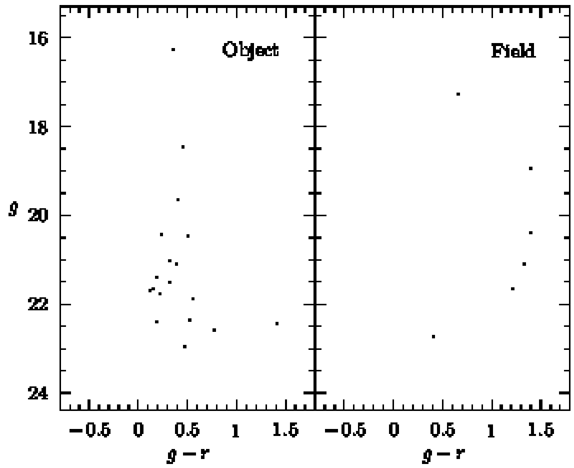

Figure 10: *Left panel (object): Colour magnitude diagram of Koposov II, it is not very easy to see the caracteristics of SSP on this CMD, but we can see a litle overdensity with the background Right panel*

Another detection very important done by this method was the tails of Sagitarius Dwarf, if we plot the list of detected objects in the data of boss we have the dispersion of figure 3, where we can see a reconstitution of the tidal streams of Sagittarius Dwarf that is a satellite of our Galaxy located on the stellar Halo. In this plot of figure 1 we can observe too, detections around the galaxies Triangulum (M33) and Andromeda (M31).

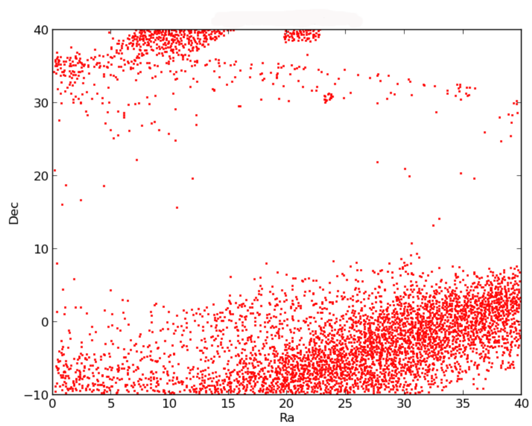

Figure 11: *Plot of the objects detected by the method of MF on BOSS, we can clearly see the tails of Sagittarius dwarf in the bottom corner. The concentration of points in the external points of the northern region correspond to M31 and M32, largely resolved, whereas M31 has to a large system of satellite galaxies.*

5 Conclusions

From the results showed in this document we can validate the method of Matched Filter using Simulated SSPs combine with Sextractor for the search of substructures. And after analyzing the results we can say that the method is very capable of detect substructures of very different sizes, we shown the detection of globular clusters, open clusters, dwarf galaxies and even tidal streams that is a considerable large structure compared with a globular cluster. There are still some difficulties on the method that can be improve. Like the decontamination of the CMD and the tool to determine the size of the detected object, sparse generally split the objects in more than one in the process of detection. We expect to use this method on the data of the project Dark Energy Survey (DES, http://www.darkenergysurvey.org/), that is a new survey deeper than SDSS, that is going to cover part of the south hemisphere equatorial. Therefore, in this data will be more probable to find new clusters, dwarf galaxies and stellar streams. In fact, there is an estimative that dozens of dwarf galaxies not discovered yet are in the area covered by DES (Tollerud et al, ApJ, 688, 277). The census of these objects and the properties of the system where they were formed (spacial distribution, luminosity function, etc), brings us import information about the historical mass aggregation of the Milk Way, further on it can be confronted with previsions of structures formation in the Universe.