A Mathematical Morphology Approach to the Characterization of Astronomical Objects

Alcione Jandir Candéas[1]
Ulisses de Mendonça Braga Neto[2] *
Edson Costa de Barros Carvalho Filho[1]

[1]UFPE - Universidade Federal de Pernambuco
Departamento de Informática, Caixa postal 7851
50732-970 Recife, PE, Brasil
{ajc,ecdbcf}@di.ufpe.br

[2]The Johns Hopkins University
Department of Electrical and Computer Engineering
3400 N. Charles St. Baltimore MD 21218
ulisses@ece.jhu.edu

Abstract. We deal in this work with a broadly studied topic in astronomical imaging, namely the star/galaxy discrimination problem. Through the use of Mathematical Morphology, we propose an original approach to the characterization of these astronomical objects, which is based on gray-level shape-size information. The main steps in our method are image pre-processing, segmentation and feature extraction, all of which employ Mathematical Morphology tools that were implemented using the MMach toolbox for the Khoros system. We present a comparison between our segmentation results, based on the watershed method, and those of the SExtractor software package. The shape-size information is extracted through the use of the gray-level morphological pattern spectrum, which yields good analysis attributes that promise to be suitable for future work in neural-network automatic classification.

1 Introduction

The Image Processing and Analysis (IPA) field has been characterized by the increasing adoption of specific solutions, where techniques that are suitable to some applications may be fully inadequate to others. On the other hand, Mathematical Morphology (MM) is a general methodology that has provided a unified approach to handle problems in Medicine, Geology, Geography, Remote Sensing, and many others. In spite of the diversity of goals, all of these problems have a common feature: the need to extract shape information from images.

We employ Mathematical Morphology in the problem of star/galaxy discrimination, which is currently a broadly examined topic in the field of Astronomical Imaging. The shape information extracted with the use of MM tools proves to be suitable for the characterization of those astronomical objects, while promising to be effective for neural-network automatic classification.

This work represents also an attempt to call the attention of the general Image Processing and Analysis community back to Astronomical Imaging. The IPA field has grown strongly in the sixties due to the american space program, carried out by NASA. Since then, the technology developed has been applied in many other fields, leading to the development of new digital techniques for filtering, pattern recognition, formal grammar, neural networks, artificial intelligence, and others. By presenting an original Mathematical Morphology approach to the star/galaxy discrimination problem, we expect to revive the interest in IPA for astronomical images, as well as motivate and support the multidisciplinary field. We also point out that Astronomical Imaging is a new promising application for MM.

In Section 2 we outline some topics on astronomy, astronomical images and the problem of star/galaxy discrimination, while in section 3 we discuss our method and the gray-level MM tools required by it. In section 4, we show the results obtained and a comparison with the segmentation results of SExtractor (Source Extractor, a software that builds a catalog of objects from astronomical images). Finally, section 5 gives some concluding remarks and directions for future work.

*The second author was supported at the time this work was performed by CNPq DCR grant 300804/95-4.

2 The Problem Context

2.1 Astronomy

The galactic structure and dynamics, environmental effects in the galactic arrangement and evolution, the diversity of galaxy morphologic types, in addition to the distribution of dark and bright matter in the whole universe, are some relevant current problems in Astrophysics, one of the numerous branches of Astronomy. These kind of problems usually require a statistical approach [9] to treat large amount of star and galaxy data collected over wide sky regions.

With the advent of new ground-based spectroscopic measurement tools, in addition to space observatories such as the Hubble Space Telescope, deeper and more accurate galaxy and stellar catalogs are needed. Some of the most complete catalogs were visually compiled from photographic surveys. The current tendency is to use more accurate acquisition instruments and faster scanners to compile galaxy catalogs over specific regions of the sky. The corresponding increase in image data calls for new automatic detection and classification methods.

2.2 Astronomical Images

The development of astronomical image acquisition methods and tools, such as CCD (Charge Coupled Device) cameras and photographic plate scanners, in addition to the increase of available data in digital form, have stressed the relevance of computer interaction in service of Astronomy, which can customize and improve results of observational programs. The digitized samples of the photographic plates have great utility in astronomical research.

A remarkable contribution to the star-galaxy surveys are the digitized images derived from the 936 pairs of photographic plates of the National Geographic-Palomar Observatory Sky Survey (POSS). They were generated using the Automated Plate Scanner (APS) of the APS Project and Catalog at the University of Minnesotta. A more detailed discussion about this project can be found in [7, 11].

A collection of image samples covering the entire sky and widely used for work in analysis and classification by the international astronomical community is the Digital Sky Survey (DSS). It consists of a set of CD-ROMs produced by the Catalogs and Surveys Branch of the Space Telescope Science Institute (STScI). In fact, DSS northern hemisphere samples are based on POSS photographic data, and the southern hemisphere imaging survey is from Siding Spring Observatory-SERC Southern Sky Survey, Australia. All samples are available as FITS (Flexible Image Transport System) images, the standard data interchange and archival format employed worldwide by researchers in Astronomy.

2.3 Star/Galaxy Discrimination

Cataloguing astronomical objects automatically calls for a solution in IPA. However, due to the increase in the accuracy of the observational and registering tools in the last decades, not all of the currently detectable objects have been catalogued accordingly. This can be explained by the time-consuming, tiresome task of visual identification required by the cataloguing process. Therefore, it is necessary to develop recognition and classification automatic procedures.

One of the most important topics in automatic cataloguing is the star/galaxy discrimination problem, which is currently greatly employed in the automatic identification of astronomical objects. This issue has been studied by many research groups and professionals, as can be seen in [3, 10, 14]. An interesting contribution toward this goal is the APS Project and Catalog, that has generated an Internet-accessible astronomical catalog of millions of stars and galaxies by using a neural network classifier.

However, all catalogues have limitations due to the adopted criteria. It is necessary to verify the catalog quality with respect to completeness and correctness in order to produce reliable analysis results. This depends on the accuracy of the acquisition instrument correspondent to the image chosen, as well as the selected features and the reliability of the classifier utilized.

3 The Proposed Method

Most works found in the literature are based on features extracted from photometric attributes. This means that these techniques employ pixel intensities, which are correlated to the object apparent brightness. In this work we propose instead a morphological approach to the characterization of astronomical objects, which is closer to the human way of performing the analysis of astronomical images.

This is accomplished by using Mathematical Morphology tools to extract gray-level shape information that can be useful for discriminating stars and galaxies in large scale galaxy-survey images, like the ones in DSS and POSS. These images typically contain thousands of astronomical objects of varying size and brightness, mostly stars and galaxies, as is shown in the image sample of fig. 1 (Note that this image is inverted, for contrast enhancement purposes).

A human viewer can perceive in the image that

stars are well-defined objects, whereas galaxies present themselves as fuzzy blobs. Further, "large" stars present a cross effect due to optical properties of the acquisition instrument. We propose to use these shape features in order to discriminate between the two classes of objects.

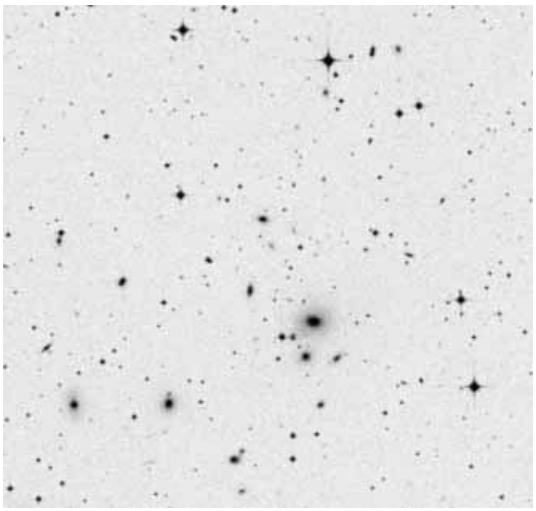

Figure 1: Large scale galaxy-survey image sample

3.1 Gray-Level Mathematical Morphology

The goal of this section is to present a brief review on the main gray-level MM notions and tools we utilize in this work. The interested reader can find a more thorough description in [5, 12, 15].

The gray-level MM operators deal with functions $f : E \rightarrow [0, 255] \in \mathbf{Z}$, where E is the usual planar digital grid. This corresponds to the usual concept of gray-level images. The basic geometrical idea behind these operators is to probe the image with a function defined on a small subset of the image domain of definition, called a *structuring element*.

The basic dual gray-level MM operators are *dilation* and *erosion*, defined respectively by [1]:

$$\delta_g(f)(x,y) = \max\{f(x',y') \dot{+} g(x-x', y-y')\}, \quad \forall (x',y') \in (B^t + (x,y)) \cap E \quad (1)$$

$$\epsilon_g(f)(x,y) = \min\{f(x',y') \dot{-} g(x'-x, y'-y)\}, \quad \forall (x',y') \in (B + (x,y)) \cap E \quad (2)$$

where f is the image, g and B denote respectively the structuring element values and domain of definition ($B \subset E$). The summation and subtraction operators used in (1) and (2) are not the usual ones, but slightly altered versions necessary to comply with algebraic constraints (for more details, see [6]). Note that the erosion operator reflects the idea of probing the image, with respect to the "fitting" of the structuring element "underneath" the image.

The composition of erosion and dilation yields the *opening* and *closing* operators, which are given respectively by:

$$\gamma_g = \delta_g \epsilon_g \quad (3)$$

$$\phi_g = \epsilon_g \delta_g \quad (4)$$

Openings and closings are morphological filters [13], and they have good noise-removal properties. The closing operator is the most often used in this work, since it acts on the gray-level "background" and we consider inverted images.

A nice enhancement tool can be obtained from openings or closings, known as the *top-hat transform*. We consider the closing case:

$$\tau_g(f) = \phi_g(f) - f \quad (5)$$

Due to the simplification properties of the closing operator, the top-hat defined above has the property of enhancing "valleys", that is, dark regions in the image, serving as a good background remover.

By using a suitable family of closings (resp. openings) one gets a powerful tool for describing shape and size in image analysis, namely the *granulometries* [4]. The method is based on "sieving" an image followed by measure of the residue left on the sieve. In practice, the granulometries consist of a sequence of closings (resp. openings) by a set of appropriately selected increasing larger structuring elements. By measuring the volume under the image after each closing, we can build a *size-distribution curve*:

$$\Phi(\lambda) = \frac{V(\lambda) - V(0)}{V(\Lambda) - V(0)}, \quad \lambda \geq 0 \quad (6)$$

where λ is the parameterization of the closing family, $V(\lambda)$ is the volume (i.e. the positive integral) of the image at each iteration and Λ is the parameter associated to the largest structuring element, which is selected to be the one large enough to "wipe out" or utterly close the object of interest. Note that Λ is itself a useful analysis attribute.

The function above is monotone increasing in the interval $[0, 1]$, so it may be viewed as a cumulative probability distribution. Its associated probability density function is called the *Pattern Spectrum*, which is given by the following discrete derivative:

$$\Gamma(\lambda) = \Phi(\lambda + 1) - \Phi(\lambda), \quad \lambda \geq 0 \quad (7)$$

Due to the properties of $\Phi(\lambda)$ mentioned above, the function $\Gamma(\lambda)$ is non-negative and ideally should become null for $\lambda > \Lambda$.

The pattern spectrum is a powerful tool for characterizing shape and size in image analysis [8]. In a manner reminiscent of the Fourier spectrum, it shows the decomposition of a given object in terms of a fundamental shape "scaled" by the increasing values of the parameter λ. We can define a useful analysis attribute, the *average size*, which is the expected value of the pattern spectrum:

$$\overline{\lambda} = \sum_{\lambda=0}^{\Lambda} \lambda \, \Gamma(\lambda) \tag{8}$$

Another very useful attribute based on the pattern spectrum is the *average roughness*:

$$\theta = -\sum_{\lambda=0}^{\Lambda} \Gamma(\lambda) \log[\Gamma(\lambda)] \tag{9}$$

Note that the average roughness is in fact the information-theoretic entropy of the pattern spectrum [8]. The larger this attribute gets, the more complex with respect to the structuring elements used the image is.

The last MM tool we need is the *watershed* transformation. The watershed is a powerful Mathematical Morphology tool for segmentation [16]. It can be defined entirely in terms of geodesic MM operators [12], but the most intuitive approach is to think of an image as a relief, and suppose that water is oozing and rising at equal speed from every regional minimum, starting from the lowest one and then from each of the others as soon as the global water level reaches its altitude. Dams are built in the places where water from different minima would merge. After the relief is completely flooded, the dams rising above the water surface constitute the watershed line, which is composed of closed contours that involve each of the regional minima and correspond to the crest lines of the relief.

3.2 The Detailed Procedure

In this section we are going to describe the steps that comprise our method for the characterization of stars and galaxies in large scale galaxy-survey images.

3.2.1 Pre-Processing

Image Inversion: This is a good practice in astronomical imaging, where the purpose is object contrast enhancement for human viewers. In our case there is another reason; although most Mathematical Morphology operators come in dual sets with respect to gray-level complementation, the inversion is necessary to get the right topographic image model (the objects of interest become "valleys") required by the watershed method in the segmentation step.

Top-hat transform: It is often necessary to enhance the image, diminishing environmental conditions (luminosity, atmospheric density, and so on) at the surface acquisition time, for instance optical effects like halos around large stars, which can affect the segmentation step. This background removal operation is performed with a top-hat transform of the inverted image, using as structuring element a flat euclidean disk of radius $r = 8$. After the top-hat is applied, the image is inverted once again and normalized back to the full dynamic range.

Open-close filter by a flat structuring element: A composition of an opening followed by a closing is used to remove background noise that is enhanced by the top-hat step. The flat structuring element is a zero-valued function having just its domain of definition, that is, a set. In our case, the set is the 3×3 square.

3.2.2 Segmentation

Region segmentation by the watershed method: We apply the direct watershed transformation to the image (with no prior change of homotopy). Using the flooding interpretation of the watershed (see section 3.1), the output is a tessellation of the image where each closed contour defines a region which contains an object inside. The watershed is able to handle most of the object overlapping occurrences (this is called "deblending" in the astronomical literature).

3.2.3 Feature Extraction

Computation of the pattern spectrum: After each region in the image is properly isolated and centered in a local neighborhood image, we compute its closing pattern spectrum and associated measures of average size and roughness. In order to minimize border effects on the computation, we pad the region with the mean value of the entire original image.

The family of structuring elements used is the non-planar digital disks of increasing radius, according to the city-block distance. The family can be seen as an increasing sequence of scaled square-based pyramids, which satisfies the requirements for a valid granulometry, since the family is generated by iterated Minkowski sums of the digital disk of unitary radius. The parameterization for the granulometry is thus the digital disk radius, so the average size attribute may be seen as in fact the object average radius according to the city-block metric [4].

As we discussed in section 3.1, the largest radius Λ should correspond to the closing that removes the object, when the sequence of closings should termi-

nate. We determine Λ by means of a homogeneity criterion, namely, the gray-level standard deviation of each closed image, and then select a minimum threshold beyond which the computation stops.

4 Experimental Results

In this section we give some results we have obtained from the application of the proposed method, as well as offer some remarks about them. We compare our segmentation results with the ones from SExtractor, a software package for astronomical imaging.

Taking into account the occurence of galaxies, we have selected for this study the following six FITS images from DSS:

- Group 1701: (J2000)[1] α[2] [02 08 32.1] δ[-55 39 6.5]
  Frame: 500x500[3] / 32 bits
- ABELL 3667: (J2000) α[20 10 50.3] δ[-56 40 23]
  Frame: 530x530 / 16 bits
- O0140: (J2000) α[01 04 34.3] δ[-23 50 08]
  Frame: 530x530 / 16 bits
- ABELL 3775: (J2000) α[21 31 24.8] δ[-43 18 38]
  Frame: 530x530 / 16 bits
- ABELL 3705: (J2000) α[20 41 55.5] δ[-35 16 06]
  Frame: 530x530 / 16 bits
- ABELL 3698: (J2000) α[20 35 08.1] δ[-25 15 34]
  Frame: 530x530 / 16 bits

We implemented our method on the Khoros image processing system, version 1.0, running in a Sun IPX workstation under Unix and X11R5. Khoros is a very popular IPA open platform developed at New Mexico University and freely available through anonymous ftp. Khoros turned out to be a very convenient tool for algorithm development and fast prototyping. We have used the operators of MMach, a Khoros toolbox for Mathematical Morphology [2], also freely available through anonymous ftp, at São Paulo University.

We would like to point out that there is a bug in the Khoros 1.0 conversion routine from FITS to VIFF, the format used by Khoros, so that we could not input the original images directly. The solution adopted was to use the XV 3.10 package to convert the images to formats readable by Khoros, such as PBM and (uncompressed) TIFF.

[1] estimated center coordinates to year 2000

[2] right ascension (α, in hours, minutes and seconds) and declination (δ, in degrees, minutes and seconds) - celestial coordinates of the equatorial system.

[3] 500x500 pixels on DSS means 15'x15' arcsec on the sky

We see in fig. 2 the sequence of operations used to obtain the segmented regions, as described in section 3.2, using one of the images (ABELL 3698).

We can see that the top-hat operation has successfully removed the halos and enhanced the objects by removing the background. The segmentation obtained with the watershed was near-optimal, where most detected regions correspond to one and only one object. We can see some overlapping objects were satisfactorily separated, but there are a few problems, like close or interacting object pairs that were not separated. There are some regions which do not correspond to any object, due to residual background noise (This is a common problem for the other deblending techniques found in the literature). In addition, the large galaxy in the bottom of the image was taken apart by the segmentation, due to the non-homogeneity of the spiral arms (in fact, a galaxy is a large set of astronomical objects). This oversegmentation is handled by many methods through post-processing techniques. SExtractor, for instance, deals with the latter case, however, it does not consider the empty-region one.

Another advantage derived from the watershed segmentation is that we have the option of including or not the regions that touch the image border (fig. 2-d depicts the "no border" case). This is done automatically in the MMach implementation of the watershed, whereas other methods do not offer this benefit, requiring visual inspection by human experts in order to accomplish it (the objects in the border in astronomical images are known to cause problems in the analysis step).

In table 1 we can see the number of objects detected by our method, with and without border exclusion, for the six images considered, as well as the results obtained by SExtractor with a satisfactory parameterization, both before and after oversegmentation post-processing (there is no post-processing in our method at this time).

Image	Proposed Method		SExtractor	
	Border	No border	Original	Post-Proc
G1701	1672	1572	1543	1457
A3667	1879	1767	1785	1700
O0140	1834	1731	1484	1413
A3775	1890	1785	1308	1257
A3705	1728	1638	1684	1620
A3698	1890	1785	1831	1722

Table 1: Comparison of segmentation results

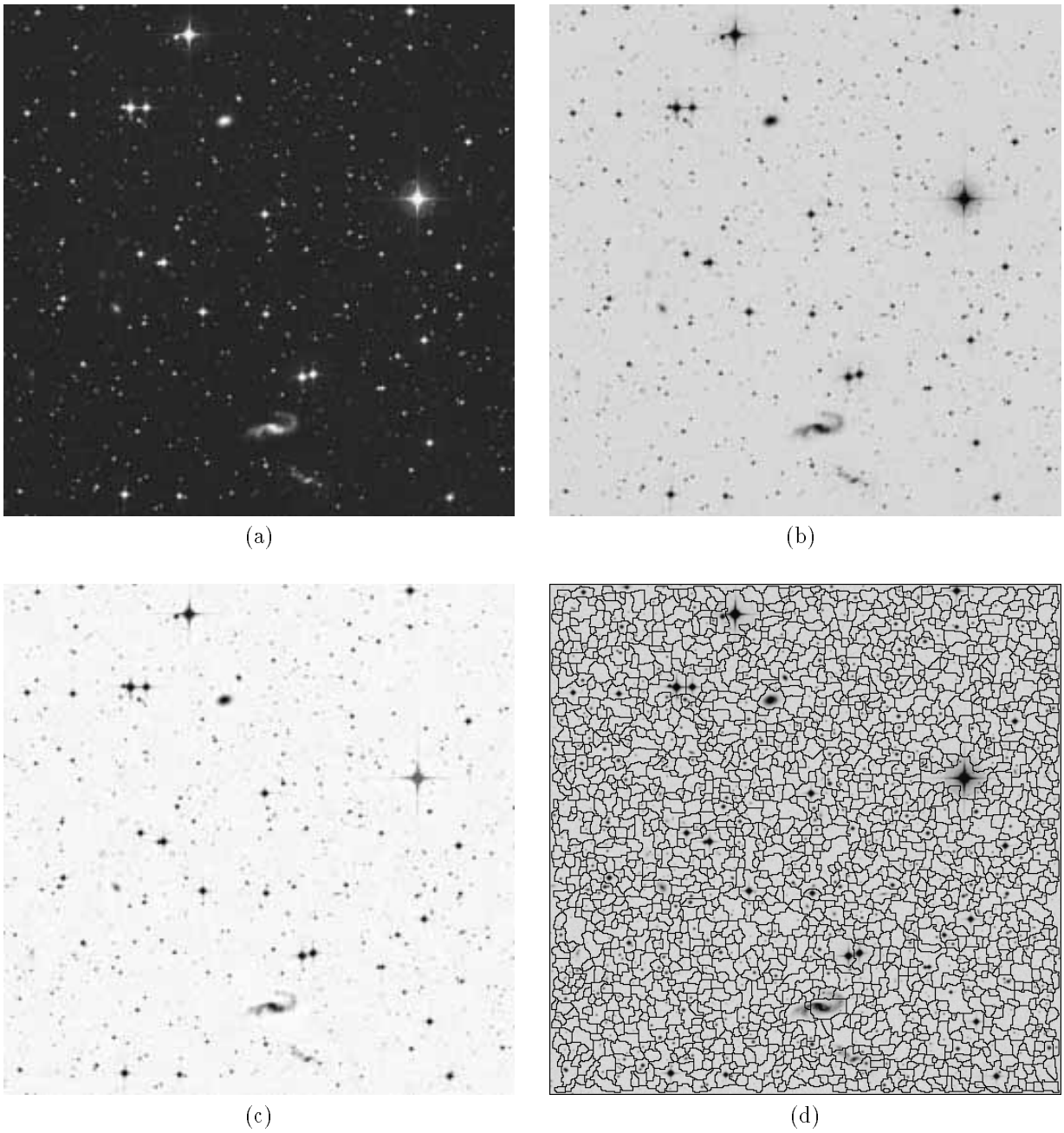

Figure 2: Sequence of operations for the object region segmentation: (a) original image (b) inverted image (c) result of top-hat transform followed by open-close filtering (d) watershed line superimposed to the original inverted image

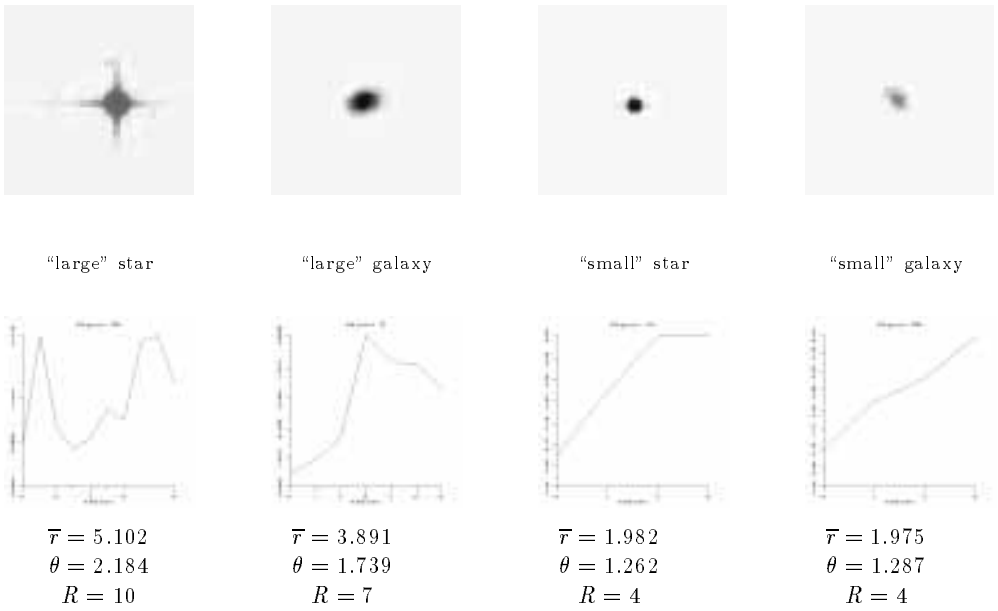

Figure 3: Pattern spectrum characterization of four representative objects

The pattern spectrum extraction requires that each segmented region be considered separately. This is done by labelling the regions (this is already the output of the watershed MMach implementation), thresholding and subsequent masking the pre-processed image. The mean value of the original image used to pad the region proved to be a good estimate of the background gray-level value.

In fig. 3 we can see four of the padded regions in image A3698, along with the respective pattern spectrum and analysis attributes. We use the symbols $\overline{r}$ and R for the average and largest size, instead of the previously defined $\overline{\lambda}$ and Λ, to stress the fact that the parameterization of the granulometry corresponds to the radius according to the city-block metric.

The regions correspond to two stars and two galaxies of different apparent sizes. For the two "largest" objects, as given by the average radius and largest radius (according to the city-block metric induced by the particular family of structuring elements used, see section 3.2), the pattern spectrum clearly differentiates between them, since the star produces a bimodal curve, due to the cross effect discussed earlier, whereas the galaxy is associated with a unimodal curve. This is also reflected in the average roughness attribute, which is larger for the star. As for the "smaller" objects, we can see that the pattern spectrum for the star rises faster, as a result of the sharp gray-level stellar profile, while the galaxy, as a non-stellar object, possesses a smoother profile, as indicated by its pattern spectrum. However, for this case the average roughness attribute was inconclusive.

5 Concluding Remarks

In this work we have introduced an original approach to the problem of star/galaxy discrimination, based on Mathematical Morphology operations. The proposed method is based on high-level gray shape-size information, which is closer to the human visual characterization of astronomical objects, whereas other approaches use pixel-based geometrical and statistical attributes.

The watershed is a simple technique that proved to be a very good tool for region segmentation. Other methods, like SExtractor, use involved and heuristical approaches to this end, while our pre-processing step followed by the watershed yielded comparable results. We have yet to consider a post-processing step for dealing with the broken-galaxy and empty-region problems, which will further improve the segmentation step. Another possibility is to consider the use of marker-based homotopy change of the image prior to the application of the watershed, as long as the markers can be obtained automatically.

We have shown that the pattern spectrum is a good shape-size descriptor. We think we can get even better results by looking for a more suitable family of structuring elements for the granulometry. Al-

though the family used in this work was able to give satisfactory results, it is very simple, not taking into account detailed structural object information. For instance, we know that a distinctive feature of galaxies is the fuzzyness and slow gray-level variation, and this could be explored better with a family of properly scaled gaussian functions as structuring elements.

Most objects can be satisfactorily treated by our method, except the very "small" ones, which can be detected but are hard to characterize by the pattern spectrum. The "small" objects which are not too faint can be handled by other methods in the literature with the use of photometric attributes. The faint ones are treated subjectively, by requiring the assistance of a human viewer to try to classify these objects visually or reject them for the classification step. But we would like to point out that even then, all the methods found in the literature are not completely automated in the classification step for all cases, while we intend to use our method to build a fully automatic neural-network classifier.

Acknowledgements

The authors would like to thank Dr. Laerte Sodré and his group at IAG-USP (Instituto Astronômico e Geofísico da Universidade de São Paulo) for kindly providing the images used in this work and their results obtained with the SExtractor software, and also for the remarks on astronomical imaging technical literature. We also thank Drs. Gerald Banon, Ivo Busko and Steve Odewahn for the incentive received.

References

[1] G.J.F. Banon. Characterization of translation-invariant elementary operators for gray-level morphology. In *Neural, Morphological and Stochastic Methods in Image and Signal Processing*, pages 68–79, San Diego, USA, Jul 1995. Proc. SPIE 2568.

[2] J. Barrera, G.J.F. Banon, and R.A. Lotufo. A Mathematical Morphology toolbox for the KHOROS system. Technical Report RT-MAC-9403, Instituto de Matemática e Estatística - Universidade Estadual de São Paulo, São Paulo, SP - Brasil, Jan 1994.

[3] E. Bertin. *SExtractor 1.0 - User's guide*. Institute d'Astrophysique de Paris, 1995.

[4] U.M. Braga Neto and R.A. Lotufo. Image analysis of porous media by 3-d mathematical morphology. In *Anais do SIBGRAPI'95 - VIII Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens*, pages 59–66, São Carlos, SP - Brasil, Out 1995.

[5] E.R. Dougherty. *Mathematical Morphology in Image Processing*. Marcel Dekker, New York, 1993.

[6] H.J.A.M. Heijmans. *Morphological Image Operators*. Academic Press, Inc., 1994.

[7] R.M. Humphreys and R.L. Pennington. Workshop on digitized optical sky surveys. Edited by C. Jaschek and H.T. MacGillivray, 1989.

[8] P. Maragos. Pattern spectrum and multiscale shape representation. *IEEE Transactions on Pat. Anal. Mach. Intel.*, 11(7):701–716, Jul 1989.

[9] S.C. Odewahn, E.B. Stockwell, R.L. Pennington, R.M. Mumphreys, and W.A. Zumach. Automated star galaxy discrimination with neural networks. *The Astronomical Journal*, 103(1), Jan 1992.

[10] S.C. Odewahn, R.M. Mumphreys, G. Aldering, and P. Thurmes. Star-galaxy separation with a neural network - II - Multiple Schmidt plate fields. *Publications of the Astronomic Society of the Pacific*, 105:1354–1365, Nov 1993.

[11] R.L. Pennington, R.M. Humphreys, and F.D. Ghigo. *Mapping the sky: Past heritage and future directions (edited by Debarbet, J.A. Eddy, H.K. Eichorn and A.R. Upgren)*, pages 437–440. Kluwer, Dordrecht, 1987.

[12] J. Serra. *Image analysis and mathematical morphology*. Academic Press, Inc., 1982.

[13] J. Serra. *Image Analysis and Mathematical Morphology. Vol. 2: Theoretical advances* (edited by J. Serra), chapter 5, *Introduction to Morphological Filters*. Academic Press, Inc., 1988.

[14] M. Serra-Ricart, X. Calbert, L. Garrido, and V. Gaitan. Multidimensional statistical analysis using artificial neural networks: astronomical applications. *The Astronomical Journal*, 104(4):1685–1695, Oct 1993.

[15] S.R. Sternberg. Grayscale morphology. *Computer Vision, Graphics and Image Processing*, 35:333–355, 1986. Special edition on Mathematical Morphology.

[16] L. Vincent and P. Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pat. Anal. Mach. Intel.*, 13(6):583–598, Jun 1991.