Astronomical Data Analysis Software and Systems XIII
ASP Conference Series, Vol. 314, 2004
F. Ochsenbein, M. Allen, and D. Egret, eds.

Advances in Automated Algorithms For Morphological Classification of Galaxies Based on Shape Features

S. Goderya and Jonathan D. Andreasen

Physics Department, Illinois State University, Normal, IL 61790-4560, Email: goderya@phy.ilstu.edu

N. S. Philip

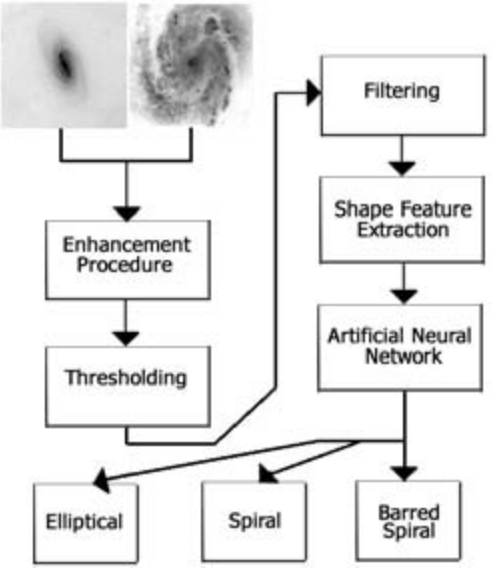
St. Thomas College, Kozhencherry, India

Abstract. Among the many celestial objects in the universe, galaxies offer insights as to how the universe was formed and is continuing to develop. The morphological classification of galaxies is important just for this reason. The challenge lies in classifying the estimated billions of galaxies that are in the universe, a very small amount of which are now being studied by various sky survey like the Hubble Deep Field and the Sloan Digital Sky Survey. The automated procedure described here uses an image enhancing technique, segmentation, shape feature extraction and a supervised artificial neural network to classify the galaxies. When trained to classify galaxies as E/S0 or S, the network is able to learn 98.3% of the galaxies correctly and identify 89.9% of galaxy images in a test set. The major challenge is in the development of robust and automated segmentation schemes. With manually threshold images and Difference Boosting Neutral Networks we were able to achieve considerable success in developing a supervised classifier capable of sorting galaxies into subclasses.

1. Introduction

When defining a galaxy's classification, a human classifier makes precise measurements. This process can take a fairly long time, and the human vision system is not especially suited for repetitive procedures such as extracting mathematical features of galaxies. Abraham et al. (1996), described how just two features of galaxies, the Asymmetry Index and Central Concentration of Light, can be used to distinguish between E/S0 and S galaxies. Properties such as these which are characteristic of galaxy types can be given to an Artificial Neural Network (ANN) (Angstenberger 1996 & Mahonen 1995), which in turn can classify the galaxies. Our scheme for classification can be seen in Figure 1. This paper focuses on all the parts of this scheme and demonstrates how an ANN is able to easily distinguish between E/S0 and S galaxies. We discuss important image processing techniques, shape extraction methods and ANN that allowed the implementation of this scheme. All the code has been implemented in C language

Figure 1. Overall Automated Galaxy Classification Scheme.



on Linux platform, with the exception of interactive thresholding module that was performed by using Khoros Image Processing software [1].

2. Image Processing

The images used in this study were obtained from the Zsolt Frei Catalog (Frei 1999) which contains approximately 113 different galaxy images. This database is often used as a benchmark for astronomical study. The images are carefully calibrated CCD images. The images downloaded from the Zsolt Frei website are processed in order to extract the necessary shape features. The first step is a window to window histogram equalization process which performs a local histogram equalization on each defined subsection of the image (Parker 1994). This is used primarily to enhance the arms of spiral galaxies. We have tried several other algorithms for image enhancement, however this algorithm produces the best results.

After enhancement, the images are thresholded to convert the gray scale information to binary scale information. Two automated methods which rely on the image having a bi-modal histogram, Isodata algorithm (Ridler 1978) and the Triangle algorithm (Zack 1977) were tried. In the case of galaxy images, the Isodata technique tends to choose a threshold too high, turning too many of the necessary pixels that carry information about the galaxy into background pixels. The triangle algorithm is quite the opposite, choosing a threshold too low which includes too many pixels from the background that are not part of the galaxy. We adopted a midway approach by taking the average of the two

[1]Khoros Software: http://www.khoral.com

values for our automated threshold value. In order to test the reliability of the shape feature extraction by automated techniques, we also decided to threshold the images interactively using Khoros Image Processing Software.

After thresholding, the images are processed further to eliminate noise and other celestial objects via erosion. Next the galaxy's foreground is made uniformly white by eliminating black pixels in it via dilation. The images are then furthered filtered via a blob coloring routine which only keeps the largest connected foreground region in the image to obtain the isolated galaxy image.

3. Shape Extraction

Five different shape features which characterize a galaxy were extracted for each galaxy and were used as inputs to the ANN. Three of these parameters (Compactness, Bounding Rectangle to Perimeter ratio and Bounding Rectangle to Fill Factor ratio) have already been previously described (Goderya 2002). The Bounding Polygon, obtained by Jarvis' March method (Parker 1994), is implemented to compute the Bounding Polygon to Perimeter (BP-P) and Bounding Polygon to Fill Factor (BP-FF) ratios. The ratios involving the Bounding Polygon help to further distinguish between ellipticals and spirals, however, because of its tight binding nature it could also potentially be used to distinguish between different subclasses of spiral galaxies. Another benefit of constructing ratios is that the parameters are then invariant to galaxy size and rotation. In order to use the five features, we must implement an Artificial Neural Network which can see the non-linear effects and work in multi-dimensional space.

4. Artificial Neural Network

We experimented with two different types of networks. The first was the standard back-propagation algorithm. Two configurations $5\times2\times2$ and $5\times2\times3$ were tried. We calculated the Mathews correlation coefficients (MC) (Clark 1999). A value of 1 for MC indicates that the network perfectly identified every galaxy, a value of -1 indicates misclassification and a value of 0 indicates that the network is unstable. For our first configuration we find MC=0.944 for the training file and MC=0.690 for the test file. For the second configuration the Mathews coefficients for the training file are MC(E/S0)=0.884, MC(S)=0.871 and MC(SB)=0.819, while for the test file they are MC(E/S0)=0.54, MC(S)=0.280 and MC(SB)=0.184. It appears that the standard back-propagation algorithm does fairly well in classifying ellipticals and spiral galaxies. However it is not able to completely learn all the general characteristics in order to discriminate between simple spirals and barred spiral galaxies.

The second network we tried is the Difference Boosting Neural Network [2]. Our results show 100% classification accuracy with our training data thereby confirming the the fact that the technique we adopted fro extracting the parameters are valid and very effective.

---

[2]See P7.10 for details: Sajeeth N. Philip, "Optimal Selection of Training Data for the Difference Boosting Neural Networks" ADASS XIII, October 12-15, 2003

5. Discussion

We believe that it is possible to develop robust and practical automated galaxy classifiers for large sky surveys. There is considerable amount of effort being put by different people in this type of work. It is also important to keep in mind that shape of the galaxy is dependent on red shift, band pass bias, dust scattering and low surface brightness. The challenge would be make classifiers that would take into consideration all these problems. To do this, it will be important to investigate more parameters that are physically meaningful and correlate with the astrophysical properties of the galaxies. This would result in a classification scheme that would be more elaborate and free of any deficiencies unlike the Hubble classification scheme. We have started looking into these directions and the results will be presented in future meetings.

Acknowledgments. This research has made use of the Zsolt Frei on-line catalog provided by the Department of Astrophysical Sciences at Princeton University. The authors would like to thank the Khoros Software people for providing their software to the academic community. Finally, thanks are due to the Department of Physics at Illinois State University for providing the computational and manuscript preparation facility for this research. Shaukat Goderya also acknowledges the financial support from American Astronomical Society, The ADASS organizing committee.

References

Abraham, R. G., Tanvir, N. R. Santiago, B. X. Ellis, R. S. Glazebrook, K. & Van den Berg, S., 1996, MNRAS, 279, L49

Angstenberger, J., 1996, in Neural Networks & Their Applications, edited by J. G. Taylor, (Wiley), pp. 143-152

Clark, J. W., 1999, in Scientific Applications of Neural Networks: Springer Lecture Notes in Physics, edited by J. W. Clark, Lindenaw, T. and Ristig, M. L., Vol. 522, (Springer-Verlag, Berlin)

Goderya, S. N. & Lolling, S. M., 2002, Morphological Classification of Galaxies Using Computer Vision and Artificial Neural Networks: A Computational Scheme, Ap&SS, 279, pp. 377-387

Mahonen, P. H. and Hakala, P. J., 1995, Automated Source Classification using a Kohonen Network, ApJ, 452, L77-L80

Parker, J. R., 1994, Practical Computer Vision using C, (Wiley)

Ridler, T. W. and Calvard, S., 1978, Picture thresholding using an interactive selection method, IEEE Trans. on Systems, Man, and Cybernetics, SMC-8(8), 1264-1291

Websource: Frei, Z., 1999, Zsolt Frei Galaxy Catalog. Retrieved 2002 from Princeton University, Department of Astrophysical Sciences Web site: http://www.astro.princeton.edu/ frei/catalog.htm

Zack, G. W., Rogers, W. E. and Latt, S. A., 1977, Automatic Measurement of Sister Chromatid Exchange Frequency, Journal of Histochemistry and Cytochemistry 25 (7), pp. 741-753