ROMAI J., **5**, 1(2009), 123–130

OBJECT CLASSIFICATION METHODS WITH APPLICATIONS IN ASTRONOMY

Corina Săraru

Department of Computer Science, University of Piteşti

corina_sararu@yahoo.com

Abstract

This paper reviews some of the methods used in practice for object classification and studies the possibility of combining a well-known object classification technique with other image processing methodologies such as edge detection. An application of this proposed process may be the classification of galaxies, based upon the Hubble classification. The feature taken into consideration is the galaxy shape.

Keywords: object classification, PCA, Hubble galaxy classification.

2000 MSC:85A99.

1. INTRODUCTION

The need of automated object recognition is of great importance nowadays not only because of the practical impact in almost all fields of information manipulation, but also because of the large quantity of data available for processing.

Many approaches have been proposed in this purpose. There are different types of artificial neural network architectures taken into consideration as well as statistical methods to discriminate between classes of objects. Also, there are efficient methods that use fractal dimensions of objects to be classified in order to assign them to a certain class.

The large amount of data received from satellites nowadays makes very difficult their analysis by the human factor manually. That is why there are many attempts to automate this process, to develop ways in order to obtain processed elements out of the information provided by observatories or satellites.

For example, Institut d'Astrophysique & Observatoire de Paris has developed a software [9] to able create catalogues of objects by analyzing images that contain different surfaces of the sky. The main feature of this package is the use of neural networks applied in classification techniques.

Other example of this type of software is YODA (Yet another Object Detection Application) [10]. This type of software represents an important methodology of astronomical image processing. It computes different shape parameters and classifies objects according to more than one approach.

There are a few developing directions regarding the classification of objects in astronomy images. First, there is the need of pre-processing the raw information received from the information capturing device (such as satellites, terrestrial and extra-terrestrial telescopes or NASA spacecrafts). The results of astronomical observations are sometimes one dimensional signals that need to be transformed into digital images. The following step is to process the image obtained, so that the result emphasizes the objects appearance. This is done taking into consideration some of object features in the picture, among which shape plays an important part. The last step is the actual classification.

2. CLASSIFICATION OF GALAXIES

There are several types of features relevant to the classification (of galaxies). They can be divided into: photometry, profile and shape features [2]. The photometric features (which represent the central concentration of light index) and the profile features (the radial distribution of surface brightness, otherwise representing asymmetry) can be used [1] to morphologically classify galaxies.

The shape features are more attractive when dealing with morphology of galaxies because the morphology itself is a visual context for classification.

Here are some important shape features which are successfully used in classification techniques: elongation (the measure of flatness of an object), a form factor which is the ratio between the area and the square perimeter of the galaxy (this is important because elliptical galaxies are more luminous while spiral ones have much broader areas and less luminosity), convexity (which is very large in spiral galaxies and very small for elliptical ones) [2]. As seen in fig. 1, there are a few main types of galaxies according to shape classification:

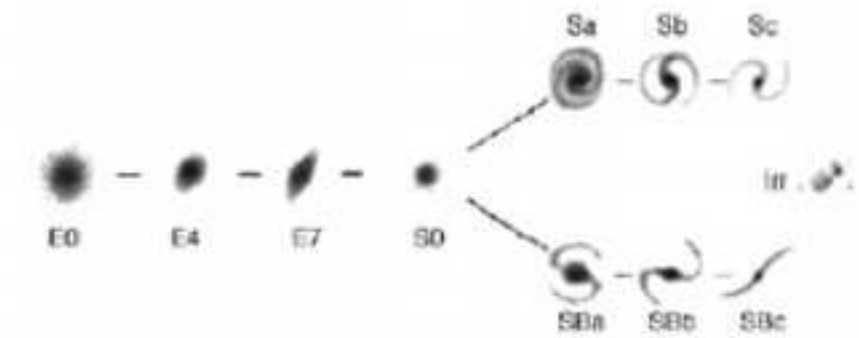

Fig. 1. Hubble's classification scheme.

elliptical galaxies (denoted by **E**), spiral galaxies (denoted by **S**). According to the shape of the spiral, the spiral galaxies can be subdivided into three subclasses: spherical or barred-like; lenticular (denoted by **S0** and characterized by a bright center); and irregular, which can not be assigned to any class mentioned above.

3. ARTIFICIAL NEURAL NETWORKS AND GALAXY CLASSIFICATION

Artificial neural networks have been successfully used in automation of galaxy classification. Both supervised (back propagation method) and unsupervised (self organizing maps such as Kohonen networks) types of neural networks were brought to light by researchers.

The first attempts used the raw image data to train the classifiers, leading to high error rates.

Nielsen and Odewahn (1995) emphasize the idea of using parameters characterizing galaxies instead of straightforward pixels of the source image. Their results showed that the use of profile features-based classifier is the most efficient (as compared to other feature-based methods or raw pixel use) [2]. A description of a back propagation application with two hidden layers is found in [4]. The method assumes the use of some parameters in order to compute the weight vectors of the neural network; first, photometric parameters (such as surface brightness, concentration index and color) are used for classification. Then, parameters are switched to profile-like features (brightness

profiles in two band passes [2]). The last experiment was made on raw pixel data obtained from images.

Another approach in this field of research is presented in [5] and it uses an algorithm called DBNN (Difference Boosting Neural Network).

Although this architecture is closely related to the naive Bayesian classifier, it gives some degree of freedom regarding the correlation of the data attributes. The instrument used to this purpose is the association of a threshold window with every attribute. This window influences the decision coefficients of the classifier. The classification system contains a boosting technique, which emphasizes the difference between training data elements.

There are other more recent papers which point to the idea of using also other mathematical tools such as the fractal signature of classified objects [6].

Another approach is the one presented in [3]. It implies the classifier Random Forrest and involves a number of decision tree classifiers. The individual decisions are combined to give the final classification result. The algorithm proposed in [3] is an efficient tool for the galaxy classification problem. The preprocessing part implies geometric transformations of the image, such that shape features of the galaxies are enhanced.

These transformations are:

1 a threshold, in order to extract the the bright part of the image and eliminate some background irrelevant elements;

2 rotation of the image with an angle given by the first principal component of the image, so that the galaxy is brought into a horizontal position (the standard position for all images, in order for the classifier to have higher efficiency);

3 then, the image is resized to standard dimension, in order to be included into the training sample matrix.

The use of the pixel information would be very expensive in terms of resources because of the dimension of the data (e. g. if all images would be standard 128x128, the training matrix would be Nx16384, where N is the number of images taken into consideration). That is why in order to reduce dimen-

sionality, the authors propose the use PCA (Principal Component Analysis). As experimental results show, this is a very effective method which enables significant reduction of computation time and resources use.

The next step is represented by the actual classification, which is done by a classifier that assumes the use of decision trees.

4. PROCESSING DATA

The algorithm reviewed above was implemented as follows:

1 preprocessing techniques described above;

2 result images were resized in standard dimension;

3 application of PCA techniques and use of them in classification at different values;

4 training of the artificial neural network by a number of N digital images. The original algorithm is using RF techniques for classification;

5 testing of the weight matrix by presenting to it new images yet to be classified.

The experimental results presented below were obtained by applying the above algorithm with the following changes:

1 in order to enhance optimally the shape and brightness distribution of the galaxies, preprocessing techniques described above were used, at different values of the threshold;

2 result images were resized in dimension 24x24 (and not 128x128, as the original algorithm suggested).

3 application of PCA techniques and use of them in classification at different values. Improvement of efficiency relative to the increasing number of principal components used in training of the artificial neural network was noticed, as it can be noticed from the results below.

4 training of the artificial neural network by a number of N digital images (N=33) in order to obtain a minimum error rate for classifying the

Table 1 Classification on training data set

number of PCs	error
8	0.2821
15	0.2308
20	0.2051
25	0.1282
30	0.0513
35	0

Table 2 Classification on 5 new images

number of PCs	error
8	80%
15	20%
20	40%
25	20%
30	40%
35	20%

images given to training. The experimental results presented below are obtained by using a standard classifier.

5. EXPERIMENTAL RESULTS

The images used were downloaded from [11]. The implementation was written in Matlab, version 7.6.0.324 and tested on configuration including a 2Ghz processor and 1G of RAM.

6. CONCLUSIONS

In automated galaxy classification good results were obtained by using several types of artificial neural networks. Experimental results show that even on a reduced dimension of space of training samples, the efficiency of the algorithm is satisfying. Future directions in this domain may include the use of different algorithms such as adaptive algorithms for different neural networks, boosting algorithms or combinations of methods generally used in object classification.

Acknowledgements. The images used in this research were taken from www.seds.org. Special thanks are due to Dr. Maria Miroiu for her important collaboration and suggestions.

References

[1] Abraham, B., Valdes, F., Yee, H., Carlberg, R., *Automated morphological classification of faint galaxies*, 1993.

[2] Goderya, S. N., Lolling, S. M., *Morphological classification of galaxies using computer vision and artificial neural networks: a computational scheme astrophysics and space science* **279**(2002), 377 - 387.

[3] Calleja, J., Fuentes, O., *Automated Classification of Galaxy Images. Lecture Notes in Computer Science*, **3215**(2004).

[4] Odewahn, S., *Galaxy Classification Using Artificial Neural Networks*,I nternational Joint Conference on Neural Networks, **6**(1999), 3824 - 3829.

[5] Philip, N. S., Wadadekar, Y., Kembhavi, A.,Joseph K. B., *A difference boosting neural network for automated star-galaxy classification*, Astronomy & Astrophysics, **385**, (2002).

[6] Lekshmi, S., Revathy1, K., Prabhakaran Nayar, S. R., *Galaxy classification using fractal signature*, Astronomy& Astrophysics, (2003).

[7] Miroiu, M., Vaduva, I., *Special Tailored Boosting Methos for Face Recognition*, SEERC, (2008), 101 -112.

[8] Turk, M. A.,; Pentland, A. P., *Face recognition using eigenfaces*, IEEE Computer Society Conference **3-6**(1991) 586 - 591.

[9] http://terapix.iap.fr/rubrique.php?id_rubrique=91

[10] http://www.aanda.org/index.php?option=article&access=standard&Itemid=129 &url=/articles/aa/pdf/2003/01/aah2400.pdf

[11] http://www.seds.org/messier/objects.html#galaxy