VISUAL AND AUTOMATIC CLASSIFICATION OF GALAXY IMAGES

Allan Wirth
Harvard-Smithsonian Center for Astrophysics

ABSTRACT

The requirements for a classification system for galaxy images are discussed and the advantages and disadvantages of both visual and automatic classification techniques are assessed. The results of some preliminary experiments in the automatic classification of images are also presented.

1. INTRODUCTION

Of all quantities associated with an astronomical object, visual appearance is the most easily measured. The chief task of the morphologist is the creation of the scale upon which the quantity "appearance" is to be measured. Until very recently the primary technique used to construct such a scale for galaxies has been the visual inspection and ordering of galaxy images in what seems to be a reasonable way. In doing so we are utilizing the great power of the human eye-brain system to detect (or imagine) regularities of structure even in very low signal-to-noise situations. The hypothesis that such regularities represent the action of physical processes is basic to all morphological research. With the rapidly increasing capability of non-human image processing technology, it is becoming possible to attempt to construct a more objective classification system. In this paper, we will examine the advantages and limitations of the present classification techniques and briefly describe some preliminary work on the automatic classification of galaxies.

2. VISUAL CLASSIFICATION OF GALAXIES

The great advantage of visual classification over all other methods of studying the structure of galaxies is the rapidity with which it can be done and the huge number of galaxies within reach of present telescopes. On sky limited fine grain plates of scales between 100 and

M. Capaccioli (ed.), Astronomy with Schmidt-Type Telescopes, 129–131.

https://doi.org/10.1017/S0252921100106414 Published online by Cambridge University Press

60 "/mm taken in good seeing, it is possible to derive detailed classifications (i.e. full Hubble or Yerkes type) for galaxies larger than ~15" diameter, which is roughly equivalent to a magnitude of 17. Over the whole sky, there are ~500000 such galaxies, which, for an experienced classifier, represents less than half a man-year of work.

The primary limitations of visual classification of galaxies are twofold. The first stems from the difficulty the human mind has in perceiving relationships in a multidimensional classification space. This is, perhaps, the reason for the lack of success we have had in providing physical explanations for the observed patterns in galaxy classifications. Stellar spectral classification has provided very clear insight into the physics of stellar structure because nature has fortuitously allowed stars only a small number of independent physical parameters. If, for instance, galaxies form sequences in a six dimensional space, such sequences may be invisible to the human classifier. The second limitation is the difficulty of selecting the parameters to use in the classification. The eye may be drawn to the most striking structure in the image as a critical parameter, but it is not clear that such structure is physically important. As an example, the spiral structure seen in photographs taken in blue light may be classified in great detail but represents only a comparatively small physical variation relative to the underlying galactic disc. There may be other equally easily determined parameters that are unappealing to our esthetic views but of greater physical significance.

3. AUTOMATIC CLASSIFICATION

There are two ways in which automation may be introduced into the classification process, each being related to one of the deficiencies of visual techniques. First, there exist a large number of algorithms that may be used to search for correlations between the classification parameters of a set objects. These routines are in fairly wide use in the fields of medicine, ecology, geology and remote sensing (see, for example Sneath and Sokal 1973). Some very promising results have been obtained from the application of these methods to the available galaxy data (Whitmore 1983). The second automatic technique which may be applied to galaxy classification is the creation of machine generated classifiers. That is, to use objective algorithms to produce classification parameters directly from digital images of galaxies. As discussed elsewhere in these proceedings, this method has been used very successfully to sort faint images into galaxy or star bins. Its application to larger images of the sort of interest here has been hindered by the large number of numerical operations required to convert the image into a set of classification parameters. With the greatly increased processing power of computers now available, it becomes realistic to consider the wholesale classification of galaxies from such plate material as the new IIIaJ sky surveys.

Several experiments designed to explore the possibilities for

https://doi.org/10.1017/S0252921100106414 Published online by Cambridge University Press

automatic classification have been begun. Using standard techniques, extended images are located in PDS digitized CTIO schmidt plates. After subtracting the sky background, a number of image parameters are calculated. These include the first three moments of the intensity distribution, the isophotal axial ratio, and the slope of the intensity distribution at a given isophote. Some of these quantities are found to correlate with visual classification parameters, e.g., the ratio of second moment to total intensity relates to the Yerkes concentration class and axial ratio to the ellipticity parameter. However, the most interesting delineation, that between spiral and elliptical, cannot easily be made using only these quantities. In order to recognize spiral and irregular galaxies, it is necessary to be sensitive to higher frequency components in the image. One technique that allows a direct estimate of the relative importance of high frequency structure in an image is the production of a spatial power spectrum. Using an FPS-120B array processor on a VAX 11/780 host, it is possible to obtain the Fourier transform of a 256X256 image in somewhat less than one second. Such transforms have been produced for a number of galaxy images, using as input both the ordinary cartesian representation of the data and, because galaxies generally have axial symmetry, the same data mapped into polar coordinates. The transforms of the latter are more suited to easy interpretation by the human mind. The power at azimuthal frequencies in the range 2-6 cycles/2 radians is very strongly related to the apparent strength of the spiral structure in the image. It appears that a quite reliable discriminant between the early and late type galaxy bins may be constructed from this transform. One great advantage of working in polar Fourier space is that the power at various azimuthal frequencies does not depend upon the radial scale or position of the galaxy in the image. Thus, the same classifier may be used for both large and small galaxy images.

In order to make the fullest use of the objective classification power of machines, rather than attempt to interpret these derived parameters in terms of present classification systems or our current ideas of galaxy structure, it is better to combine both phases of automated classification. If, for a large sample of galaxies, a number of arbitrary parameters are generated and then the distribution of these points in the classification space is examined for clusterings and correlations, it is possible to allow the machine to select the most useful parameters. This process has been used to successfully create a device that rapidly searches images of blood samples for diseased cells. If a similar technique also proves successful for galaxy images, we may, for the first time, be able to obtain a truly objective classification system free of all prejudices, both physical and esthetic.

REFERENCES

Sneath,P.H.A., and Sokal,R.R.: 1973,"Numerical Taxonomy" (Freeman:San Fransisco)
Witmore,B.: 1983, Private communication.