

# Mathematical Morphology: Star/Galaxy Differentiation & Galaxy Morphology Classification

Jason A. Moore<sup>A,B</sup>, Kevin A. Pimbblet<sup>A</sup> and Michael J. Drinkwater<sup>A</sup>

<sup>A</sup> Department of Physics, University of Queensland, Brisbane QLD 4072, Australia

<sup>B</sup> Corresponding author. E-mail: jmoore@physics.uq.edu.au

Received 2006 March 9, accepted 2006 October 11

**Abstract:** We present an application of Mathematical Morphology (MM) for the classification of astronomical objects, both for star/galaxy differentiation and galaxy morphology classification. We demonstrate that, for CCD images,  $99.3 \pm 3.8\%$  of galaxies can be separated from stars using MM, with  $19.4 \pm 7.9\%$  of the stars being misclassified. We demonstrate that, for photographic plate images, the number of galaxies correctly separated from the stars can be increased using our MM diffraction spike tool, which allows  $51.0 \pm 6.0\%$  of the high-brightness galaxies that are inseparable in current techniques to be correctly classified, with only  $1.4 \pm 0.5\%$  of the high-brightness stars contaminating the population. We demonstrate that elliptical (E) and late-type spiral (Sc-Sd) galaxies can be classified using MM with an accuracy of  $91.4 \pm 7.8\%$ . It is a method involving fewer ‘free parameters’ than current techniques, especially automated machine learning algorithms. The limitation of MM galaxy morphology classification based on seeing and distance is also presented. We examine various star/galaxy differentiation and galaxy morphology classification techniques commonly used today, and show that our MM techniques compare very favourably.

**Keywords:** techniques: image processing — methods: data analysis — methods: miscellaneous

## 1 Introduction

The bulk of modern astronomical observations are performed with charge-coupled devices (CCDs) and hence are, almost by default, digital in nature. Also, many pre-existing sky surveys compiled with photographic plates have been digitized using plate-measuring facilities such as SuperCOSMOS (Hambly et al. 2001a; Hambly, Irwin, & MacGillivray 2001b; Hambly et al. 2001c) and the Automatic Plate Measuring (APM) machine (Kibblewhite et al. 1984). This digitization of astronomical data has provided the opportunity for computational solutions to many image analysis problems. Over the past four decades, many digital techniques have been developed for filtering, pattern recognition, neural networks, artificial intelligence, and others. Unsurprisingly, many of these techniques have found applications in the area of astronomical imaging.

In particular, accurately and uniformly classifying objects in astronomical images is of great importance. One way for this to be achieved is by performing the classification by eye; however, there are issues with this. Firstly, eye-ball classification is extremely subjective and it is difficult to ensure uniform classification. As an example, Shaver (1987) presented evidence for a possible large-scale structure in the distribution of quasars out to redshift of about 0.5, concentrated in the direction of the cosmic microwave background dipole. However, follow up study by Drinkwater & Schmidt (1996) concluded that there is no evidence for such a

concentration, and that the earlier result was probably biased by the use of non-uniform image classifications. Secondly, most astronomical observations, particularly those for cosmological studies, seek to compile object catalogues over large regions of the sky. The corresponding increase in image data (a single photographic plate can contain hundreds of thousands of detectable objects) calls for the development of fast image processing, recognition, and classification by automated means. This problem of finding a robust method to cleanly distinguish between all astronomical object types is one of the most challenging in astronomical image analysis.

### 1.1 Star/Galaxy Differentiation

The problem of separating the galaxy population from the stellar population within an astronomical image is one with a variety of proposed solutions. In broad terms, stars and galaxies can be distinguished from each other on the basis of their light profiles. Stars are expected to possess a highly peaked point-like profile (the size of which is given by the point spread function of the given image), whilst galaxies tend to be more extended in nature. One basic technique is to take two parameters which describe an object in an image, plot them against each other, and then use the line segment that optimally separates the stars from the galaxies. Parameters which have been commonly used (plotting against the magnitude of the object) include isophotal area (e.g. Reid & Gilmore

1982), peak flux (e.g. Jones et al. 1991), or surface brightness (e.g. Harmon & Mamon 1993).

As an example to consider, to separate stars from galaxies in the Las Campanas/AAT Rich Cluster Survey photometric catalogue (LARCS), Pimbblet et al. (2001) followed the technique of Reid et al. (1996), by calculating the magnitude difference between 4.0- and 2.0-arcsecond diameter apertures on *B*-band exposures. Using these, Pimbblet et al. (2001) tested various other parameters of merit, including a concentration index (Abraham et al. 1994), full width at half maximum (FWHM), SExtractor's own estimate of stellarity (Bertin & Arnouts 1996), and ellipticity. Of these parameters, they found that both FWHM and SExtractor possess strong reliability and accuracy in separating stars and galaxies. SExtractor is very competitive as it offers not only a rapid dissemination of results, but also has the ability to reliably deblend sources and perform star/galaxy differentiation in a highly automated fashion, thus making the package well-suited to the reduction of large-scale galaxy survey data.

A plethora of additional parameters exists which can also be helpful, including core magnitude, intensity weighted first moment radius, radius of gyration, the Seaborn (1979) classifier, the  $r^{-2}$  'moment' of Kron (1980), the *Q* classifier of Lefèvre et al. (1986), and the  $\psi$  parameter of Maddox et al. (1990).

A more sophisticated, and less subjective, approach is to automate the classification, using any number of the above parameters. The classifier can be trained on a set of pre-classified objects by performing an automated 'machine learning' algorithm such as decision trees (e.g. Weir, Fayyad, & Djorgovski 1995), self organizing maps (e.g. Miller & Coe 1996), artificial neural networks (e.g. Odewahn et al. 1992, 1993; Bertin & Arnouts 1996; Andreon et al. 2000; Philip et al. 2002), or fuzzy set reasoning (Mähönen & Frantti 2000). SExtractor uses a trained neural network for star/galaxy classification. The main problem with this technique is the large number of 'free parameters'. Each node, the number of nodes, and the layering of the nodes are all parameters that need to be set or tweaked within an artificial neural network. Additionally, the 'machine learning' algorithms must be extensively trained on pre-classified images which may introduce further biases.

### 1.2 Galaxy Morphology Classification

Another challenging problem in astronomical image analysis is finding a robust method to cleanly distinguish between all major classes of galaxies. While most classifications of galaxy morphology have been performed by eye using the Hubble tuning-fork system, and its subsequent extensions (see Hubble 1926, 1936; van der Bergh 1960; Sandage 1961), there still exist a few automated classification techniques.

The most common technique for classifying the morphology of a galaxy is through light decomposition. This

involves fitting a model to the surface brightness profile of the galaxy, typically based on the Exponential, de Vaucouleurs (1948, 1959), and/or Sérsic (1963, 1968) profiles. These models effectively decompose the galaxy profile into bulge (de Vaucouleurs) and disk (Exponential) components, making them handy techniques for classifying galaxies into morphological types defined by the Hubble tuning-fork system. As an example, GALFIT (Peng et al. 2002) is a galaxy and point source fitting algorithm that fits two dimensional parameterized, axisymmetric functions (including those mentioned above) directly to images.

There are several issues, however, with using light decomposition. Firstly, the de Vaucouleurs profile does not produce the best fit for all bulges (see Andredakis & Sanders 1994; MacArthur, Courteau, & Holtzman 2003). Secondly, even assuming a more general Core-Sérsic model (Graham & Driver 2005, and references therein) for the bulge, the minimization of at least five parameters must be used to determine the 'best' solution, which is not always unique.

Conselice (2003) presents an alternative solution for classifying galaxy morphology using a three dimensional 'CAS' volume: the concentration (*C*), asymmetry (*A*), and clumpiness (*S*). He further argues that these three parameters correlate with important modes of galaxy evolution: gross form, major merger activity, and star formation. The definitions and formalism for the 'CAS' parameters can be found in Appendix A.

### 1.3 Outline

The plan for the remainder of the paper is as follows. In Section 2, the quantitative image analysis technique Mathematical Morphology (MM) is presented, including an overview of the operators used in this work. In Section 3, the problem of star/galaxy differentiation is investigated, including a comparison of MM with the results of current techniques. In Section 4, the problem of galaxy morphology classification is investigated, including a comparison of MM with the results of current techniques. We summarize our findings in Section 5.

## 2 Mathematical Morphology

MM is a branch of digital image processing and analysis originating from the work of Matheron (1975) and Serra (1982), who worked on a number of problems in mineralogy and petrology. They laid down the foundations of MM, a new approach to quantitative image processing. MM has now achieved a status as a powerful method for image processing, with applications in material science, microscopic imaging, pattern recognition, medical imaging, and even computer vision.

The International Society for Optical Engineering (SPIE) now holds an annual conference devoted to morphology applications, but astronomy applications

have been limited. Some of these applications include image smoothing (Lea & Kellar 1989), removing cirrus emission (Appleton, Siqueira, & Basart 1993; He 1996), astronomical object extraction (Candeas, Braga Neto, & Carvalho Filho 1996, 1997), and quantifying simulated galaxy distributions (Ueda 1999).

The central idea of MM is the process of examining the geometrical structure of an image by matching it with smaller patterns at various locations in the image. By varying the size and shape of the matching patterns, called structuring elements<sup>1</sup>, useful shape information can be obtained about the different features inside images and their interrelations. Although MM was originally designed to work on only binary images, subsequent modification led to its application to grey-scale images.

### 2.1 Structuring Elements

An essential part of the MM morphological operations is the structuring element used to probe the input image. The structuring element is a pattern of discrete points relative to some origin, called the active centre. It is often chosen to be a circle or line segment, but other choices are possible depending on the particular application. This however should not be viewed as a limitation, since it actually leads to additional flexibility in algorithm design.

One of the most common structuring elements for best representing a circle is the diamond shape, created using the city-block metric<sup>2</sup>. We adopt this as the most appropriate structuring element for the astronomical applications in this paper, with the features of interest being approximately circular in shape. Figure 1 contains three examples of the diamond structuring element, illustrating the start of an increasing series. The usefulness of a monotonically increasing series of structuring elements is described in Section 2.3.

## 2.2 Operators

In MM there are two elementary morphological operators: dilation and erosion, from which the other operators and tools can also be formed. Here, we give just a brief summary of each operator, but a more detailed description of both the operation of, and the theoretical basis for, these basic operators can be found in the work of Heijmans (1992, 1994, 1995).

### 2.2.1 Dilation Operator

Dilation is the process of brightening, expanding, and growing regions around the positive pixels of an image,

$I$ , using a structuring element,  $S$ . To perform this operation, the structuring element slides over the image. For each position that the structuring element is at, the maximum value contained within the structuring element (superimposed on the input image) is used as the pixel value for the output image:

$$(I \oplus S)_{x,y} = \max_{(i,j) \in S} (I_{x+i,y+j}) \tag{1}$$

The bottom diagram in Figure 2 shows a dilated image. Note the outward expansion of the shape that results from the operation.

### 2.2.2 Erosion Operator

Erosion is complementary to, but the opposite of dilation. Instead of expanding and growing regions of the image, erosion shrinks and wipes them out. The process is exactly the same as for the dilation operator, save that instead of calculating the maximum pixel value contained within the structuring element, it calculates the minimum:

$$(I \ominus S)_{x,y} = \min_{(i,j) \in S} (I_{x+i,y+j}) \tag{2}$$

The top-right diagram in Figure 2 shows an eroded image. Note the inward contraction of the shape that results from the operation.

### 2.2.3 Opening Operator

Opening is the process of separating and isolating small regions of an image from larger ones, and simultaneously eliminating information contained on small scales. By sequencing an erosion followed by a dilation, the opening operator can be formed:

$$I \circ S = (I \ominus S) \oplus S \tag{3}$$

The opening operator also has the property of being idempotent:

$$(I \circ S) \circ S = I \circ S \tag{4}$$

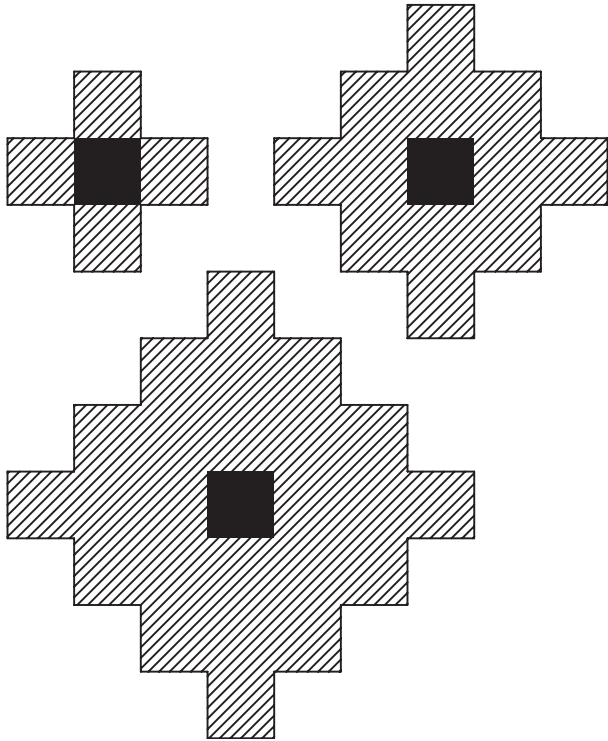
This means that no more changes will occur to an image, even if the opening operation is repeated multiple times. The opening operator is called a morphological filter from this property, as well as it being an increasing morphological operator; that is,  $X \subseteq Y$  implies that  $X \circ S \subseteq Y \circ S$ . The middle panels in Figure 3 shows two opened astronomical images, one of a spiral galaxy and one of a group of elliptical galaxies. Note the elimination of small scale features, of the same size scale as the structuring element used, from the original images.

### 2.2.4 Closing Operator

The closing operator is the opposite of the opening operator, in a similar way as dilation is to erosion. Closing is the process of filling in holes in the pixel distribution, and joining together previously separate

<sup>1</sup> In other disciplines, the structuring element is often referred to as a 'kernel'. In this work, however, we will stick with the original term.

<sup>2</sup> A diamond can be defined on a grid by the city-block metric,  $d(x,y) = |x| + |y|$ , where  $x$  and  $y$  are the grid coordinates relative to the active centre. Positions with  $d(x,y)$  less than a defined radius are included in the shape.



**Figure 1** This figure illustrates an increasing series of diamond structuring elements, created using the city-block metric. These  $3 \times 3$  (top-left),  $5 \times 5$  (top-right), and  $7 \times 7$  (bottom) structuring elements are the first three in the series. The shaded regions (both solid and hatched) indicate the grid contained within the structuring elements, with the solid fill representing the active centres.

image features. Mathematically, it is a dilation followed by an erosion:

$$I \bullet S = (I \oplus S) \ominus S \tag{5}$$

The closing operator has the same basic features as the opening operator such as being idempotent and working as a morphological filter. The bottom panels in Figure 3 show two closed astronomical images, one of a spiral galaxy and one of a group of elliptical galaxies.

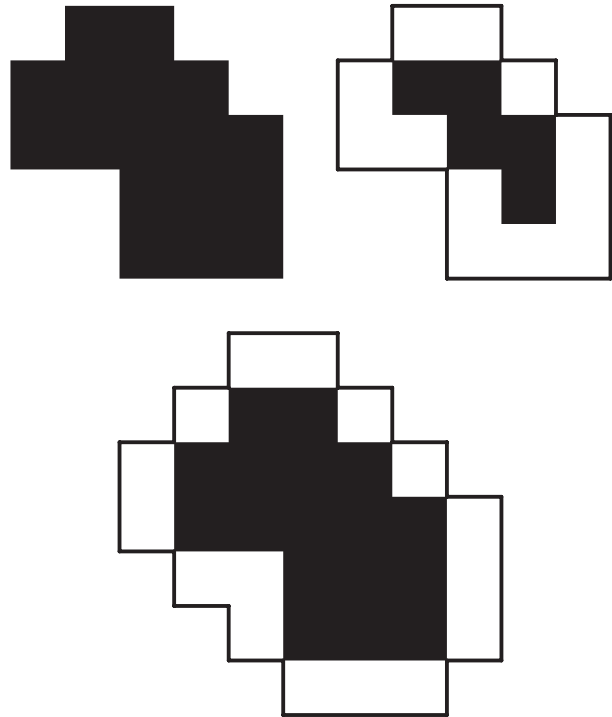
### 2.2.5 Gradient Operator

The elementary morphological operators can be combined together to detect any edges (i.e. sharp gradients) around shapes in an image, as well as the curvature of these objects. The gradient operator is formed by combining the dilation and erosion operators:

$$\nabla_S I = (I \oplus S) - (I \ominus S) \tag{6}$$

### 2.3 Granulometry

Granulometry uses the property of morphological operators which can be used to remove (or enhance) artifacts in an image of a certain size and shape, analogous to the sieving of rocks by sequentially using sieves of progres-



**Figure 2** The result of performing the dilation and erosion operations on the same shape, using a  $3 \times 3$  diamond structuring element. The top-left diagram shows the original shape. The bottom diagram shows the dilated shape as the outline and the original shape as the solid fill. The top-right panel shows the eroded shape as the solid fill and the original shape as the outline.

sively larger or smaller sizes. Granulometry consists of a sequence of closing (or opening) operations using an increasing series of structuring elements. By measuring the volume under the image after successive closings, a size distribution curve can be built:

$$\Phi(\lambda) = \frac{V(\lambda) - V(0)}{V(A) - V(0)}, \lambda \geq 0 \tag{7}$$

where  $\lambda$  is the parameterization of the series of closing operations,  $V(\lambda)$  is the volume of the image at each iteration, and  $A$  is the parameter associated with the largest structuring element (selected to be the one large enough to ‘wipe out’ the object of interest).

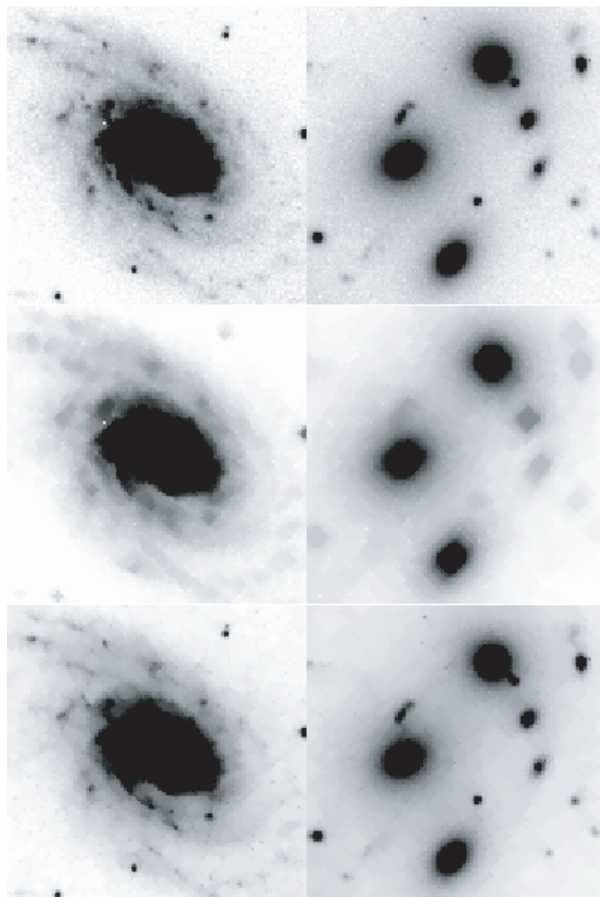
The size-distribution curve is monotonically increasing, so it may be considered to be a cumulative probability distribution. The associated probability density function is called the pattern spectrum, which is given by the following discrete derivative:

$$\Gamma(\lambda) = \Phi(\lambda + 1) - \Phi(\lambda), \lambda \geq 0 \tag{8}$$

We can define a useful shape analysis attribute, the average size, which is the expected value of the pattern spectrum:

$$\bar{\lambda} = \sum_{\lambda=0}^A \lambda \Gamma(\lambda) \tag{9}$$





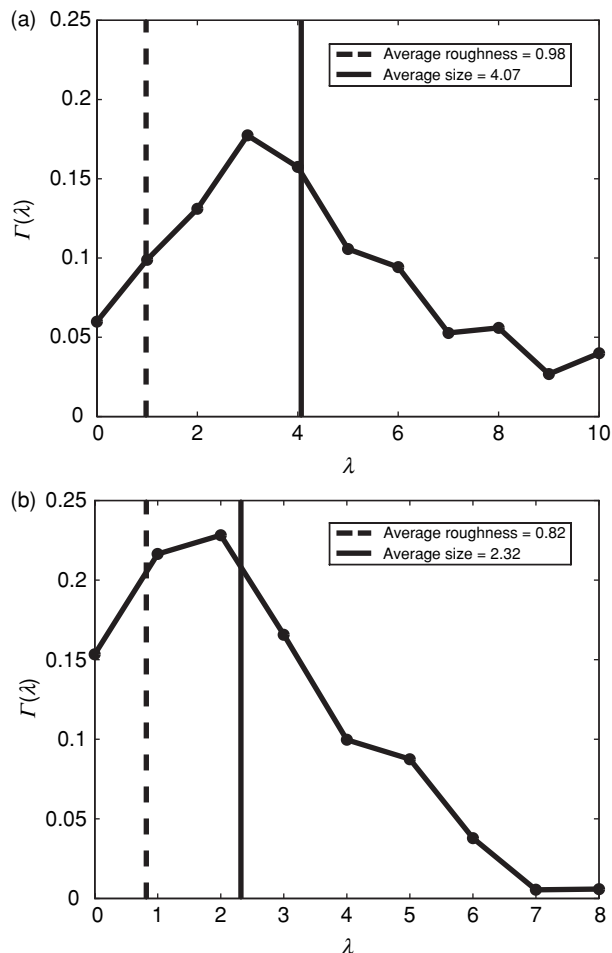
**Figure 3** The result of performing opening and closing operations on a spiral galaxy (left) and a group of elliptical galaxies (right). The top row shows the original astronomical images. The middle row shows the opened images, with small scale structures — on the order of the structuring element — being eliminated in both. The bottom row shows the closed images, with gaps between features being filled in.

Another very useful attribute is the average roughness of the pattern spectrum, which is identical to the equation for average uncertainty (entropy) in information theory:

$$\theta = - \sum_{\lambda=0}^A \Gamma(\lambda) \log [\Gamma(\lambda)] \quad (10)$$

Figure 4 illustrates how the pattern spectrum will look for typical spiral and elliptical galaxies. Also included on the plots are the values of average size and average roughness.

Of the MM tools derived, the average size,  $\lambda$ , and the average roughness,  $\theta$ , are a powerful combination for characterizing shape and size in image analysis (see Maragos 1989), providing statistical information about the shape content of the image. In this work, we present an implementation of basic MM operators and applications of MM in two astronomical object classification problems: star/galaxy differentiation and galaxy morphology classification. We expect MM to be superior to many of the current techniques for astronomical object classification, with the benefit of only using one ‘free



**Figure 4** Pattern spectrum plots for CCD images of typical (a) spiral and (b) elliptical galaxies. The pattern spectrum represents the distribution of size within an image, based on the shape of the structuring element. The dashed vertical lines on the plots indicate the average roughness, and the full vertical lines represent the average size.

parameter’, the structuring element, compared with at least five (ellipse fits) for the techniques mentioned in Sections 1.1 and 1.2.

### 3 Star/Galaxy Differentiation

To test the advantage of using MM to differentiate between stars and galaxies, both CCD and photographic plate images were used.

CCD chips possess many qualities that make them superior to photographic plates, including a large dynamic range, higher quantum efficiency, a higher linear response, and relatively low noise levels. Photographic plates can have a larger surface area, making them perfect for capturing a wide field-of-view, although they have drawbacks such as the low sensitivity and non-linear response of the plates, saturation of bright objects, high surface brightness limits, and possible inconsistencies that exist between different plates in large surveys (such as the SuperCOSMOS Sky Survey or the APM Sky Catalogue).

The major disadvantage of using photographic plates in astronomy is the saturation of bright objects such as stars or bright galaxies. The profile of a saturated star becomes extended, similar to that of a galaxy. This makes it difficult to differentiate stars from galaxies based on their light profiles. CCD images, however, maintain most of the original light profile, giving them an advantage in star/galaxy differentiation. Using both astronomical image types allows us to test the limitations of the various classification techniques.

### 3.1 Classifier Performance

We focus on the separation of the galaxy and star populations, which is most important in, for example, galaxy surveys and cosmological studies. Therefore, we rate the performance of a star/galaxy classifier based on the following three criteria:

1. Completeness — the percentage of galaxies correctly classified as such.
2. Contamination — the percentage of stars incorrectly classified as galaxies.
3. Correctness — the overall percentage of objects correctly classified as stars or galaxies.

Rating a star/galaxy classifier by these criteria, particularly the first two, gives an indication of how well the classifier extracts the galaxy population (e.g. for follow up spectroscopic study) without having too many stars contaminate the sample.

In this work, a multi-layer perceptron (MLP) is used as the automated classifier to separate the stars from the galaxies for a given star/galaxy classification technique. MLP (see Haykin 1998) is the most common neural network, and has been applied to many complex pattern recognition problems. It is composed of a layering of nodes (or perceptrons) whose output is determined by a linear combination of the previous nodes put through a nonlinear activation function. The weighting factors within the linear combination are varied until the optimal solution for a given set of training data is found. Based on the number of layers, and nodes in each layer, any function can be modelled. In this work, however, we have reduced the number of network nodes to the minimum to perform basic linear classification.

### 3.2 CCD Images

The CCD images were taken from the data archive of Smail et al. (1997), who produced a classified catalogue of morphological types of 1857 astronomical objects within ten intermediate-redshift ( $0.37 < z < 0.56$ ) rich galaxy cluster fields, observed using the Wide-Field Planetary Camera 2 (WFPC2) on the Hubble Space Telescope (*HST*). They classified each detectable object within the field by eye, according to the Revised Hubble scheme, as either a star, spheroidal (E/S0), early-type spiral (Sa-Sb), late-type spiral (Sc-Sd), barred spiral (SBa-SBd), or irregular (Irr/Sm) galaxy. The reliability

of these classifications is hard to quantify, but was considered by the authors to be accurate to within one class or better in at least 80% of cases. The objects in the sample range in magnitude from 23.5 at the faint end to 15.5 at the bright end, over the filter bands *B* (F450W), *V* (F555W), *R* (F702W), and *I* (F814W).

We used this catalogue, both the *HST* images and the object segmentation generated using SExtractor, to test the effectiveness of the various classification techniques in differentiating between stars and galaxies in CCD images. We limit our results to those taken with the *R*-band filter (725 classified objects).

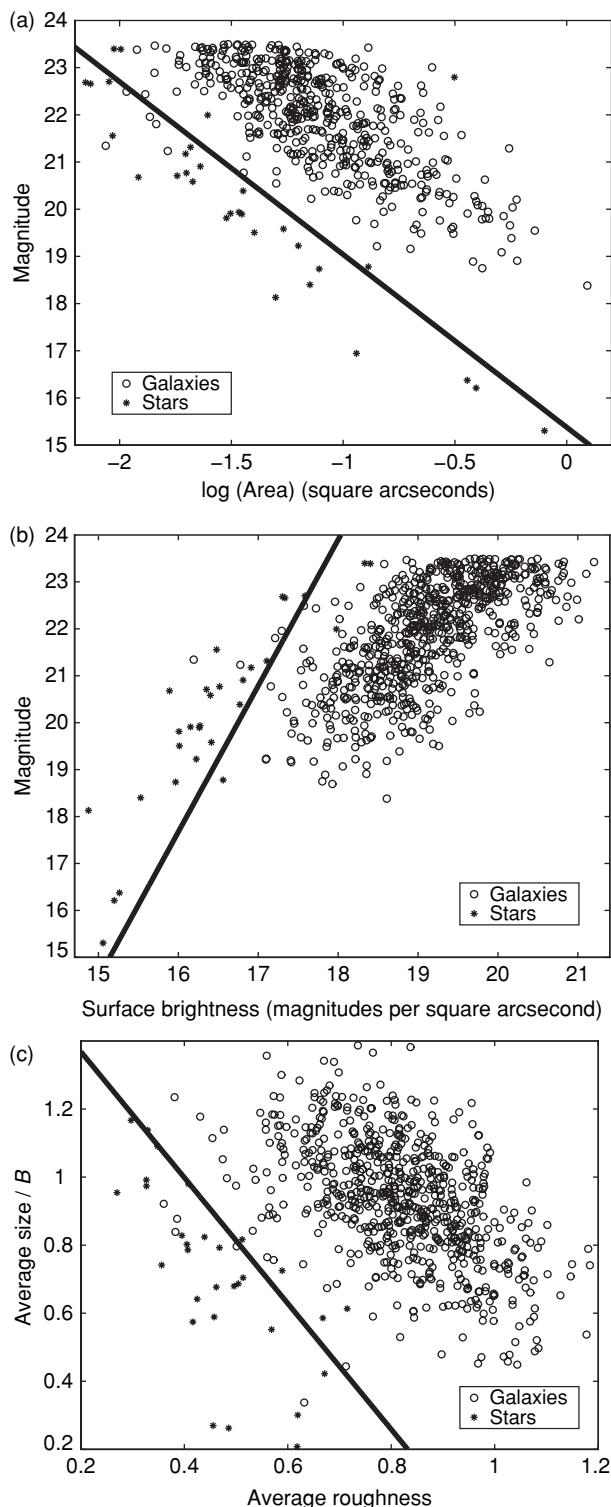
For these images, magnitude, aperture area, surface brightness (SB), and the MM tools average size and average roughness, were used to differentiate between pre-classified stars and galaxies. Before calculating average size and average roughness, an MM gradient operation was performed to enhance sharp edges, increasing the difference between star and galaxy light profiles. To calculate average size and average roughness, an increasing series of diamond structuring elements was used for reasons given in Section 2.1.

Figure 5 illustrates how magnitude, aperture area, and surface brightness can be used to differentiate between stars and galaxies in CCD images, and the results of using MM. In each plot, the line segment used to separate stars from galaxies is included, the line having been constructed using an MLP. These results are summarized in Table 1.

### 3.3 Photographic Plate Images

Doyle et al. (2005) used 4315 fields from the SuperCOSMOS Image Archive (Hambly et al. 2001a; Hambly, Irwin, & MacGillivray 2001b; Hambly et al. 2001c) in their work on finding optical counterparts for the HI Parkes All Sky Survey (HIPASS). We used the same photographic plate images, along with the object segmentation generated using SExtractor, to test the effectiveness of using the various classification techniques to differentiate between stars and galaxies in photographic plate images. We classified each detectable object within the field by eye in a similar way to Smail et al. (1997). The objects in the sample ranged in *B*-band magnitude from 20.0 at the faint end to 13.0 at the bright end. For these images, magnitude, aperture area, surface brightness (SB) and peak flux (PF) are used to differentiate between the pre-classified stars and galaxies.

Figure 6 illustrates how magnitude, aperture area, surface brightness and peak flux can be used to differentiate between stars and galaxies in photographic plate images. The magnitude is calculated using the integrated pixel flux over the aperture, the exposure time, and the photometric zeropoint. Hence, objects which have saturated to the maximum pixel value will have their magnitudes under-estimated. In each plot, the line segment used to separate stars from galaxies is included and has been constructed using an MLP neural network. These results are summarized in Table 2.



**Figure 5** Star/galaxy classification plots for CCD image objects between *R*-band magnitudes 23.5 and 15.5 using current and MM techniques. (a) Classification using magnitude and aperture area. (b) Classification using magnitude and surface brightness. (c) Classification using two parameters from MM—average size and average roughness. The classification line in each panel has been constructed using an MLP neural network. The classification performance of MM in separating stars and galaxies is comparable to that of current techniques. These results are summarized in Table 1.

**Table 1.** Star/Galaxy differentiation results for CCD images<sup>A</sup>

Method	Comp. [%]	Cont. [%]	Corr. [%]
Mag–Area	99.4 ± 3.8	16.1 ± 7.2	98.8 ± 3.7
Mag–SB	99.4 ± 3.8	16.1 ± 7.2	98.8 ± 3.7
MM	99.3 ± 3.8	19.4 ± 7.9	98.5 ± 3.7

<sup>A</sup> Using 725 objects—31 stars and 694 galaxies.

For the photographic plate images, MM cannot be used to classify stars and galaxies in the same way that it can for CCD images. The ‘peakiness’ that MM is able to detect in the CCD images does not exist in the photographic images, due to the non-linear response of the plate.

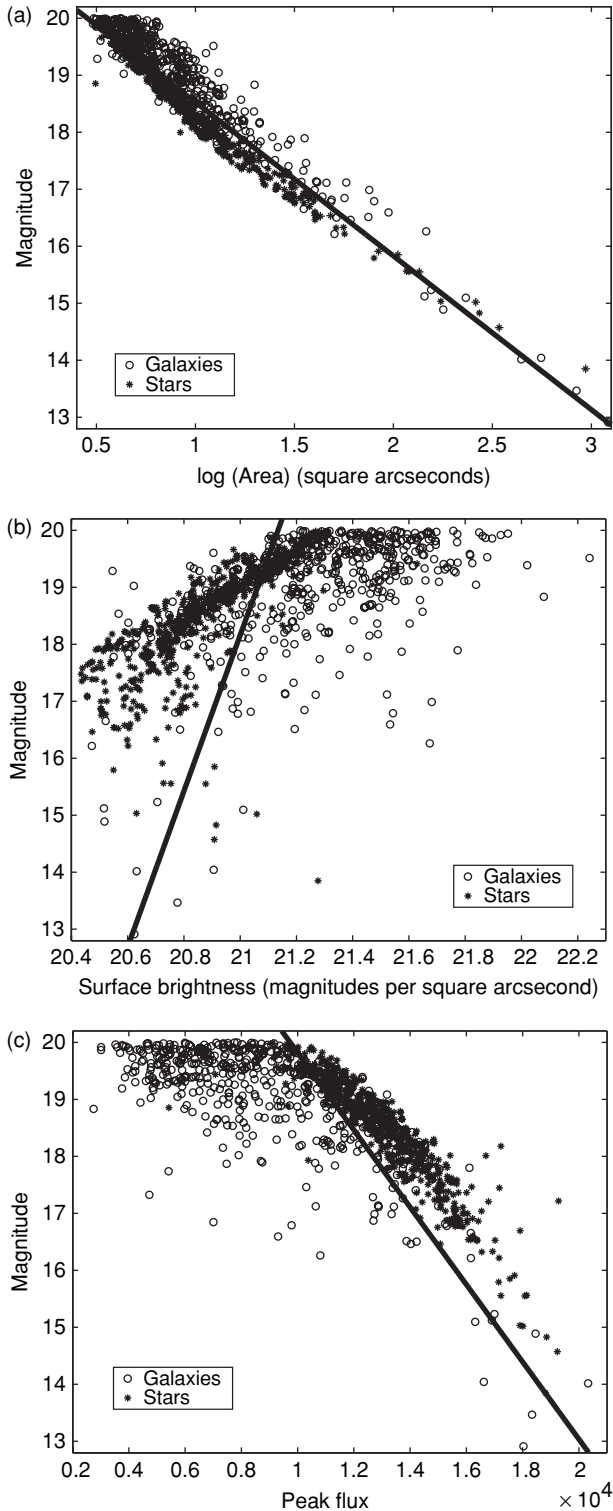
The biggest difficulties for current techniques are differentiating between stars and galaxies at the high-brightness limit and low brightness limit. This high-brightness limitation, best illustrated in Figure 6 (middle panel), results from the saturation of the stars’ profiles, causing them to become extended like the profile of a galaxy. The low brightness limitation results from the resolution of the image, where there just are not enough pixels to accurately characterize the object.

We use the tools of MM to calculate the sizes of the diffraction spikes, which become substantial in the stars brighter than magnitude 16.5. The power of the structuring element in MM is the ability to match against features within an image. We use an increasing series of structuring elements in the shape of a diffraction spike (plus sign), and the technique of granulometry, to calculate the diffraction spike length. This feature presents a method for differentiating galaxies from bright stars. Figure 7 illustrates how this has been achieved, with the objects in this plot taken from the region in Figure 6 (middle panel) where distinction between stars and galaxies is difficult, and with magnitudes brighter than 17.0 and surface brightnesses brighter than 20.9 magnitudes per square arcsecond. Again, the line segment used to separate stars from galaxies was constructed using an MLP neural network. These results are included in Table 2.

### 3.4 Results

For CCD images, 99.3 ± 3.8% of galaxies can be separated from stars using MM, with 19.4 ± 7.9% of the stars being misclassified, comparing very favourably with current techniques. This result, along with the fact that MM requires only one ‘free parameter’ compared with at least five for techniques involving apertures, presents MM as a powerful method for star/galaxy differentiation in CCD images.

For photographic plate images, conventional techniques fail to differentiate between stars and galaxies at the high-brightness limit. A solution to this will either involve human intervention or a complex algorithm. MM provides a far less complex solution, involving fewer ‘free

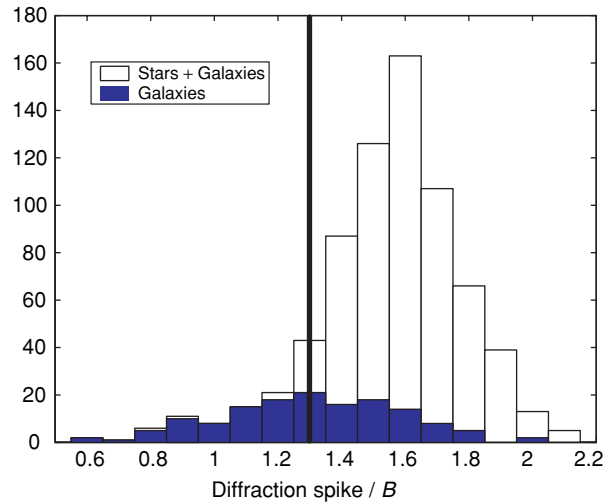


**Figure 6** Star/galaxy classification plots for photographic plate objects, between  $B$ -band magnitudes 20.0 and 13.0. (a) Classification using magnitude and aperture area. (b) Classification using magnitude and surface brightness. Notice that the classification line has rotated due to low-brightness and high-brightness mixing. (c) Classification using magnitude and peak flux. The classification line in each panel has been constructed using an MLP neural network. In these three plots, stars brighter than magnitude 16.5 are completely inseparable from the galaxy population. These results are summarized in Table 2.

**Table 2.** Star/Galaxy differentiation results for photographic plate images<sup>A</sup>

Method	Comp. [%]	Cont. [%]	Corr. [%]
Mag–Area	$22.4 \pm 4.0$	$1.6 \pm 0.5$	$83.2 \pm 3.4$
Mag–SB	$22.4 \pm 4.0$	$1.6 \pm 0.5$	$83.2 \pm 3.4$
Mag–PF	$21.4 \pm 3.9$	$1.2 \pm 0.5$	$83.4 \pm 3.4$

<sup>A</sup> Using 713 high-brightness objects — 570 stars and 143 galaxies.



**Figure 7** Diffraction spike length, calculated using MM, for the bright stars and galaxies found in the inseparable region of Figure 6 (middle panel). The classification line has been constructed using an MLP neural network. Galaxies that are unable to be separated from bright stars (using current techniques alone) can be recovered using the MM diffraction spike tool. These results are included in Table 2.

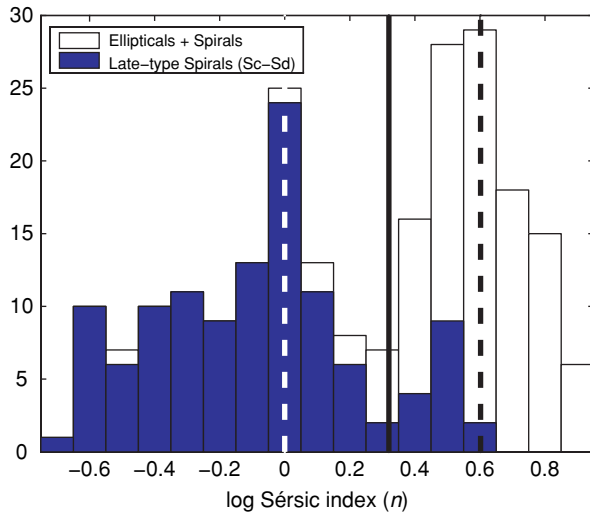
parameters’ than other complex alternatives. Using the MM diffraction spike tool allows  $51.0 \pm 6.0\%$  of the galaxies that are inseparable in Figure 6 (middle panel), with magnitudes brighter than 17.0 and surface brightnesses brighter than 20.9 magnitudes per square arcsecond, to be correctly separated from the bright stars.

#### 4 Galaxy Morphology Classification

To test the advantage of using MM to classify galaxy morphology we exclusively used CCD imaging. The CCD images were again taken from the data archive of Smail et al. (1997). It must be noted that peculiar and merging galaxies were not included in this catalogue, so the ability of classification techniques to discriminate between such galaxy types is not able to be determined. For the images, the techniques of Sérsic fitting (using the GALFIT software package<sup>3</sup>), the ‘CAS’ parameters, and the MM tools of average size and average roughness, were used to differentiate between pre-classified elliptical (E) and late-type spiral (Sc–Sd) galaxies. Before calculating

<sup>3</sup> <http://zwicky.as.arizona.edu/~cyp/work/galfit/galfit.html>





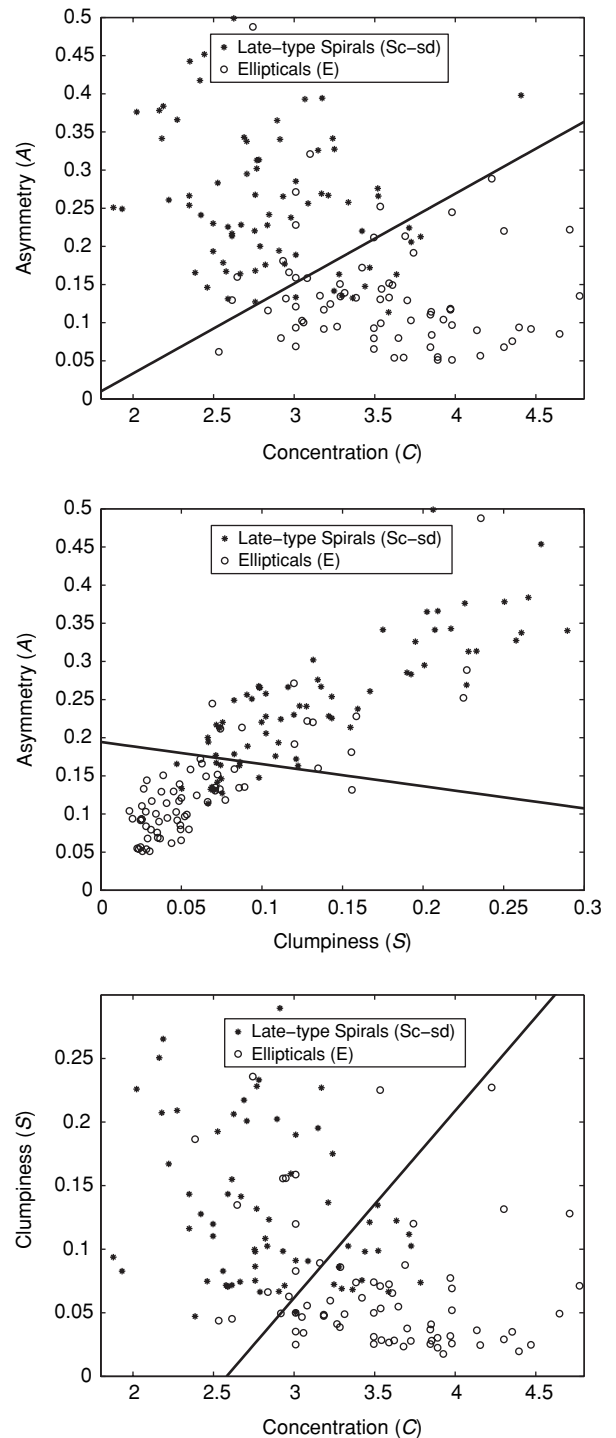
**Figure 8** Galaxy morphology classification plot for Sérsic fitting, using the elliptical (E) and late-type spiral (Sc-Sd) galaxies in the data archive of Smail et al. (1997). The two dashed lines represent the de Vaucouleurs (bulge,  $n = 4$ ) and Exponential (disk,  $n = 1$ ) profiles. The full line represents the classification line, which was constructed using an MLP neural network. This plot demonstrates the classification performance of Sérsic fitting, using the GALFIT software package, in separating elliptical and late-type spiral galaxies between magnitude 23.5 and 16.0. These results are summarized in Table 3.

average size and average roughness, an MM close operation was performed to average any spiral arm features into a disk component. To calculate average size and average roughness, an increasing series of diamond structuring elements was used for reasons given in Section 2.1.

Figures 8 and 9 illustrate how the Sérsic index,  $n$ , and ‘CAS’ parameters, respectively, can be used to classify elliptical and late-type spiral galaxies. In each plot, the line segment included and used to separate elliptical galaxies from late-type spiral galaxies has been constructed using an MLP neural network. The galaxies with a core size less than four pixels were excluded from the calculation. We found that a severely under-sampled core can create difficulties with the clumpiness and asymmetry parameters, not providing a fair comparison with the other techniques. These results are summarized in Table 3.

Figure 10 shows how MM can be used to classify elliptical and late-type spiral galaxies. The plot includes line segment used to separate elliptical galaxies from late-type spiral galaxies, the line having been constructed using an MLP neural network. These results are included in Table 3.

To test the effect of atmospheric smoothing (or seeing) on the performance of the MM galaxy morphology classification, the *HST* images were artificially blurred using the IRAF Gaussian task from space-based (FWHM of 0.1 arcsec) to ground-based (FWHM of 2.12 arcsec) seeing conditions. The upper panel of Figure 11 shows how the classification line is altered as the seeing increases, shifting toward higher average size values as



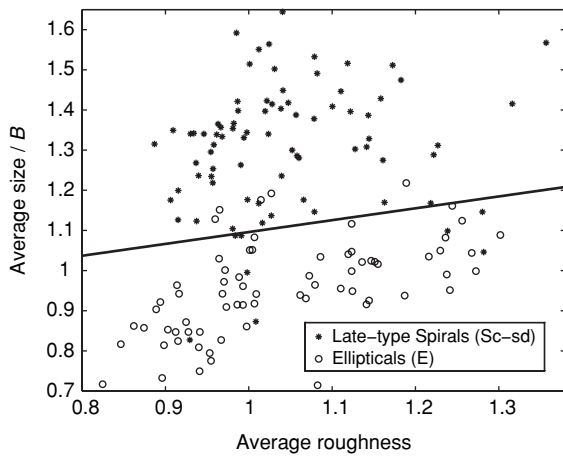
**Figure 9** Galaxy morphology classification plots for the ‘CAS’ parameters, using the elliptical (E) and late-type spiral (Sc-Sd) galaxies in the data archive of Smail et al. (1997). The classification line in each panel has been constructed using an MLP neural network. The plots demonstrate the classification performance of the three ‘CAS’ parameters in separating elliptical and late-type spiral galaxies between magnitudes 23.5 and 16.0. These results are included in Table 3.

the galaxy profiles extend due to the blurring. The lower panel of Figure 11 shows how the classification performance of MM is reduced as the seeing increases.

**Table 3.** Galaxy Morphology classification results for CCD images<sup>A</sup>

Method	Correctness [%]
Sérsic Fitting (GALFIT)	89.8 ± 6.3
Asymmetry–Concentration	84.2 ± 7.4
Asymmetry–Clumpiness	80.9 ± 7.3
Clumpiness–Concentration	79.6 ± 7.2
MM	91.4 ± 7.8

<sup>A</sup> Using 152 objects — 71 elliptical (E) and 81 late-type spiral (Sc-Sd) galaxies.

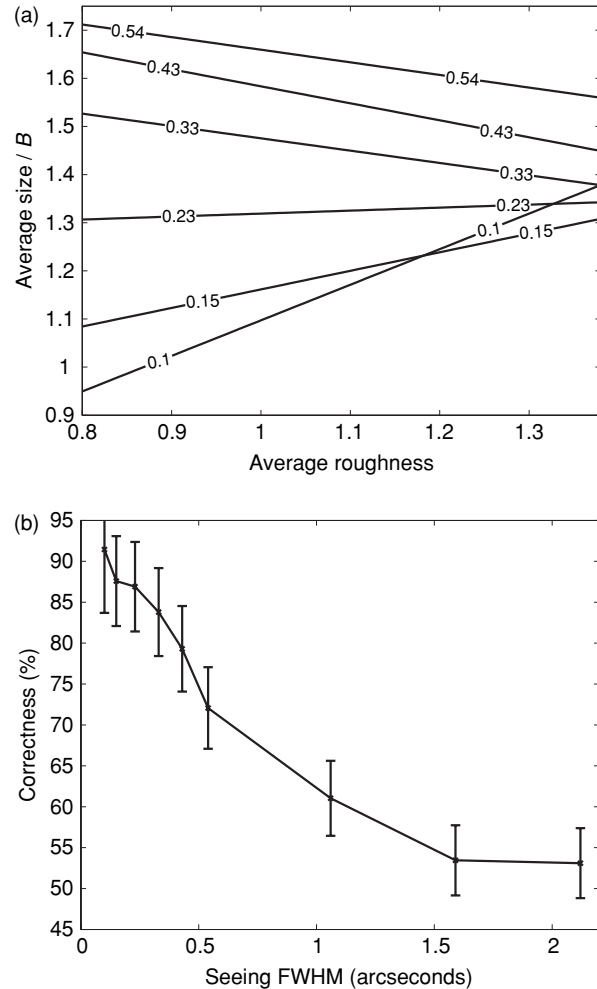


**Figure 10** Galaxy morphology classification plot for MM, using the elliptical (E) and late-type spiral (Sc-Sd) galaxies in the data archive of Smail et al. (1997). The plot shows galaxy morphology classification using two parameters from MM — average size and average roughness. The classification line has been constructed using an MLP neural network. The classification performance of MM is comparable to those of current techniques in separating elliptical and late-type spiral galaxies between magnitudes 23.5 and 16.0. These results are included in Table 3.

#### 4.1 Results

For CCD images,  $91.4 \pm 7.8\%$  of elliptical and late-type spiral galaxies can be correctly classified using MM, comparing very favourably with current techniques. These results present MM as a powerful alternative for galaxy morphology classification in CCD images.

Atmospheric seeing effects the MM galaxy morphology classification by shifting the classification line toward higher average size values and reducing the correctness of the classification. The performance of the classification is acceptable (at least 80% correct) for seeing FWHM equal to or better than 0.4 arcseconds. For a redshift of  $0.465 \pm 0.095$  ( $D_A = 1193.1^{+122.4}_{-152.1}$  Mpc; for  $H_0 = 71$ ,  $\Omega_M = 0.3$ ,  $\Omega_A = 0.7$ ) this corresponds to a physical size of  $2.19^{+0.24}_{-0.29}$  kpc. This size is of the same order as the central bulge of each galaxy and, therefore, the Gaussian seeing profile will become a significant element of the galaxy profiles. The difference between the profiles of elliptical and spiral galaxies is reduced with even larger seeing sizes, preventing MM from being able to distinguish between the



**Figure 11** Effect of atmospheric smoothing on the MM galaxy morphology classification. (a) Change in the classification line between space-based (FWHM of 0.1 arcsec) to ground-based (FWHM of 0.54 arcsec) seeing conditions. (b) Reduction in the correctness of the classification between space-based (FWHM of 0.1 arcsec) to ground-based (FWHM of 2.12 arcsec) seeing conditions. The performance is acceptable (at least 80% correct) for seeing better than 0.4 arcsec.

two classes. From these results we can write the following seeing-distance relationship: MM will classify galaxy morphology at a performance of 80% correctness or greater if

$$s \times d < 477.24^{+48.96}_{-60.84} \quad (11)$$

where  $s$  is the seeing FWHM (in arc seconds) and  $d$  is the cosmological angular size distance (in Mpc). The consequence of this is that, for a ground-based telescope with a 1.0 arcsecond seeing FWHM, the MM classification is acceptable to a maximum distance of  $477.24^{+48.96}_{-60.84}$  Mpc ( $z = 0.13 \pm 0.02$ ).

The difficulty remains in finding what physical processes will correlate with the average size and average roughness. For example, using the ‘CAS’ framework, Conselice (2003) use some quantitative reasoning to show that the concentration index correlates with gross form (bulge to disk ratio), asymmetry correlates with

recent merger/interaction activity, and clumpiness correlates with star formation (using the  $H\alpha$  equivalent widths). We are currently investigating the relationship between average size and average roughness with respect to star formation rate in galaxies. This will provide correlation with both internal and external galaxy mechanisms, and will be presented in a future publication.

## 5 Summary

We have demonstrated that, for CCD images,  $99.3 \pm 3.8\%$  of the galaxy population is separated from the stellar population using MM, with  $19.4 \pm 7.9\%$  of the stars being misclassified, comparing very well with current techniques involving aperture magnitudes. We have also demonstrated that, for photographic plate images, the MM diffraction spike tool allows  $51.0 \pm 6.0\%$  of the bright galaxies, inseparable from the bright stars in current techniques, to be correctly classified, with only  $1.4 \pm 0.5\%$  of the high-brightness stars contaminating the population. Lastly, we have demonstrated that elliptical (E) and late-type spiral (Sc-Sd) galaxies can be classified using MM at an accuracy of  $91.4 \pm 7.8\%$ , which compared well with the techniques of Sérsic fitting and the use of ‘CAS’ parameters.

We have presented here an automated method, alternative to current techniques, for both star/galaxy differentiation and galaxy morphology classification. It is a method involving fewer ‘free parameters’ than current techniques, especially automated machine learning algorithms, and doesn’t require a machine learning algorithm for the training of the classifier. The classification tools developed in this work could, however, also be used as an additional layer of parameter space for any current machine learning approaches.

We intend to use this method to study the relationship between galaxy morphology and star formation rate, and how each varies within the supercluster environment. This environment allows us to probe the complete range of densities at which galaxies are affected by star formation suppression and morphology alteration mechanisms. We shall study how these galaxy properties vary between dense cluster cores, inter-cluster filaments, and the low-density ‘field’ regions.

## Acknowledgements

We thank the anonymous referees for their careful reading and thoughtful comments which have improved the quality of this work. We also kindly thank Marianne Doyle and David Rhode for providing us with the SuperCOSMOS images, along with the SExtractor segmentation and magnitude calibration for each field. For their useful conversations, we also thank various members of the Astronomy Group at the University of Queensland. J.A.M. acknowledges financial support from an EPSA University of Queensland Joint Research Scholarship. K.A.P. acknowledges financial support from an EPSA University of Queensland Research Fellowship.

## References

- Abraham, R. G., Valdes, F., Yee, H. K. C. & van den Bergh, S. 1994, *ApJ*, 432, 75
- Andredakis, Y. C. & Sanders, R. H. 1994, *MNRAS*, 267, 283
- Andreoni, S., Gargiulo, G., Longo, G., Tangliaferri, R. & Capuano, N. 2000, *MNRAS*, 319, 700
- Appleton, P. N., Siqueira, P. R. & Basart, J. P. 1993, *AJ*, 106, 1664
- Bertin, E. & Arnouts, S. 1996, *A&AS*, 117, 393
- Candeas, A. J., Braga Neto, U. & Carvalho Filho, E. 1996, *Anais do IX SIBGRAPI*, 235
- Candeas, A. J., Braga Neto, U. & Carvalho Filho, E. 1997, *J. Braz. Comp. Soc.*, 3
- Conselice, C. J. 2003, *ApJS*, 147, 1
- de Vaucouleurs, G. 1948, *AnAp*, 11, 247
- de Vaucouleurs, G. 1959, *Handb. der Physik*, 53, 275
- Doyle, M. T., et al. 2005, *MNRAS*, 361, 34
- Drinkwater, M. J. & Schmidt, R. W. 1996, *PASA*, 13, 127
- Graham, A. W. & Driver, S. P. 2005, *PASA*, 22, 118
- Hambly, N. C., et al. 2001a, *MNRAS*, 326, 1279
- Hambly, N. C., Irwin, M. J. & MacGillivray, H. T. 2001b, *MNRAS*, 326, 1295
- Hambly, N. C., Davenhall, A. C., Irwin, M. J. & MacGillivray, H. T. 2001c, *MNRAS*, 326, 1315
- Harmon, R. & Mamon, G. 1993, In *ASP Conf. Ser. 43: Sky Surveys: Protostars to Protogalaxies*, Soifer, B. T., San Francisco, ASP, 15
- Haykin, S. 1998, *Neural Networks – A Comprehensive Foundation*, 2nd ed., Englewood Cliffs, Prentice-Hall
- He, L. X. 1996, Ph.D. Thesis, Iowa State University
- Hubble, E. 1926, *ApJ*, 64, 321
- Hubble, E. 1936, *The Realm of the Nebulae*, New Haven, Yale University Press
- Heijmans, H. J. A. M. 1992, *Nieuw Archief voor Wiskunde*, Vierde Serie, 10, 237
- Heijmans, H. J. A. M. 1994, in *Shape in Picture: Mathematical Description of Shape in Grey-level Images*, 147
- Heijmans, H. J. A. M. 1995, in *SIAM Review*, 37, 1
- Jones, L. R., Fong, R., Shanks, T., Ellis, R. S. & Peterson, B. A. 1991, *MNRAS*, 249, 481
- Kibblewhite, E. J., Bridgeland, M. T., Bunclark, P. & Irwin, M. J. 1984, in *NASA Conf. Publ. 2317: Astronomical Microdensitometry Conference*, Klinglesmith, D. A., Washington, D.C., NASA Scientific and Technical Information Branch, 277
- Kron, R. G. 1980, *ApJS*, 43, 305
- Lea, S. M. & Kellar, L. A. 1989, *AJ*, 97, 1238
- Lefèvre, O., Bijaoui, A., Mathez, G., Picat, J. P. & Lelièvre, G. 1986, *A&A*, 154, 92
- MacArthur, L. A., Courteau, S. & Holtzman, J. A. 2003, *ApJ*, 582, 689
- Maddox, S., Sutherland, W., Efstathiou, G. & Loveday, J. 1990, *MNRAS*, 243, 692
- Mähönen, P. & Frantti, T. 2000, *ApJ*, 541, 261
- Matheron, G. 1975, *Random Sets and Integral Geometry*, New York, John Wiley and Sons
- Maragos, P. 1989, *IEEE Trans. Pattern Anal. Mach. Intell.*, 11, 701
- Miller, A.S. & Coe, M. J. 1996, *MNRAS*, 279, 293
- Odehahn, S., Stockwell, E., Pennington, R. M., Humphreys, R. & Zumach, W. 1992, *AJ*, 103, 318
- Odehahn, S., Humphreys, R. M., Aldering, G. & Thurnes, P. 1993, *PASP*, 105, 1354
- Philip, N. S., Wadadekar, Y., Kembhavi, A. & Joseph, K. B. 2002, *A&A*, 385, 1119
- Pimblett, K. A., Smail, I., Edge, A. C., Couch, W. J., O’Hely, E. & Zabludoff, A. I. 2001, *MNRAS*, 327, 588
- Peng, C. Y., Ho, L. C., Impey, C. D. & Rix, H-W. 2002, *AJ*, 124, 266
- Reid, N. & Gilmore, G. 1982, *MNRAS*, 201, 73
- Reid, I. N., Yan, L., Majewski, S., Thompson, I. & Smail, I. 1996, *AJ*, 112, 1472
- Sandage, A. 1961, *The Hubble Atlas of Galaxies*, Washington, D.C., Carnegie Institution of Washington

- Shaver, P. A. 1987, *Nature*, 326, 773  
 Sebok, W. 1979, *AJ*, 84, 1526  
 Serra, J. 1982, *Image Analysis and Mathematical Morphology*, London, Academic Press  
 Sérsic, J. L. 1963, *BAAA*, 6, 41  
 Sérsic, J. L. 1968, *Atlas de Galaxias Australes*, Cordoba, Observatorio Astronomico  
 Smail, I., Dressler, A., Couch, W. J., Ellis, R. S., Oemler, A. (Jr), Butcher, H. & Sharples, R. M. 1997, *ApJS*, 110, 213  
 Ueda, H. 1999, *PASJ*, 51, 435  
 van der Bergh, S. 1960, *ApJ*, 131, 215  
 Weir, N., Fayyad, U. M. & Djorgovski, S. 1995, *AJ*, 109, 6

### Appendix A: 'CAS' Volume

Here, we give just a brief summary of each parameter, but a more detailed description of the operation of these can be found in the work of Conselice (2003).

#### Concentration of Light

The concentration index,  $C$ , is defined as (Bershady, Jangren, & Conselice 2000) the ratio of the 80% to 20% curve of growth radii, normalized by a logarithm:

$$C = 5 \times \log \left( \frac{r_{80\%}}{r_{20\%}} \right) \quad (12)$$

where  $r_{x\%}$  represents the size of the aperture which contains  $x\%$  of the total light flux of the galaxy (in some implementations the 75% and 25% curve of growth radii are used instead).

#### Asymmetry

The asymmetry index,  $A$ , is defined as (Conselice, Bershady, & Jangren 2000) the volume of the image of a

galaxy which has been rotated  $180^\circ$  around its center and then subtracted from its pre-rotated image, then normalized to the original image volume:

$$A = \frac{\sum |I - R|}{\sum |I|} \quad (13)$$

where  $I$  is the original image and  $R$  is the rotated image. Conselice (2003) first reduce the effective resolution of the image,  $I$ , using a filter of size  $\frac{1}{6} \times r(\eta = 0.2)$ , to have asymmetry only sensitive to large-scale stellar distributions.

#### High Spatial Frequency

The clumpiness parameter,  $S$ , is defined as (Conselice 2003) the amount of light contained in high spatial frequency structures to the total amount of light in the galaxy. Computationally, this is defined as the volume of the image of a galaxy which has been blurred (smoothed) using a filter and then subtracted from its pre-smoothed image, then normalized to the original image volume:

$$S = \frac{\sum(I - B)}{\sum I} \quad (14)$$

where  $I$  is the original image and  $B$  is the blurred (smoothed) image. Conselice (2003) use a smoothing filter of size  $\frac{1}{5} \times 1.5 \times r(\eta = 0.2)$  and exclude the region inside  $\frac{1}{20} \times 1.5 \times r(\eta = 0.2)$ , as its high-frequency power is unrelated to the stellar light distribution. They also force any negative pixel values in the difference image,  $I - B$ , to zero before computing the clumpiness.