

# SExtractor: Software for source extraction

E. Bertin<sup>1,2</sup> and S. Arnouts<sup>1</sup>

<sup>1</sup> Institut d'Astrophysique de Paris, 98bis Boulevard Arago, 75014 Paris, France

<sup>2</sup> European Southern Observatory, Casilla 19001, Santiago 19, Chile

Received July 18; accepted August 17, 1995

**Abstract.** — We present the automated techniques we have developed for new software that optimally detects, deblends, measures and classifies sources from astronomical images: SExtractor (*Source Extractor*). We show that a very reliable star/galaxy separation can be achieved on most images using a neural network trained with *simulated* images. Salient features of SExtractor include its ability to work on very large images, with minimal human intervention, and to deal with a wide variety of object shapes and magnitudes. It is therefore particularly suited to the analysis of large extragalactic surveys.

**Key words:** methods: data analysis — techniques: image processing — galaxies: photometry

## 1. Introduction

A large fraction of today's scientific work done on astronomical images is not done on the images themselves, but rather on catalogs of objects produced from these images. This includes, for example, studies of number counts, clustering properties, or colour/magnitude distributions. Obviously, in all cases, the reliability of the source extraction process is of major importance.

Since the late 70's and beginning of the 80's, various computer programs have been developed to create automatically galaxy catalogs from astronomical images, e.g. FOCAS (Jarvis & Tyson 1981), the APM software (e.g. Maddox et al. 1990a), the COSMOS system (e.g. Beard et al. 1990) and the PPP package (Yee 1991). SExtractor (*Source Extractor*) is new software especially designed to process in batch mode large digital images (up to  $60\,000 \times 60\,000$  pixels) similar to the ones that will be produced by future CCD arrays. As SExtractor is primarily intended to process large amounts of survey data, special attention has been paid to speed and robustness in the extraction of objects in the image, regardless of their shape or size.

SExtractor has been used in the reduction of several recent photometric galaxy surveys (e.g. Smail et al. 1995; Arnouts et al. 1995; Bertin et al. 1995). This paper describes the algorithms and the performance of the software (note: SExtractor does have a specific "PHOTO" mode for photographic scans that we will not describe here).

The complete analysis of an image is done in six steps: estimation of the sky background, thresholding,

deblending, filtering of the detections, photometry, and star/galaxy separation. In the following, a section is devoted to describing each step.

## 2. Background estimation

Each pixel value is the sum of a "background" signal and light coming from the objects we are interested in. To be able to detect the faintest objects and also to measure accurately their fluxes, we need to have a precise estimation of the background level in any place of the image, i.e. a "background map"<sup>1</sup>. To construct a background map, SExtractor makes a first pass through the pixel data, computing an estimator of the local background in each mesh of a grid which covers the whole frame. Several background estimators have been described in the literature (Newell 1983; Bijaoui 1980; Beard et al. 1990; Almoznino et al. 1993). The Bijaoui estimator is, in principle, the most unbiased, but it proves to be very noisy with small samples, and requires excessive computing time. Other estimators are based on the determination of the mode of the histogram. From extensive simulations done with various mesh-sizes and object crowding, we have adopted for

<sup>1</sup>Strictly speaking, there should be one background map per object; that is, what would the image look like if this specific object was absent. Therefore using a unique background map for all the objects implies that we assume the background does not vary too much within the zone over which the photometric analysis is performed, i.e.  $\approx 1$  full-width at half-maximum (FWHM) of the profiles. Such an assumption is no longer valid in very crowded fields: the photometry becomes confusion-limited.

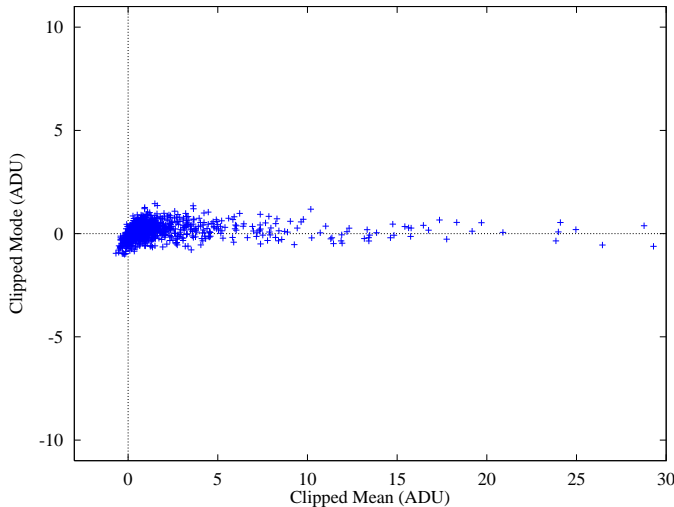
SExtractor an estimator which is a combination of  $\kappa\sigma$ -clipping and mode estimation, similar to the one employed in Stetson’s DAOPHOT program (see e.g. Da Costa 1992). Briefly, the local background histogram is clipped iteratively until convergence at  $\pm 3\sigma$  around its median. If  $\sigma$  is changed by less than 20% during that process, we consider the field to be uncrowded and we simply take as a value for the background the mean of the clipped histogram. Otherwise we estimate the mode with:

$$\text{mode} = 2.5 \times \text{median} - 1.5 \times \text{mean} \quad (1)$$

This expression is different from the usual approximation

$$\text{mode} = 3 \times \text{median} - 2 \times \text{mean} \quad (2)$$

(e.g. Kendall & Stuart 1977), but was found, from the simulations, to be more accurate with our clipped distributions. Figure 1 shows that the expression (1) is considerably less affected by crowding than a simple clipped mean – like the one used in FOCAS or by Infante (1987) – but is  $\approx 30\%$  noisier. This is why we turn back to the mean for uncrowded fields.



**Fig. 1.** Monte-Carlo simulations of  $32 \times 32$  pixels background meshes polluted by random gaussian profiles. Each point corresponds to a background estimation inferred from a different realization. The true background lies at 0 ADU. While being slightly noisier, the “clipped mode” gives a more robust estimate than a clipped mean in crowded regions

Once the grid is set up, a median filter can be applied to it, in order to suppress possible local overestimations due to bright stars. The resulting background map is then simply a bilinear interpolation between the meshes of the grid. The choice of the mesh size is an important step. Too small, the background estimation is affected by the presence of objects and random noise. Too large, it cannot reproduce the small scale variations of the background. On most images, a width of 32 to 128 pixels works fine.

### 3. Detection

With no assumption on the shape of the objects, one can consider two classical detection techniques: peak finding and thresholding. In both cases a second pass through the data is required to glue (for peak finding) or split (after thresholding) close detections. Peak finding is more appropriate for stars, but is not suited to the detection of Low-Surface-Brightness (LSB) objects (Yee 1991). This is why we have chosen thresholding. SExtractor uses Lutz’s one-pass algorithm (Lutz 1979) to extract 8-connected contiguous pixels from a “template frame”. The template frame results from the convolution “on-the-fly” of the original image with some appropriate convolution mask. Any convolution mask can be used with SExtractor; the choice depends on which type of objects one wants to detect best. For faint unresolved sources, the Point Spread Function gives optimum results (Irwin 1985). A larger mask is more adapted for detecting LSB objects. A “wavelet” filter can also be used to detect object at a specific scale, or point-sources over a chaotic background (e.g. Coupinot et al. 1992).

### 4. Deblending merged objects

With the detection method described above, it is necessary to separate neighbours that have been extracted as a single source.

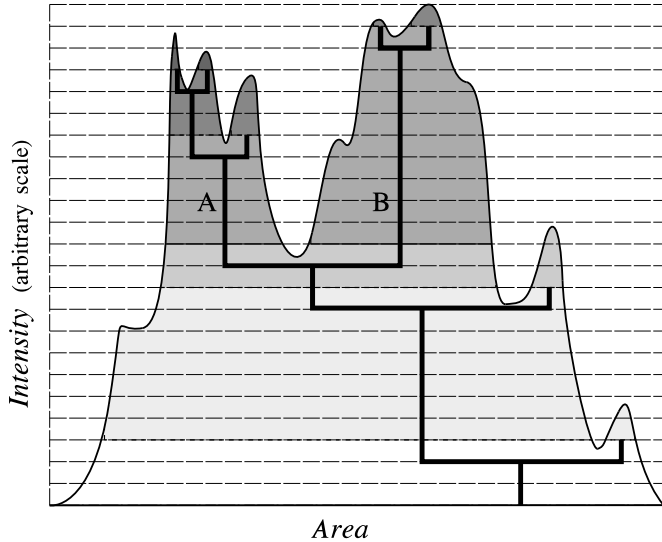
Great care has been taken in designing the deblending routine for merged objects. On survey images, SExtractor has often to deal with both patchy, extended Sc-d galaxies (which have to be considered as single entities), and close or interacting pairs of optically faint galaxies (which have to be considered as separate objects). This is traditionally a point where many extraction programs fail, leading to catalogs containing extended galaxies split into several smaller detections.

#### 4.1. Finding the real components

Basically, our algorithm employs a multiple isophotal analysis technique similar to the one designed for the COSMOS machine (Beard et al. 1990). Each extracted set of connected pixels is re-thresholded at 30 levels exponentially spaced between its primary extraction threshold and its peak value. This gives us a “model” of the light distribution within the object(s), which is stored in the form of a tree structure (Fig. 2). Then the algorithm goes downwards, from the tips of branches to the trunk, and decides at each junction whether it shall extract two (or more) objects or continue its way down. To meet the conditions described earlier, the following simple decision criteria were adopted: at any junction threshold  $t_i$ , any branch will be considered as a separate component if

- (1) the integrated pixel intensity (above  $t_i$ ) of the branch is greater than a certain fraction  $\delta_c$  of the total intensity of the composite object.
- (2) condition (1) is verified for at least one more branch at the same level  $i$ .

Note that ideally, condition (1) is both flux and scale-invariant; however for faint, poorly resolved objects, the efficiency of the deblending is limited mostly by seeing and sampling. From the analysis of both small and extended galaxy images, we find a good value for  $\delta_c$  of  $5 \cdot 10^{-3}$ .



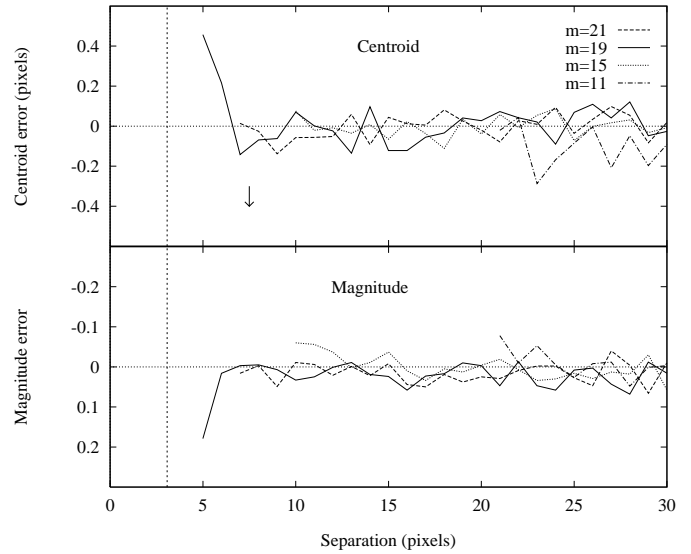
**Fig. 2.** A schematic diagram of the method used to deblend a composite object. The areal profile of the object (the smooth curve) can be described as a tree-structure (thick lines). The decision to regard a branch as a distinct object is determined according to its relative integrated intensity (tinted area). In the case above, the original object is split into two components A and B. Pixels lying below the separation threshold are assigned to their most credible “progenitor” afterwards

#### 4.2. Sticking the pieces again

The outlying pixels with fluxes lower than the separation thresholds have to be reallocated to the proper components of the merger. To do so, we opted for a *statistical* approach. At each pixel we compute the contribution that is expected from each sub-object using a bivariate gaussian fit to its profile. This is turned into a probability for that pixel to belong to the sub-object. For instance, a faint pixel lying halfway between two close bright stars having the same magnitude will be appended to one of these with equal probabilities. One big advantage of this technique is that the morphology of any object is completely defined simply through its list of pixels.

#### 4.3. Tests

To test the effects of deblending on photometry and astrometry measurements, we made several simulations of photographic images of double stars with different separations and magnitudes under typical observational conditions (Fig. 3). It is obvious that multiple isophotal techniques fail when there is no saddle point present in the profile (i.e. for separations  $< 2\sigma$  in the case of gaussian images). In the very worst cases, we measure a magnitude error  $\lesssim 0.2$  mag and a shift in the centroid  $\lesssim 0.4$  pixels for the fainter companion, but no other systematic effects are noticeable.



**Fig. 3.** Centroid and corrected isophotal magnitude errors for a simulated 19<sup>th</sup> magnitude star blended with a 11, 15, 19 and 21<sup>th</sup> mag. companion as a function of distance (expressed in pixels). Lines stop at the left when the objects are too close to be deblended. The dashed vertical line is the theoretical limit for unsaturated stars with equal magnitudes. In the centroid plot, the arrow indicates the direction of the neighbour. The simulation assumes a 1 hour exposure with a 1 m Schmidt telescope and a Moffat PSF with a seeing FWHM of 3 pixels (2"). The results are of course applicable to fainter objects on deeper exposures

#### 5. Filtering the detections

When using low thresholds, spurious detections are often made in the wings of objects with shallow profiles (for example, elliptical galaxies). It comes from the fact that the background is locally higher there, leading to a lower relative threshold and thus in a higher detection rate of noise peaks. One solution to this problem is to verify whether or not there could have been a detection if there were no neighbours. This is what the “cleaning” procedure does. As detections are made, objects are put in a *FIFO* (*First-In First-Out*) stack. When the stack is full, objects are

examined one by one before being sent to the catalog. SExtractor computes the contribution to the mean surface brightness of each object from its neighbours. This is then subtracted, and if the mean surface brightness still falls above the detection threshold, the object is accepted into the catalogue. The contribution from the wings of neighbours is computed assuming a gaussian extrapolation of their profiles. As real profiles have, in general, broader wings than a pure gaussian, it is often necessary to expand their estimated width by a certain factor. Values ranging between 1.0 and 2.0 are usually sufficient.

## 6. Photometry

In addition to isophotal and circular aperture magnitudes, SExtractor has the possibility to estimate the “total” magnitude of an object.

The total apparent magnitude is the most physically relevant photometric parameter to extract directly from pixel data. For non-stellar objects, isophotal and aperture photometry have both their advantages and disadvantages when used to estimate total magnitudes (e.g. Irwin & Hall 1983). Aperture photometry is known to be generally less biased than isophotal photometry, but it only works in non-crowded regions. These constraints, and the need of keeping the processing as fast as possible, led us to adopt the following scheme. For each object, 2 types of total magnitude are computed, one uses an adaptive aperture, and the other an isophotal correction.

*Adaptive aperture magnitudes* Our adaptive aperture photometry routine is inspired by Kron’s (1980) “first moment” algorithm. (1) The second order moments of the object profile are used to define an equivalent bivariate gaussian profile with mean standard deviation  $\sigma_{\text{iso}}$  (2) An elliptical aperture whose ellipticity  $\epsilon$  and position angle  $\theta$  are defined by these moments is scaled to  $6\sigma_{\text{iso}}$  (which corresponds roughly to 2 isophotal “radii”). (3) Within this aperture we compute the “first moment”:

$$r_1 = \frac{\sum rI(r)}{\sum I(r)} \quad (3)$$

Kron (1980) and Infante (1987) have shown that for stars and galaxy profiles convolved with gaussian seeing an almost constant fraction of the flux is expected to lie within a circular aperture of radius  $kr_1$ , independently of their magnitude. This picture remains unchanged if we consider an ellipse with  $\epsilon kr_1$  and  $kr_1/\epsilon$  as the principal axes. A balance between systematic and random errors is achieved for  $k \approx 2$ . With  $k = 2.5$ , the mean fraction of flux lost is about 6%.

*Corrected isophotal magnitudes* . If we make the assumption that the intensity profiles of the faint objects recorded on images are roughly gaussian because of atmospheric

blurring, then the fraction  $\eta = \frac{I_{\text{iso}}}{I_{\text{tot}}}$  of the total flux enclosed within a particular isophote reads (see Maddox et al. 1990b):

$$\left(1 - \frac{1}{\eta}\right) \ln(1 - \eta) = \frac{A.t}{I_{\text{iso}}} \quad (4)$$

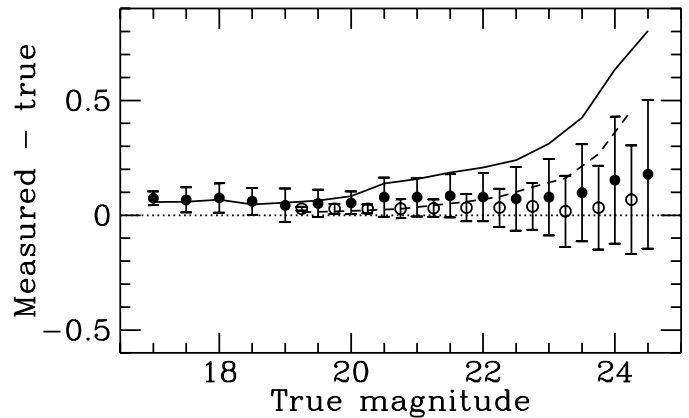
where  $A$  is the area and  $t$  the threshold related to this isophote. Equation (4) is not analytically invertible, but a very good approximation to  $\eta$  (error  $\lesssim 10^{-3}$  for  $\eta > 0.4$ ) can be made with the second-order polynomial fit:

$$\eta \approx 1 - 0.1961 \frac{A.t}{I_{\text{iso}}} - 0.7512 \left(\frac{A.t}{I_{\text{iso}}}\right)^2 \quad (5)$$

A “total” magnitude  $m_{\text{tot}}$  estimate is then

$$m_{\text{tot}} = m_{\text{iso}} + 2.5 \log \eta \quad (6)$$

Clearly this simple correction works best with stars; but it proves to give quite accurate results with disk galaxies. Still, the broad wings of spheroidal galaxy profiles can lead to large errors when a high isophotal threshold is used.



**Fig. 4.** Lost flux (expressed as a mean magnitude difference) with SExtractor’s total magnitudes as a function of true total magnitude on simulated  $R$ -band CCD frames. The parameters of the simulations are the same as in Sect. 7.6.1; the completeness limit arises at  $R \approx 24$ . Open and filled circles are for (unsaturated) stars and galaxies, respectively. The dashed and continuous curves represent the difference with isophotal magnitudes at the extraction threshold ( $26 \text{ mag.arcsec}^{-2}$ ) for stars and galaxies

SExtractor uses the following procedure to give the best estimate of the total magnitude for an object: the adaptive aperture method is taken, except if a neighbour is suspected to bias the magnitude by more than 0.1 mag. In these cases (which are normally less than 20% in high galactic latitude fields on moderately deep exposures), the corrected isophotal magnitude is taken. The behaviour of SExtractor total magnitudes, estimated on simulated CCD images, is shown in Fig. 4. One can see by this example that the fraction of lost flux is remarkably constant for both stars and galaxies (less than 2% variations rms),

nearly down to the completeness limit. With  $k = 2.5$ , we measure a mean offset of 0.06 mag for galaxies and 0.03 mag for stars (although the latter depends on the exact shape of the PSF, and doesn't lead to any improvement compared to the standard, fixed aperture magnitude also provided by SExtractor).

## 7. Star-galaxy separation

A good discrimination between stars and galaxies is essential for extragalactic studies. The common assumption is that galaxy images look more extended or fuzzy than those of stars (or QSO's). The reliability of such an approach has been spectroscopically confirmed by Colless et al. (1990) for COSMOS data from faint objects ( $b_j \approx 22$  mag). A great number of different star/galaxy classifiers has been described and applied so far. Simple estimators in a 2-parameter space like magnitude-isophotal area (e.g. Reid & Gilmore 1982), magnitude-peak intensity (e.g. Jones et al. 1991), or magnitude-surface brightness (e.g. Harmon & Mamon 1993) provide the simplest way of separating the two classes. These simple estimators do not use all the information available in the intensity profile. Other, more sophisticated estimators have been developed to do so. This includes the Sebok (1979) classifier, the  $r^{-2}$  “moment” of Kron (1980), the  $Q$  classifier from Lefèvre et al. (1986) and the  $\psi$  parameter of Maddox et al. (1990b). All these estimators share one common default: a lack of robustness, especially when confronted with merged objects, or objects having close neighbours (see e.g. Slezak et al. 1987).

If one sees an object as a vector of parameters, classifying stars and galaxies optimally is nothing more than finding the *best* frontier hypersurface between the two classes in parameter-space. The decision hypersurface can be made of hyperellipsoids (Jarvis & Tyson 1981). As shown in Odewahn et al. (1992) and Bertin (1994) for photographic images, neural networks provide an elegant and extremely robust tool for determining the best frontier between stars and galaxies. They can easily handle problematic cases like merged objects or close neighbours, as long as they have *learned* to recognize them. However, these neural networks need to be trained with a set of prototypes before being used. Selecting and labelling prototypes from real images is a manual and rather tedious task that can only be justified for a large amount of homogeneous data (a whole Schmidt plate for instance). Unfortunately, most surveys consist of image sets of variable quality (seeing, depth, ...) for which each would in principle require a different training. A solution to this may be to perform the training with some homogeneous set of images, and to apply an appropriate transformation to the input parameters of other sets, so as to recover a similar distribution of objects in parameter-space (Odewahn et al. 1993). This is not satisfying, since it brings back the necessity of a painful parametrization, and above all

there is no assurance that the transformation is accurate enough to recover the full potential of neural classification, especially for the “outliers”.

In fact, astronomical images produced by modern detector arrays (CCDs, infrared arrays) share many similarities: the intensity scale is linear with a good precision over a large range until a sharp saturation appears, the background noise is essentially white noise, and atmospheric blurring makes the core of the point spread function (PSF) gaussian. Hence correctly sampled images ( $\text{FWHM} \gtrsim 2$  pixels) can roughly be described by three numbers<sup>2</sup>: pixel scale, depth ( $S/N$  at a given magnitude), and seeing FWHM. The first two can be eliminated by a convenient choice of classification parameters (see Sect. 7.3.1), and we shall use the seeing information as a “tuning-button” and add it as input to our neural network.

Assuming that today's astronomical images share many common properties, one can think of using similar *artificial* images, accurate enough to train the neural network, instead of using real images. Let's now see how we can train a neural network with such simulations, and what performances we can obtain from the resulting classifier on real data.

### 7.1. Neural networks as classifiers

A neural network (NN) is basically a group of connected units called *neurons* whose behaviour is inspired from real biological neurons. One of the most interesting feature of NNs is their ability to “learn” and to “generalize” in a given context. This has led to use them, during the last ten years, in a growing number of difficult pattern recognition problems going from spoken words or hand-written character identification to the discrimination of particles in high-energy physics. When not too many neurons are involved, as it is the case here, it is possible to simulate the working of the neural network with a computer program. For an introduction to neural computational techniques see Wasserman (1989), Beale & Jackson (1990) or Hertz et al. (1991). Lahav (1994) gives an overview of possible applications of neural networks in statistical astronomy.

### 7.2. Principle

In the case of star/galaxy separation, we have chosen a multi-layered network architecture and a learning procedure known as *backpropagation*. Of all the NNs, it is probably one of the best-studied, and has been intensively applied with success for many classification tasks. Such networks have one input layer, one output layers, and one or more “hidden” layers between them. Neurons from one

<sup>2</sup>Doing so, we neglect of course the change in appearance of galaxies from UV to NIR, from large to narrow-band filters; but the effect of this change is not larger than the one due to the natural spread of galaxy types in the  $B$  band.

layer are connected to all neurons of the previous layer, but there is no connection between neurons of a same layer. In a classifier, the input nodes are given the parameters defining the object, and the output nodes contain “affiliation indices” to each class. The NN is used in two modes:

- a “training” mode in which the network “learns” to associate inputs and desired outputs which are repeatedly presented to it (supervised learning).
- a “play-back” mode in which it simply responds to new patterns according to prior training.

In the multi-layer network we consider, a neuron  $i$  in layer ( $l$ ) has the following output

$$\mathbf{o}_i^{(l)} = g(\mathbf{w}_i^{(l)} \cdot \mathbf{o}^{(l-1)} - b_i^{(l)}) \quad (7)$$

where  $\mathbf{w}_i^{(l)}$  is the weight vector of the connections between the neurons from the previous layer  $l - 1$  and neuron  $i$ , and  $b_i^{(l)}$  is a bias term which may be considered as the weight of a connection with a steady neuron

output. The activation function  $g$  is the usual sigmoid function:  $g(u) = 1/(1 + e^{-u})$ . The training procedure is organized as follows. First of all, weights and biases are initialized to random values. Then for each pattern, of a “training set” loaded into the input layer at iteration  $t$ , a forward propagation through the network is done using Eq. (7). Weights and biases are adjusted layer by layer backward, according to the error backpropagation learning rule (Rumelhart et al. 1986):

$$\Delta \mathbf{w}_j^{(l)}(t+1) = -\eta \nabla E(\mathbf{w}_j^{(l)}(t)) + \alpha \Delta \mathbf{w}_j^{(l)}(t) \quad (8)$$

$E$  is a quadratic cost function which measures the output error of the network:

$$E = \frac{1}{2} |\mathbf{O} - \mathbf{o}|^2 \quad (9)$$

where  $\mathbf{O}$  is the desired output vector and  $\mathbf{o}$  the response of the network to the training pattern. The “learning rate” of the network is determined by  $\eta$ , while the *momentum parameter*  $\alpha$  may be set to non-zero (but  $< 1$ ) to speed up the convergence in flat regions of the cost surface, and prevent the network being fooled by local fluctuations. All learnings here have been carried out with  $\eta = 0.1$  and  $\alpha = 0.2$ .

As learning progresses,  $E$  decreases slowly. With experimental data, a million learning steps are often needed for  $E$  to reach a stable minimum, which means that many passes through the training set are necessary (faster learning algorithms do exist, but as we only learn once, speed is not here a matter of concern).

### 7.3. Optimization for the star/galaxy classification

#### 7.3.1. Input parameters

Choosing the right input parameters is a crucial step in star/galaxy separation. The following points were taken

into account in our attempt to choose “optimal” parameters. They should

- i) discriminate efficiently the two classes over the whole magnitude range accessible with any linear detector.
- ii) be invariant under translation and rotation (scale invariance is not demanded because of seeing).
- iii) bear some robustness concerning noise, image distortions, and influence of close neighbours.
- iv) be as independent as possible from the characteristics of the exposure (depth, scale).

Simple parameters that fulfil these requisites are isophotal areas. A good depiction of image profiles can be done by combining some isophotal areas  $A_0, A_1, \dots$  and the maximum pixel value above the sky  $I_{\max}$ . Taking more isophotal areas than only  $A_0$  (corresponding to the lowest isophote) allows the classifier to work better with dim objects (a high relative uncertainty on  $I_{\max}$  weakens the contrast between the two classes), bright stars ( $I_{\max}$  reaches the saturation level), and deblended images (the faint pixel allocation procedure distorts the wings of the profile). This set of parameters is very similar to the one adopted by Maddox et al. (1990a) for the APM survey, the difference being in the distribution of the isophotes: a steady scale for the APM, and normalized to each profile in our case. We did not include any elongation measurement in the parameter set in order to comply with iii); in fact, as it has been remarked by Odewahn et al. (1992), ellipticity and most sophisticated geometrical parameters are not of major importance in star/galaxy separation, particularly for faint objects.

Thus, we end with 10 parameters: 8 isophotal areas, 1 peak intensity, and 1 “control-parameter” which is the seeing. Isophotal areas are given to the network in units of “squared seeing FWHM”, which eliminates the need for a pixel scale information (we assume that the images are correctly sampled). There should be no difference concerning the extracted profiles (except noise), between a deep exposure analysed at a high threshold, and a less deep one analysed at a lower threshold, so we can get rid of the depth information by expressing the peak intensity in units of extraction threshold. Because of the wide brightness range of the objects that may appear in the image, the neural network is fed with the logarithms of the parameters.

Each of these is scaled in order to have the input distribution falling in the interval  $[-1, +1]$  within  $3\sigma$ . Weights and biases are initialized with random values between  $-N_l^{-1}$  and  $+N_l^{-1}$ , where  $N_l$  is the number of nodes in the layer  $l$ . These precautions ensure at the beginning of training that the net input of each neuron is in the domain of maximum sensitivity of the activation function.

### 7.3.2. Topology of the network

The optimal architecture of the NN in terms of performance depends on the distribution of the pattern vectors in parameter space. A multi-layer network is characterized by the number of layers and the number of neurons in each layer.

*Number of layers.* The frontier between stars and galaxies in parameter space is assumed to be continuous. In this case, it can be shown that one hidden layer is enough to separate the two classes (e.g. Cybenko 1989).

*Number of neurons in each layer.* The input layer contains as many units as there are parameters. The output layer contains only one neuron, as “star” and “galaxy” are two classes mutually exclusive. It will define a “stellarity index”: 0 for a galaxy, 1 for a star, or any intermediate value for more ambiguous objects.

Note that we have not created a class for defects such as cosmic ray impacts, bad column or glitches. The major reason being that on “correctly sampled” images (FWHM  $\gtrsim 2$  pixels) they can easily be discarded by standard techniques<sup>3</sup>. We want the classifier to have a sufficient number of degrees of freedom to separate classes, and few enough to be able to “generalize” and not be too tightly bounded to the training data. In networks like the one used here, this feature is determined by the number of neurons in the hidden layer. In fact, our training set is large enough that we could use a large number of neurons (a few tens) without loss of generalization (see Bertin 1995). But, in order to save processing time, we restricted the number of “hidden neurons” to the reasonable value of 10 (the same as the number of inputs). The resulting neural network is shown in Fig. 5.

### 7.4. Training

Six hundred  $512 \times 512$  simulation images containing stars and galaxies were generated to train the network (the simulations are described in the appendix). They were done in the blue band, where galaxies present very diversified aspects. The two PSF parameters (seeing FWHM and Moffat  $\beta$  parameter) were chosen randomly with  $0.025 \leq \text{FWHM} \leq 5.5''$  and  $2 \leq \beta \leq 4$ . The pixel scale was always taken less than  $\approx 0.7$  FWHM to warrant correct sampling of the image. Bright galaxies are simply too rare in the sky to constitute a significant training sample on such a small number of simulations. So, keeping a constant comoving number density, we increased artificially the number of nearby galaxies by making the volume element proportional to  $z.dz$ . The stars were given a number-magnitude distribution identical to that of galaxies. Thus any pattern presented to the network has 50% chance to

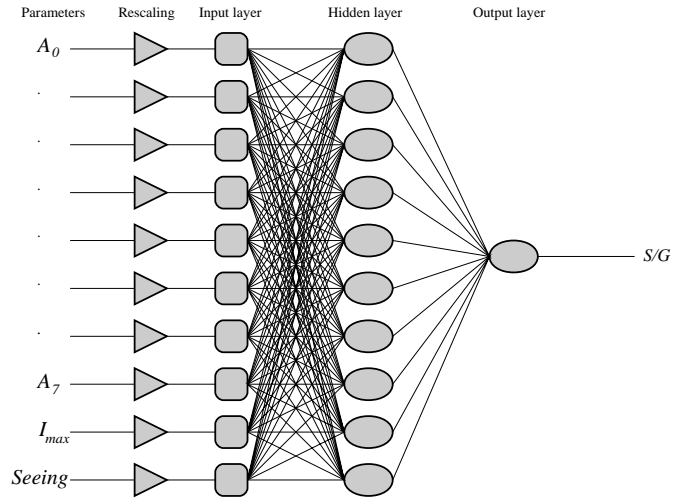


Fig. 5. Diagram of the neural star/galaxy classifier

correspond to a star or a galaxy, whatever its magnitude<sup>4</sup>. This weighting is certainly not optimum for classifying real object with  $V \lesssim 18$  (where stars outnumber galaxies) or inversely  $V \gtrsim 22$  (galaxies outnumber stars), but is preferable here, as all depth information have been withdrawn from the input parameters for the sake of generality. In Sect. 7.5 we indicate how this problem can be easily alleviated. As one can see in Fig. 6, the crowding in the simulated images is higher than what one sees on real images of the field, allowing for the presence of many “difficult cases” (close double stars, truncated profiles, etc...) that the neural network classifier will have to deal with.

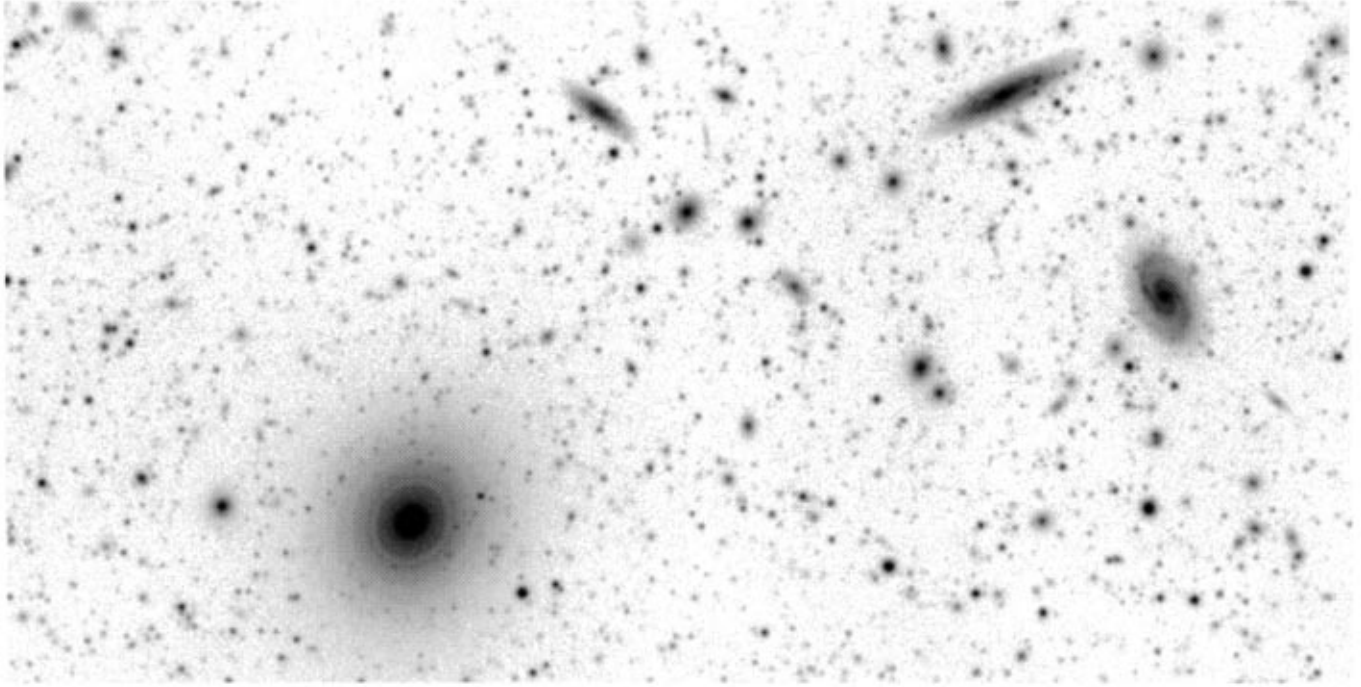
SExtractor was run on each image with 8 different extraction thresholds. It produced a catalog with about  $10^6$  entries, for each of which the 10 classification parameters are available. Backpropagation learning took about 15 min on a SUN SPARC20 workstation. The corresponding set of synaptic weights was then saved to a file, ready to be used in “feed-forward only” mode during source extraction.

### 7.5. Confidence estimation of the classification

It has been proven, both theoretically and experimentally (Richard & Lippmann 1991 and references therein), that the cost function of a multilayered neural net as ours is minimized when outputs estimate bayesian a posteriori probabilities. Although the present classifier does not output exact bayesian probabilities (because real data unavoidably differ a bit from simulated ones), it provides at least a confidence estimate in the classification. This is obvious in Fig. 9, where the stellarity index tends to “degenerate” to intermediate values as magnitude increases

<sup>3</sup>Optical artifacts like spikes or scattered light rays would certainly be worth identifying, as they are found to pollute all survey catalogs, especially around bright stars; but the classification parameters used here are simply inadequate.

<sup>4</sup>Faint galaxies have less chance of being detected than faint stars, but it would have little effect since the ones that are lost at a given magnitude are predominantly the most extended and consequently the easiest to classify.



**Fig. 6.** Example of a simulated CCD field (displayed using a logarithmic scale) for training the star/galaxy classifier

and classification becomes harder. Besides, qualitative estimates of confidence done by eye are in good agreement with those of the classifier.

Thus, the problem of contamination mentioned in Sect. 7.4, arising when the classification becomes difficult and the two classes are in unbalanced proportions, can be straightforwardly solved by adjusting the decision boundary between “stars” and “galaxies” within the stellarity index range  $[0,1]$ .

#### 7.6. Testing the neural network classifier

We checked the performance of the classifier both on simulated and real images. Simulated images are required for very faint objects for the obvious reason that their nature cannot be certified from ground-based images.

##### 7.6.1. Simulation images

A new set of images were simulated, but this time with the purpose of testing the trained classifier. The characteristics of these images match those of typical 20 mn CCD exposures in the  $R$  band, at high galactic latitude, with a 3.5 meter telescope, a pixel size of  $0.35''$  and a seeing FWHM of  $0.9''$ . The classification of extracted sources was done with SExtractor, entering different values for the seeing parameter. Objects with a stellarity index less than 0.5 were identified as “galaxies”, and “stars” otherwise. As expected, the classifier gets its peak performance when the seeing parameter is set to  $0.9''$  (Fig. 7), with a

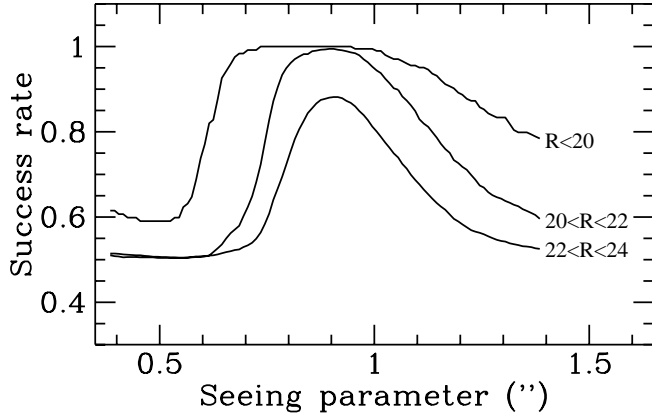
tolerance depending on magnitude. The tolerance in seeing is quite large for bright objects ( $\approx 20\%$ ), which means that an approximate seeing parameter can be entered to pick up the brightest stars in an image and determine a more accurate estimate for classifying fainter detections. Figure 8 shows the degradation of the classification quality with magnitude, using the optimum seeing parameter value ( $0.9''$ ). More errors are made for stars than galaxies at low flux, because of crowding: a faint star has a significant probability to catch wings added to its profile by a background galaxy and get misclassified.

##### 7.6.2. Real images

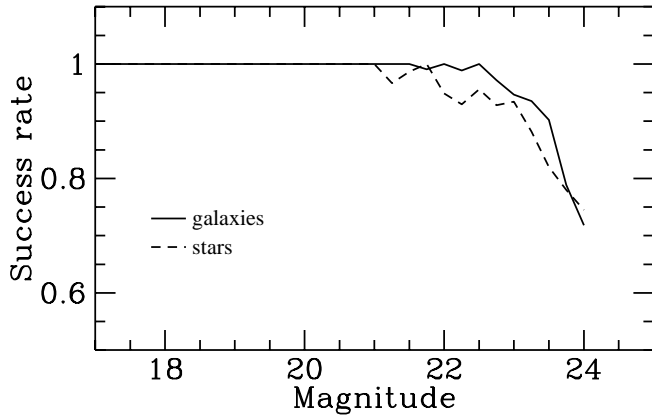
CCD images from several ground-based telescopes with different depth, seeing and pixel scale were processed by SExtractor (Fig. 9), and examined by eye. All isolated and non-saturated objects identifiable by eye were found to have been correctly classified. Saturation features differ from one instrument to another, and lead to misclassifications for very bright stars on some images. Close blended and undeblended stellar pairs are classified properly, but some objects that might be compact members (a few percent) of deblended galaxy-galaxy pairs were found to be classified as stars.

In conclusion, if one excludes heavily saturated stars, the classification of “bright” objects appears to be reliable at least at a conservative  $\approx 95\%$  level, on both simulated and real images with a good sampling. Figure 10 shows an





**Fig. 7.** Global success rate of the classifier (on sets containing the same proportion of simulated stars and galaxies) as a function of the seeing parameter, for three  $R$  magnitude ranges



**Fig. 8.** Success rate as a function of  $R$  magnitude for the same simulated stars and galaxies as in Fig. 7

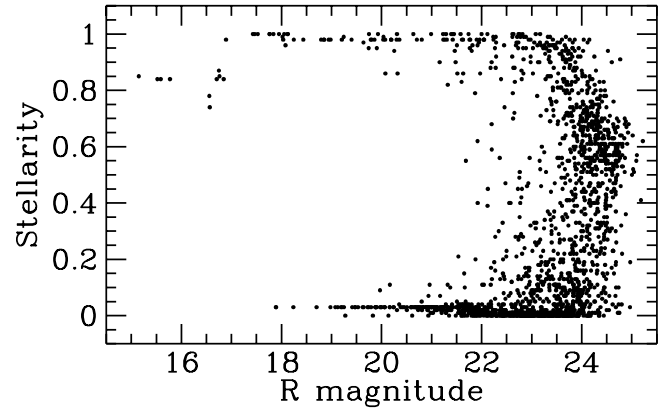
example of a CCD image and its interpretation by SExtractor.

## 8. Speed

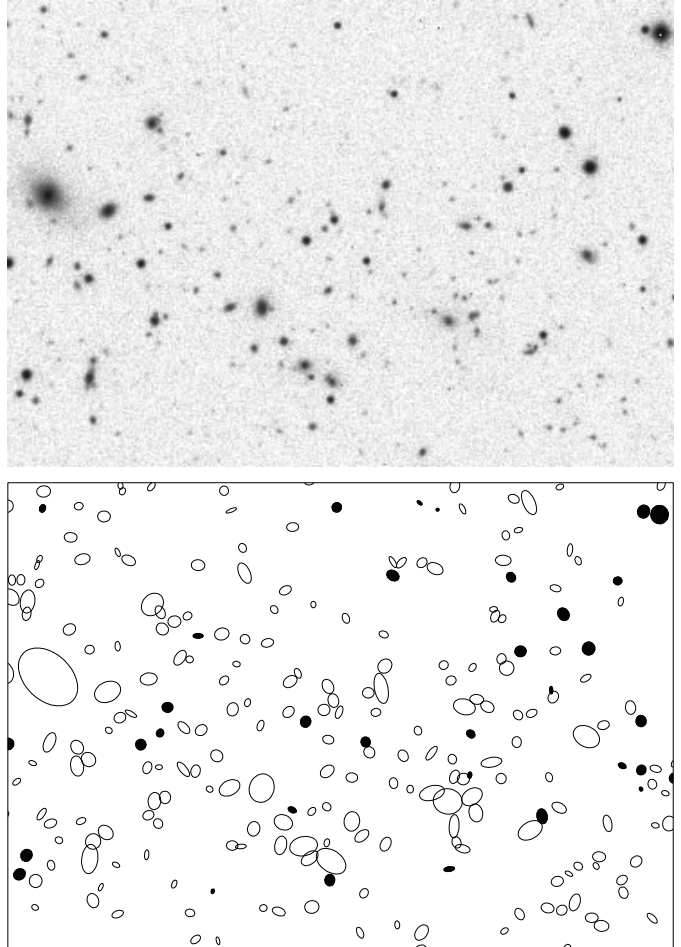
Together with the background estimation, the detection/deblending/filtering process is the most time-consuming part of the image analysis. With a SUN SPARC20, we measure a global processing speed of about  $40 \text{ kpix.sec}^{-1}$  without convolution, and  $\approx 30 \text{ kpix.sec}^{-1}$  with small convolution masks.

## 9. Summary

We have described SExtractor, a programme for the automatic analysis of moderately crowded astronomical images. Speed, robust deblending and estimation of “total” magnitudes make of SExtractor a tool particularly well-suited to batch processing of large survey data.



**Fig. 9.** Classifier output (“stellarity index”) as a function of  $R$  magnitude for an NTT image from the photometric sample of the ESO-Sculptor survey (Arnouts et al. 1995). The seeing FWHM is  $0.8''$ . Note the decrease in “stellarity” for saturated stars



**Fig. 10.** Interpretation by SExtractor of a CCD frame (ESO 3.6 m,  $R$  band, 3 min) containing a distant galaxy cluster. Filled ellipses are objects with stellarity index  $> 0.8$  (stellar objects)

We demonstrated that reliable star/galaxy classification can be achieved using a tunable neural network trained on realistic simulation images. Because the essential features of modern images can be described by only a few parameters, this approach (simulation of data + tunable neural network) could be advantageously extended to other classes of objects (like Hubble types), or even to the measurement process itself (optimal determination of positions, magnitudes, etc.).

The SExtractor package and its technical documentation are electronically available through World Wide Web at: [ftp://ftp.iap.fr/pub/from\\_users/bertin/](ftp://ftp.iap.fr/pub/from_users/bertin/)

*Acknowledgements.* We thank all the users of the beta-version of SExtractor for their comments and suggestions, P. Fouqué and C. Lidman for comments on the manuscript of this paper.

### A. Appendix: simulation of artificial galaxy images

The simulated images we have used to test deblending, photometry and star/galaxy separation contain galaxies and stars with  $B$  magnitudes ranging from 10 to 27. Most of the difficulty concerns the simulation of galaxies. In order to simulate the large number of images needed for the neural network training, we have tried to find a compromise between realism and speed. Our concern was not to build a cosmological tool, but simply a fast code capable of producing convincing sky images.

Fundamental parameters defining the morphology of a galaxy are its Hubble type  $T$  and its absolute blue magnitude  $M^5$ . The Hubble type  $T$  is randomly chosen between  $-5$  and  $+10$ , with a global type distribution complying with the one measured by Spieckermann (1992), and a detailed one taken from Lauberts & Valentijn (1989). The blue-band absolute magnitude  $M$  distribution follows a Schechter's (1976) law

$$\phi(M)dM = \phi^* \exp(0.92(\alpha + 1)(M^* - M) - \exp(0.92(M^* - M))) dM \quad (A1)$$

Values like  $M^* = -20.9$ ,  $\alpha = -1.1$  and  $\phi^* = 2.3 \cdot 10^{-3}$  (for  $H_0 = 50 \text{ km.s}^{-1}.\text{Mpc}^{-1}$ ) match real counts around  $B = 19^{\text{th}}$  magnitude (e.g. Yoshii & Takahara 1988). No dependence on redshift  $z$  or morphological type was introduced in the luminosity function; that is, we did not include luminosity or density evolution of the galaxies (this would have been beyond the scope of this simple simulation).

Each galaxy contains a spheroidal component, which we assumed to follow a de Vaucouleurs law (in  $\text{mag.pc}^{-2}$ )

$$\mu_S(r) = M_S + 8.3268 \left( \frac{r}{r_e} \right)^{\frac{1}{4}} + 5 \log r_e + 16.6337, \quad (A2)$$

<sup>5</sup>So far, most of the characteristics and scaling laws of galaxies have been established in the  $B$  band; that's why we shall do all the computations in  $B$  and do a rough conversion to another bandpass by using color indices.

where  $M_S$  is the  $B$  absolute magnitude of the spheroid and  $r_e$  its effective radius in parsecs. For ellipticals, from the work of Binggeli et al. (1984), we adopt the following mean relation between  $r_e$  (expressed in pc) and  $M_S$

$$r_e = \begin{cases} \left( \frac{h}{0.5} \right)^{0.5} 10^{3.5-0.3(M_S+20.5)} & \text{if } M_S < -20.5 \\ \left( \frac{h}{0.5} \right)^{-0.5} 10^{3.5-0.1(M_S+20.5)} & \text{otherwise} \end{cases} \quad (A3)$$

Bulges generally appear to be more diffuse than ellipticals (see e.g. Kent 1985), but in the context of this simulation, no distinction has been made between the two. The intrinsic flattening  $q$  of spheroids is taken between 0.3 and 1, and within this range follows a normal distribution with  $\langle q \rangle = 0.65$  and  $\sigma_q = 0.18$  (Sandage et al. 1970). Once again, we assume the same distribution for bulges and ellipticals, even if there is some controversy about this (see Boroson 1981). The apparent axis ratio is then  $\sqrt{q^2 \sin^2 i + \cos^2 i}$ , where  $i$  is the inclination of the (oblate) spheroid with respect to the line of sight.

The disk component, when present, is given an exponential profile:

$$\mu_D(r) = \mu_0 + 1.0857 \left( \frac{r}{r_h} \right), \quad (A4)$$

where  $r_h$  its scale-length, and  $\mu_0$  its central surface brightness. We refer to the studies of Freeman (1970), van der Kruit (1987), and Binggeli (1994) and adopt  $\langle \mu_0 \rangle = 21.65 \text{ mag.arcsec}^{-2}$  for non-dwarf galaxies ( $M \leq -17$ ), and  $\langle \mu_0 \rangle = 21.65 + 0.7(M + 17)$  otherwise<sup>6</sup>.  $\mu_0$  follows a gaussian distribution with  $\sigma_{\mu_0} = 0.35$  around these means. The scale length  $r_h$  (in pc) of the disk comes with the normalization

$$r_h = \text{dexp}(-4.713 + 0.2(\mu_0 - M_D)), \quad (A5)$$

where  $M_D$  is the  $B$  absolute face-on magnitude of the disk. We take into account the truncation in the light profile visible for many disk galaxies (see van der Kruit 1987) by introducing a sharp parabolic decline in surface brightness magnitude at  $r = 5r_h$ . A logarithmic spiral pattern, consistent with the measurements of Kennicutt (1981) and Schweizer (1976), is used to modulate the disk of late-type galaxies ( $T > 0$ ), chiefly to introduce some realistic irregularities in their profile. The disk profile is then given an axis ratio of  $\cos i$ .

Galaxies with  $T \leq -4$  have of course  $M_S = M$  and  $M_D = +\infty$ . For galaxies with  $-4 < T < 8$  the fractional luminosity of the bulge is a monotonic function of the Hubble type  $T$ , as shown by Simien & de Vaucouleurs (1986). We used the following empirical fit (their Eq. 4) to obtain  $M_S$  from  $M$ :

$$M_S = M + 0.80 + 0.145T + 0.0284T^2 + 0.00267T^3 \quad (A6)$$

<sup>6</sup>Although there are claims (McGaugh 1995, and references therein) that these mean surface brightnesses only reflect selection effects.

This parametrization can be extended to later types ( $T \geq 8$ ), because in that case  $M_S - M > 5$  mag, i.e.  $M_D - M \approx 0$ . Internal absorption  $A_i$  of the disk (in the blue band) is modeled by

$$A_i = -\alpha(T) \log \cos i \quad (\text{A7})$$

with (de Vaucouleurs et al. 1991)

$$\alpha(T) = \begin{cases} 1.5 - 0.03(T - 5)^2 & \text{if } T \geq 0 \\ 0 & \text{if } T < 0 \end{cases} \quad (\text{A8})$$

Note that we apply this correction only for the disk. To avoid excessively large  $A_i$  the inclination angle  $i$  is kept  $< 80^\circ$ .  $R$  band galaxy parameters are crudely derived from the  $B$  band parameters by assuming  $B - R = 1.5$  for the spheroidal component and  $B - R = 0.7$  for the disk (except for S0 galaxies). The internal absorption in  $R$  is assumed to be 0.56 times what it is in  $B$  (Cardelli et al. 1989). K+e corrections are applied separately for the two components (except once again for S0 galaxies) by using the polynomial fits of Metcalfe et al. (1991), which are expected to be valid at least to  $z \approx 1.5$ . Finally, each galaxy is “placed” at its redshift  $z$ . A Friedmann universe with zero cosmological constant is assumed. The (bolometric) luminosity and the angular diameter distance are respectively (Weinberg 1972):

$$d_L = \frac{c}{H_0 q_0^2} \left( q_0 z + (q_0 - 1)(\sqrt{2q_0 z + 1} - 1) \right) \quad (\text{A9})$$

and

$$d_A = \frac{d_L}{(1+z)^2} \quad (\text{A10})$$

After proper scaling and dimming, the objects are all placed in a temporary array to be convolved with the point spread function

$$I(r) = \frac{I_0}{\left(1 + \left(\frac{r}{R}\right)^2\right)^\beta} \quad (\text{A11})$$

which is the Moffat (1969) function. Stellar image profiles can be represented with a good accuracy for most instruments with any seeing by adjusting  $R$  and  $\beta$ . Taking  $\beta = 3$  yields  $\text{FWHM} \approx R$ .

An oversampling of  $3 \times 3$  is made on each pixel; but the convolution is applied only with the final pixel size. One may normally fear some distortion of the images due to binning; however, comparaisons made with finer pixel grids have convinced us that these effects affect the profiles negligibly in the conditions of our simulations, while reducing considerably the time to produce the images.

Figure 11 shows some examples of simulated galaxy images as described above.



**Fig. 11.** Examples of 15<sup>th</sup> magnitude simulated galaxy images. From left to right: E, S0, Sb and Sd/Irr types

## References

- Almoznino E., Loinger F., Brosch N., 1993, MNRAS 265, 641
- Arnouts S., de Lapparent V., Mathez G., Mazure A., Mellier Y., Bertin E., Kruszewski A., 1996, (accepted for publication in A&AS)
- Beale R., Jackson T., 1990, Neural Computing: an Introduction, Adam Hilger, Bristol
- Beard S.M., McGillivray H.T., Thanisch P.F., 1990, MNRAS 247, 311
- Bertin E., 1994, Ap&SS 217, 49
- Bertin E., Dennefeld M., 1996, (accepted in A&A)
- Bertin E., 1996, Thèse de Doctorat, Université Paris VI
- Bijaoui A., 1980, A&A 84, 81
- Binggeli B., Sandage A., Tarenghi M., 1984, AJ 89, 64
- Binggeli B., 1994, in Dwarf Galaxies. In: Meylan G., Prugniel P. (eds.), ESO Conference and Workshop Proceedings No. 49, 13
- Boroson T., 1981, ApJS 46, 177
- Cardelli J.A., Clayton G.C., Mathis J.S., 1989, ApJ 345, 245
- Colless M., Ellis R.S., Taylor K., Hook R., 1990, MNRAS 244, 408
- Da Costa G.S., 1992, in Astronomical CCD Observing and Reduction Techniques. In: Howell S.B. (ed.)
- Coupinot G., Hecquet J., Aurire M., Futauly R., 1992, A&A 259, 701
- Cybenko G., 1989, Mathematics of Control, Signals and Systems 2, 337
- Freeman K.C., 1970, ApJ 160, 811
- Harmon R., Mamon G., 1993, in Sky surveys: Protostars to Protogalaxies. In: Soifer T. (ed.), PASP
- Hertz J., Krogh A.S., Palmer R.G., 1991, Introduction to the Theory of Neural Computation. Addison-Wesley
- Infante L., 1987, A&A 183, 177
- Irwin M.J., 1985, MNRAS 214, 575
- Irwin M.J., Hall P., 1983, in Proc. Workshop Astronomical Measuring Machines. In: Stobie R.S. & McInnes B. (eds.) ROE, Edinburgh, 111
- Jarvis J.F., Tyson J.A., 1981, AJ 86, 476
- Jones L.R., Fong R., Shanks T., Ellis R.S., Peterson B.A., 1991, MNRAS 249, 481
- Kendall M., Stuart K., 1977, The Advanced Theory of Statistics, Vol. 1. Charles Griffin & Co., London
- Kennicutt R.C., 1981, AJ 86, 1847

- Kent S.M., 1985, *ApJS* 59, 115
- Kron R.G., 1980, *ApJS* 43, 305
- van der Kruit P.C., 1987, *A&A* 173, 59
- van der Kruit P.C., 1988, in *The World of Galaxies*. In: Corwin G.C., Bottinelli L. (eds.). Springer-Verlag, New York, 256
- Lahav O., 1994, *Vistas in Astronomy* 38
- Lauberts A., Valentijn E.A., 1989, *The Surface Photometry Catalogue of the ESO-Uppsala Galaxies*, European Southern Observatory
- Lefèvre O., Bijaoui A., Mathez G., Picat J.P., Lelièvre G., 1986, *A&A* 154, 92
- Lutz R.K., 1979, *The Comp. J.* 23, 262
- McGaugh S.S., Bothun G.D., Schombert J.M., 1995, *AJ* 110, 573
- Maddox S.J., Sutherland W.J., Efstathiou G., Loveday J., 1990a, *MNRAS* 243, 692
- Maddox S.J., Efstathiou G., Sutherland W.J., 1990b, *MNRAS* 246, 433
- Metcalfe N., Shanks T., Fong R., Jones L.R., 1991, *MNRAS* 249, 498
- Moffat A.F.J., 1969, *A&A* 3, 455
- Newell E.B., 1983, in *Proc. Workshop Astronomical Measuring Machines*. In: Stobie R.S. & McInnes B. (eds.) ROE, Edinburgh, 15
- Odewahn S.C., Stockwell E.B., Pennington R.L., Humphreys R.M., Zumach W.A., 1992, *AJ* 103, 318
- Odewahn S.C., Humphreys R.M., Aldering G., Thurmes P., 1993, *PASP* 105, 1354
- Reid N., Gilmore G., 1982, *MNRAS* 201, 73
- Richard M.D., Lippmann R.P., 1991, *Neural Comp.* 3, 461
- Rumelhart D.E., Hinton G.E., Williams R.J., 1986, *Nat* 323, 533
- Sandage A., Freeman K.C., Stokes N.R., 1970, *ApJ* 160, 831
- Schechter P., 1976, *ApJ* 203, 297
- Schweizer F., 1976, *ApJS* 31, 313
- Sebok W.L., 1979, *AJ* 84, 1526
- Simien F., de Vaucouleurs G., 1986, *ApJ* 302, 564
- Slezak E., Bijaoui A., Mars G., 1987, *A&A* 201, 9
- Smail I., Hogg D.W., Yan L., Cohen J.G., 1995, *ApJ* 449, L105
- Spiekermann G., 1992, *AJ* 103, 2103
- de Vaucouleurs G., de Vaucouleurs A., Corwin H.G., Buta R.J., Paturel G., Fouqué P., 1991, *Third Reference Catalogue of Bright Galaxies*. Springer, New York
- Wasserman P.D., 1989, *Neural Computing: Theory and Practice* Van Nostrand Reinhold, New York
- Weinberg S., 1972, *Gravitation and Cosmology*, Wiley, New York
- Yee H.K.C., 1991, *PASP* 103, 396
- Yoshii Y., Takahara F., 1988, *ApJ* 326, 1