Mining pixels: The Extraction and Classification of Astronomical Sources

Emmanuel Bertin[1,2]

[1] Institut d'Astrophysique, 98bis Bd Arago, F-75014 Paris
[2] DEMIRM, Observatoire de Paris, 61 Av. de L'Observatoire, F-75014 Paris

Abstract. In most imaging surveys, detecting sources and classifying them in an automatic way may be seen as the very first stage of the data-mining process. After presenting the general aspects of image detection and classification in the astronomical context, I review the different techniques used in current projects. Emphasis in put on promising developments such as artificial neural networks and vision models.

1 Introduction

The history of automatic source extraction goes back to the middle of the 60's, with the first computerized counts of radiosources and projects to exploit the enormous amount of information accumulating on photographic plates, the so-called "Schmidt problem" (Fellgett 1970 [23]). The first fully automatic measuring machines put into service, the GALAXY machine in Edinburgh (Stoy 1970 [84]) and the APMS in Minneapolis (La Bonte 1970 [37]), already allowed simple flux and position measurements at a rate much higher than what could be done by hand (Luyten 1974 [44]). Since then, detection and classification software has evolved, although at a relatively low pace. This is mainly because simple techniques are already able to meet the specifications required by most scientific programs.

The extraction of deep-sky astronomical sources seems indeed a rather straightforward task as compared to many other image vision problems: one generally deals with light-emitting sources on dark backgrounds. Perspective and deformation effects on object shapes can be ignored, except in the rare cases of strong gravitational lensing. Moreover, on images coming from linear detectors, the Point Spread Function (PSF) is readily available in the form of star images. There are some special difficulties associated with the processing of astronomical images, though. Sources have no clear-cut boundaries. Within the same image, algorithms must deal with a huge dynamic range in object size (reaching $\times 10,000$) and signal-to-noise ratio (> 80 dB with modern instruments). They must cope with varied problems such as a changing PSF in ground-based images, light pollution due to optical reflections in wide-field exposures, or low photon statistics with narrow-band-visible, X-ray and

gamma-ray detectors. And most importantly, the systematics in the extracted catalogs have to be well-controlled, as most scientific studies rely on the combination of thousands of measurements.

This review is somewhat biased in favor of the processing of optical/near-infrared images since those contribute the most, by far, to the ever growing quantity of pixels acquired around the world. Because of space constraints, "pre-detection" image processing like image-subtraction (e.g. Alard 2000 [1], and references therein) and catalog-based filtering (e.g. Vandame, this conference) or classification (e.g. Szalay, this conference) are not described here. The image classification part focuses on neural networks; a review of other classification techniques can be found in the contributions of Bijaoui, Djorgovski and Lahav (this conference). The review is organized as follows. In section 2.1 we give an overview of the basic detection techniques, including the problems of deblending and background determination. Section 2.2 is devoted to multiscale analysis. 3.1 presents basic star/galaxy classification techniques. The different kinds of trainable classifiers used in astronomy are mentioned in 3.2. Section 3.3 introduces neural networks, and 3.4 describes applications for star/galaxy separation and the identification of defects. Finally, we discuss the current trends and identify future developments in Section 3.6.

2 The detection of astronomical images

2.1 Basic detection algorithms

Under the assumption that background noise is stationary at least in the wide sense, optimum detection of isolated sources with known profile $\phi(\boldsymbol{x})$ is obtained by correlating the data with the matched filter

$$h(\boldsymbol{x}) = \phi * \mathcal{F}(\frac{1}{P}), \tag{1}$$

where P is the noise power-spectrum, and $\mathcal{F}$ the Fourier transform. As we shall see later, low-frequency noise components can be absorbed in a "background model" subtracted prior to detection. In addition, on co-added CCD images, noise is close to white down to the Point Spread Function (PSF) scale, hence the $\frac{1}{P}$ term can be dropped; the detection filter is then simply the PSF.

Local peak search Following or while performing matched filtering, the most intuitive method to make the actual detection, is searching for a local peak (Newell & O'Neil 1976 [55] , Herzog & Illingworth 1977 [27], Kron 1980 [36], Buonanno et al. 1983 [16], Stetson 1987 [82], Mighell 1989,99 [49],[50], Yee 1991 [90], Vikhlinin et al. 1995 [86]). This is generally done by comparing, on the filtered image, the value of the current pixel with that of its neightbours. Local peak search detects sources at a single scale (that of

the matched filter) and is therefore fairly insensitive to inhomogeneities in the background on larger scales. Deblending of close, unresolved sources is generally good in one pass, and for heavily blended point-sources, several passes can be done through the data with iterative detection/subtraction. All these features make the local peak search more appropriate for the detection of stars and other unresolved sources (especially in crowded fields), yielding magnitude-limited catalogs. However it is not well suited to detecting complex objects like resolved late-type galaxies, which have a tendency to break-up components.

Thresholding and segmentation The detection problem with extended objects can be partly solved if instead of looking for a local peak, one looks for connected sets of pixels above a given threshold in the background-subtracted, match-filtered image (Deutschman 1971 [21], Lutz 1979 [43], Stobie et al. 1979 [83], Jarvis & Tyson 1979,81 [32], [33], Hall & Mackay 1984 [25], Irwin 1985 [30], Le Fèvre et al. 1986 [41], Slezak et al. 1988 [75], Beard et al. 1990 [5], Bonnet & Mellier 1994 [13], Bertin & Arnouts 1996 [8]). The match-filter — in most cases — acts as a low-pass filter, and although detection is not as optimal at scales larger than the PSF as it is for point-sources, it behaves much better than peak-search on extended objects. The thresholding technique has the disadvantage that it requires a more accurate background-subtraction, but this is needed nowadays for image-mosaicing anyway. Besides, to maximize detectability, the threshold has to be set as low as possible, at a level where the wings of neighbor object profiles often merge. This makes a deblending routine mandatory to separate individual sources that make up the primary detection.

Various deblending algorithms have been proposed and used in surveys: local maxima searches (e.g. Lupton & Gunn 1986 [42], Lasker et al. 1990 [39]), cuts along principle axes (Jarvis & Tyson 1979 [32], Valdes 1982 [85]), multi-thresholding (Irwin 1985 [30], Beard et al. 1990 [5]). A "contrast threshold" can be applied to avoid splitting noise peaks of patchy galaxy features (Bertin & Arnouts [8]). After the components have been identified, pixels in the object wings are re-associated on a statistical basis to their most probable contributors.

Despite its complications and limitations, thresholding/segmentation is nevertheless fast and efficient when properly tuned. It is certainly the most commonly used detection technique in current optical/near-infrared imaging surveys. Catalogs generated with this method are primarily surface-brightness-limited for extended objects (resolved galaxies).

Background estimation As we saw before, the thresholding/segmentation technique relies a lot on the quality of the background map subtracted prior to detection. Background maps are made from interpolated local background values. The area over which the background value is estimated defines the

upper limit in detection scale. Unfortunately, a "local background" cannot always be uniquely defined; especially in crowded fields: formally, there should be one background map per source. With this limitation in mind, sophisticated model-dependent background estimators were proposed (e.g. Newell 1979,83 [56],[57], Bijaoui 1980 [10]), but were not widely adopted, as they were computationally too expensive. In fact, in all but the most severe cases, estimates of the mode of the pixel intensity distribution have been shown to yield satisfactory results, at least for detection purposes (Lasker et al. 1990 [39], Bertin & Arnouts 1996 [8]).

Pixel weighting and filtering Earlier we assumed the background noise was wide-sense stationary over the image. In fact this is almost never the case with images obtained with detector arrays, because of vignetting, pixel-to-pixel sensitivity variations, interference fringes, or glitches. With mosaic cameras, dithered co-addition makes things much worse by introducing abrupt changes in coverage over the image. This problem is solved to a large extent by introducing a weight $\propto 1/\sigma^2$ (where σ^2 designs the local variance) assigned to each science pixel. A zero-weight indicates a bad pixel. "Weight maps" are propagated throughout the reduction pipeline in parallel to the scientific data (see Nonino et al. 1999 [58]). "Co-added" weight-maps can then be used to modulate the detection threshold of the final co-added science images. Note that pixel-pixel covariances are ignored; they would just take too much space to store. Hopefully they are generally negligible at the detection scale.

Detection of LSB (Low-Surface-Brightness) objects The possibility that a large number of low surface-brightness galaxies, "missed" with traditional detection criteria, might make up a significant fraction of the galaxy population has motivated the search for LSB features over wide fields in the visible (see e.g. Bothun et al. 1997 [14] and references therein). The diffuse light from unresolved galaxies in distant clusters is another source of low surface brightness fluctuations of the sky background (Dalcanton 1996 [19]). Searches for these kinds of objects require images with a very homogeneous background on intermediate scales, like digitized Schmidt plates or CCD drift-scans. Here again the matched filter technique gives good results using an exponential profile, at the condition that "standard" sources are detected and masked beforehand (see for instance Testa et al. , this conference). A slightly less efficient, but simpler technique is to use a local background estimator (median filter or mode).

2.2 Multiscale approaches

The basic detection techniques described in the previous section have a major shortcoming: they perform well only in a narrow range of object scales. This can lead to poor results on some data, like X-ray cluster images, where a

small number of photons are disseminated over overlapping sources having a wide range of sizes, or images of HII regions, where points sources are superposed to a complex structure of nebulae and dark molecular clouds. Multiscale detection tries to solve this problem by extracting the signal at multiple scales. A tree graph of detections can then be built, where individual objects are identified (and reconstructed if necessary).

Most multiscale analyses use wavelet decomposition (Coupinot et al. 1992 [18], Bijaoui 1994 [11], Rosati et al. 1994 [68], Bijaoui & Rué 1995 [12], Grebenev et al. 1995 [24], Damiani et al. 1997 [20], Lazzati et al. 1999 [40], Starck et al. 1999 [80], Starck et al. 2000 [81]) although other decompositions like the pyramidal median transform (e.g. Starck et al. 1995 [78]) or more empirical bandpass filters (e.g. Kaiser et al. 1995 [34]) have been experimented with success.

Wavelet decomposition Briefly, a basis of function ψ_a is derived from the analyzing wavelet ψ (in 2D):

$$\psi_a = \frac{1}{a}\psi(\frac{\boldsymbol{x}}{a}). \tag{2}$$

ψ is localized, isotropic and has zero mean. It is generally used on dyadic scales. Most popular 2D wavelets approximate the 2-dimensional "Mexican hat" $(2 - |\boldsymbol{x}|^2)e^{-|\boldsymbol{x}|^2/2}$ (e.g. Slezak et al. 1990 [76]). The wavelet transform $w(\boldsymbol{x}, a)$ of the input image vector $I(\boldsymbol{x})$ is obtained by correlating I with the ψ_a.

Object detection The detection process is roughly similar to that described in §2.1, except that background-subtraction and deblending are now implicit. At each scale a significance level is computed, based on pure Gaussian, Anscombe-corrected Poisson statistics (see Starck et al. 1998 [79]), or using Monte-Carlo methods for low-counts. Then thresholding/segmentation takes place, and a tree graph is built, that links detections from scale to scale (Fig. 1). The procedure might be concluded simply by the identification of source positions and sizes at their scales of maximum significance, as in the `Imcat` software (Kaiser et al. 1995 [34]). However to recover properly the actual structure of objects requires a more complex reconstruction procedure, as in Bijaoui & Rué's 1995 "vision model" ([12]).

This is no question that compared to the basic approach, multiscale techniques have a higher potential in deblending capabilities. Yet, they have not been used much for detection outside the domain of high-energy imaging; in particular their requirements in processing-time has been so far prohibitive for large surveys ($\approx$ 10 times slower than a "standard" extraction). Hopefully, multiscale analyses of large volumes of imaging data will be in reach of the forthcoming generation of computers.

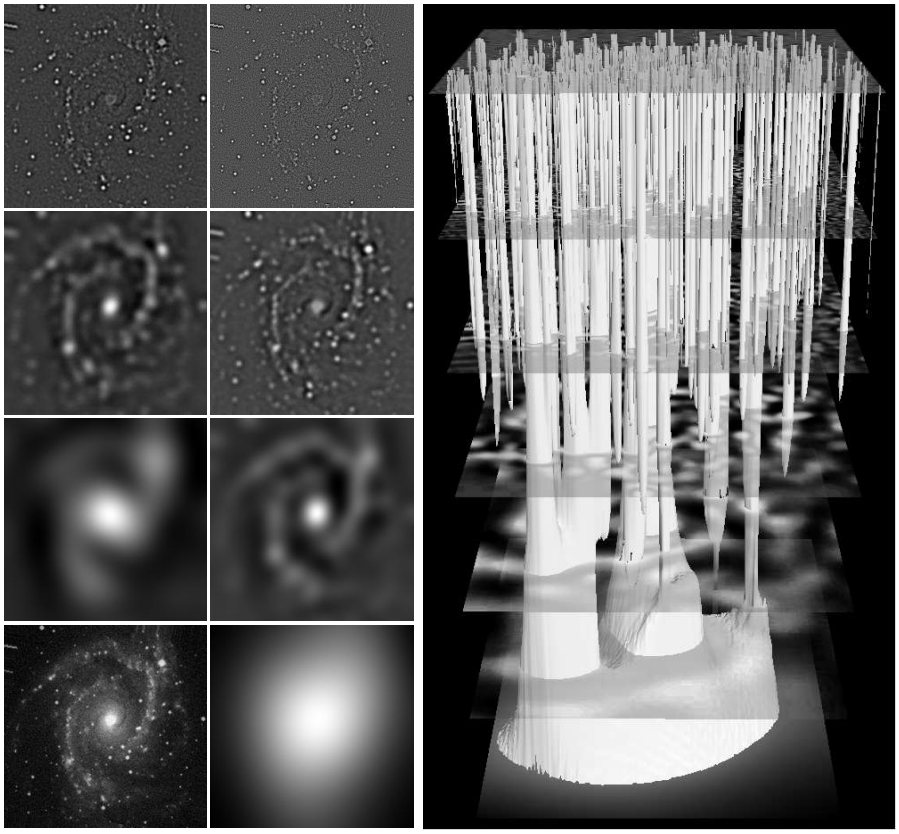

Fig. 1. *Left:* Wavelet decomposition of a photographic image of the spiral galaxy NGC 2997 (from `www.multiresolution.com`). The original image is at the bottom-left. *Right:* Detection tree derived from segmentation of the wavelet planes on the left. The signal at larger scales (bottom) is dominated by the galaxy profile, while individual stars stand out at the smaller scale (top)

.

3 The classification of astronomical images

Optimally, one would like to see classification as a part of the detection process, working directly on pixel data. This has not been possible in real-world applications up to now, because of insufficient computing power available. Classification is therefore done after detection, but this impacts on reliability for blended or low-surface-brightness objects. In the following we will more particularly discuss the problem of star-galaxy separation.

3.1 Basic approaches for star/galaxy separation

One can distinguish between two kinds of star/galaxy separation techniques. The first approach is to perform a direct comparison between the object profile and a template of the PSF through some distance measure (Carter & Godwin 1979 [17], Le Fèvre et al. 1986 [41], Lasker et al. 1990 [39], Maddox et al. 1990 [46], Yee 1991 [90], Jarrett et al. 2000 [31]). Making simple assumptions about the shape of diffuse objects, it is possible to build "by hand" a Bayesian classifier (Sebok 1979 [71], Valdes 1982 [85]).

Unfortunately these methods are easily confused by close multiple stars. The PSF-fitting can be extended to 2 or more components, but this makes the processing more computationally expensive.

A more general technique consists in partitioning a low-dimensional space defined by estimators sensitive to the shape of profiles (Herzog & Illingworth 1977 [27], Peterson et al. 1979 [65], Jarvis & Tyson 1979 [32], Kron 1980 [36], Reid & Gilmore 1982 [66], Shanks et al. 1984 [74], Le Fèvre et al. 1986 [41], Heydon-Dumbleton et al. 1989 [28], Doi et al. 1995 [22], Kaiser et al. 1995 [34]). Contrary to what one may think, elongation is poorly discriminative (McGillivray et al. 1976 [45]).

2-dimensional parameter-space diagrams, where the stellar locus can be isolated by eye, allow quick-and-dirty classification (Fig. 2). They are still in common usage in small projects. But for more reliable performance, a fully automatic, multi-dimensional analysis is required, and must be treated as a standard pattern-recognition problem.

3.2 Trainable classifiers

Generalities Trainable classifiers have two operating modes:

- a learning (or "training") mode in which a subset of data called "training set", representative of the whole sample, is used to adjust internal parameters, and
- an exploitation mode, in which the patterns fetched in input are actually classified.

The learning can be "supervised", i.e. the members of the training set are preclassified (manually of through some external mean). It can also be "unsupervised", in which case the learning procedure identifies a finite number of clusters or patterns of training set samples in input parameter-space, and each cluster or set of clusters is subsequently assigned a given class ("labeled"). With unsupervised classifiers training and exploitation may coincide.

The choice of the input parameters is important. In the ideal case, with a perfect learning tool and a training set containing an infinite number of elements, all possible parameters (for instance all the object pixels) could be fetched. However, learning algorithms are imperfect, iterative, global optimization procedures, and training sets are of limited size. Optimizing in a

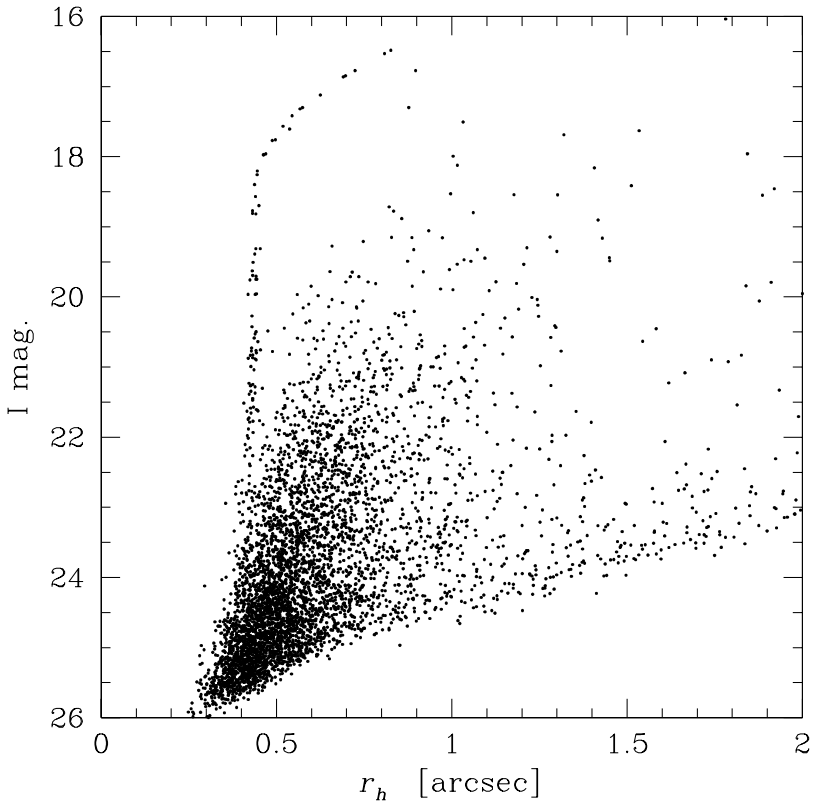

Fig. 2. A half-light radius – magnitude diagram obtained with the SExtractor software. Half-light radius is defined as the radius of the circle that includes 50% of the total flux received from the source (it is equal to half the Full Width at Half Maximum for Gaussian profiles). This kind of diagram has been popularized by Kaiser et al. (1995 [34]). The thin stellar locus is obvious on this deep CCD image with good image-quality. Note the detector saturation which enlarges the half-light radius of stellar images at magnitudes brighter than 18, and the detectability at faint magnitudes limited by surface-brightness (flux limit $\propto r_h^{-2}$).

high dimensional input-space that include many irrelevant inputs increases a lot training time, and more importantly, makes the possibility for the classifier to get trapped in local minima much more likely. The selection of relevant features can be manual, or automatic by adding a "dimensional-reduction" layer to the classifier. We will come back to this later, and will assume from now on that the classifier is provided with a limited number of relevant parameters.

Applications to astronomical image classification Automatic classifier techniques that have been used to classify astronomical images include clustering algorithms (Jarvis & Tyson 1979 [32], Connolly, this conference), nearest neighbours (Murtagh 1992a [52]), fuzzy-set theory (Spiekermann 1992 [77], Mähönen & Frantti 2000 [48]), decision trees (Weir et al. 1992,1995 [87],[88], Salzberg et al. 1995 [70], Owens et al. 1996 [63], Jarrett et al. 2000 [31], White 2000 [89]), and neural networks. Among classifiers with supervised training, decisions trees and neural networks are currently the most commonly used. In the following we will focus on applications of neural network classifiers. Decision trees are described elsewhere (e.g. Djorgovski, this conference). A specific reason driving the interest in neural networks for detection and classification of images is that one knows for sure that the biological implementations work perfectly well!

3.3 (Artificial) neural networks

Artificial neurons are cellular automata that have a very simple behavior, originally inspired from their biological counterparts. Interconnecting these automata leads to systems able to exhibit complex responses to input stimuli. In theory, some neural networks can perform arbitrarily complex mappings, provided that enough neurons are interconnected. Neural networks are able to learn, by adjusting their internal states to minimize some cost function. They are also able to "generalize" by reducing the dimensionality of the input space.

In the zoo of all the neural networks described in the literature, astronomical applications have so far essentially made use of a very small number of well-known, proven models. In the following we will focus on two of them: the Multi-Layer Perceptron (MLP), and the Self-Organizing Map (SOM).

The formal neuron . The output o of the formal neuron is a scalar function (which can be linear) of the inputs and synaptic-weights. The most widely used model is that of McCulloch & Pitts (1943):

$$o = g(\sum_j \mathbf{w}.\mathbf{x} - b), \tag{3}$$

where $\mathbf{w}$ and $\mathbf{x}$ are the weight and input vectors, respectively. g is a saturating activation function with threshold b. A convenient and popular activation function is the sigmoid:

$$g(u) = \frac{1}{1 + e^{-u}}. \tag{4}$$

As can be seen, each McCulloch & Pitts neuron basically partitions the input space in two linearly-separated half-spaces.

Network topology The most commonly-used network topologies are depicted in Fig. 3. Neurons in the Hopfield (1982 [29]) model are interconnected. Layered networks like Multi-Layer Perceptrons have no connections among neurons of a same layer. In feature-mapping networks like Kohonen's (1987 [35]) SOM, one forces a correlation between the weight vectors or neighboring neurons.

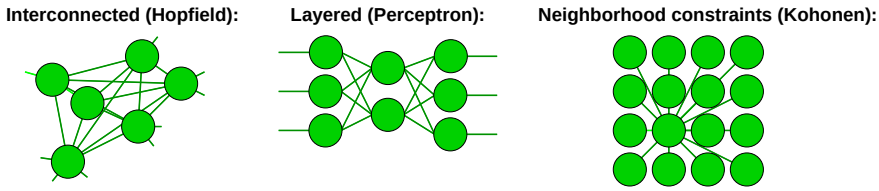

Fig. 3. Commonly-used network topologies. Hopfield's associative-memory networks are seldom employed in astronomical applications. Multi-Layer Perceptrons are mostly used for supervised learning while the Kohonen networks are designed for unsupervised feature-mapping.

.

Learning Given a "training" data set, learning proceeds through the iterative minimization of a cost function $E(\mathbf{w})$. The simplest cost function is the squared error in output:

$$E = |\mathbf{O} - \mathbf{o}|^2. \tag{5}$$

With supervised learning, the network learns to associate the provided input and output patterns (e.g. multi-layer Perceptron). With unsupervised learning, internal parameters are adjusted to minimize the differences between input patterns and the response of the network to these patterns (e.g. Vector Quantization, SOMs, Component Analysis).

The minimization of the cost function operates at the neuron level in the form of a learning rule. Many learning rules have been proposed in the literature, based on mathematical, empirical, or neurophysiological arguments. The Generalized Delta Rule (Rumelhart et al. 1986 [69]) and its derivatives applied to the MLP, have certainly been the most popular for supervised learning. The Generalized Delta Rule updates the synaptic weights by "retropropagating" through all the successive layers the gradient of the error cost function (Fig. 4a):

$$\mathbf{w}_j(t+1) = \mathbf{w}_j(t) - \eta \nabla E(\mathbf{w}) \tag{6}$$

On the side of unsupervised learning, Hebb's (1949 [26]) finding that the connection between two biological neurons linked together is strengthened when they are firing simultaneously, has been most influential. Applications of the basic model and its variations include auto-associative memory and Principal Component Analysis (PCA). Another approach, more popular

among astronomers, is that of Kohonen's Self-Organizing Maps. SOM neurons have a linear activation function: their output is just the weighted sum of inputs. SOM learning is based on competitive learning: at each training step, a "winner" is selected (the neuron whose weight vector is "closer" to the current training sample vector, in terms of the Euclidean distance), and its weight vector is updated so as to bring it closer to the training vector. The original aspect of SOM learning is that the winner's weight vector update is propagated to its neighbors. For neuron j, the weight update rules writes

$$\mathbf{w}_j(t+1) = \mathbf{w}_j(t) + h_{cj}(t)(\mathbf{x} - \mathbf{w}_c), \tag{7}$$

where c is the "winner" index, $\mathbf{x}$ the current training pattern vector, and $h_{cj}(t)$ a compact kernel with decaying width. Because of the finite range of h_{ci}, "pockets" rapidly emerge in the lattice, where groups of similar patterns can be found. These groups tend to self-organize their relative positions in such a manner that a large-scale ordering appears. The result is a lattice of weight vectors that provides a discrete, but ordered, mapping of input vector features.

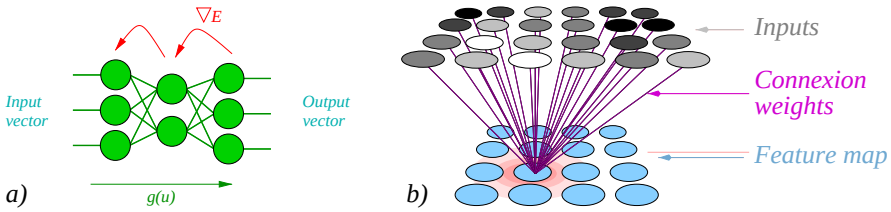

Fig. 4. *a)* Error gradient backpropagation in the Multi-Layer Perceptron. *b)* Update of the output neurons in the Self-Organizing Map.

3.4 Application to image classification

Star/galaxy separation The first application of supervised neural networks for astronomical image classification has been star/galaxy separation on digitized photographic surveys (Odewahn et al. 1992 [59], Bertin 1994 [7], Bazell & Peng 1998 [4]), Mähönen & Frantti 2000 [48]). The measured photographic plate transmission has a strongly non-linear response to incoming flux, which produces a winding decision boundary in parameter space; neural networks are particularly efficient at managing such non-trivial mappings. The neural network of choice is a single-output multilayer Perceptron fetched with images parameters, such as isophotal measurements, and trained to derive the class of the current object; point source (star) or diffuse object (galaxy). Each photographic plate has a slightly different response; to prevent tedious learning on each survey plate, one may apply a normalization to

the input parameters (Odewahn et al. 1993 ([60]) or use rank statistics (e.g. White 2000 [89]).

As illustrated in Fig. 5, one must be cautious while training the classifier. The MLP in its original version does not offer automatic increase/pruning of its degrees of freedom. A loss of performance occurs if the frontier between classes is too tortuous for the capabilities of the classifier. Conversely, too small a training sample and/or too many degrees of freedom in the classifier leads to "overlearning" (learning "by heart"). Possible selection effects in the making of the training set must be carefully balanced, particularly in the regime where classes are overlapping in parameter space. Those constraints are not specific to neural networks; they apply to all kinds of trainable classifiers. Unfortunately these aspects are often neglected, and may lead to unexpected results, which are somethimes wrongly interpreted as failures from neural networks to provide reliable results.

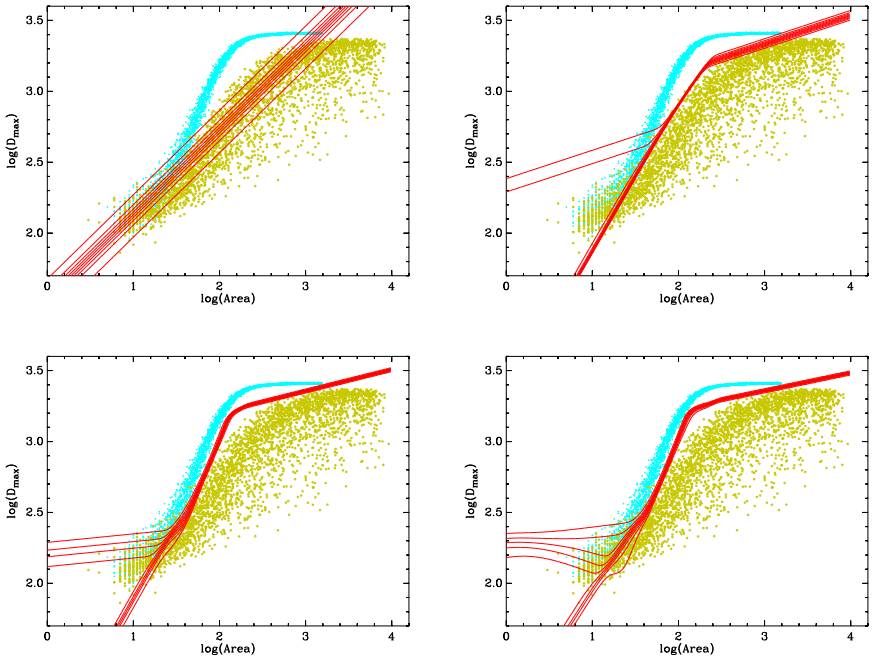

Fig. 5. Star/galaxy separation on photographic plates with an MLP. The diagrams show the distribution of simulated stars (upper cloud) and galaxies (lower cloud) in a 2D subspace of typical input parameters used for star/galaxy separation. The response of the MLP output after training is displayed as isocontours in the]0,1[range with a step of 0.1 (0=galaxy, 1=star). The number of neurons in the hidden layer determines the number of degrees of freedom of the classifier. From the upper left to the lower right: 1,2,4, and 8 hidden neurons.

The linear behavior of modern detectors brings on a simpler, more stable context for classification, as one can now define a PSF. Still, the frequent contamination by neighbors makes the trainable classifier a solution potentially superior to predefined estimators of the "fuzziness" of detections (Serra-Ricart et al. 1992,1996 [72],[73]). The bright part of the PSF in ground-based images is generally dominated by the atmospheric seeing component, which has a fairly uniform profile. To first order, and for properly sampled data, the seeing FWHM (in arcsec) is therefore the only free parameter for all image profiles. This makes it possible to train the classifier on realistic sky image simulations. This is what has been done for training SExtractor's `CLASS_STAR` neural network classifier (Bertin & Arnouts 1996 [8]). The architecture of the network is that of a classical MLP star/galaxy classifier (Fig. 6), but one of the inputs acts as a control knob that makes the neural network tuned to a particular seeing width. This input must be fetched with the actual seeing FWHM of the image. The output yields a "stellarity index" (0 for a diffuse object, 1 for a star) (Fig. 7).

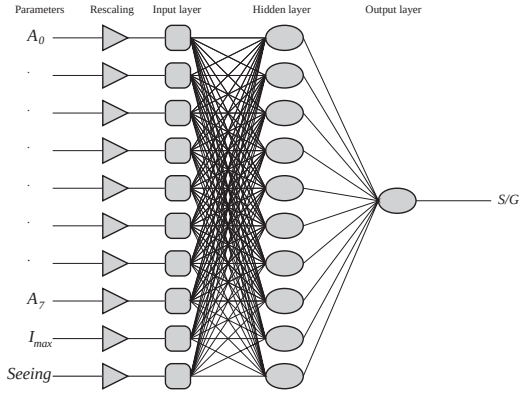

Fig. 6. SExtractor's star/galaxy neural network classifier.

Identification of image defects In astronomy like in several other experimental sciences, neural networks have been successfully applied to the identification of image defects (e.g. Murtagh 1992b [53], Rogers & Riess 1994 [67]). Instead of having to painstakingly define classification rules or algorithms to handle each kind of artifact, a simple training of a neural network on calibration or real data suffices. Still, a detection process — which may not be optimized to extract defects — and an appropriate choice of input parameters are both necessary. Fortunately, with many detectors, most defects can be identified by analysing simultaneously only a small number of pixels (e.g. cosmic ray impacts, electronic glitches, charge bleeding features).

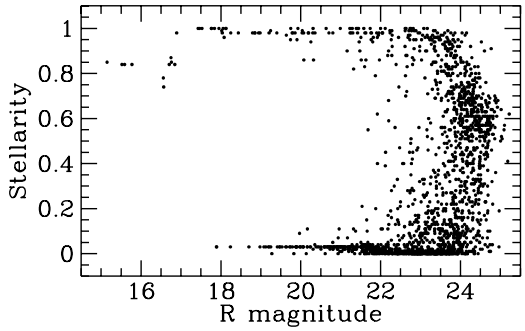

Fig. 7. SExtractor's star/galaxy classifier output as a function of magnitude for objects detected on a deep CCD image.

It is therefore possible to fetch pixel values directly to the neural network, as in an "artificial retina". A single-output MLP connected to a small compact pixel window (Fig. 8) can be trained to filter data, acting a bit like a "non-linear convolution" operator. The EIS-WIDE survey (Nonino et al. 1999 [58]) was the first to benefit from an automatic masking of small defect with an artificial retina based on an MLP, the EyE[1].

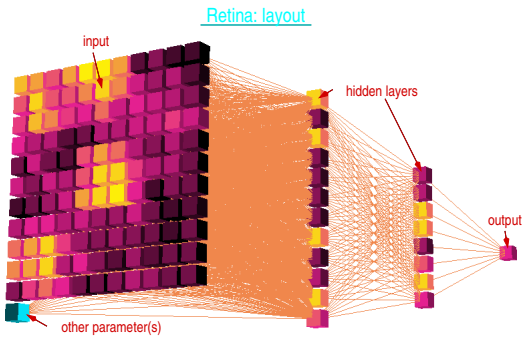

Fig. 8. Example of an artificial retina based on a Multilayer Perceptron, with a single output.

3.5 Dimensional reduction and feature extraction of pixel data

In practice, MLPs can drive directly artificial retinae with up to a few tens of pixels. Above this size, the number of input dimensions is too large to ensure proper convergence in a reasonable amount of time with existing learning

[1] `ftp://ftp.iap.fr/pub/from_users/bertin/eye`

algorithms. To allow classifying larger images, one must add a "dimensional reduction" layer that compresses the input vector.

A first approach to dimensional reduction of the input vector is to decompose it as a sum of a subset of basis vectors. Possible decompositions incluse the DCT (Discrete Cosinus Transform), on which the well-known JPEG compression standard is based (e.g. Pan et al. 2000 [64]), PCA, commonly used in astronomy for compression and classification (e.g. Lahav, this conference) and ICA (Independant Component Analysis), another linear decomposition with the additional constraint that the resulting bases are mutually independent up to higher order statistics (not only to second-order, as in PCA). ICA is more computationally intensive than PCA, and obviously has a lower compression efficiency in terms of residual *rms* error, but it provides a better separation of mixed components (e.g. Baccigalupi et al. 2000 [3] or Bijaoui, this conference). Contrarily to Principal Components, Individual Components are sensitive to localized, oriented, and bandpass-selective features when trained with "natural" image data (Bell & Sejnowski [6]), a property shared with the spatial receptive fields of simple cells in mammalian cortex.

One may also try to compress the information provided to the classifier by mapping their features. In most cases, the linear mapping techniques mentioned above provide a rather inefficient representation of the subspace defined by all possible input patterns. The purpose of feature-mapping is to provide a non-linear mapping of the input space preserving as much as possible the topological relations between the training set patterns. Feature-mapping capabilities are an essential property of Self-Organizing Maps. Several image classification experiments with SOMs have been carried out so far, either directly from pixel data (Mähönen & Hakala 1995 [47], Bertin 1998 [9], Bringer & Boër 2000 [15]), or more conservatively on measured object parameters (Miller & Coe 1996 [51], Naim et al. 1997 [54], Mähönen & Frantti 2000 [48]). Yet none of them has been used in survey production. Although SOMs perform rather well for unsupervised classification on detected sources, in practice it is not really efficient as a dimensional reduction layer for detection: the limited number of nodes makes it rather inappropriate to handle at reasonable processing speed the richness of patterns found at random locations in an image.

3.6 What's next?

A first practical implementation of a two-stage, fully neural detector, NExt, is described in Andreon et al. (2000 [2], see also Longo et al. , this conference). The second layer is a neural-gas neural network (unsupervised classifier) that separates background from object pixels. The first layer performs a non-linear PCA on a window of 3×3 or 5×5 pixels. Not surprisingly the 3 first components return a smoothed version of the (compressed) input image, and the derivatives along y and x, respectively. Fed with these 3 components, the second network cannot perform significantly better than a

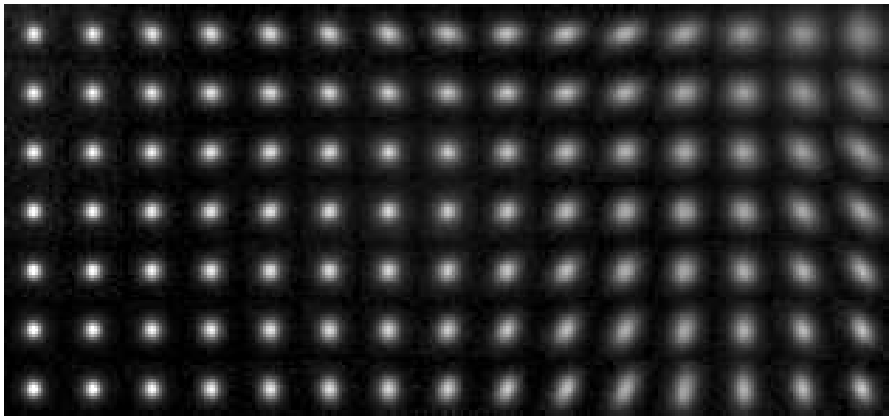

Fig. 9. Two-dimensional Self-Organizing Map applied to flux-normalized detections in a deep I-band image. The postage stamps show the weight-vector associated to each node at the end of training, coded as grey-level images. Note the locus of spheroids (center), of point-sources (lower left), and the "malicious" location of saturated stars among face-on disk galaxies (upper right)

simple matched filter+thresholding combination; hence the application may seem a little academic. Nevertheless NExt is certainly a first step towards more intelligent and more robust detection routines, that are able to adapt themselves to the quality of the data.

Finally, a short note on the current situation in terms of software development. It is a fact that among imaging survey developers, the current priority is rather on the side of image processing (co-addition, image subtraction, separation of components, homogenization of image quality), measurement (morphological parameters, weak-lensing measurement), and data handling (information retrieval, catalog-data mining), most of which are still the limiting factor for the scientific return of the final data. In parallel, as we have seen, several extraction techniques have been developed during the past years. However, extraction tools based on simple, standard algorithms like DAOPhot, FOCAS or SExtractor are still the most popular. This will probably continue until new software shows evident superiority and equivalent speed and stability.

References

1. Alard C. (2000) in "The Impact of Large-Scale Surveys on Pulsating Star Research", ASP Conference Series **203**, 50
2. Andreon S., Gargiulo G., Longo G., Tagliaferri R., Capuano N. (2000) MNRAS **319**, 700
3. Baccigalupi C., Bedini L., Burigana C., De Zotti G., Farusi A., Maino D., Maris M., Perrotta F., Salerno E., Toffolatti L., Tonazzini A. (2000) MNRAS **318**, 769

4. Bazell D., Peng Y. (1998) ApJS **116**, 47
5. Beard S.M., McGillivray H.T., Thanisch P.F. (1990) MNRAS **247**, 311
6. Bell A.J., Sejnowski T.J. (1996) in "Advances in Neural Information Processing Systems" **9**, MIT press
7. Bertin E. (1994) Ap&SS **217**, 49
8. Bertin E., Arnouts S. (1996) A&A **117**, 393
9. Bertin E. (1998) in "The Impact of Near-IR Sky Surveys on Galactic and Extragalactic Astronomy", Ed. N. Epchtein, Kluwer, 267
10. Bijaoui A. (1980) A&A **84**, 81
11. Bijaoui A. (1994) in "Astronomy from wide-field imaging", Eds. H.T. McGillivray, E.B. Thomson, B.M. Lasker, I.N. Reid, D.F. Malin, R.M. West and H. Lorenz, IAU Symposium **161**, Kluwer, 213
12. Bijaoui A., Rué F. (1995) Signal Processing **46**, 229
13. Bonnet H., Mellier Y. (1994) A&A **303**, 331
14. Bothun G., Impey C., McGaugh S. (1997) PASP **109**, 745
15. Bringer M.,Boër M. (2000) in "Astronomical Data Analysis Software and Systems IX", ASP Conf. Series **216**, 640
16. Buonanno R., Buscema G., Corsi C.E., Ferraro I., Iannicola G. (1983) A&A **126**, 278
17. Carter, D., Godwin J.G. (1979) MNRAS **187**, 711
18. Coupinot G., Hecquet J., Aurière M., Futaully, R. (1992) A&A **259**, 701
19. Dalcanton J.J. (1996) ApJ **466**, 92
20. Damiani F., Maggio A., Micela G., Sciortino S. (1997) ApJ **483**, 350
21. Deutschman W.A. (1971) in "Automation in Optical Astrophysics (IAU Colloq. 11)", Eds. H. Seddon, M.J. Smyth, Pub. Roy. Obs. Edinburgh **8**, 194
22. Doi M., Fukugita M., Okamura S. (1995) ApJS **97**, 59
23. Fellgett P.B. (1970) Optics Technology **2**, 61
24. Grebenev S.A., Forman W., Jones C., Murray S. (1995) ApJ **445**, 607
25. Hall P., Mackay C.D. (1984) MNRAS **210**, 979
26. Hebb D.O. (1949) "The Organization of Behavior: A neuropsychological theory", Wiley
27. Herzog A.D., Illingworth G. (1977) ApJS **33**, 55
28. Heydon-Dumbleton N.H., Collins C.A., MacGillivray H.T. (1989) MNRAS **238**, 379
29. Hopfield J.J. (1982) in Proceedings of the National Academy of Sciences, 2554
30. Irwin M.J. (1985) MNRAS **214**, 575
31. Jarrett T.H., Chester T., Cutri R., Schneider S., Skrutskie M., Huchra J.P. (2000) AJ **119**, 2498
32. Jarvis J.F., Tyson J.A. (1979) in "Instrumentation in Astronomy III", SPIE **172**, 422
33. Jarvis J.F., Tyson J.A. (1981) AJ **86**, 476
34. Kaiser N., Squires G., Broadhurst T. (1995) ApJ **449**, 460
35. Kohonen T. (1987) "Self-Organization and Associative Memory", 2nd edition, Springer
36. Kron R.G. (1980) ApJS **43**, 305
37. La Bonte A.E. (1970) in "Proper Motions" (IAU Colloq. **7**), Ed. W.J.Luyten, University of Minnesota, Minneapolis, 26
38. Lahav O., Naim A., Buta R.J., Corwin H.G., de Vaucouleurs G., Dressler A., Huchra J.P. (1995) Science **267**, 859

39. Lasker B.M., Sturch C.R., McLean B.J., Russell J.J., Jenker H., Shara M.M. (1990) AJ **99**, 2019
40. Lazzati D., Campana S., Rosati P., Panzera M.R., Tagliaferri, G. (1999) ApJ **524**, 414
41. Le Fèvre O., Bijaoui A., Mathez G., Picat J.P., Lelièvre G. (1986) A&A **154**, 92
42. Lupton R.H., Gunn J.E. (1986) AJ **91**, 317
43. Lutz R.K. (1979) The Computer Journal **23**, 262
44. Luyten W.J. (1974) Science **18**, 351
45. MacGillivray H.T., Martin R., Pratt N.M., Reddish V.C., Seddon H., Alexander L.W.G., Walker G.S., Williams P.R. (1976) MNRAS **176**, 265
46. Maddox S.J., Efstathiou G., Sutherland W.J., Loveday J. (1990) MNRAS **243**, 692
47. Mähönen P., Hakala P.J. (1995) ApJ **452**, 77
48. Mähönen P., Frantti T. (2000) ApJ **541**, 261
49. Mighell K.J. (1989) MNRAS **238**, 807
50. Mighell K.J. (1999) in "Astronomical Data Analysis Software and Systems VIII", ASP Conf. Series **172**, 317
51. Miller A.S, Coe M.J. (1996) MNRAS **279**, 293
52. Murtagh F.D. (1992a) in "Astronomical Data Analysis Software and Systems I", ASP Conf. Series **25**, 265
53. Murtagh F.D. (1992b) in "Data Analysis in Astronomy", 103
54. Naim A., Ratnatunga K.U., Griffiths R.E. (1997) ApJS **11**, 357
55. Newell B., O'Neil,Jr E.J. (1977) PASP **89**, 925
56. Newell B. (1979) in "International Workshop on Image Processing in Astronomy", Eds. G. Sedmak, M. Capaccioli, R.J. Allen., Osservatorio Astronomico di Trieste, Trieste, 100
57. Newell B. (1983) in "Workshop on Astronomical Measuring Machines", Eds. R.S. Stobie, B. McInness, ROE, Edimburgh, 15
58. Nonino M, Bertin E., da Costa L., Deul E., Erben T., Olsen L., Prandoni I., Scodeggio M., Wicenec A., Wichmann R., Benoist C., Freudling W., Guarnieri M.D., Hook I., Hook R., Mendez R., Savaglio S., Silva D., Slijkhuis R. (1999) A&AS **137**, 51
59. Odewahn S.C., Stockwell E.B., Pennington R.L., Humphreys R.M., Zumach W.A. (1992) AJ **103**, 318
60. Odewahn S.C., Humphreys R.M., Aldering G., Thurmes P. (1993) PASP **105**, 1354
61. Odewahn S.C. (1995) PASP **107**, 770
62. Odewahn S.C., Windhorst R.A., Driver S.P., Keel W.C. (1996) ApJ **472L**, 13
63. Owens E.A., Griffiths R.E., Ratnatunga K.U. (1996) MNRAS **281**, 153
64. Pan Z., Rust A.G., Bolouri H. (2000) in "Proc. of The IEEE-INNS-ENNS International Joint Conf. on Neural Networks", vol. **3**, 149
65. Peterson B.A., Ellis R.S., Kibblewhite E.J., Bridgeland M.T., Hooley T., Horne D. (1979) ApJ **233**, L109
66. Reid N., Gilmore G. (1982) MNRAS **201**, 73
67. Rogers R.D.,Riess A.G. (1994) PASP **106**, 532
68. Rosati P., Burg R., Giacconi R. (1994) in "The Soft X-ray Cosmos", Eds. E.M. Schlegel,R Petre, AIP Conf. Proceedings **313**, 260
69. Rumelhart D.E., Hinton G.E., Williams R.J. (1986) Nature **323**, 533

70. Salzberg S., Chandar R., Ford H., Murthy S.K., White R. (1995) PASP **107**, 279
71. Sebok W.L. (1979) AJ **84** 1526
72. Serra-Ricart M., Gaitan V.,Delgado S., Perez-Fournon I. (1992) in "Astronomical Data Analysis Software and Systems I", ASP Conf. Series **25**, 254
73. Serra-Ricart M., Gaitan V., Garrido L., Perez-Fournon I., 1996, A&AS **115**, 195
74. Shanks T., Stevenson P.R.F., Fong R., McGillivray H.T. (1984) MNRAS **206**,767
75. Slezak E., Bijaoui A., Mars G. (1988) A&A **201**, 9
76. Slezak E., Bijaoui A., Mars G. (1990) A&A **227**, 301
77. Spiekermann G. (1992) AJ **103**, 2102
78. Starck J.-L., Murtagh F., Bijaoui A. (1995) in "Astronomical Data Analysis Software and Systems IV", ASP Conf. Series **77**.
79. Starck J.-L., Murtagh F., Bijaoui A. (1998) Image Processing and Data Analysis, Cambridge University Press
80. Starck J.-L., Aussel H., Elbaz D., Fadda D.,Cesarsky C. (1999) A&A **138**, 365
81. Starck J.-L., Bijaoui A., Valtchanov I., Murtagh F. (2000) A&AS **147**, 139
82. Stetson P.B. (1987) PASP **99**, 191
83. Stobie R.S., Smith G.M., Lutz R.K., Martin R. (1979) in "International Workshop on Image Processing in Astronomy", Eds. G. Sedmak, M. Capaccioli, R.J. Allen., Osservatorio Astronomico di Trieste, Trieste, 48
84. Stoy R.H. (1970) in "Proper Motions" (IAU Colloq. **7)**,Ed. W. J. Luyten, University of Minnesota, Minneapolis, 48
85. Valdes F. (1982) in "Instrumentation in Astronomy IV", SPIE **331**, 465
86. Vikhlinin A., Forman W., Jones C., Murray S. (1995) ApJ **451**, 542
87. Weir N., Fayyad U.M., Djorgovski S., Roden J.C., Rouquette N. (1993) in "Astronomical Data Analysis Software and Systems II", ASP Conf. Series **52**, 39
88. Weir N., Fayyad U.M., Djorgovski S. (1995) 1995 AJ **109**, 2401
89. White R.L. (2000) in "Astronomical Data Analysis Software and Systems IX", ASP Conf. Series **216**, 577
90. Yee H.K.C. (1991) PASP **103**, 396