Exp Astron (2006) 21:151–168
DOI 10.1007/s10686-007-9079-0

ORIGINAL ARTICLE

Removing cosmic-ray hits from CCD images in real-time mode by means of an artificial neural network

Wacław Waniak

Received: 5 November 2006 / Accepted: 14 August 2007 /
Published online: 29 September 2007

Abstract A feed-forward artificial neural network has been implemented to the problem of removing cosmic-ray hits (CRH) from CCD images. The results of a number of tests demonstrate the effectiveness of this method especially for undersampled stellar profiles. The problem of optimal and low price preparing of training data, which could enable real-time or at least fast post-processing filtering out of CRH is discussed. The training and test ensembles were composed of a number of synthetic stellar profiles involving different S/N ratios and CRH images taken from real data. Certain aspects of the network's architecture and its training efficiency for different modes of the back-propagation procedure as well as for the pre-process normalization of data have been examined. It is shown that for training set composed of stellar images and CRH at a ratio of 1:2 recognition can reach 99% in the case of stars and 96% for CRH. To determine the extent to which the cognition power of a network trained using an ensemble of circular symmetric stellar profiles of a given radius can be generalised the test data included stellar profiles of different radii, as well as elongated profiles. The goal was to mimic temporal changes in seeing as well as such problems as image defocusing, the lack of isoplanatism and improper sideral tracking of a telescope. The experiments provided us with the conclusion that for S/N>10 excellent classification property is maintained in cases where the change in the radius of a circular profile is up to 30%, as well as for elongated profiles where the longest dimension is almost double that of the shortest one. Moreover, the generalization capability has been investigated for test images of synthetic pairs of overlapping stars with different distances between components. Almost 99% recognition efficiency was achieved even if the separation was nearly three times the radius of the stellar profile, a case when two stars could be analyzed by appropriate software as separate objects. The example of removal of CRH from real CCD images is presented to give an idea of how an algorithm based on a neural network can work in practice. The result of such an experiment appears fully consistent with the conclusions drawn from the tests made on synthetic data.

Keywords Image processing · Stars/cosmic-ray hits discrimination · Artificial neural networks

W. Waniak (✉)
Astronomical Observatory, Jagiellonian University, ul. Orla 171, 30-244 Cracow, Poland
e-mail: waniak@oa.uj.edu.pl

1 Introduction

Algorithms based on the cognitive capabilities of an artificial neural network have been used for a dozen or so years to solve as many problems of observational astrophysics as possible. The main fields where such methods can be applied are the following: object recognition, separating different classes of items, prediction regarding the behaviour of cosmic objects or observational equipment. Beginning with galaxy/star discrimination in sky survey images (see e.g.[1, 16, 17]), neural networks can be used to provide a correct analysis of stellar spectra (see e.g. [7, 10, 23], examine the light curves of variable stars (see e.g. [3, 6, 22]) and make morphological classification of galaxies (see e.g. [2, 11, 21]).

The problem of discriminating between cosmic-ray hits (CRH) and images of point-like or almost point-like objects recorded on CCD frames is of crucial importance both for ensuring accurate results of bulk photometry of stellar fields as well as for deconvolution of astronomical images of faint objects. Approaches to solving this question are generally based either on image arithmetic using multiple co-aligned photographs of the same sky region or on a comparative parametric analysis of individual CRH and objects of interest. The situation deteriorates when only a few exposures of a given region can be utilized, the exposure time increases or when stellar profiles are sampled using an insufficient number of detector pixels. The latter two problems were characteristic of the Hubble Space Telescope (HST) wide field/planetary camera (WFPC), hence a number of papers have been written on the problem of the statistical properties of CRH registered by this instrument, as well as on the methods of their removal (see e.g. [9, 15, 24]).

Visual inspection and flagging offer a far more certain but unfortunately less effective and, in addition quite subjective approach to discriminating between small point-like sources and CRH recorded in astronomical images. The natural way was modifying this method to become faster and more objective, while at the same retaining its flexibility and recognition properties. The obvious solution was to apply of an artificial neural network as a simple but still fairly "smart" model of human cognition. [15] implemented such an approach to the CRH discrimination in HST WFPC frames. He chose as the input data to be analyzed by the network a number of parameters outlining individual items, parameters produced by photometry packages such as DAOPHOT, FOCAS, INVENTORY and so on. [5] also described an automated neural network approach to recognising stars, CRH and noisy picks on CCD images.

To ensure effective working of an artificial neural network (i.e., number of false alarms is reduced to a minimum) it is vital to prepare adequate training data, select the suitable net architecture and provide effective training. The aim of the present work is to demonstrate how a proper training set of stellar profiles and CRH images can be prepared, and illustrate how tolerant a neural network can be in the case of quite substantial alterations in analyzed point-spread-function (PSF) profiles. Such an approach makes it possible to implement suitably taught network for treatment of astronomical images obtained under changing seeing, focussing, imperfect telescope tracking etc.

The issues concerning training data, net architecture and the learning phase will be discussed in the next section. The following section presents the results of a number of experiments, which demonstrate that the approach ensures a high degree of effectiveness and recognition capability. The generalization properties of the network were also analyzed for stellar profiles that differ quite significantly from those presented to the net during the training phase. Section 4 provides sample results of the operation of the star/CRH classifier on real CCD images of stellar fields. The final section presents a number of conclusions summarizing the results.

2 The training set, neural network and learning procedure

The most plausible way to implement the artificial neural network for star/CRH separation procedures would be to use a unique universal set of neuron weights, which would make it possible to properly reduce data gathered under different observing conditions. Thus, we could reject CRH from astronomical images in real-time mode (immediately after exposure and before writing to storage memory) without the need for repetitive network training.

A list of factors, which can alter the profile of a stellar image during an observation night is so long, that I will only enumerate here variable seeing, defocusing, aberrations involved by flexure of the optical parts or bending of the mechanical structure of a telescope, and imperfect sideral tracking. At least the first factor depends on the effective wave length of the photometric band used. Where the optical path ensures perfect, stable focussing and negligible aberrations, total PSF is almost entirely controlled by seeing. Figure 1 illustrates such a case by showing the correlation between atmospheric mass and the full width at half maximum (FWHM) radius of the stellar profile referred to the R band. For two nights with almost excellent photometric conditions a series of 10 min exposures was obtained using a 0.5-m telescope of the Astronomical Observatory of the Jagiellonian University equipped with thermoelectrically cooled Photometrics CCD with 24 μm square pixels. The mean scale was 0.736 arcsec/pixel. The FWHM radius of PSF was derived by taking 20 stellar profiles evenly situated on the frames. Any trace of anisoplanatism was noticed. The general trend of the relation presented in Fig. 1 can be easily reproduced by a simple model of atmospheric turbulence with Kolmogorov distribution. Unfortunately, short term variations in the seeing existed. They had a typical time scale of 0.5 h and an amplitude comparable to the mean level of seeing itself. This circumstance suggests that the appropriate way to train the neural network should be to select trial data from each individual image such as, e.g., its sub-frame containing a reasonable number of exemplary profiles. Obviously, the neural network should be relearned by referring to each exposure separately. Such an approach is extremely time consuming. On the other hand, CRH profiles recorded on astronomical frames determine the individual property of the CCD detector. They depend on

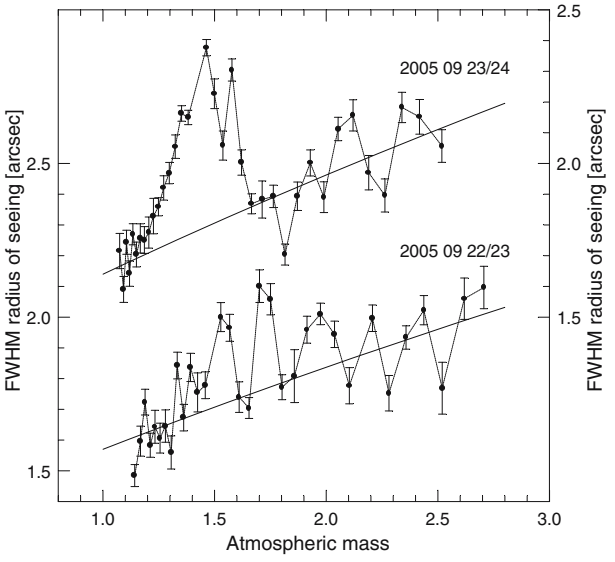

Fig. 1 The relationship between the radius of the seeing disc for the R filter and the atmospheric mass for arbitrarily selected nights of September 2005. *Smooth continuous lines*—a simple model of atmospheric turbulence occurring in two layers (the ground level layer surrounding a telescope and the other one at an altitude of 3 km). The *left Y axis* is valid for the *lower plot*, and the *right Y axis* for the *upper plot*

the pixel size and the architecture, as well as on its quantum efficiency. An intrapixel distribution of response (see e.g. [12]) can play an important role, especially if the charge from a given pixel leaks to adjacent ones, thereby broadening the CRH profile.

For the purpose of examining the discriminative capability and generalization property of the neural network, both, training and testing ensembles used, consisting of synthetic stellar profiles and real CRH patterns taken from the CCD frames. Such an approach is quite natural, as it would be difficult (although not impossible) to create a synthetic CRH or select PSF profiles with an assumed width and S/N ratio from real data. The CRH profiles were retrieved from dozens of "dark" frames obtained using integration times equal to 50, 250 and 1,000 s, respectively. After the bias subtraction the average "dark" frame was computed from multiple images using the iterative $k\sigma$ approach to avoid any CRH influence. The mean "dark" was subtracted from the individual frames. Next, the CRH were sought automatically using a cut-off level equal to $n\sigma$ of the noise referred to local annular vicinity of each pixel. When there was a positive detection, the position of the local maximum brightness was retrieved and taken as the new center of the annulus. The position of the photo centre of the detected CRH's was computed after the local background was subtracted. The coordinates of this center become the midpoint of the small square pattern that was used as a training object. Taking into account the stellar profiles and CRH observed in the real data I concluded that 15×15 pixel pattern was quite acceptable. Due to the local character of the searching procedure the results are negligibly dependant on the global behaviour of the sky background and tolerant to weak extended sources.

Although the influence of hot pixels should have been negligible once the mean "darks" are subtracted from the "dark" images, the intense dark signal resulted from insufficient cooling makes this supposition doubtful. Moreover, if the $n\sigma$ cut-off level is too low it can dangerously enhance a number of one-pixel noisy events, which mimics single-point CRH occurrences. With the aim of testing which value of n is acceptable I have analyzed the correlation between the time of integration and the number of CRH found for different values of an n multiplier. The result was visualized in Fig. 2. When the level of detection

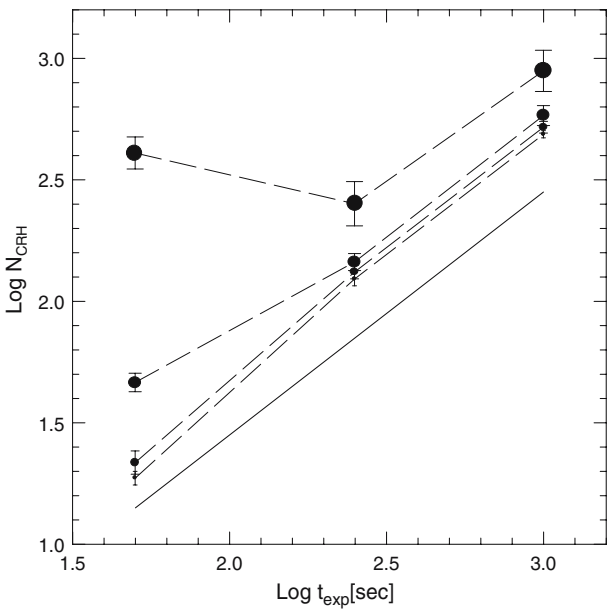

Fig. 2 The logarithmic dependency between the exposure time of the "dark" frames and the number of detected CRH. The cut-off level which was used by detecting software rises from 4σ (the *largest circles*) by 1σ up to 7σ (the *smallest circles*). By way of comparison, the linear dependency is represented by the *thick solid line*

was underestimated and the integration time short the number of CRH clearly increases in comparison with the expected value. Thus, I accepted a 6σ cut-off level as quite optimal.

Figure 3 presents the logarithmic distribution of equivalent radii of CRH detected in the "dark" frames obtained at different exposure times. The equivalent radius of a given pattern was defined as the radius of a cylindrical profile containing the total signal of this pattern and an elevation equal to the maximum observed value. I think that this quantity more adequately determines the effective size of CRH than the FWHM measure. The latter parameter is difficult to determine for irregular objects of a few points extend. As one can see, the results for 50, 250 and 1,000 s exposure duration gave similar results, whereas 4,000 s exposures are characterised by a relatively smaller amount of larger CRH. This phenomenon can result from the destruction of cluster events by substantial noise in dark signal. The profile of a CRH would be broken-up into a number of single peaks. The statistic of all detected CRH shows that their mean equivalent radius was equal to 0.7 pix, whereas the maximum value was close to 1.5.

Stellar profiles were synthesized using the Moffat function [14]. Although originally created to model stellar profiles recorded on astronomical photographic plates it appears to be a quite acceptable tool in modern CCD photometry of stars. As was shown by [19] the Moffat profile with a β exponent equal to 4 adequately describes the brightness distribution of a point-like object produced by seeing. A similar conclusion can be drawn from our observations. The discrepancies between theoretical and actual brightness distributions are of the order of a few percent. For the reader's convenience I have given here the conversion rate between equivalent and FWHM radii valid for the Moffat profile:

$$r_{equ} \approx 1.327 r_{FWHM}$$

Stellar images of different "brightness" were placed on a constant background and biased by Poisson noise. The appropriate value of S/N ratio (valid for the maximum signal in a profile) was achieved by normalising the Moffat profile. I took into account the

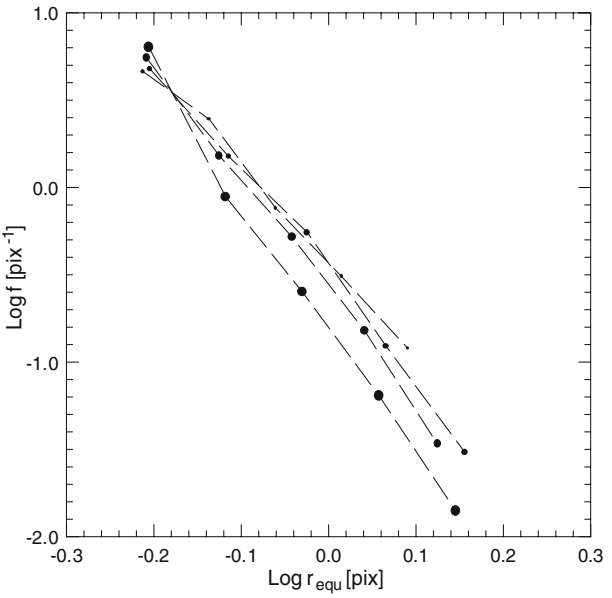

Fig. 3 The logarithmic plot of the distribution of the equivalent radii of detected CRH. The *filled circles* of increasing size show the results for the exposure time of "dark" frames equal to 50, 250, 1,000 and 4,000 s, respectively

pixelation effect for CCD (see, e.g., [4]), which forms the registered brightness distribution of the stellar image, especially if substantial undersampling takes place. Thus, theoretical profiles were randomly positioned onto a net of pixel centers and integrated within the borders of each pixel assuming constant intrapixel sensitivity. The latter assumption however may not be valid. The results given by [9] and [12] provide us with a clear but weaker boost of quantum efficiency toward the center of a pixel as well as with a slight overflow of the charge beyond the pixel boundary. Frames with synthetic, noisy biased stellar images were analyzed using a detection algorithm based on the $n\sigma$ cut-off principle, but this time the cut-off level was applied accordingly to the S/N ratio in stellar profiles. The 15×15 pixels sub-frames, each containing a stellar pattern, were used as training data.

In practice, a properly prepared training set composed of real stellar profiles is almost as easy to achieve as in the case of CRH data. All that is needed are short time exposures of the stellar field. To avoid CRH any classical multi-frame approach can be utilized and then any star detection procedure should be implemented. Two sets of exemplary profiles, one with CRH and one with star images, as well as a fragment of the training ensemble composed of randomly shuffled sub-frames containing both classes of objects, are presented in Fig. 4. The goal was to give the reader some idea about training and testing data. Thus, it tends to show the cluster effects of cosmic ray hits rather than trivial one-pixel events.

The architecture of the actual neural network makes use of a single hidden layer, which appears to be sufficient enough to solve the problem of discrimination between objects of two classes and ensures little time consuming during learning and operating phases. The input layer contains 225 neurons and the output layer has only one neuron, which provides the decision whether a stellar profile or CRH has been identified. Signals from pixels belonging to the training and testing sub-frames were presented to input neurons starting from the top left corner along the x axis and successively row by row. As was shown by [1] the way in which signals from consecutive pixels are presented to the input layer has little impact on the operating properties of a neural network. On the other hand, adequate

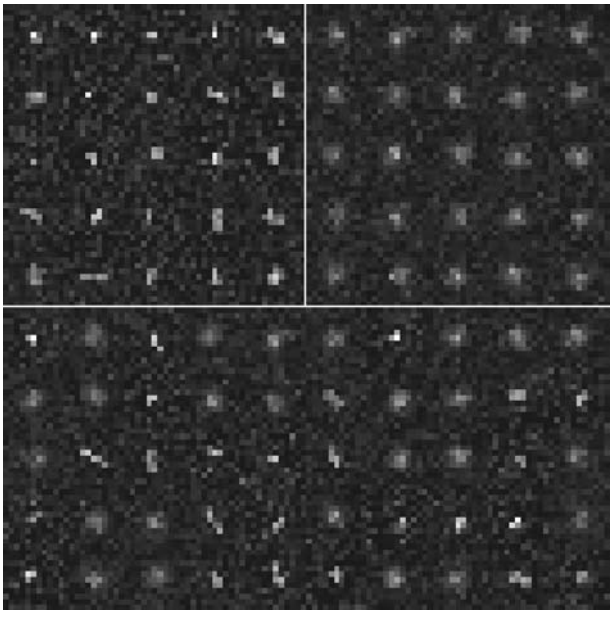

Fig. 4 Example images of CRH detected in "dark" frames (*top left panel*) and synthetic stellar profiles with an FWHM radius of 1.761 pix and an S/N ratio close to 6 (*top right panel*). The *bottom part* presents randomly shuffled images of 25 CRH and 25 stellar profiles with the same parameters as described above. Much the same frames were used during the learning phase. The objects visible in the *lower panel* are not the same as those depicted in the upper two ones

normalization of the input data can be of crucial importance. I applied normalization referred to the total signal observed in a given sub-frame of the training set. All the neurons have a logistic activation function. Although I tested the network with linear activation for the first two layers of neurons the results I obtained were decidedly worse.

According to [13], the number of neurons in the hidden layer should be close to the square root of the product of the numbers of the input and output units. To determine adequate quantity, neural networks with 2, 5, 15, 30 and 60 neurons respectively were tested. The training ensemble consists of 200 CRH (taken from the "dark" frames with different exposure times to enable various S/N ratios) and 100 stellar profiles with S/N ranging from 4 to 200 and an equivalent radius equal to 1.761 pix (a bit higher than the maximum for CRH). It was presented to the network 100 times and the weights of the neurons were updated using a back propagation algorithm [20]. Testing set contained 200 CRH and 200 star images. Figure 5 (bottom part) shows the recognition efficiency of these networks as a fraction of false alarms for training and testing data. Error bars visible in this image and in the following ones present dispersion originating from seven independent training phases, each starting from different, randomly initiated weights. As could be expected, along with growing number of neurons in hidden layer, discriminative capability increases both for training ensemble and for testing stellar profiles. In the case of CRH, however, 30 neurons appears to be quite a good number, because if the quantity is smaller the network is not so "smart" and if the number is too large it can lose its generalization capability. A similar situation take a place when the neural network is taught using an increasing number of presentations of the training ensemble (see Fig. 5 top part). The fraction of recognition errors decreases for training data and grows for test data containing both CRH and stars. This is a typical example of overtraining. Thus, further on I used 255-30-1 network architecture and 100 iterations during the training phase.

For the kind of neural network and training/testing data described above certain modifications of the training algorithm were examined. These included tests using different but constant or changing during training phase gain of weights actualisation, different but

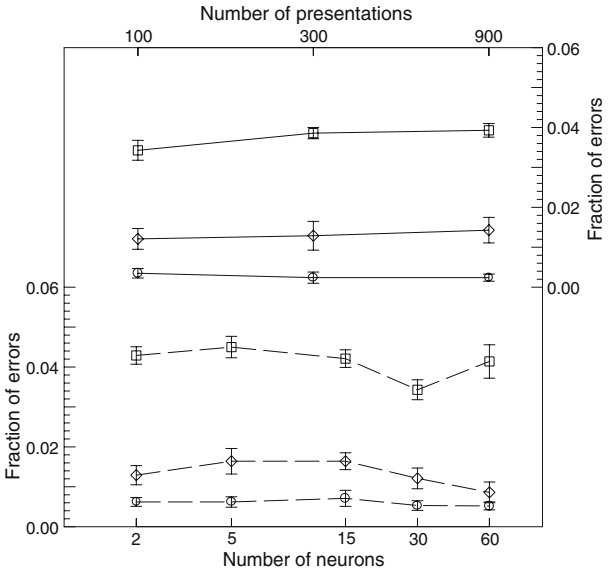

Fig. 5 *Bottom part*—the relationship between the number of neurons in the hidden layer and the efficiency of recognition expressed by the fraction of classification errors is depicted by the *dashed lines* referred to the *bottom X* and *left Y axis*. The *circles* represent the result for the training set, the *diamonds* correspond to the test ensemble of stellar images, whereas the *squares* depict the network efficiency for the test set of CRH. *Top part*—three plots (*solid lines*) relating to the *top X* and *right Y axis* display the correlation between the number of iterations in the learning process and the recognition capability of a network with 30 neurons in the hidden layer. The symbols have the same meaning as previously

constant versus varying momentum values (see e.g. [18]), as well as simulated annealing with descending temperature (see e.g.[8]). Anticipated enhancement of the training efficiency or the recognition capability of the network were very merely visible comparing with the results given by the pure unmodified algorithm. What is more, some drawbacks appeared. Accordingly with global ideas of implemented modifications rise in gain induced faster but less stable solutions, while an increase in momentum gave more stable solutions but less effective training and poorer discriminative capability. Simulated annealing resulted in slightly better learning efficiency and fewer false alarms for the training data (solution easily abandons local minima of network energy), but caused clearly inferior recognition for test data. Finally, I accepted as the quite optimal approach a constant actualization gain equal to 1 and a training algorithm without momentum and simulated annealing.

The next item that should be tested is the method used to normalize the input data. All the tests described above were conducted for an input signal normalized to total brightness. The method using normalization to the maximum value in a given sub-frame resulted in a distinct improvement in discriminative efficiency for CRH but in a disastrous deterioration in results for stellar images. So, I decided to use the former type of normalization in further tests.

One other important question to consider is how the results of the operation of the neural network depend on the relative content of CRH and stellar profiles in the training ensemble. Intuition suggests that better generalization properties can be achieved if CRH are more numerous than star images. A CRH profile is quite arbitrary, as it can be a single-point event or a cluster with varying prolateness and irregular brightness distribution. The trial results presented in Fig. 6 fully confirm such a presumption. The number of false alarms for the training set do not depend on its relative content, which is understandable. The recognition efficiency for the test objects of a given class increases with their abundance in the training ensemble, but a proper balance between them can be easily deduced. In our case the optimal situation arises when the ratio of CRH to stellar profiles stands at 2:1. In this case, the percentage of false diagnoses should be close to 1 for stars and less than 4 in the case of CRH. Taking different relative abundances of training objects for both classes

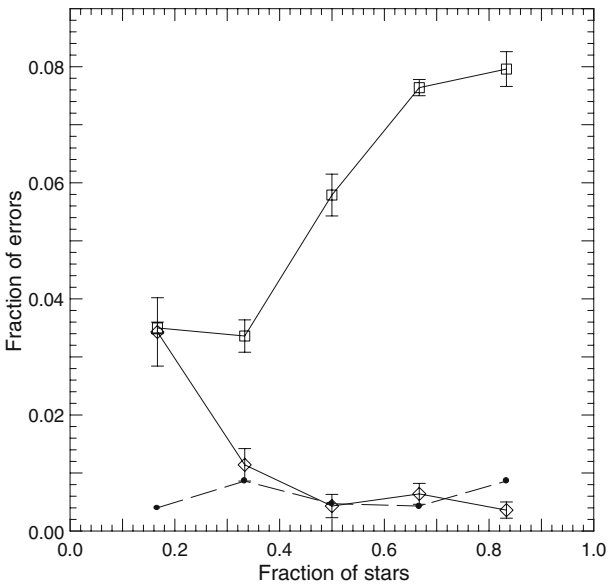

Fig. 6 The recognition efficiency of the network (as represented by the fraction of false alarms) versus the relative abundance of stellar images and CRH in the training and test ensembles. The *dots* represent the result for the training set, while the *open diamonds* and *open squares* correspond to the test sets composed purely of stellar images and CRH, respectively

we can tune the neural network along with our requirements. It can be made more restrictive for CRH if our aim is to reject as many hits as possible while allowing a simultaneous loss of stars. On the other hand, the network can be made more sensitive to stellar profiles if we want to minimize drop of their number.

3 Generalization ability of the network

At the beginning of the previous section it was indicated that the global character of CRH profiles is specific for a given CCD chip. On the other hand, the brightness profile of point-like source can vary from one position to another on a CCD frame (lack of isoplanatism due to optical aberrations, curved focal plane) or from one time moment to another (unstable seeing conditions, the bending of the optical parts or the mechanical structure of a telescope). Having a neural network with an appropriate architecture, an adequate method for preparing training data and optimal training algorithm, as well as knowing the general behaviour of the network (see previous section) we can start to examine its generalization capability.

Firstly, I focused on unstable seeing, which can change the FWHM radius of the PSF, but not the course of a profile. Much the same situation takes place in the case of weak defocusing and moderate curvature of the focal plane. As was the case previously the training ensemble contained CRH patterns and stellar profiles with S/N ranging from 4 to 200, but this time the FWHM radius of the profiles was equal to 0.754, 1.000, 1.327 and 1.761 pix, respectively. That means equivalent radii changing from 1.0 (a bit higher than the mean value for CRH) up to 2.337 (one and a half the largest value obtained for CRH) According to the conclusion made at the end of the previous chapter, the relative content of CRH and star images was 200:100 for the two largest FWHM radii and 150:150 for the two smallest ones. The training ensemble was presented to the network 100 times for seven sets of randomly initiated values of neuron weights. Thus, we had indeed seven neural networks each of them trained by the same data but with different starting points. This procedure enables us to quote a mean result, which is weakly dependent on the starting values of the weights. The standard deviation was generally a few times smaller than the mean result itself.

The network then operated on the test data of 200 patterns of CRH. The result is shown in the top panel of Fig. 7 by filled-squares connected by a thick line. As can be expected in the situation of increasing undersampling of the stellar profiles (a declining FWHM radius) the discriminative capability decreases, but not so markedly. The fraction of false alarms varies from 3 to 10%. Next, the neural network trained by stellar images with a given FWHM radius was used to analyze the number of trial ensembles, each prepared using profiles of selected radii (close or equal to the value utilized during the learning phase) and S/N ratio decreased from 30 to 10, 6, 5 and then to 4. A number of conclusions can be drawn from this experiment (see Fig. 7).

For decreasing S/N ratio and PSF equivalent radius approaching the mean value for CRH (a case of serious undersampling) the neural network completely loses its recognition capability. However, this is quite usual. The achievement of a proper level of discrimination in such circumstances would be a miracle. Additional tests that have been carried out using students as “animated” neural networks have produced quite similar results.

On the other hand, for S/N ratio better than 5 entirely good recognition can be achieved even for less plausible cases involving the narrowest possible stellar profiles. As the PSF radius increases, the network begins to operate effectively for lower S/N ratios. Moreover, discriminatory power is not strictly limited to stellar profiles that have radii such as the profiles used for training. Generally, if the S/N ratio increases, the degree of tolerance

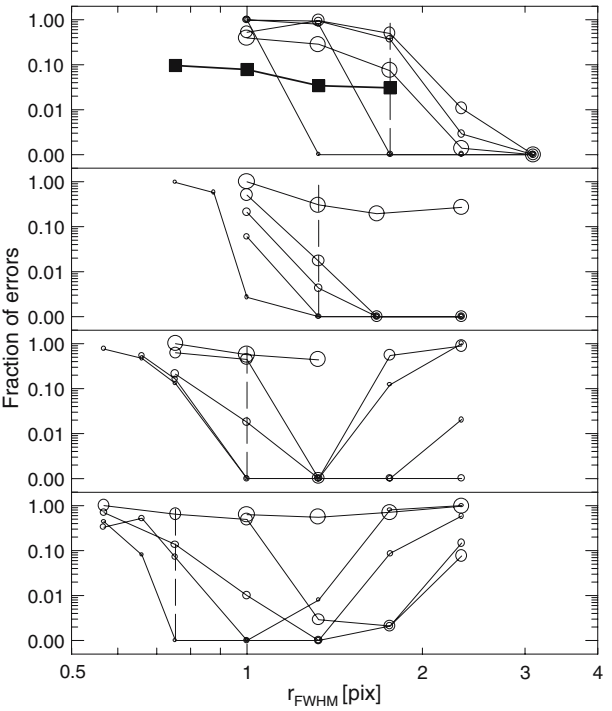

Fig. 7 The relationship between the fraction of recognition errors and the FWHM radius of the profile of stellar images contained in the test ensembles. Four cases are presented, each for a specific FWHM radius of the profile used during the training phase (this value was indicated by a *vertical dashed line*). The radii of the *circles* increase with the descending value of the S/N ratio. They are equal to 30, 10, 6, 5 and 4, respectively. In addition the *top panel* presents (by *filled squares* connected by a *thick line*) the correlation between the proportion of false alarms for CRH and the FWHM radius of stellar profiles belonging to the training set. For the sake of clarity, error bars showing the dispersion of the results produced by different sets of initial weights of neurons have not been drawn. They appeared to be roughly comparable to the symbols depicted

exhibited by the network grows, but unfortunately in different ways for both "sides" of the value of the profile radius involved during the training phase. In the case of narrower profiles contained in the test ensembles, generalization is less explicit, especially for low S/N ratios. On the other hand, wider profiles imply quite good recognition capability for a major range of radii even for such an S/N ratio. This asymmetry becomes more distinct if the equivalent radius of stellar profiles used as training data approaches the mean value for CRH.

To sum up, neural networks trained by the PSF profiles of a given radius accurately recognize such objects as well as slightly narrower images, but have a tendency to prefer profiles of markedly larger radii. Thus, it is obvious that a neural network trained once using a representative sample of stellar profiles should effectively discriminate between CRH and stars even when changes in seeing approach 30%, especially if the seeing radius increases.

The behaviour of the network seems to resemble, to some extent, human cognition of stars and CRH in astronomical images. In the case of properly sampled stellar patterns with a sufficient S/N ratio, their photometric structure is easily recognized as being stellar. By contrast, if the profile radius drops to values characteristic of CRH and, additionally, the S/N ratio declines, the chances of properly classifying an object decrease. Moreover, if CRH and the stellar profiles of different radii are recognised manually, the narrower profiles would more frequently be assumed to be an effect of cosmic rays.

Next, I looked at the tolerance of the neural network towards a limited degree of asymmetry appearing in the PSF profile The network was trained on the basis of an ensemble of 200 CRH and 100 circular stellar profiles with FWHM radii equal to 1.327 pix. It was then applied to the test sets, each one containing elongated stellar images of assumed ellipticity. This parameter was defined as the ratio of the biggest to the smallest FWHM widths. As the narrowest cross-section of the profile was kept unchanged, increasing

ellipticity involved an increase in the equivalent radius. Although the aim of such an approach is to imitate PSF deformation through inaccurate sideral tracking of a telescope, to make the test more global, the directions of the longest axis of the profiles were randomised. Additional tests, not presented here, showed that the behaviour of the neural network is statistically independent on the direction. This was not a surprise, as the training stellar profiles were characterised by circular symmetry.

Figure 8 presents the discriminative property of the network when ellipticity increases from 1 to almost 3 and the S/N ratio is equal to 30, 10, 6, 5 and 4. Both cases with minimal noise biasing yield perfect recognition independently of the elongation of the test profiles. S/N ratios of 5 and 6 induce better results for increasing ellipticity. This is quite normal when we consider that it implies an upward movement in the equivalent radius. The most extreme case of noise biasing shows additional loss of discriminative capability. Here the structure of the brightness distribution is seriously impaired by noise which results in a decrease in the equivalent radius (actual, not assumed). The course of this test offers evidence that the neural network worked as a cut-off filter that identified objects with appropriate equivalent radii.

The question now arises of how the above result will change if an increase in the elongation of the stellar profile is not paralleled by increase in the equivalent radius. In real observations such a situation is charactertic of astigmatic aberration for astronomical optics, which can result in a lack of isoplanatism. Thus, when preparing test ensembles, elliptic patterns were synthesized in such way that the shortest and longest FWHM widths gave assumed ellipticity, which increased from 1 to 3 and thereby simultaneously provided a constant equivalent radius equal to 1.327 pix. The effect is shown in Fig. 9 as the percentage of recognition errors correlated with the elongation measure and S/N ratio. The actual plot is quite different from the previous one. Perfect recognition is achieved only for

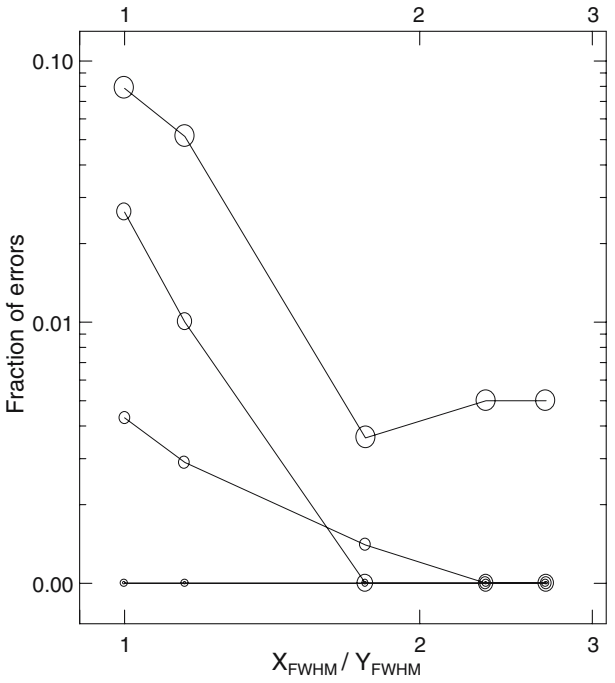

Fig. 8 Discriminative efficiency represented by the proportion of false alarms as a function of the elongation of the profile of stellar images. The FWHM width has remained unchanged, whereas the FWHM length has increased. The radii of *open circles* rise as the S/N ratio diminishes from 30 through 10, 6, 5 down to 4. The training ensemble was composed of circular symmetric stellar profiles where r_{FWHM}=1.327 pix

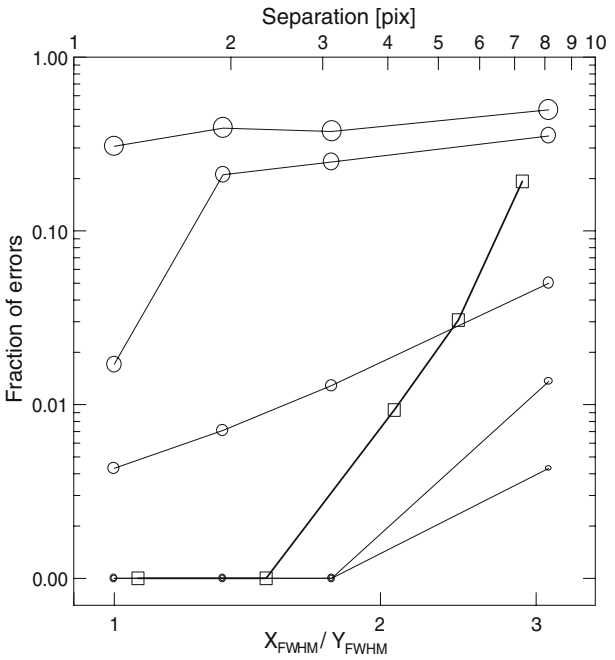

Fig. 9 The correlation between classification efficiency and elongation of the profile of stellar images (bottom X axis) is depicted by *open circles* of radii that increase as the S/N ratio diminishes from 30 through 10, 6, 5 down to 4. The PSF profiles were broadened in one direction but narrowed in the perpendicular one to keep their equivalent radii unchanged. The training ensemble contained circular symmetric stellar profiles of an effective radius comparable with the radii of the elongated profiles (r_{FWHM}= 1.327 pix). A *thick solid line* joining the *open squares* represents the result for double stellar profiles of given separation (*top X axis*). Pairs were oriented arbitrarily and the S/N ratio was greater than 4. As was the case previously, the training set contained circular symmetric stellar profiles with an FWHM radius of 1.327 pix

the most limited noise impact and for an ellipticity smaller than 2. An S/N ratio equal to 6 is a transient case, beyond which even an almost round profile does not assert proper discrimination. Such behaviour is evidence that the network is sensitive to brightness distribution inside the object to be classified.

One other test might explain whether the fairly extensive possibility of generalization described above would ensure that the neural network operated properly on real images. In this case a network trained using single patterns should be used to analyse multiple stellar profiles overlapping each other. Unfortunately, parameters describing such profiles (the number of adjacent stars, their relative brightness, positions and S/N ratio) are so numerous that careful testing of all the possibilities would be quite a tremendous task. I have carried out a highly restricted analysis here that looks only at double stellar profiles as trial objects. In general, if the number of merged components increases, the equivalent radius of the resultant pattern grows and its photometric profile becomes more symmetric compared with a doublet or triplet. A similar situation can be encountered in a case of atmospheric seeing, that produces a transient pattern involving a couple of speckles if a telescope with a moderate aperture is used (slightly exceeding the Fried parameter). After a long exposure time (longer than the characteristic time for frozen atmospheric turbulence) consecutive patterns overlap and PSF controlled by seeing finally becomes more regular with a width much greater than each individual speckle. Tests yet conducted showed that in such a situation a neural network should discriminate multi profile objects quite well.

As a parameter which could be settled, separation between centers of stars was accepted. Further parameters; position angle of components and their relative brightness, were randomised. The S/N ratio of individual profiles (also randomly selected) varied between 4 and 100. An FWHM radius was equal to 1.327. The result of the experiment is presented in Fig. 9 as a thick line connecting open squares. The separation between components started from a measure comparable to the FWHM radius of PSF profile. Thus, both merged

patterns created a single profile with only one maximum. If the separation is less than 2 pix almost perfect recognition can be achieved. This is quite normal, considering that such a situation resembles the case with elongated stellar profiles. Recognition level deteriorates as the distance increases, but even with a separation of 4–5 pix quite good results are achieved. Distinctly poor discrimination in extreme cases of 6–7 pix distance between centers of patterns should not be a cause for concern, because in such a situation stars become almost completely resolved and can be treated as separate objects by an appropriate software.

All the tests described above provide us with a fairly accurate diagnosis of the behaviour of the neural network used to identify stellar profiles and CRH in astronomical images. Recognition capability is achieved partially by analyzing the global characteristics of objects such as equivalent width and the asymmetry of their profiles and in part, by using information based on the pixel-to-pixel distribution of signal, as e.g. gradient, texture and so forth. Because of the variety of parameters representing a given object, the neural network has generalization properties and widespread tolerance with regard to unexpected classes of patterns. Of course, we should bear in mind that this is the case with objects whose shapes are radically uncomplicated and, as a consequence, a network with extremely simplified architecture can recognize them fairly well.

4 Experiments with real data

A number of tests described in the previous section shows that an optimally designed neural network is capable of properly distinguish between patterns produced by cosmic-ray impacts on CCDs and regular stellar profiles resulting from seeing, optical defects and inaccurate telescope tracking. The margin of satisfactory recognition efficiency appeared to be broad enough to ensure that proper treatment of images obtained in varying conditions by a network trained only once. All the tests have been performed using data that have been partly synthesized mathematically. The CRH profiles were taken from real exposed images, whereas the stellar profiles were purely theoretical, yet quite close to the ones derived from observations.

Finally it is time to confront the neural network based algorithm of image filtering with purely real data. Such an approach has one obvious advantage in that it can provide an answer to the question of how effective the algorithm is in practice. On the other hand, the most serious drawback to the method is that it is extremely difficult to control certain parameters of training and testing patterns. Moreover, achieving suitable ranging for the shapes and sizes of stellar profiles is almost impossible.

First of all I treated the R-filtered image of a fragment of open cluster NGC 7790 using a neural network trained on data containing real CRH profiles taken, as previously, from a number of “dark” auxiliary frames. On this occasion, however, the stellar profiles were extracted by the detecting software from the selected sub-frame of the treated image.

The architecture of the network and the learning procedure were chosen optimal taking into account conclusions drawn in the previous sections. Figure 10 presents the result of the filtering out of CRH. The image depicted in the top panel of this figure is only a fragment of the whole frame, fragment void of training stars. The frame used in this experiment was a stack of ten individual photographs each exposed during 30 s. Due to such co-addition a number of likely appeared CRH rises. On the other hand, with a couple of photographs of the same field it is relatively easy to remove patterns resulting from CR impacts. By comparing the outputs from the neural network and the median filtering I was able to determine that the efficiency of star/CRH separation in the former approach fully conforms

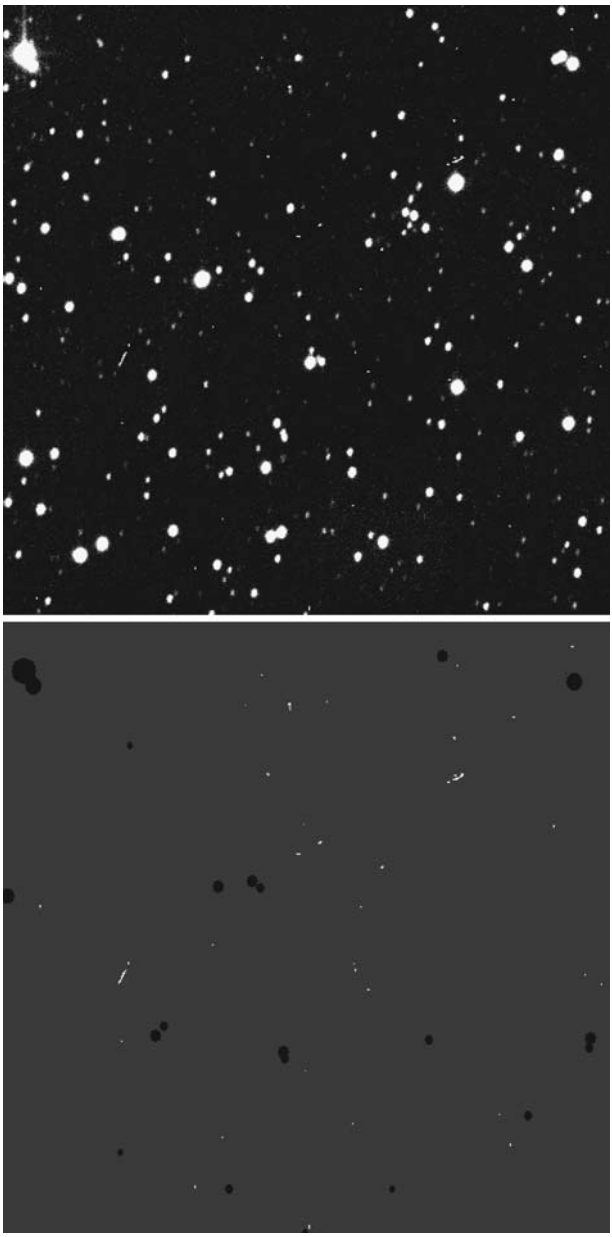

Fig. 10 The result of filtering out of CRH by means of a neural network algorithm. The *upper frame* (450×450 pix) shows a 5.5×5.5 arcmin photograph of the region of NGC7790 open cluster. Ten R filtered exposures, each lasting 30 s, have been co-added. The *lower panel* shows the result of the treatment procedure. The images of correctly recognized CRH have been imposed on a flat grey background. The *black circles* represent incorrectly classified stellar images. The radii of such circles increase as the stellar brightness ascends

to the results presented in Fig. 7. A detailed analysis of all individual cases of false alarms (stellar profiles detected as CRH) revealed that the most frequent reason for their occurrence was the over-exposure or substantial weakness of a star, the overlapping of a couple of profiles and the close vicinity of the brightest stars. Such behaviour in the case of the neural network classifier is quite understandable.

One test which might prove even more valuable than the one previously presented would be to investigate how exactly the neural network algorithm separated stars and CRH if a

long series of images obtained under varying conditions is treated by the net trained using a highly restricted trial sample. I returned to the data from the night of 23/24 September 2005 (see Fig. 1 and the description in the second section). The top six photographs, which have seeing radii close to one other, were used to establish the training profiles of stellar images. CRH patterns were obtained as previously. Moreover, the network architecture and learning process were the same as in the previous case.

Figure 11 (filled symbols connected by solid lines) shows the extent to which results of image processing depend on the seeing radius. The behaviour of the neural network was described using a correlation between air mass and the number of detected CRH as well as the number of erroneous recognition acts. The fraction of false alarms, widely used variable in the previous sections, was not presented here, as the total number of stars recorded on the frames is constant. Moreover, the main goal was to examine the dependency between recognition capability and seeing but not to present the value of the former parameter itself. A comparison with Fig. 1 shows that recognition efficiency is merely dependent on the seeing radius, although the last parameter varies by about 40 percentage points. Such a conclusion should come as no surprise bearing in mind the discussion presented in the third section together with Fig. 7.

Finally, I adopted one more approach to the problem of recognising stellar images and CRH. In the past a number of trials have been carried out using either synthetic or real stellar profiles during both the learning and testing phases. In section 2 I pointed out that the Moffat function with a β parameter equal to 4 can reproduce the brightness distribution of seeing-induced PSF almost perfectly. If this is the case, it would be uniquely interesting to learn the neural network using synthetic stellar profiles and process the real data with it. Figure 11 (open symbols) presents the result of such an experiment. The first six frames obtained under the smallest and almost constant seeing were used to determine the mean stretch of the Moffat function. Such a synthetic profile was utilized to produce training data. Patterns with noise-biased images of "stars" were generated by taking into account the pixelation effect. Next, the trained network processed all the real images giving an outcome almost merely differing from the result of the previous experiment. What is more, this

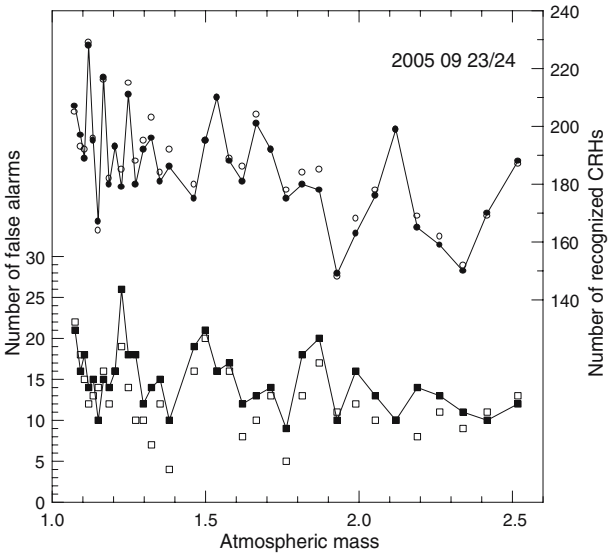

Fig. 11 Recognition efficiency achieved for real images versus atmospheric mass. For each frame of the night 2005 09 23/24 the figure shows the number of detected CRH (*top plot*) as well as the number of stars erroneously identified as CRH (*bottom plot*). The *filled symbols* connected by a *solid line* show the result when a real profile of PSF has been used, while the *open symbols* represent a correlation obtained for a synthetic PSF

approach indicates slightly better behaviour for the neural network. The number of recognized CRH increased on average while the amount of erroneous star detections systematically declined. This apparent paradox is understandable when we notice that using a unique average synthetic profile instead of a number of real ones biased by individual effects ensures a higher degree of network generalization.

The result of the last experiment suggests a new way of processing of astronomical images in real-time mode. Due to the fairly flexible functioning of the network a number of sets of neuron weights can be prepared beforehand, each for a given seeing radius. To carry out this task all we need are the learning sets containing both synthetic stellar profiles referring to given PSF and CRH profiles typical for a specific CCD model, profiles derived from auxiliary frames. It can not be ruled out that for specific class of PSF the architecture of the network, the learning method as well as the number of presentations used during the training phase may have to be adjusted to make the recognition process more effective (cf. the second section). Once we have determined the radius of the PSF disk characteristic of a given night an appropriate set of weights of neurons can be chosen. Then real time processing of consecutive exposed frames can be carried out.

5 Conclusions

The present paper dealt with the problem of removing CRH from CCD frames, a problem which appears to be of crucial importance for processing of astronomical images, especially in the case of undersampling of the PSF. An artificial neural network with one hidden layer was implemented to resolve this issue. A number of experiments were carried out to select the correct number of neurons acting in the hidden layer, the relative abundance of CRH and stellar profiles in the training data, as well as the learning mode. The image patterns produced by cosmic ray hits were taken from auxiliary "dark" frames, whereas the circular stellar profiles were synthesized using a Moffat function.

The input layer of the network had 225 neurons. Each of them took a signal from one of 225 pixels belonging to the 15×15 square sub-frame, including a profile of the object of interest. I adopted as the optimal approach, i.e. one offering superior recognition effectiveness and simultaneously a less time-consuming, a network with 255-30-1 architecture trained using the back-propagation method void of modifications, which appeared to be less effective. The relative abundance of CRH and stars in the training data, at a ratio of 2:1, as well as normalization based on a total signal of the profile ensured that the network achieved optimal behaviour. As a consequence, separation efficiency reaches 99% for stars and 96% for CRH.

A series of tests was carried out to determine whether it was possible to use one specific set of weights of neurons to reduce images obtained with a changing seeing radius. A neural network trained on data composed of the CRH and stellar profiles of given radius was used to process trial data with star patterns of different radii and S/N ratios. As might have been expected recognition efficiency increases when the S/N was higher and for a value markedly greater than 10 a fairly good result was achieved for radii ranging from about 0.7 up to 2 times the radius adopted when preparing the training data. In addition, elongated profiles as well as synthetic "double stars" were used as test objects and each time the generalization value of the network was very high, especially when the S/N ratio was itself high. Even when separation between components was three times the radius of the stellar profile almost 99% recognition efficiency was obtained. Such a distance between stars ensures that they can be properly analysed as separate objects. All the tests revealed

that the network behaved satisfactorily on synthetic stellar profiles and remained quite flexible in relation to variable seeing conditions, optical distortions, defocusing and imprecise telescope tracking.

The final step I took to confirm the usefulness of the proposed method in practice was to train and test the network using real astronomical images as a source of stellar profiles. Recognition efficiency appeared to be comparable with the values derived from the test based on synthetic stellar images. Moreover, no correlation was observed between the output of processing a series of real astronomical images and the variable radius of the seeing disc. A slightly better result was achieved by a network trained on a theoretical stellar profile involving actual seeing. This means that it is possible to establish a quite simple and effective algorithm for filtering out CRH from photographs. The neural network trained once on CRH taken from auxiliary "dark" frames together with stellar profiles synthesized to match the PSF profile of a radius similar to (but not necessarily the same as) the actual value can almost perfectly reduce a series of images obtained under moderately varying observing conditions.

References

1. Bazell, D., Peng, Y.: A comparison of neural network algorithms and preprocessing methods for star–galaxy discrimination. Astrophys. J. Suppl. Ser. **116**, 47–55 (1998)
2. Bazell, D., Miller, D.J.: Class discovery in galaxy classification. Astron. J. **618**, 723–732 (2005)
3. Belokurov, V., Evans, N.W., Le Du, Y.: Light-curve classification in massive variability surveys—I. Microlensing. Mon. Not. R. Astron. Soc. **341**, 1373–1384 (2003)
4. Buonanno, R., Iannicola, G.: Stellar photometry with big pixels. Publ. Astron. Soc. Pac. **101**, 294–301 (1989)
5. El-Bassuny Alawy, A., et al.: CCD image identification: An artificial neural approach. In: HIPPARCOS and the Luminosity Calibration of the Nearer Stars, 24th meeting of the IAU, Joint Discussion 13, Manchester, August 2000
6. Feeney, S.M., Belokurov, V., Evans, N.W., An, J., Hewett, P.C., Bode, M., Darnley, M., Kerins, E., Baillon, P., Carr, B.J., Paulin-Henriksson, S., Gould, A.: Automated detection of classical novae with neural networks. Astron. J. **130**, 84–94 (2005)
7. Gulati, R.K., Gupta, R., Rao, N.K.: A comparison of synthetic and observed spectra for G-K dwarfs using artificial neural networks. Astron. Astrophys. **322**, 933–937 (1997)
8. Hinton, G.E., Sejnowski, T.J.: Learning and relearning in Boltzmann machines. In: Rumelhart, D.E., McClelland, J.L. (eds.) Parallel Distributed Processing: Exploration in the Microstructure of Cognition, pp. 282–317. MIT, Cambridge (1986)
9. Holtzman, J., Hester, J.J., Casertano, S., Trauger, J.T., Watson, A.M., Ballester, G.E., Burrows, C.J., Clarke, J.T., Crisp, D., Evans, R.W., Gallagher III, J.S., Griffiths, R.E., Hoessel, J.G., Matthews, L.D., Mould, J.R., Scowen, P.A., Stapelfeldt, K.R., Westphal, J.A.: The performance and calibration of WFPC2 on the Hubble Space Telescope. Publ. Astron. Soc. Pac. **107**, 156–178 (1995)
10. Klusch, M., Napiwotzki, R.: HNS: a hybrid neural system and its use for the classification of stars. Astron. Astrophys. **276**, 309–319 (1993)
11. Lahav, O., Naim, A., Sodré, Jr, L., Storrie-Lombardi, M.C.: Neural computation as a tool for galaxy classification: methods and examples. Mon. Not. R. Astron. Soc. **283**, 207–221 (1996)
12. Lauer, T.R.: The photometry of undersampled point-spread functions. Publ. Astron. Soc. Pac. **111**, 1434–1443 (1999)
13. Martens, T.: Practical neural networks recipes in C^{++}. Academic, New York (1993)
14. Moffat, A.F.J.: A theoretical investigation of focal stellar images in the photographic emulsion and application to photographic photometry. Astron. Astrophys. **3**, 455–461 (1969)
15. Murtagh, F.D.: Cosmic ray discrimination on HST/PC images: Object recognition-by-example. In: Worrall, D.M., Biemesderfer, C., Barnes, J. (eds.) Astronomical Data Analysis Software and Systems I. ASP Conf. Ser. 25, pp. 265–273 (1992)
16. Odewahn, S.C., Stockwell, E.B., Pennington, R.L., Humphreys, R.M., Zumach, W.A.: Automated star/galaxy discrimination with neural networks. Astron. J. **103**, 318–331 (1992)

17. Philip, N.S., Wadadekar, Y., Kembhavi, A., Joseph, K.B.: A difference boosting neural network for automated star–galaxy classification. Astron. Astrophys. **385**, 1119–1126 (2002)
18. Plaut, D., Nowlan, S., Hinton, G.: Experiments on learning by back propagation. Technical Report CMU-CS-86-126, Department of Computer Science, Carnegie Mellon University, Pittsburgh (1986)
19. Racine, R.: The telescopic point-spread function. Publ. Astron. Soc. Pac. **108**, 699–705 (1996)
20. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. Nature **323**(9), 533–536 (1986)
21. Storrie-Lombardi, M.C., Lahav, O., Sodré, Jr, L., Storrie-Lombardi, L.J.: Morphological classification of galaxies by artificial neural networks. Mon. Not. R. Astron. Soc. **259**, 8–12 (1992) (Short Communication)
22. Tagliaferri, R., Ciaramella, A., Milano, L., Barone, F., Longo, G.: Spectral analysis of stellar light curves by means of neural networks. Astron. Astrophys. Suppl. Ser. **137**, 391–405 (1999)
23. Willemsen, P.G., Hilker, M., Kayser, A., Bailer-Jones, C.A.L.: Analysis of medium resolution spectra by automated methods—application to M55 and ω Centauri. Astron. Astrophys. **436**, 379–390 (2005)
24. Windhorst, R.A., Franklin, B.E., Neuschaefer, L.W.: Removing cosmic-ray hits from multiorbit HST wide field camera images. Publ. Astron. Soc. Pac. **106**, 798–806 (1994)