Pattern Analysis & Applications (2002)5:15–22

Finding Hidden Events in Astrophysical Data using PCA and Mixture of Gaussians Clustering

Maria Funaro[1,2], Maria Marinaro[1,2,3], Alfredo Petrosino[2,3] and Silvia Scarpetta[1,2]

[1]Dipartimento di Fisica 'E. R. Caianiello', University of Salerno, Baronissi (SA), Italy; [2]INFM – University of Salerno, Baronissi (SA), Italy; [3]IIASS, Vietri sul Mare (SA), Italy

Abstract: The Principal Component Analysis (PCA) is applied to a set of astronomic data to obtain a separation between variations of luminosity and noisy fluctuations. A clustering with the Mixture of Gaussians method, performed in the principal subspace, allows us to classify the data according to the features of interest. Our results are compared with those obtained by the AGAPE (Andromeda Galaxy and Amplified Pixels Experiment) collaboration.

Keywords: Astrophysical data; Classification; Data analysis; Mixture of Gaussians; Pixel lensing; Principal Component Analysis

1. Introduction

Einstein's theory of general relativity provides for light deflection by gravitational field. This has the consequence that the image of the light source becomes split or deformed due to the presence of a gravitational mass (lens) between the source and the observer. In general, the images of the light source are multiple (standard gravitational lensing); when the angular separation is very small so that the images cannot be resolved, we speak of microlensing [1,2].

Many collaborations (EROS [3], OGLE [4]], MACHO [5,6], DUO [7], AGAPE [8,9], Columbia/VATT [10]) are involved in microlensing, because it is a powerful method to detect dark matter present in the universe: when a compact object with substellar mass passes near the line of sight of a star, the luminosity of this star will be varied. So, the analysis of luminosity variations in astronomical images is an important topic, and it requires the implementation of computational techniques which need be fast and accurate.

In particular, a new technique, the so-called pixel lensing, has been proposed [11–14] for the study of microlensing events. The essence of this technique is to register the luminosity of a whole group of stars on a single pixel element of a CCD camera. Within this framework, the availability to distinguish luminosity variations due to a physical event from noisy fluctuations is in high demand in real time astronomical data analysis.

The aim of this paper is to present a neural network-based method to perform a pre-selection of 'interesting' pixels, i.e. pixels whose luminosity variation could be related to physical events, and not to a noisy fluctuation. The method proposed combines the Principal Component Analysis (PCA) [15,16] of the luminosity temporal series, which are commonly unevenly sampled and are characterised by a low signal-to-noise ratio, and clustering realised by the Mixtures of Gaussians (MOG) technique [15]. Specifically, the clustering is applied in the principal subspace obtained by the PCA analysis. This strategy makes us able to classify the data according to the features we are interested in. In particular, from the analysis of the eigenvalues spectrum, we note that the dominant structure is determined by the first four eigenvectors. Of these, the first one is not useful to reach our aim, being related to variations of luminosity present in the data essentially due to resolved stars, i.e. stars which sharply emerge from the luminosity background[1]. On the contrary, the other three eigenvectors allow a fine separation of the interesting luminosity variations from noise.

This method is applied to a set of pixel-lensing data toward M31, taken at the MDM observatory, in order to

Received: 22 December 2000
Received in revised form: 26 March 2001
Accepted: 20 April 2001

[1] The resolved stars are characterised in luminosity, magnitude and colour with respect to the background.

do a comparison with the results obtained by the AGAPE [8,9] collaboration. We show that our approach has the same efficiency as the AGAPE method for the pre-selection of interesting pixels.

2. METHOD DESCRIPTION

It is our aim to uncover as much as possible about the unknown state $\mathbf{s}_t$ of the generator system (the sky) from the observations which have been made through a measurement function $X_t = X(\mathbf{s}_t)$ (measurement of luminosity) which is typically affected by noise [17,18]. We cannot invert the measurement function, but we can use the observations of the system, made at different times, to construct vectors with the same information content as the original system state $\mathbf{s}_t$, or at least with a sufficient information content about what we are interested in the physical system. We associate to each pixel of CCD images a d-dimensional vector $x_i(t)$ containing the measurements of luminosity of the pixel i in the temporal window. Then, to extract the most relevant features present in our data, we apply the PCA. Finally, we use the Mixture of Gaussians algorithm to classify the data according to the features we are interested in.

2.1. Principal Component Analysis

A common problem in statistical pattern recognition is that of feature selection or feature extraction. Feature selection refers to a process whereby a data space is transformed into a feature space. The transformation could be designed in such a way that the data set may be represented by a reduced number of effective features. In other words, given a data set of d-dimensional vectors $\mathbf{x}_i$, we want to characterise them using p numbers, where $p < d$, preserving as much of the relevant information as possible. Each d-dimensional vector can be written as a linear combination of d orthonormal vectors, or as a smaller number of orthonormal vectors plus a residual error. PCA selects the orthonormal basis which minimises the residual error average, or equivalently, which maximises the variance [15,16].

The standard method to perform Principal Component Analysis consists of the eigendecomposition of the covariance matrix of the vector data $\mathbf{x}_i$ and the projection of the data, orthogonally, onto the subspace spanned by the eigenvectors belonging to the dominant eigenvalues (principal components). So, we reduce the number of features needed for effective data representation. This method of representation is commonly referred to as *subspace decomposition.*

The number of operations necessary to compute the covariance matrix is $O(Nd^2)$, where N is the number of d-dimensional data vectors. The cost of the direct diagonalisation of the $d \times d$ covariance matrix is $O(d^3)$.

Difficulties can arise in the form of computational complexity when N or d are in the order of several thousand. In this case, to compute the principal directions it is better to use the *EM* algorithm (Expectation Maximisation), because it does not require us to compute the covariance matrix, and has a complexity limited by $O(kNd)$ operations, where k is the number of leading eigenvectors to be learned. Moreover, the EM algorithm for PCA [19–21], compared to the standard PCA, defines a proper probability model in the space of inputs, and allows an *on-line analysis:* new data can be added to the model efficiently without repeating the calculus of the covariance matrix.

To understand how the EM algorithm works for the selection of the principal components, consider linear Gaussian models, according to which, a d-dimensional variable $\mathbf{y}$ can be produced as a linear transformation of some k-dimensional latent variable $\mathbf{x}$ plus a d-dimensional Gaussian noise variable $\mathbf{v}$ ($k \leq d$), i.e.:

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{v} \quad \mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})^2 \quad \mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}). \tag{1}$$

where $\mathbf{C}$ ($d \times k$) denotes the transformation matrix. The variables $\mathbf{x}$ are assumed to be independent and identically distributed according to a Gaussian distribution with zero mean value and unit variance. Also, $\mathbf{v}$ are independent and normally distributed with zero mean and covariance matrix $\mathbf{R}$.

Considering that the sum of two independent Gaussian distributed quantities is also Gaussian distributed, we have

$$\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}\mathbf{C}^{\mathrm{T}} + \mathbf{R}) \tag{2}$$

The EM algorithm is realised in two steps: the E-step and the M-step.

The problem of *state inference* is resolved in the E-step. Given fixed model parameters $\mathbf{C}$ and $\mathbf{R}$, we are interested in the posterior probability $P(\mathbf{x}|\mathbf{y})$ over a single hidden space, given the corresponding single observation. This can be computed with the help of Bayes' theorem, taking into account Eqs (1) and (2), and considering that the multiplication of Gaussian distributions is also a Gaussian, although not normalised:

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y})} = \frac{\mathcal{N}(\mathbf{C}\mathbf{x},\mathbf{R})\big|_{\mathbf{y}}\mathcal{N}(\mathbf{0},\mathbf{I})\big|_{\mathbf{x}}}{\mathcal{N}(\mathbf{0},\mathbf{C}\mathbf{C}^{\mathrm{T}} + \mathbf{R})\big|_{\mathbf{y}}} = \mathcal{N}(\boldsymbol{\beta}\mathbf{y},\mathbf{I} - \boldsymbol{\beta}\mathbf{C})\big|_{\mathbf{x}} \tag{3}$$

where $\boldsymbol{\beta} = \mathbf{C}^{\mathrm{T}}(\mathbf{C}\mathbf{C}^{\mathrm{T}} + \mathbf{R})^{-1}$.

The M-step consists of the *learning* (i.e. in the identification of matrices $\mathbf{C}$ and $\mathbf{R}$ that maximise, in the presence of hidden states, the likelihood of the observed data given by $P(\mathbf{y}) = \int P(\mathbf{x},\mathbf{y})\mathrm{d}\mathbf{x}$, where $P(\mathbf{x},\mathbf{y})$ is the joint probability ($P(\mathbf{x},\mathbf{y}) = P(\mathbf{y}|\mathbf{x})P(\mathbf{x})$).)

PCA is a limiting case of the linear-Gaussian model as the covariance of the noise $\mathbf{v}$ becomes infinitesimally small and equal in all directions. Mathematically, PCA is obtained by taking into account the limit $\mathbf{R} = \lim_{\epsilon \to 0} \epsilon \mathbf{I}$. This oper-

[2] The symbol ~ means 'distributed according to'. A multivariate normal (Gaussian) distribution with mean μ and covariance matrix Σ is written as $\mathcal{N}(\mu, \Sigma)$, and its expression is

$$\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{k/2}(\det(\boldsymbol{\Sigma}))^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^{\mathrm{T}}\,\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}$$

ation makes it likely that a point **y** will be dominated solely by the squared distance between it and its reconstruction **Cx**.

In this case, using the Bayes' theorem, it is possible to prove that the E-step reduces to:

$$P(\mathbf{x}|\mathbf{y}) = \delta\left(\mathbf{x} - (\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{y} \right) \tag{4}$$

As consequence, the posterior over states collapses to a single point:

$$\mathbf{X} = (\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{Y} \tag{5}$$

due to the infinitesimal value of noise.

Given **X** obtained in the E-step, and with the aim of maximising the likelihood of the observed data, we obtain a new estimate for the model parameters:

$$\mathbf{C}^{\mathrm{new}} = \mathbf{Y}\mathbf{X}^{\mathrm{T}}\,(\mathbf{X}\mathbf{X}^{\mathrm{T}})^{-1} \tag{6}$$

where **Y** is a $d \times N$ matrix of all the observed data, and **X** is a $k \times N$ matrix of the unknown states. The columns of **C** span the space of the first k principal components.

This algorithm can be performed online giving only a single data-point at a time; this is a big advantage in this pixel lensing preselection application, since new data, registered by the telescopes, are continuously collected. For each new data point we do not need to restart the procedure for all the data points, as in standard batch PCA algorithms, but simply to iterate the adaptive online algorithm.

2.2. Clustering Algorithms: Mixture Models

In the following, to obtain the *pdf* of our data, we will use a method called the Mixture of Gaussians (MOG) [15].

Mixture models are semi-parametric techniques that combine the advantages of both parametric methods (such as Gaussian distribution) and non-parametric techniques (such as k-means). In fact, they are not restricted to specific functional forms, and their complexity does not grow with the size of the data set, but only with the complexity of the problem being solved. The price to be paid is the setting up of the model using the data set (i.e. the training of the model), which is computationally intensive compared with the simple procedure needed for parametric or non-parametric methods.

In mixture models, the density function can be written as a linear combination of M basis functions[3] in the form

$$p(\mathbf{x}) = \sum_{j=1}^{M} p(\mathbf{x}|j)P(j) \tag{7}$$

The coefficients $P(j)$ are called the *mixing parameters*. They represent the prior probability that the data point **x** has been generated from component j of the mixture ($j = 1,\ldots,M$).

The component density functions are class-conditional densities, normalised so that

$$\int p(\mathbf{x}|j)\mathrm{d}\mathbf{x} = 1 \tag{8}$$

Using Bayes' theorem, the posterior probability that the jth component is responsible for generating the data point **x** is

$$P(j|\mathbf{x}) = \frac{p(\mathbf{x}|j)P(j)}{p(\mathbf{x})} \tag{9}$$

Particular mixture models are 'Gaussian mixture models', where the individual component densities are given by Gaussian distribution functions.

Assuming that each Gaussian has a covariance matrix which is some scalar multiple of the identity matrix, so that $\Sigma_j = \sigma_j^2\mathbf{I}$, we have:

$$p(\mathbf{x}|j) = \frac{1}{(2\pi)^{d/2}\sigma_j} \exp\left\{ -\frac{\|\mathbf{x} - \bar{\mathbf{x}}_j\|^2}{2\sigma_j^2} \right\} \tag{10}$$

Model parameters ($P(j)$, $\bar{\mathbf{x}}_j$, σ_j) can be determined using the EM algorithm. For each data point $\mathbf{x}^n$ ($n = 1,\ldots,N$), a variable z^n is introduced as an integer in the range $(1, M)$ specifying which component of the mixture has generated the data point. We guess some value for the parameters of the mixture model, and we use these, together with Bayes' theorem, to find the probability distribution of the $\{z^n\}$. We then compute the expectation of the negative log-likelihood, given by $\mathrm{E} = -\sum_{n=1}^{N} \ln p(\mathbf{x}^n, z^n)$. This is the E-step. The new parameter values are found by maximising the likelihood, or by minimising the expected error, with respect to the parameters (M-step). The two steps are iterated during training until the convergence is reached.

3. EXPERIMENTAL RESULTS

3.1. Data Structure Analysis

The data we have examined has been taken at the MDM observatory with telescope McGraw-Hill and a CCD camera of 2048×2048 pixels. This field has been divided into four smaller fields of size 1100×1100 pixels. The origins of these fields are (0,0), (948,0), (0,948), (948,948). The images, taken at different times, have been already calibrated (geometrical alignment, photometric alignment, seeing variations [8,13]) by the AGAPE collaboration, so that we can directly test our method on the calibrated data.

To check the validity of our method, we have considered windows of 101×101 pixels, randomly taken in each of the four fields, and applied the PCA decomposition using both the standard and the EM algorithms.

We report the experimental results obtained on one of the areas examined (fourth field $100 \leq x \leq 200$ and $100 \leq y \leq 200$).

For each of the 10,201 pixels, we have 35 values of luminosity, unevenly sampled on a period of 270 days. To each pixel, we have associated a vector whose components are the percentage variations of the 35 values of luminosity with respect to their own mean value.

[3] In our problem, M represents the number of classes into which we want to subdivide our data distribution.

After this preliminary operation, we applied the standard PCA on the luminosity values matrix composed by 10,201 rows and 35 columns.

To understand which eigenvectors determine the dominant structure of our data set, we have plotted the eigenvalues λ_i, i = 1,…,35 (Fig. 1). By analysing the plot, we understand that the projection of the data set onto the first four eigenvectors captures the effective features of the data, because the eigenvalues related with the remaining eigenvectors are very small compared to the first four. This consideration is reinforced by the results we have obtained from the calculus of the information content for each eigenvector. We have found that the first eigenvector contributes to 51.5% of the information, the second eigenvector increases the information by 12.2%, the third eigenvector is responsible for 10.9% of the information, the fourth eigenvector gives a contribution of 6.6%, giving the information content obtained in the first four eigenvalues as a total of 81.2%. If we consider the fifth eigenvector, the information increases by 2.4% only. This suggests we should use only the first four eigenvectors.

Before analysing our data in terms of PCA, we report the behaviour of the first four eigenvectors in Fig. 2. It is evident that the components of the first eigenvectors (Fig. 2(a)) show large fluctuations around a constant value, while the components of the other three eigenvectors (Figs 2(b–d)) show a regular increase or decrease, which could be related to physical variation of the luminosity.

To better characterise the physical events in the feature space, because of the lack of labelled real data, we have introduced some vectors which simulate, respectively, Gaussian noise and variations of luminosity associated with significant astrophysical events, and we look at their projections onto each of the first four eigenvectors.

Specifically, being interested in comparing our results with those obtained by the AGAPE approach, we have taken as the significant variation of luminosity the Paczynski curve [1] (see Fig. 3). This is the curve which represents the simplest lensing event.

In particular, we have simulated 90 Gaussian noisy curves, 90 Paczynski light curves centred on different values of the

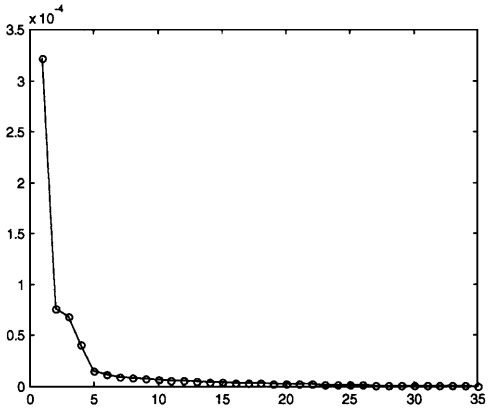

Fig. 1. Eigenvalues spectrum: in abscissa the index of the ordered eigenvalues.

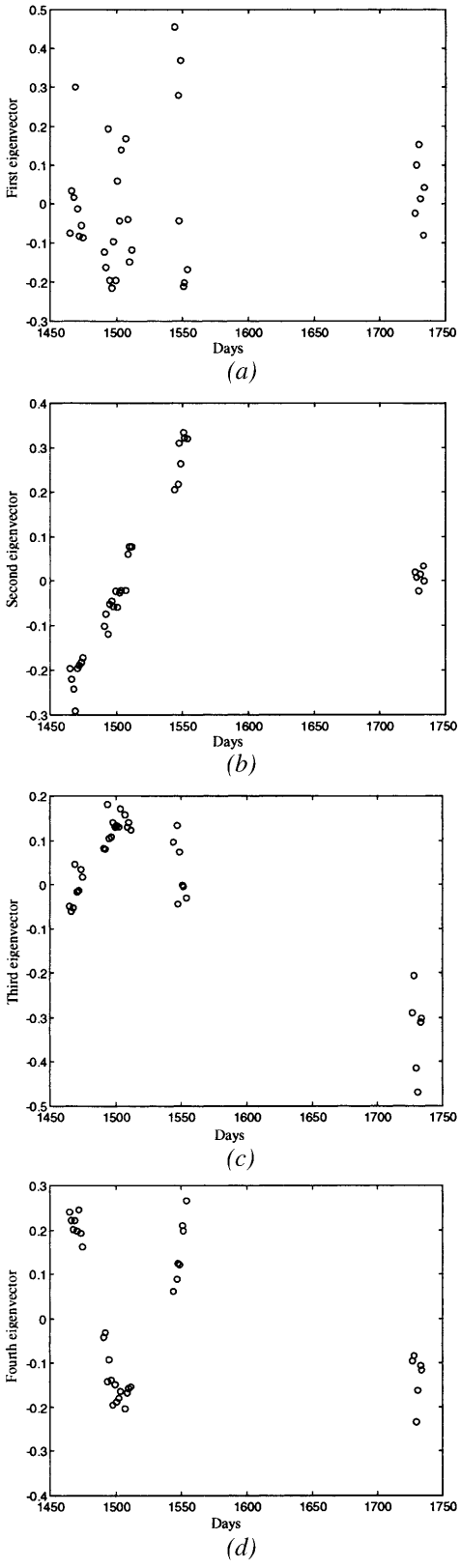

Fig. 2. Values of the first four eigenvectors against the observation times.

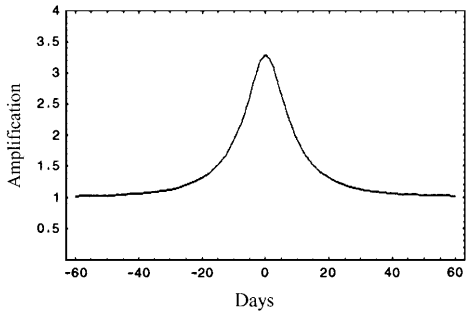

Fig. 3. The Paczynski curve. A massive object passing sufficiently near the line of sight of a star induces a characteristic variation of the star's luminosity, acting as a lens. The Paczynski curve describes the simplest microlensing event in which both the lens and the source of light are considered point-like.

temporal axis and whose amplitude was taken in accordance with the astrophysical data, and 90 other Paczynski light curves to which we have added Gaussian noise.

We see that the simulated curves, either the Gaussian noisy curves or the Paczynski curves, have projections onto the first eigenvector around the zero value. This means that the first eigenvector is not useful for selecting Paczynski light curves from noise.

This becomes more evident by representing the data set in the subspace spanned by the first and second eigenvectors. There is (Fig. 4) an overlapping between Gaussian noisy curves (yellow circles) and astrophysical signals (blue points = Paczynski light curves, red circles = Paczynski light curves with Gaussian noise) onto the first eigenvector. This is a further confirmation that the first eigenvector is not useful to reach our target.

On the contrary, for the simulated Paczynski curves we have found non-zero projection values onto the other three eigenvectors, while the simulated noisy curves have projection values near to zero. This means that the second, third and fourth eigenvectors select the interesting luminosity variations.

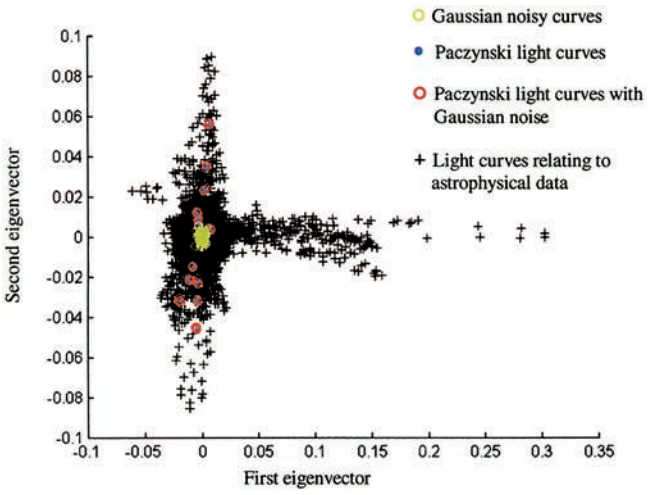

Fig. 4. Simulated and astrophysical light curves in the subspace spanned by the first and second eigenvectors. The astrophysical data with $x > 0.1$ are related to resolved stars.

Then, we represented the data set in each of the three subspaces spanned by the other three eigenvectors, taken two by two. We always found a good separation between the Gaussian noise and Paczynski curves (see, as an example, Fig. 5). Thus, the second, third and fourth components allow a separation between noise and signal variables.

Note that the method does not distinguish between blue points (noiseless Paczynski) and red circles (noisy Paczynski) (Fig. 5); this is an advantage, because it means that the performance is robust against the simulated noise. This property is useful considering that possible events will not be described exactly by an analytic curve. In practice, luminosity values are characterised by error bars because of the measurement process.

We conclude by mentioning that the standard PCA approach and EM algorithm for PCA have been applied to the data set; both methods have given similar results.

3.2. Clustering

The previous analysis has shown that interesting pixels are those whose light curves correspond to vectors with large values of the second, third and fourth components in the feature space, while the background luminosity corresponds to vectors with very small components.

To have a classification of data according to their intrinsic features, we apply a clustering algorithm based on a Mixture of Gaussians technique to the three-dimensional distribution given by the projection values of the light curves onto the second, third and fourth eigenvectors. The clustering algorithm gives a good performance with 15 classes. To each class we assign a symbol with a different colour (cyan crosses, blue circles, blue crosses, green circles, magenta circles, red circles, yellow crosses, cyan circles, blue points, magenta rhombi, red crosses, yellow crosses, black crosses, cyan points, green rhombi).

In Fig. 6 we project the results coming from the three-dimensional clustering into the plane spanned by the second and third eigenvectors.

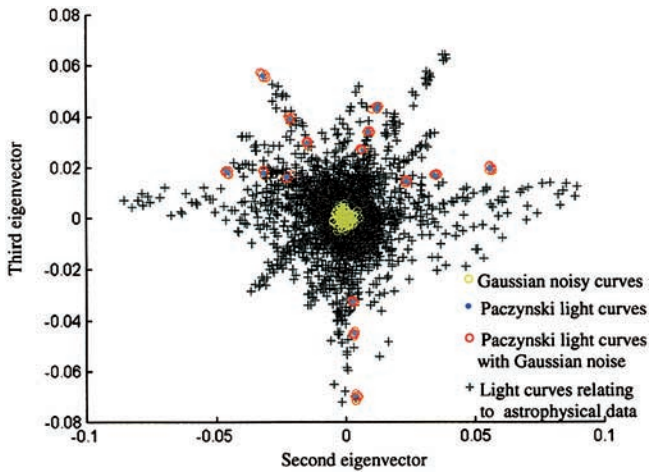

Fig. 5. Simulated and astrophysical light curves in the subspace spanned by the second and third eigenvectors.

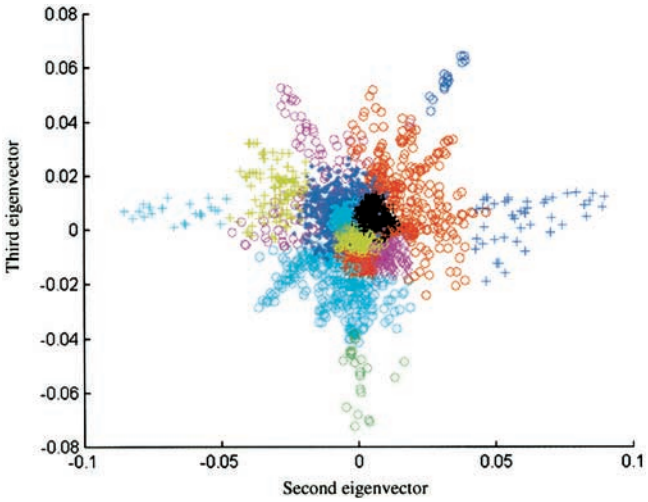

Fig. 6. Projection of the clustering results in the plane spanned by the second and third eigenvectors. Different symbols and colours represent different classes.

To label the classes, the simulated vectors have been treated like the vectors containing luminosity values, and so it has been determined at which class of the mixture model the simulated curves belong.

We have found Gaussian noisy curves in the internal classes, indicated with green rhombi, black crosses, yellow circles and cyan points (Fig. 6). These classes represent 78.9% of all data.

Simulated Paczynski curves belong to the classes indicated with green, magenta, red, cyan circles, blue, cyan, yellow crosses and blue points. These classes represent 14.0% of all data. Each class is characterised by the position of the peaks on the temporal axis. Finally, we checked (on a random choice of pixels) that real light curves belong to the appropriate class.

In conclusion, the approach adopted allows us to select the interesting light curves which belong to the classes sited in the external part of the distribution, while noisy curves belong to the classes placed in the internal region.

The pre-selection is very efficient; in fact 78.9% of the data have been eliminated.

4. COMPARISON WITH THE AGAPE APPROACH

The AGAPE collaboration [8,9] has developed a pixel lensing technique based on the monitoring of pixel light curves in order to find microlensing events in the analysis of pixel lensing data. One of the steps of their approach is the selection of interesting pixels, which is performed by choosing the pixels whose luminosity variation satisfies task-specific requirements. The AGAPE approach calculates, for each light curve, some quantities on which suitable cuts are imposed. Their results are, however, greatly affected by the cuts' values.

On the contrary, we have developed an unsupervised technique able to distinguish noise from significant variations of luminosity without using any task-specific information. We do not pre-determine any quantity, and so we do not have the problem of choosing what cuts have to be imposed on them. Furthermore, our approach classifies interesting light curves according to their shape (in particular, according to the position of the luminosity peak on the temporal window).

In Fig. 7, we compare the results obtained with both methods on the luminosity values of 10,201 pixels.

As we can deduce from the comparison between Fig. 7(a) and Fig. 7(b), the pixels selected with the AGAPE approach are those which belong to our interesting classes, i.e. classes sited in the external part of Fig. 6.

We can see that our approach individualises some more clusters which do not appear into AGAPE selection. These are clusters relative to pixels where resolved stars are present (regions labelled with black circles in Fig. 7(b)). Indeed, our approach has been applied to all the pixels in the field, i.e. both pixels registering luminosity of a single resolved star and pixels registering luminosity of a whole group of stars, while in the database used by the AGAPE approach, the pixels corresponding to resolved stars have been eliminated. Therefore, the difference is a consequence of the fact that our database is larger then that used by AGAPE.

Finally, it is worth stressing that, in order to decide if the interesting pixels, selected by the AGAPE preselection or by our approach, are effectively lensing events, it will be necessary to carry out some more tests (symmetry test, achromaticity test, uniqueness test) and develop fitting models. At the moment this is outside the scope of our aims. Tests in this direction are in progress.

5. RESOLVED STARS ANALYSIS

Further validation of our approach has been obtained by studying the variations of luminosity induced by the resolved stars. The separation of the variations of luminosity due to resolved stars from the background is accomplished by looking at the projection of the data along the first eigenvector (Fig. 4).

We have found that the light curves induced by the presence of resolved stars correspond to vectors whose first component is greater than 0.1 (astrophysical data in the right part of Fig. 4). This becomes more evident by looking at Fig. 8, where the pixels with large fluctuations of luminosity, due to the presence of resolved stars (red and yellow pixels), are well separated by the background (blue pixels), and are circularly distributed around the image of two resolved stars present in the examined area.

In conclusion, our approach gives good results in the cases examined, and has the same efficiency as the AGAPE approach for the preselection of pixels interesting in microlensing events; besides, it does not require an *a priori* knowledge of the light curve behaviour. Therefore, it is also adapted for those cases where the light curve behaviour is unknown.

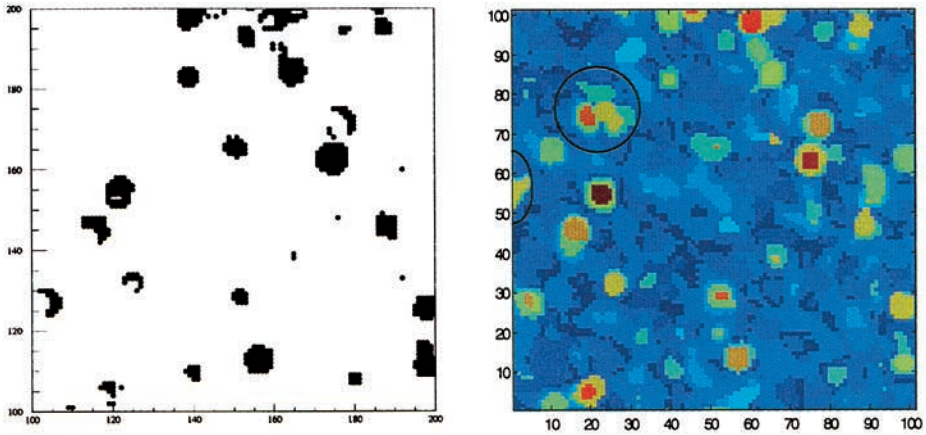

Fig. 7. (A) Pixels selected with AGAPE approach. (B) Image concerning the positions of all pixels. We have used the same colour for pixels which belong to the same class. Blue tonalities are referred to pixels belonging to classes in the central part of Fig. 7 (noise). The 85.2% of all data are in this region.

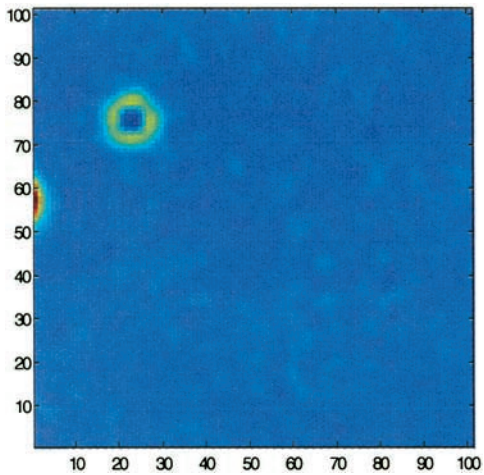

Fig. 8. Image concerning the distribution of data along the first eigenvector. The light curves of red and yellow pixels, which are influenced by resolved stars, correspond to vectors whose first component is greater than 0.1 (region on the right of Fig. 4).

Acknowledgements

The collaboration with the AGAPE team is acknowledged; in particular, Yannick Giraud-Héraud for having provided us with the data. We greatly thank Roberto Tagliaferri and Sebastiano Calchi Novati for useful discussions and suggestions.

References

1. Paczynski B. Gravitational microlensing by the Galactic halo. Astrophysical Journal 1986; 304:1–5
2. Petrou M. Dynamical Models for Spheroidal System. PhD thesis, Cambridge University, 1981
3. Aubourg E et al. Evidence for gravitational microlensing by dark objects in the galactic halo. Nature 1993; 365:623
4. Udalski A, Szymanski M, Kaluzny J et al. The optical gravitational lensing experiment: Discovery of the first candidate microlensing event in the direction of the Galactic Buldge. Acta Astronomic 1993; 43:289–294
5. Alcock C, Akerloff CW and Allsman RA et al. Possible gravitational microlensing of a star in the Large Magellanic Cloud. Nature 1993; 365:621
6. Alcock C, Allsman RA and Alves D et al. The MACHO project: 45 candidate microlensing events from the first-year galactic buldge data. Astrophysical Journal 1997; 479:119
7. Alard C, Guibert J. The DUO lens candidates. Astronomy & Astrophysics 1997; 326:1–12
8. Ansari R, Aurière M and Baillon P et al. AGAPE: a search for dark matter towards M31 by microlensing effects on unresolved stars. Astronomy & Astrophysics 1997; 324:843–857
9. Ansari R, Aurière M and Baillon P et al. AgapeZ1: a large amplification microlensing event or an odd variable star towards the inner bulge of M31. Astronomy & Astrophysics 1999; 344:L49-L52
10. Crotts APS, Tomaney AB. Results from a survey of gravitational microlensing toward M31. Astrophysical Journal 1996; 473:L87–L90
11. Baillon P, Bouquet A and Giraud-Heraud Y et al. Detection of brown dwarfs by the micro-lensing of unresolved stars. Astronomy & Astrophysics 1993; 277:1–9
12. Crotts AP. M31: a unique laboratory for gravitational microlensing. Astrophysical Journal 1992; 399:L43-L46
13. Kaplan J. Pixel Lensing. In: AA Marino, L Grenacher, S Capozziello, G Iovane (eds) Topics on Gravitational Lensing, Bibliopolis, 1998
14. Gould A. The theory of pixel lensing. Astrophysical Journal 1996; 470:201–210
15. Bishop CM. Neural Networks for Pattern Recognition. Clarendon Press, Oxford, 1997
16. Haykin S. Neural Networks: A comprehensive foundation. Prentice Hall, 1999
17. Lowe D. An approach to dynamic modelling and topographic feature extraction of wake EEG. In International Conference on Advances in Pattern Recognition, Springer-Verlag, 1998, pp 145–153
18. Lowe D. Extracting structure from wake EEG using neural networks. SPIE's Aerospace/Defence Sensing and Controls Conference: Applications and Science of Artificial Neural Networks III, 1998; 3077:17–26

19. Roweis S. EM algorithms for PCA and SPCA. In: MI Jordan, MJ Kearns, SA Solla (eds), Advances in Neural Information Processing Systems 10, MIT Press, 1998, pp 626–632
20. Roweis S, Ghahramani Z. A unifying review of linear Gaussian models. Neural Computation 1999; 11(2):305–345
21. Tipping M, Bishop CM. Probabilistic principal components. Journal of the Royal Statistical Society B 1999; 61(3):611–622

Maria Funaro is a PhD student at the University of Salerno, Italy. She received the Laurea (cum laude) in physics in 1998 at the University of Salerno. Her research interests are in the areas of unsupervised learning and clustering algorithms. She is member of SIREN (Italian Society of Neural Networks) and INFM (Italian Institute of Condensed Matter).

Maria Marinaro is Professor of Theoretical Physics at the University of Salerno, Italy. She has been Dean of the Faculty of Science from 1983 to 1995. She received the laurea in physics in 1956 at the University of Naples, and the libera docenza in 1957. From 1972 to 1976 she directed the Department of Theoretical Physics at the University of Salerno, and she founded and for 10 years directed the section of Salerno of INFM (Italian Institute of Condensed Matter). Professor Marinaro is the author of over 100 publications in the areas of theoretical physics, statistical mechanics and neural networks. She is member of SIF (Italian Society of Physics), SIREN (Italian Society of Neural Networks) and of the Accademia Nazionale di Scienze, Lettere e Arti, Naples section.

Alfredo Petrosino received the 'Laurea' degree (cum laude) in computer science at the Faculty of Sciences of the University of Salerno in 1989. He has been a fellow researcher of the Italian National Research Council (CNR) from 1989 to 1994, and of the International Institute for Advanced Scientific Studies (IIASS) in 1995. Since 1996 he has been a Researcher of the National Institute for Physics of Matter (INFM) at the University of Salerno. From 1991 he taught courses in mathematics and image processing at the University of Salerno, and on operating systems at the University of Siena. He is the author of more than 40 papers, and in 1994 he received the Academic Prize for Cybernetics from the Italian Academy of Science, Arts and Letters. He is also co-editor of the volumes *New Trends in Fuzzy Logic* and *New Trends in Fuzzy Logic II*, both published by World Scientific, Singapore. He is a member of the International Association for Pattern Recognition (IAPR) and IEEE. His research interests include image processing, pattern recognition, multimedia systems, parallel computing and neural networks.

Silvia Scarpetta is a post-doctorate researcher at the University of Salerno, Italy. She received a laurea in physics in 1995 at the University of Naples, and a PhD in physics in 1999 from the University of Salerno. She is the author of several publications in the area of statistical mechanics and neural information processing. She is a member of SIF (Italian Society of Physics), SIREN (Italian Society of Neural Networks) and INFM (Italian Institute of Condensed Matter).

Correspondence and offprint requests to: Dr. Maria Funaro, Dipartimento di Fisica 'E. R. Caianiello', University of Salerno, Via S. Allende, I-84081 Baronissi (SA), Italy. Email: maria@sa.infn.it