Automated clustering method for point spread function classification

Weinan Wang,[1] Peng Jia,[1,2]★ Dongmei Cai[1] and Huigen Liu[3]

[1]*College of Physics and Optoelectronics, Taiyuan University of Technology, Taiyuan 030024, China*
[2]*Department of Physics, Durham University, South Road, Durham DH1 3LE*
[3]*School of Astronomy and Space Science, Nanjing University, Nanjing 210023, China*

ABSTRACT

The point spread function (PSF) plays a very important part in image post-processing and high-precion astrometry and photometry. It is necessary to analyse the properties of the PSF before we use it to process data. However, in real observations, the PSF is affected by many different factors and the shape of it has inevitable spatial and temporal variations that can be hard to describe. In this paper, we propose a clustering method to evaluate the shape variations of PSFs. We analyse the performance of this method with simulated PSFs under different observation conditions. Then, we process two observational data sets with this method. The PSF clustering results can provide a reference for checking observation conditions and can be used for astrometry based on PSF fitting. In general, our method can reveal the morphologic similarities of different PSFs and can provide a reference for observations. The cluster revealed by our method can provide a reference for the evaluation of observation conditions and for the post-processing of astronomical observation data.

Key words: methods: data analysis – techniques: image processing.

1 INTRODUCTION

For optical observations, the point spread function (PSF) describes the reponse of the whole optoelectronic system to a point source. For a given telescope, during the design procedure, the PSF can be calculated using some mature optical design software, such as ZEMAX or Code V. Through suitable spatial sampling and noise adding, we can simulate the images of stars that would be obtained by photosensors. These are relatively stable and can be used as a reference to evaluate the performance of telescopes or to design post-processing algorithms (Gai & Cancelliere 2007; Krist, Hook & Stoehr 2011; Perrin et al. 2014). However, during the observation stage, the PSFs of a telescope have an inevitable spatial and temporal variation. In the temporal domain, for space telescopes, PSFs vary because of thermal deformation; for ground-based telescopes, the variations of PSFs are mainly induced by atmospheric turbulence, thermal deformation, gravitational deformation, pointing and tracking error. In the spatial domain, for wide-field telescopes or telescopes with refractive elements, field-dependent aberrations (such as coma, astigmatism, distortion) and chromatic aberrations couple with dynamical aberrations and make the shape of the PSFs very complex and hard to describe.

The reasons discussed above make it hard to evaluate the variations in PSF and limit further increments of the performance of image processing algorithms (Desiderà & Carbillet 2009), astrom-etry based on PSF fitting and the photometry method (Diolaiti et al. 2000; Gai, Busonero & Cancelliere 2013; Bradley et al. 2016). As a first-order approximation, a commonly accepted assumption is that the PSF is stable during some time-scales and a mean PSF or an artificial PSF can be used as PSF template (Schechter, Mateo & Saha 1993; Kuijken 2008; Li et al. 2016). However, the validity of this assumption depends significantly on the observation tactic and observation site. Because space telescopes have very stable observation conditions, the PSF can be generated from observation data (Anderson & King 2000). However, for ground-based telescopes, the temporal and spatial variations mean that it is hard to apply this method. In these circumstances, mean PSFs obtained directly from an epoch of all the observation data are widely used (La Camera et al. 2015). As the mean PSF is generated from observation data and reflects the common property of all the PSFs in the relevant epoch, it is necessary to analyse the property of these PSFs before we apply this method. However, because of the huge number of stars that can be used as a PSF reference, it is better to use an unsupervised classification method to analyse PSFs.

In this paper, we propose a new PSF classification method based on a cluster algorithm to reveal the temporal and spatial variations of PSFs. The PSF clusters obtained by our method can help us to analyse the distribution of PSFs with different morphological patterns and can further improve PSF-based post-processing methods. In Section 2, we give the basic theory and process of our method. In Section 3, we analyse the performance of our classification algorithm with PSFs generated from Monte Carlo simulations under different observation conditions, such as PSFs obtained from a tele-

★ E-mail: robinmartin20@gmail.com

scope with field-dependent aberrations, PSFs with different spatial sampling rates and PSFs with different signal-to-noise (S/N) ratios. In Section 4, we use observation data from two different telescopes to show the improvement of the observation results obtained using our classification algorithm. In Section 5, we make conclusions and anticipate our future work.

2 BASIC THEORY AND METHOD

PSFs obtained from observational data can be seen as images with particular structures. These structures are important for data processing and analysis. However, during observations, they are always affected by the presence of photon–electronic noise and finite spatial discrete sampling of detectors. The noise smears the wings of PSFs and the finite discrete sampling reduces the details of the PSFs. To be able to classify PSFs according to their structural variations, it is necessary to design a classification algorithm that includes data cleaning and pre-processing, feature selection, classification, cluster evaluation and representation of classification results in a data cube. Before applying this classification algorithm, we need to define what we mean by a PSF.

2.1 PSF definition

A PSF is the response of an imaging system to a point source. The imaging process can be expressed by

$$I(x, y, t_0) = \int_{t_0}^{t_0+\delta_t} [O(x, y, \tau) * g(x, y, \tau)] \, d\tau \qquad (1)$$

where $O(x, y, \tau)$ is the original image, $g(x, y, \tau)$ is the spatial and temporal variable PSF and $*$ represents convolution. In general, stars are point sources, so that $O(x, y, \tau)$ is unit pulse function. $I(x, y)$ represents the PSF of the whole imaging system. If the CCD is considered as part of the imaging system, then the imaging process will also include the effects brought about by the photon–electronic noise and the finite spatial discrete sampling of detectors as defined by

$$I(x, y, t_0) = \int_{t_0}^{t_0+\delta t} \left\{ [O(x, y, \tau) * g(x, y, \tau)]_{pixel(x,y)} \right. $$
$$\left. + N(x, y, \tau) \right\} d\tau. \qquad (2)$$

Here, $[\ ]_{pixel}$ represents the point response of the CCD, $N(x, y, \tau)$ is the noise from the CCD and the sky background and $I(x, y, t_0)$ is the observation image in t_0 with an exposure time of δ_t. For astronomical observations, images of stars can be viewed as noisy realizations of PSFs with a particular pixel sampling rate. In this paper, we use flux-normalized star images with enough S/N ratio and a moderate shape as adequate representations of PSFs. This is because with ground-based wide-field observations, we find it harder to reconstruct PSFs directly from observation data than with space telescopes (Anderson 2016). It should be mentioned that these PSFs will include both the optical imaging and the CCD sampling effects, which is different from the commonly accepted definition.

2.2 Data cleaning and pre-processing

The PSF classification results are affected by many different factors and we propose our data cleaning and pre-processing steps as follows.

First, images with different S/N ratios have strong effects on the PSF classification results, so we need to restore the PSF along with

Table 1 The information for two PSFs in the PSF data cube. The grey-scale images are modified with the zscale algorithm in DS9 (Joye et al. 2003) for a clearer demonstration.

Parameters	PSF A	PSF B
PSF image		
PSF size	9 pixels	9 pixels
Position on X-axis	253	246
Position on Y-axis	546	235
Eccentricity	0.9	0.6
S/N ratio	33	46
Exposure time	5 s	5 s
Observation time in UTC	2017/02/10	2017/02/10
Observation time (to seconds)	13:06:23	15:08:24

the S/N ratio of the original star's image. The S/N ratio we use in this paper is defined by

$$SNR = \frac{signal - noise}{\sigma_{noise}} \qquad (3)$$

where *noise* represents the mean value of the background, σ_{noise} is the standard deviation of the background and *signal* is the mean value of the star.

Secondly, to remove the images of moving targets or overexposed images, we extract images of stars with a predefined size (in general, we set the size to be 9 pixels) using SEXTRACTOR (Bertin & Arnouts 1996) and we delete images with extraordinary eccentricity as defined by

$$eccentricity = \frac{length_X}{length_Y}. \qquad (4)$$

Here, $length_X$ and $length_Y$ are the length of the star's image along the X-axis and Y-axis, respectively. We fit an ellipse to the star's image to calculate its $length_X$ and $length_Y$. The stars whose images have eccentricity between 0.5 and 2 are accepted as ordinary stars.

Then, the images of ordinary stars that we select are normalized by the total flux as shown in

$$StarImg_{normalized} = \frac{StarImg}{sum(StarImg)}, \qquad (5)$$

where *sum(StarImg)* is the summation of grey values for each pixel from the extracted star image. The normalized star image will be treated as an acceptable PSF representation and used for further PSF classification and analysis.

Finally, we also save the spatial and temporal positions of the extracted star images and we store all this information in a data cube. The PSF data cube is a sparse cube with three dimensions to mark the position of PSFs in the spatial–temporal domain. These three dimensions are the coordinates of the PSF in one frame of the images and the number or exposure time of that frame. One point in the data cube stands for a PSF with the information as shown in Table 1. After classifying PSFs in the data cube, we can draw the visibility graph of the data cube to observe the distribution of PSFs.

2.3 Feature selection of a PSF

The PSFs we have obtained are continuous grey images with a relatively large size and are contaminated by noise. These problems will increase computation intensity and reduce classification efficiency

Table 2 The parameters of the principal components of PSFs. Each PSF can be presented by fewer dimesions. We show the first four coefficients of the principal components in this table.

	PCA basis			
	$N = 0$	$N = 1$	$N = 2$	$N = 3$
PSF1	0.00287	−0.04781	0.01961	0.00733
PSF2	−0.00728	0.01216	−0.03250	0.00410
PSF3	0.01019	0.03281	0.00517	0.01198
PSF4	−0.03896	−0.00431	−0.01705	−0.02793
PSF5	0.00585	−0.05870	0.03324	−0.00128
⋮	⋮	⋮	⋮	⋮

if we directly classify these images. Besides, the complex structure of the PSFs makes it hard to describe them with an analytical function (Piotrowski et al. 2013; Janout et al. 2016). We need to design a proper presentation method for the PSFs according to their data properties.

Direct classification in the original data cube will cost huge amount of memory and computation resources, when the data dimension increases, which is called dimension curse. We need to reduce the data size through feature selection. Feature selection refers to a process whereby a data space is transformed into a feature space that has almost the same information as the original data space, yet the transformation is designed in such a way that the data set can be represented by a reduced number of effective features.

In this paper, we use the principal component analysis (PCA) method to generate PSF features (Jee et al. 2007; Bailey 2012). PCA is a statistical procedure that uses orthogonal transformation to convert a set of possibly correlated observational variables into a set of linearly uncorrelated variables called principal components (Hotelling 1933). After obtained the data cube, which contains N PSFs, each of which has $m \times m$ pixels, we can transform the PSF data cube into a two-dimensional (2D) PSF matrix with a size $N \times M$, where N is the number of PSFs and M is the size of the image of the PSF after being stretched into vector ($M = m \times m$) as shown in

$$x_i = [p_1 \cdots p_M]^T, \quad i = 1, \ldots, N. \tag{6}$$

First, we subtract the mean PSF from all the PSFs. The mean PSFs can be calculated by

$$mean = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{7}$$

and we define w_i as the mean-subtracted PSF as shown in

$$w_i = x_i - mean. \tag{8}$$

In PCA, we need to find the eigenvectors e_i and eigenvalues λ_i of the covariance matrix C defined in

$$C = W W^T \tag{9}$$

through singular value decomposition (SVD), where W is a matrix composed of the column vectors w_i placed side by side. The eigenvectors e_i correspond to non-zero eigenvalues of the covariance matrix. They produce an orthogonal basis for the subspace within which most PSF data can be represented with fewer dimensions and small errors. The eigenvectors are stored according to their corresponding eigenvalues in descending order. The eigenvector corresponding to the largest eigenvalue reflects the greatest

variance of PSFs. Through this step, PSFs can be projected on to spaces with $N'(<<N)$ dimensions as shown in

$$\Omega = [v_1 v_2 \cdots v_{N'}]^T, \tag{10}$$

where $v_i = e^T_i w_i$ and v_i is the ith coordinate of the PSFs in the new feature space, which is the value of the principal component e_i.

By decomposing PSFs with PCA, it is possible to reduce the noise effects, because the noise in the final images is a random variable and does not have a fixed structure in all PSFs. Besides, PCA can generate an orthogonal basis that can represent PSFs with fewer dimensions, which makes it easier to design the following classification algorithm. The PSFs represented by principal components are shown in Table 2.

With this method, we can generate features that will enable us to investigate PSF clustering according to the morphological variations of PSFs, as defined in Section 2.4. Generally, we calculate the sum of the eigenvalues and choose these v_i corresponding to more than 85 per cent of the sum of eigenvalues. According to our experience, for wide-field refractive telescopes (e.g. the flat-field Schmit telescope) with a diameter of 450 mm, a field of view of 2 deg and a pixel scale of 6.8 arcsec, the number of effective principle components N' is less than 10 when the observation band is 500 nm.

2.4 Feature analysis of PSFs using self-organized maps

After PCA decomposition, the PSF features are transformed to a space with fewer dimensions. We classify these PSFs in this space. For a well-behaved telescope, there is morphological similarity between PSFs with adjacent spatial or temporal coordinates and we classify these PSFs as one cluster. We can use the features presented by PCA components to find these clusters, if these clusters between parameters (or combinations of parameters) exist. These clusters can be used to divide the whole data cube into small slices that have very similar PSFs.

Because there are a large number of PSFs that need to be classified, compared with traditional unsupervised learning algorithms such as K-means clustering, we use the Kohonen self-organizing map (SOM; Kohonen 1982) for PSF classification in this paper, because the SOM is faster and needs less computer memory.
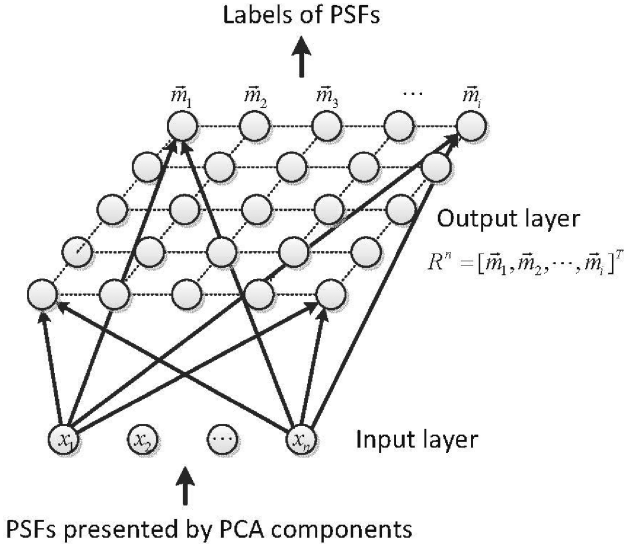
The SOM is a fully connected single-layer neural network, as shown in Fig. 1, where the network defines a mapping from the input data space (PCA coefficients for each PSF) on to a regular 2D array of nodes (map grid). A parametric reference vector $m_i \in R^n$ is associated with every node i in the map, as shown in

$$R^n = [m_1, m_2, \ldots, m_i]^T. \tag{11}$$

The array of nodes can be projected on to a rectangular or hexagonal lattice. Every input vector x, which corresponds to the PCA coefficients of a PSF, will be compared with m_i. According to the results, the inputs will be mapped to a node that has maximal similarity between the PCA coefficients of the PSF. The SOM works with the following steps.

First, the newtwork is initialized with synaptic weights m_i. This can be done by assigning them random values. After the network has been initialized, the spatial locations of neurons in the lattice are indicative of the essential statistical features contained in input patterns. For each input pattern x, we calculate the values of these synaptic weights between different neurons to find a winner neuron c, as shown in

$$\|x - m_c\| = \min_i \|x - m_i\|. \tag{12}$$

Figure 1. The structure of the SOM, which is a fully connected single-layer neural network.

After we select the winning neuron, we repeat the above process for the rest of the PSFs, as defined by

$$m_i(t + 1) = m_i(t) + h_{c,i}(t)[x(t) - m_i(t)], \tag{13}$$

where t is the discrete-time coordinate and $h_{c,i}$ is the function defining the neighbourhood. The neighbourhood function is defined over the lattice points in the map and represents the stiffness or elasticity between different data points, as shown in

$$h_{c,i} = \exp\left[-\frac{||r_c - r_i||^2}{2\sigma^2(t)}\right], \tag{14}$$

where r_c is the location of unit c on the map grid and the σ_t is the neighbourhood radius at time t. The SOM repeats the above process in several iterations until t up to a predefined value t_{max} (in general, we set t_{max} to 200).

Finally, the SOM outputs the specific topological feature map of the data set and returns different labels that represent different PSF clusters. In general, the SOM can generate PSF clusters from the PSF data cube. We can use PSFs with the same label as indicators of the PSF uniform area inside which the PSFs are similar to each other in the PCA space.

The number of groups and the learning rate are super parameters for the SOM that should be manually defined. In this paper, we classify PSFs with different numbers of clusters and we compare distributions of PSFs in the SOM with different numbers of clusters. For a well-behaved data set of PSFs, as the number of clusters increases, one cluster has a greater probability of breaking into small pieces and these small pieces have a very low probability of joining together. Therefore, we observe the behaviour of the classification results as the cluster number increases and we select the cluster number when it reaches the maximal number of clusters that we want or when the small clusters begin to merge again, as shown in Section 2.6. For the learning rate, we classify the same data set with different learning rates and we find the optimal value according to the evaluation factor that we define in Section 2.5.

2.5 Evaluation of clusters

To evaluate classification results with the same number of clusters, we use the Davies–Bouldin index (Davies & Bouldin 1979) defined
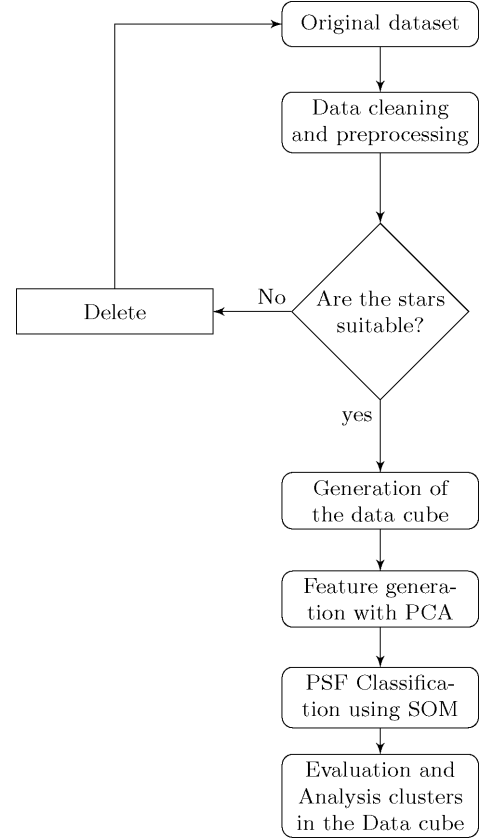
Figure 2. A flowchart of the PSF classification method.

in

$$DB = \frac{1}{k} \sum_{i=1}^{k} \max_{j \neq i} \left[\frac{avg(C_i) + avg(C_j)}{d_{cen}(\mu_i, \mu_j)}\right], \tag{15}$$

where k is the number of clusters, $avg(C)$ represents the average distance for each data point in group C and $d_{cen}(\mu_i, \mu_j)$ is the distance between two centres in the C_i and C_j groups. For the same number of clusters, the Davies–Bouldin index is lower, the difference of each cluster is larger and the difference inside clusters is lower, so the effectiveness of clustering is better. We observe each result of the DB index with different learning rates to find the optimal learning rate of the SOM, which has the lowest DB index. Moreover, because of the randomness of the SOM, we classify the same data set with the same number of clusters many times and we choose the classification results with the minimal DB index.

2.6 Data processing and representation

PSFs can be classified with different clusters according to their morphological variations, as shown in the flowchart in Fig. 2. There are two points that should be noted during the application.

First, after extracting PSFs from images, we should perform background subtraction so that the background is not a part of the PSFs. For our data sets, however, we do not subtract the background.

To better explain this issue, we draw the cross-sectional view along the X-axis of stars with different magnitudes from the same image. We use an annulus that is far from the position of the stars and we calculate the average grey values of these areas as the background of the PSF. As shown in Fig. 3(a), the area we use to calculate the background is located outside both sides of the dashed
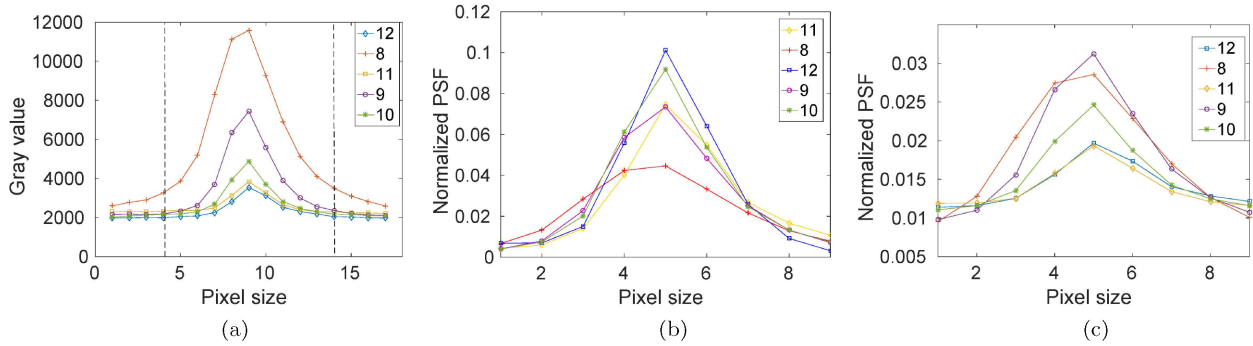
Figure 3. One-dimensional curve from images of stars with different magnitudes. Panel (a) is the cross-sectional view along the *X*-axis of stars with magnitudes between 8 and 12.Panels (b) and (c) represent flux-normalized PSFs with and without background subtraction, respectively.
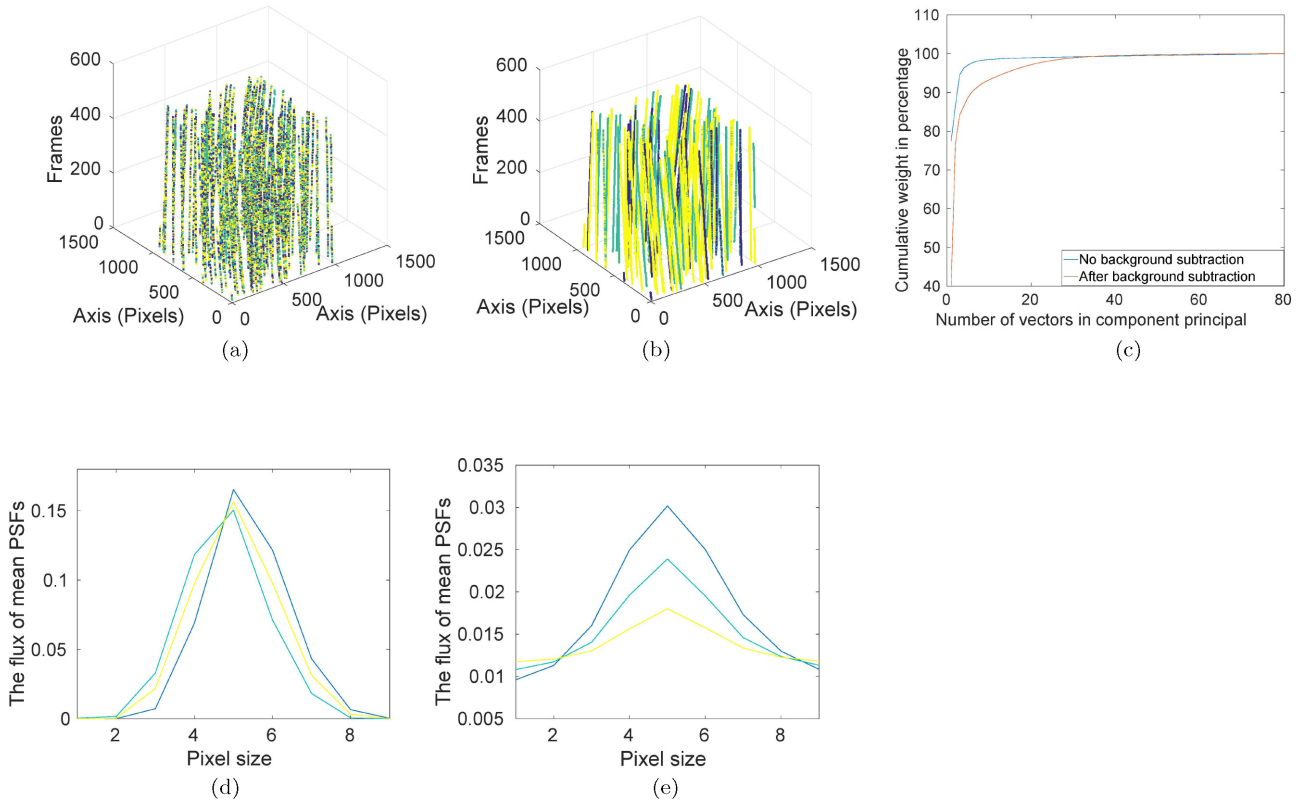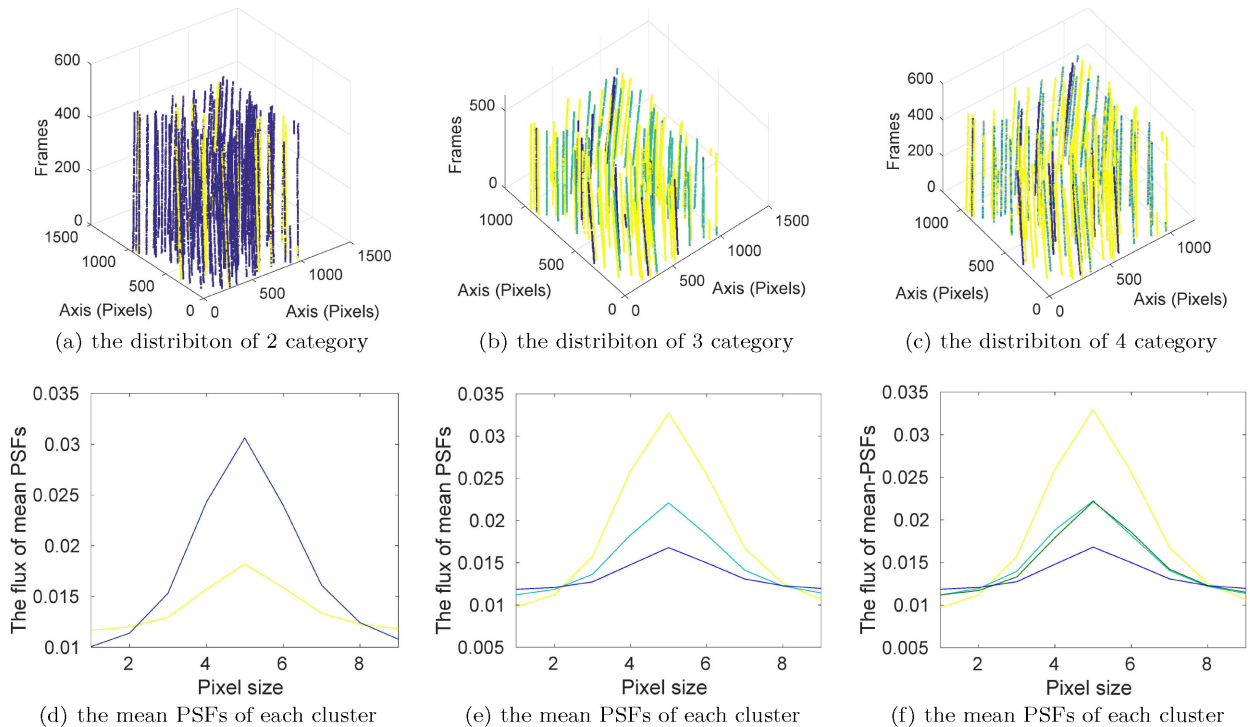

Figure 4. Panels (a) and (b) show the PSF distributions with and without background subtraction, respectively, and also the visibility graphs in the data cube. The *X*-axis and *Y*-axis are the coordinates of the PSFs in one frame of image, and the *Z*-axis is the frame number. Panel (c) represents the cumulative weight percentage of the principal components for the PSFs obtained by two methods. It is evident that background subtraction will bring addtional noise and the percentage of the first primary component will decrease. Panels (d) and (e) show the mean PSFs with and without background subtraction, respectively. Each mean PSF corresponds to the cluster with the same colour in panels (a) and (b).

line. The sections of these PSFs are shown in Figs 3(b) and (c) with and without background subtraction, respectively. It is important to emphasize that subtraction of the background would be able to exclude the influence of the magnitude if the PSF is well sampled and has a high S/N ratio. However, PSFs from our data sets have low sampling rates and S/N ratios, so the background subtraction will introduce additional errors and the PSF differences that can be revealed by the PCA components will be affected by the background subtraction method.

As shown in Figs 4(a) and (d), for the same data set, the cluster will break into small pieces because of the randomness created by

background subtraction. Besides, according to our experience, the lower the percentage of the high-order principal components, the lower the noise level. In Fig. 4(c), we find that after background subtraction, the number of high-order principal components will increase, which indicates that the noise level will increase. Although the classification results are related to the S/N ratio, the distribution of PSFs can be excavated by continuous classification of stars with the same S/N ratio.

Secondly, as the number of clusters is hard to define before classification, we choose two categories at first and increase the number of clusters. As the number of clusters increases, the number of PSFs

(a) the distribiton of 2 category

(b) the distribiton of 3 category

(c) the distribiton of 4 category

(d) the mean PSFs of each cluster

(e) the mean PSFs of each cluster

(f) the mean PSFs of each cluster

Figure 5. This figure shows how we choose the number of clusters according to the PSF distribution. Panels (a)–(c) show PSF distributions when we classify PSFs as 2, 3 and 4 clusters separately. Different colours represent different clusters. Panels (d)–(f) show the cross-section along the *X*-axis of each mean PSF, corresponding to each cluster in panels (a)–(c), respectively.

Table 3. Parameters of the flat-field Schmidt telescope.

Parameter	Value
Diameter of primary mirror	0.75 m
Wavelength	500 nm
Field of view	
Pixel size	0.1–1.7 arcsec
Spherical aberration coefficient	1.3065
Coma coefficient	−0.0044
Astigmatism coefficient	−0.0162
Field curvature coefficient	−0.0328
Distortion coefficient	−0.0332

in each cluster decreases. We check the mean PSF and the number of PSFs of each cluster. When the difference between the mean PSFs of two clusters is small and the number of PSFs in one of the clusters is small, we merge these two clusters into one cluster and stop increasing the cluster number, as shown in Fig. 5. When we increase the number of clusters from 3 to 4, only a small number of PSFs are classified as a new cluster and the difference between the mean PSF of this new cluster and that of another cluster is very small. Therefore, three is the proper number of clusters for this data set.

3 THE PERFORMANCE OF THE PSF CLASSIFICATION ALGORITHM

The SOM is a clustering algorithm and it belongs to unsupervised learning. For unsupervised learning, one of the most important problems is to restrict the classification results according to our needs, or at least to understand when the classification results can satisfy our demand. In real applications, PSF variations are interrupted by

finite CCD spatial sampling rates and stars with different S/N ratios, as described in Section 2.2. To understand our classification algorithm performance, in this section we generate different sets of PSFs with different parameters using Monte Carlo simulations for algorithm tests.

3.1 Simulation of PSFs

Our simulations are divided into two separate parts. For the optical part, we simulate all physical propagation (Cain & Watts 2016) within the telescope optic system to generate optical PSFs, which have very high sampling rates and different aberrations (Perrin et al. 2012). For the electronic part, we simulate the photosensor at a high level and we under-sample the optical PSFs according to real observation conditions (Konnik & Welsh 2011).

PSFs from the flat-field Schmidt telescope are simulated. The flat-field Schmidt telescope is widely used for sky surveys and has a wide field of view, as shown in Table 3. We add different levels of wavefront aberrations to its primary mirror to generate PSFs with different aberrations and sampling rates.

3.2 Classification performance with different CCD sampling rates

To analyse the classification performance with PSFs in different CCD sampling rates, we simulate the PSFs of the wide-field Schmidt telescope with different CCD sampling rates and we set all the noise values to be zero to avoid other influences on the classification results. The flat-field Schmidt telescope has very small field-dependent aberrations and in most parts of its field of view, the PSFs can be assumed to be one category. The shape of the cluster should
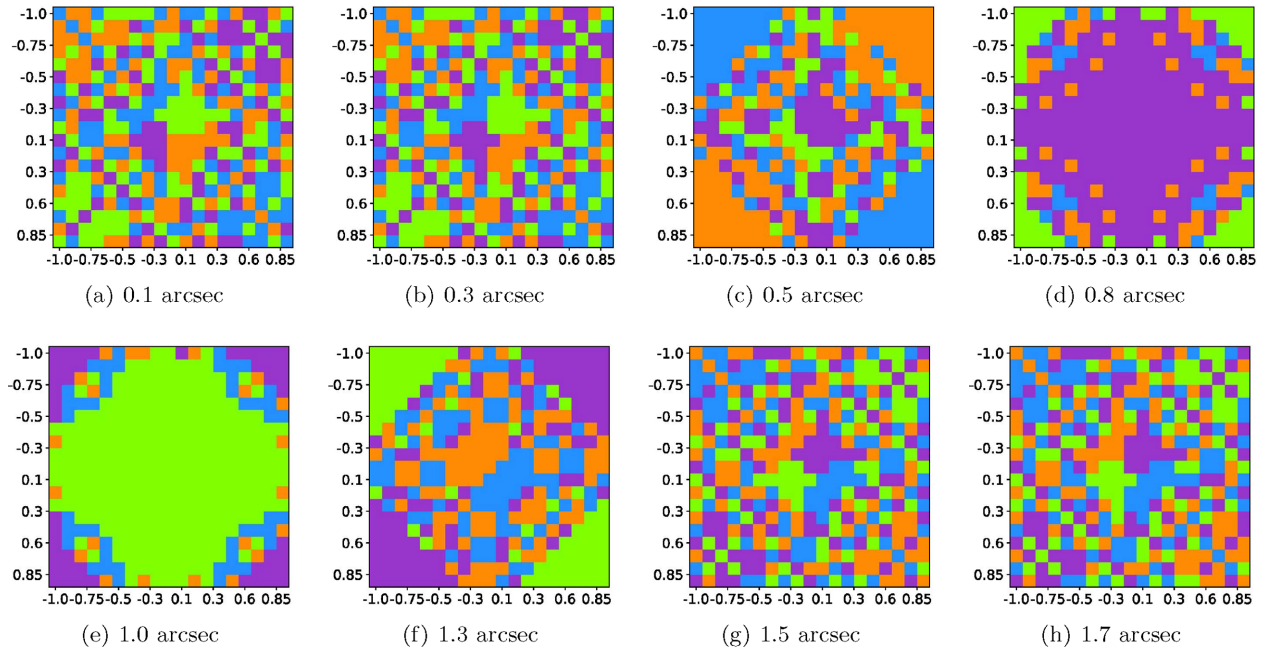
Figure 6. The distribution of PSFs in the whole field of view with different sampling rates. Different colours represent different categories and these colours are arbitrary. One block in this figure represents the area sampled by one PSF, whose size is 1×1 pixel^2. The size of each pixel varies from 0.1 to 1.7 arcsec. The *X*- axis is the percentage of the whole field of view between -100 and 90 per cent, and the *Y*- axis is the same. The centre of the focal plane is in the centre of the diagram. As the sampling rate increases, the PSF clusters will merge into a big cluster and then break into small pieces. For the low and high sampling rates, the clusters of different PSFs are very similar.
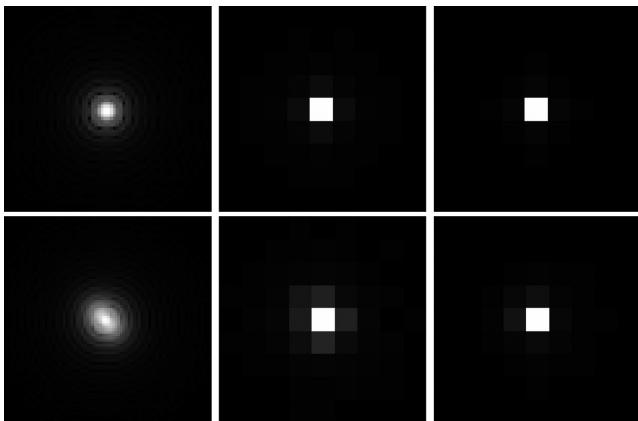
Figure 7. PSF with different sampling rates. Top: simulation PSFs are in centre of the field with pixel scale of 0.3, 1.0 and 1.3 arcsec respectively. Bottom: these PSFs are in the edge of the field with the same sampling rate of that in the top frames. It can be seen that with the high sampling rate (less than 0.5 arcsec in this telsescope), the wings of the PSFs dominate the difference between them and PSFs are clustered mainly depending on their wings. In contrast with the low sampling rate (bigger than 1.5 arcsec in this telescope), the disparity of these PSFs in different spatial field is so unclear that these PSFs are classified as one cluster.

be a circle and the centre of this cluster should be in the centre of the field of view.

In Fig. 6, as the sampling rate increases from 0.1 to 1.7 arcsec, the PSFs merge into a cluster in the centre of the field of view and then break into small pieces. Between 0.5 and 1.3 arcsec, our algorithm will give the same results as predicted by our assumptions at the start of this section. When the sampling rate is high or low, the cluster will be different from our prediction. This is for the following reasons.

For a high sampling rate, the wings of the PSFs introduce large differences and the clusters exist only in very small areas, as shown in two left panels in Fig. 7. For a low sampling rate, the differences of the PSFs are smoothed by the spatial discrete sampling. The cores of the PSFs are alike and only their wings are considered for classification, as shown in two right panels in Fig. 7. For this reason, the clusters of PSFs in low and high sampling rates are alike. Besides, for the low sampling rate, when the aberrations increase, the PSFs will merge into a big cluster again, as shown in Section 3.3. Hence, there exists a range inside which our classification algorithm can reveal the clusters best, and we need to consider this problem during real applications of our method.

### 3.3 Classification performance with different levels of aberrations

According to Section 3.2, we know that for a particular telescope, there is a limitation of the classification algorithm: we cannot effectively extract the clusters when the CCD sampling rate is too low or too high. In this subsection, we fix the sampling rate to the critical value for the flat-field Schmidt telescope (0.3 and 1.5 arcsec for high and low sampling rates, respectively) and we change the values of aberrations to test our classification algorithm performance with different levels of aberrations.

We separately set coefficients of the aberrations to 1.3, 1.5, 2 and 3 times the default values of the flat-field Schmidt telescope for PSFs in 1.5 arcsec and we decrease the values of the aberrations to 0.8, 0.6, 0.4 and 0.2 times the default values for PSFs in 0.3 arcsec. The variations of the clusters of PSFs are different, as shown in Fig. 8. The aberrations and sampling rates are shown below each figure and the initial aberration coefficients are defined in Table 3.

For a low sampling rate, as the coefficients of the aberrations increase, our method can identify PSF clusters to some extent. The
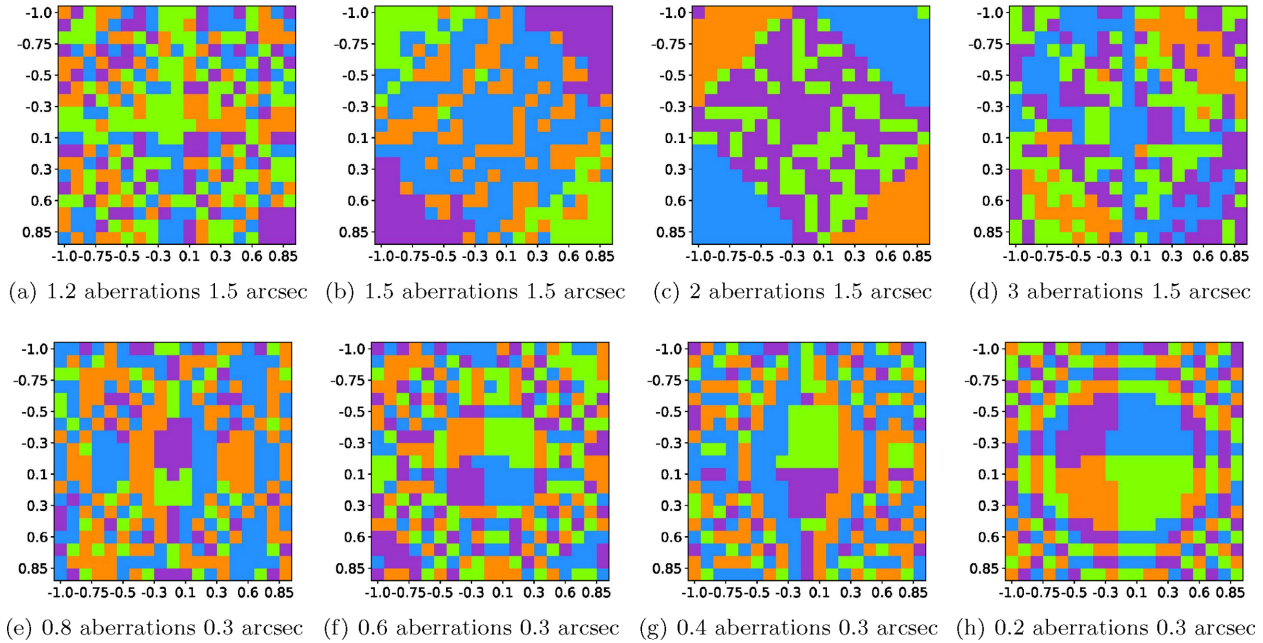
Figure 8. PSF distributions in the whole of field of view with different aberration coefficients for low and high sampling rates. Different colours represent different categories and these colours are arbitrary. One block in this figure represents the area sampled by one PSF, whose size is 1×1 pixel^2. The size of each pixel lies between 0.3 and 1.5 arcsec. The X-axis is the percentage of the field of view between −100 and 90 per cent, and the Y-axis is the same. The centre of the focal plane is in the centre of the diagram.
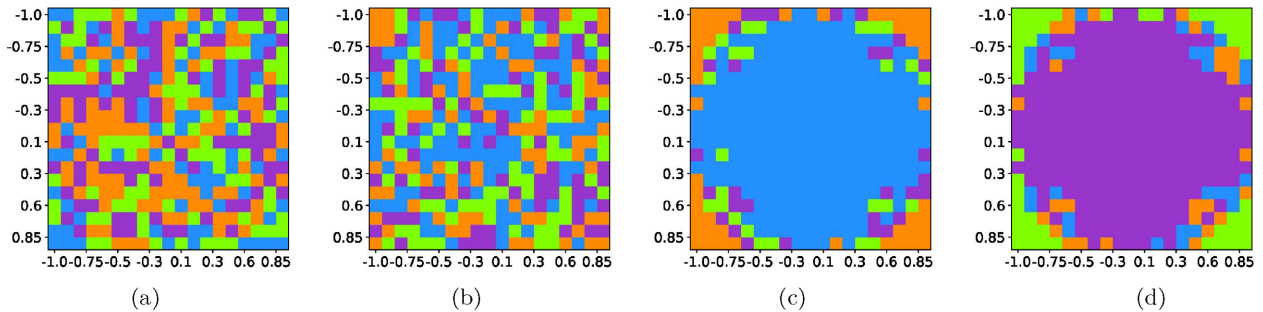
Figure 9. The panels show that various noise levels influence PSF classification and different (arbitrary) colours represent different categories. The S/N ratio is (a) 20, (b) 30, (c) 40 and (d) 49. The centre of the focal plane is in the centre of the diagram. One block represents the area sampled by one PSF, whose size is 1×1 pixel^2. Each pixel corresponds to 1.0 arcsec. The X-axis is the percentage of the field of view between −100 and 90 per cent, and the Y-axis is the same.

main variation of clusters occurs first in the corner of the field of view. However, when the values of the aberrations are too big, the difference of PSFs is so large that the cluster breaks into small pieces again, which indicates high spatial variations. We infer that this phenomenon might result from variations of the wings of the PSFs. However, for a high sampling rate, the variation of PSF clusters occurs mainly in the centre of the field of view. As the aberrations decrease, the small pieces of clusters join to make a bigger cluster, which indicates small spatial variations.

Above all, our classification algorithm can reveal the relationship between different aberrations and sampling rates. In real applications, we need to pay special attention to PSFs with low sampling rates and high aberrations or high sampling rates and small aberrations, because the boundaries of clusters will change significantly when the aberrations and sampling rates are in that range.

3.4 Classification performance with different S/N ratios

During real observations, the images of PSFs are affected by the S/N ratio of the images. We need to investigate the classification performance of our methods for images of PSFs with different S/N ratios. To fully test the classification performance, we simulate the CCD with a Monte Carlo simulation by adding different types of noise including Poisson photon noise, dark current noise, readout noise, etc.

We set the pixel scale to be 1.0 arcsec, as shown in Fig. 6(d). With this sampling rate, PSFs can be classified and a major cluster exist in the centre of the field of view with the default value of the aberrations. We generate PSFs with different S/N ratios. The PSFs with the same S/N ratio are classified to different clusters as shown in Figs 9(c) and (d). When the S/N ratio is lower than 30, the shapes
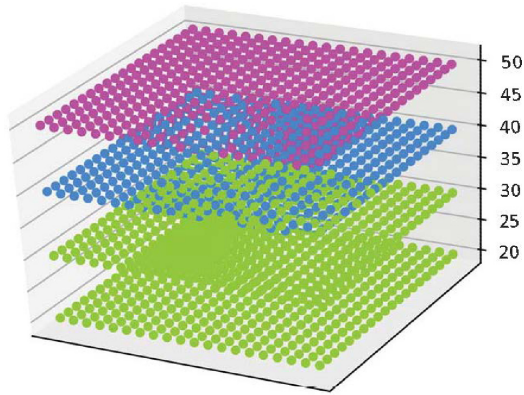
Figure 10. The figure represents the classification results of PSFs in the whole of field of view with different levels of CCD noise. The centre of the focal plane is in the centre of the diagram. The *Z*- axis is the level of the S/N ratio between 20 and 50.

Table 4. Parameters of telescopes.

	Telescope A	Telescope B
Diameter	300 mm	450 mm
Field of view	$8° \times 8°$	$1.9° \times 1.9°$
Image size	1024×1024	1024×1024
Pixel scale	7.2 arcsec	6.68 arcsec
Exposure time	30 s	4 s

of the PSFs are significantly affected by the random noise and the cluster breaks into small clusters, as shown in Figs 9(a) and (b).

During real applications, stars with different S/N ratios would be used as PSF references. To test the robustness of the classification algorithm, we classify all the PSFs with different levels of background, as shown in Fig. 10. We can see that PSFs with the same S/N ratio are classified as one cluster, which indicates that stars with different magnitudes introduce high bias to the classification. Therefore, we need to strictly distinguish the S/N ratio of stars that are used for PSF classification.

4 APPLICATION OF PSF CLASSIFICATION WITH OBSERVATION DATA

4.1 Observation instruments and data set

To test the performance of our algorithm, we obtained two series of images from two different small-aperture telescopes. Telescope A is operated by the School of Astronomy and Space Sciences, Nanjing University, and it is focused on exo-planet photometry. All the PSFs can be extracted directly from images of stars. We use these data to check the variations in observation conditions and to provide reference for further data processing and observation strategy planning for photometry. Telescope B is used for space debris astrometry (Sun & Zhao 2013; Sun, Zhao & Zhang 2014). Space debris consists of fast moving objects and we can observe these objects using the same observation strategy as that used for observation of asteroids and comets (Sun et al. 2015). This telescope is operated by Purple Mountain Observatory (PMO) and works under sidereal-tracking mode. The images of space debris are line-like and the images of celestial objects are point-like. We are able to obtain PSFs from the images of celestial objects and to use these data for PSF-fitting astrometry. The details of two telescopes are listed in Table 4.

4.2 Evaluation of observation conditions

With our classification algorithm, we classify 20 000 PSFs obtained by telescope A. These PSFs are randomly selected from all the observational data. We use the methods discussed above and find that the optimal number is 2. As shown in Fig. 11(a), images from one slice (from 2017 February 14–28) are classified as a different

cluster and the rest of the images from the whole data cube are classified as another cluster. After analysing these PSFs, we find that these PSFs have a lower S/N ratio than those from the other clusters. In Fig. 11(d), it can be seen that the flux of each mean PSF is different. Additionally, we choose images on the specific date (i.e. on February 10 and 23) to classify them independently and we find that their distributions are very different. The PSFs obtained on February 10 can be regarded as the same cluster in Figs 11(b) and (e), compared with the data obtained on February 23 in Figs 11(c) and (f). According to the classification results, this could indicate that the observation conditions have changed between February 14 and 24. We have checked the weather conditions, the phase of the moon and the telescope observation catalogue. During that period, we suspect that the full moon in the sky introduces very strong bias to the observation data. For differential photometry, if using the PSF model to fit different stars, we need to use two different models for these two days. If we use that telescope for transient detection through image difference (Alard & Lupton 1998; Zackay, Ofek & Gal-Yam 2016), we have to convolve the reference image with the mean PSF of different clusters for observation data obtained on different days.

4.3 Evaluation of PSF-fitting astrometry performance

We classify PSFs using all the data from telescope B, as shown in Figs 12(a) and (d). As we discussed in Section 3.4, the classification results are correlated with the S/N ratio of stars. Then, we classify PSFs from stars with different magnitudes separately, as shown in Figs 12(b) and (c).

PSFs from stars with a magnitude of 10 have apparently spatial–temporal related distributions and, for different clusters, the mean PSFs are apparently different, as shown in Fig. 12(e). This can be caused by the aberrations and we anticipate good astrometry results with the mean PSFs of each cluster. However, PSFs from stars with a magnitude of 11 have small and almost randomly distributed clusters. Although the mean PSFs of two different clusters are almost the same in Fig. 12(f), this might be induced by the average of many different PSFs.

To fully reflect the classification performance, we perform PSF-fitting astrometry in these data, using the Fourier-based registration method (Stone et al. 2001) with the mean PSFs obtained from our method. This method assumes that the PSFs in the same cluster are morphologically similar and only have a different subpixel shift. We can obtain this subpixel shift by maximizing the correlation coefficients between the stamp images of stars and the mean PSFs in an oversampled Fourier domain. The subpixel shift and the position of the brightest pixels in the images of celestial objects can then be used to calculate the star's position in the image.

For comparison, we also use the Gaussian function as a PSF template and we calculate the positions of stars with the least-squares fitting method (Stone 1989). The Gaussian function has a FWHM of 2 pixels. The positions we calculated with the above two
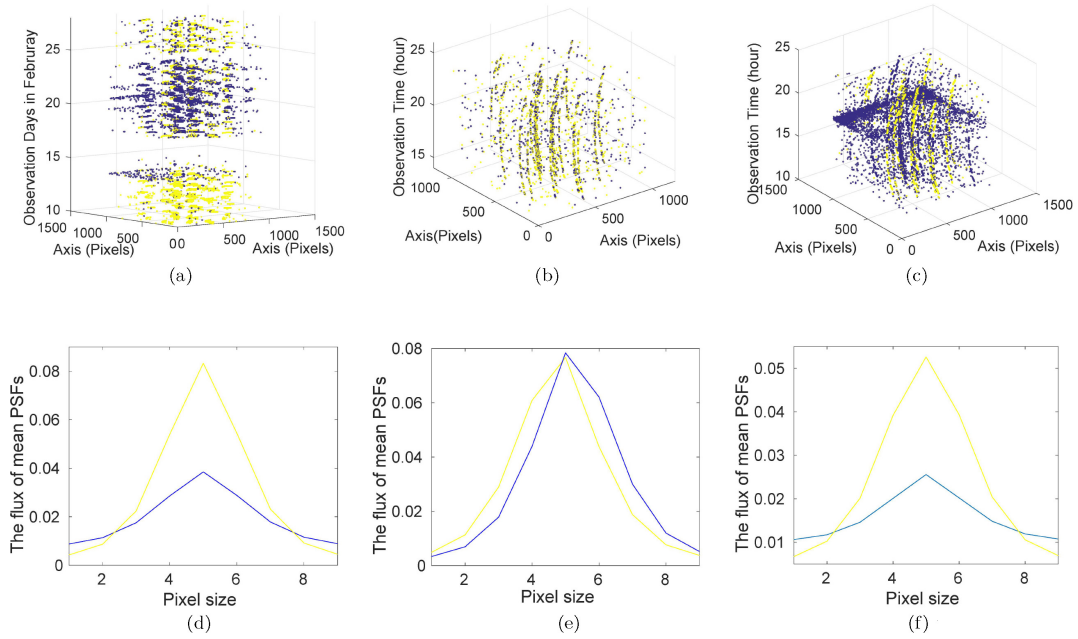
Figure 11. PSF classification results of telescope A. Panel (a) shows the result of 20 000 PSFs in all the observation data from 2017 February 10–28 and panel (d) shows the mean PSFs corresponding to each cluster in panel (a). Panels (b) and (c) show the PSF classification results on February 10 and 23, respectively. Panels (e) and (f) show the cross-section of each mean PSF corresponding to each cluster in (b) and (c), respectively. The colours also correpond to each cluster in the upper panels.
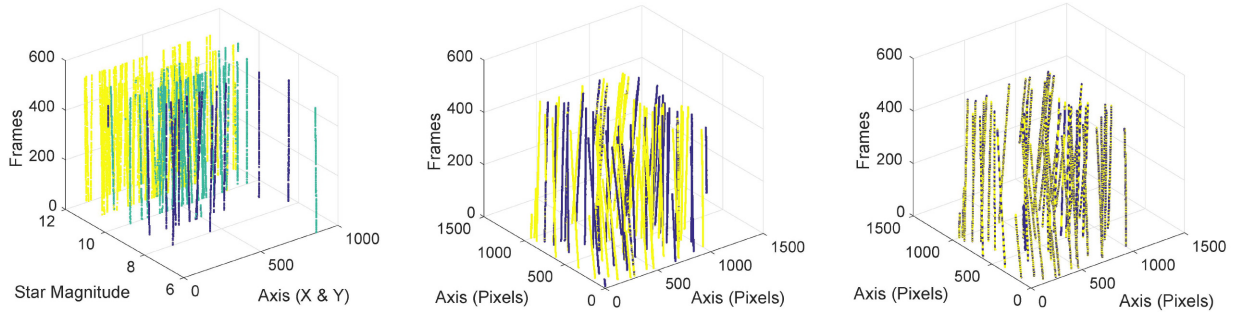
(a) the distribiton of 3 category PSFs according to stars with different magnitudes

(b) the distribution of PSFs in the magnitude of 10

(c) the distribution of PSFs in the magnitude of 11
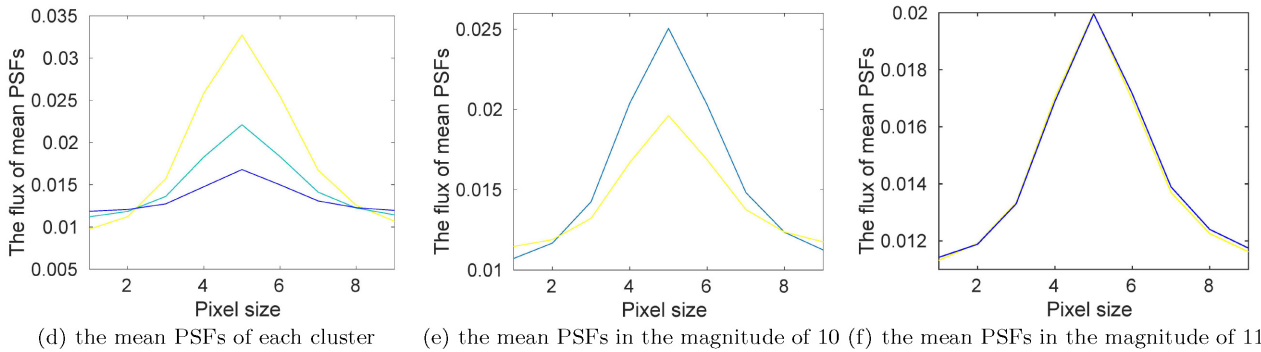
(d) the mean PSFs of each cluster

(e) the mean PSFs in the magnitude of 10

(f) the mean PSFs in the magnitude of 11

Figure 12. PSF classifcation results from telescope B. Panel (a) gives the results for all stars with PSFs and panel (d) shows the mean PSFs corresponding to each cluster. Panels (b) and (c) are the distributions of stars with magnitudes of 10 and 11 (results of other stars are the same as for stars with a magnitude of 11) and panels (e) and (f) correspond to each mean PSF. The mean PSF is the mean PSF of a cluster with the same colour. It can be seen that their stars relate to position differences for the magnitude of 10, while other stars with magnitude of 11 do not relate apparently.

methods are compared with the Tycho-2 catalogue (Høg et al. 2000) as follows.

We obtain the star positions (α_i, ξ_i) from the Tycho-2 catalogue as the theoretical positions of these stars and we construct the association between that position and the position of these stars in the image using

$$\xi = \frac{\cos\delta \sin(\alpha - \alpha_0)}{\sin\delta \sin\delta_0 + \cos\delta \cos\delta_0 \cos(\alpha - \alpha_0)}$$
$$\zeta = \frac{\sin\delta \cos\delta_0 - \cos\delta \sin\delta_0 \cos(\alpha - \alpha_0)}{\sin\delta \sin\delta_0 + \cos\delta \cos\delta_0 \cos(\alpha - \alpha_0)} \quad (16)$$

$$\xi = a + bx + cy$$
$$\zeta = d + ex + fy \quad (17)$$

as proposed in Sun, Zhao & Zhang (2013). Here, (α_0, δ_0) is the corresponding equatorial position of the frame centre. With this association, the observed position (α_c, ξ_c) of these stars can be obtained from the pixel position (x_c, y_c), and the residual position errors of each star $(\alpha_c - \alpha_i, \delta_c - \delta_i)$ are obtained.

The astrometry error of these two methods from the Tycho-2 catalogue is shown in Fig. 13. The systematic errors of Tycho-2 are nearly 0.01 arcsec so this does not have an effect on the calculation. Compared with the Gaussian fitting method, we can see that stars with a magnitude of 10 have better astrometric accuracy with the mean PSF and stars with a magnitude of 11 have almost the same astrometric accuracy.

It is worth noting that PSF classification is a clustering problem and for stars with the magnitude of 11, we do not know why the PSFs are not clustered distinctly. There are two possibilities: (i) these PSFs are nearly the same essentially, so applying the average of these PSFs for subsequent processing is a sound method (Sun & Jia 2017); (ii) the classification algorithm failed for stars with a magnitude of 11 so we should attempt to optimize the method.
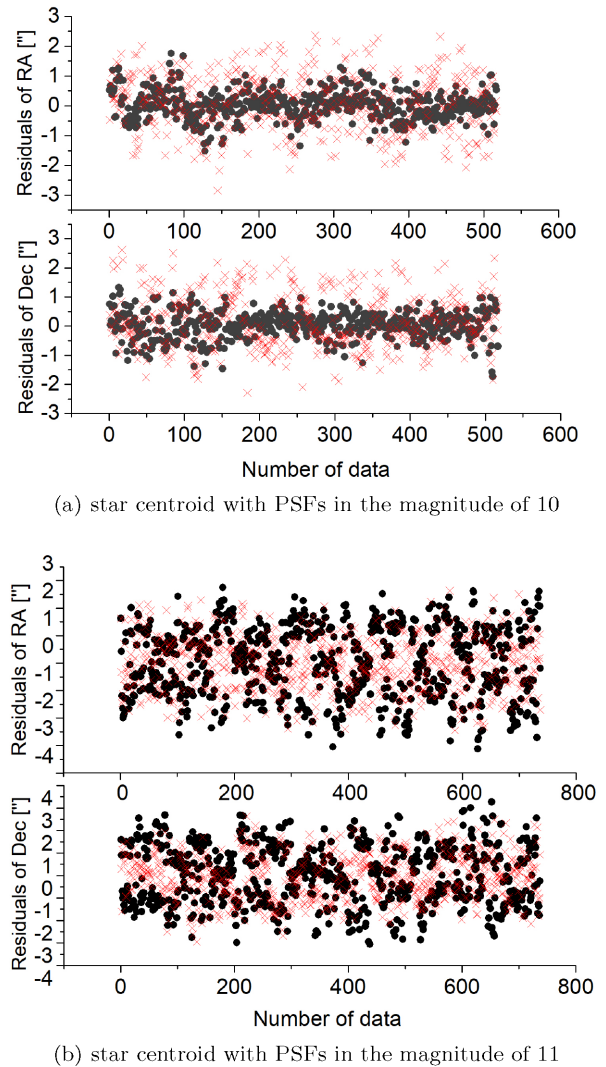
5 CONCLUSIONS AND FUTURE WORK

In this paper, we propose a clustering-based PSF evaluation method and we analyse the performance of our method with PSFs under different CCD sampling rates, levels of aberrations and S/N ratios. Moreover, we analyse data from two different telescopes and we use the classification results to evaluate observation conditions and PSF-fitting based astrometry.

For almost all observations, PSFs exhibit subtle differences, and we suggest that it is possible to find particular spatial and temporal regularities in order to increase scientific output. However, it is not always possible to find that distribution with the PSF clustering method. In the future, we should try to use a stronger neural network to classify PSFs and to generate the PSF model from observational data directly.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous referee for comments and suggestions that greatly improved the quality of this manuscript. We would like to thank the Key Laboratory of Space Object and Debris Observation, Purple Mountain Observatory, Chinese Academy of Sciences, for providing observation data for this paper. We would also like to thank the School of Astronomy and Space Sciences, Nanjing University, for providing observation data. We are grateful to Danfen Liu and Liwen Wang from Taiyuan University of Technology for providing suggestions for this paper. This

(a) star centroid with PSFs in the magnitude of 10

(b) star centroid with PSFs in the magnitude of 11

Figure 13. Astrometric residuals of stars with magnitudes of 10 and 11 of telescope B. The number of stars with the magnitude of 10 is 517 and the number with the magnitude of 11 is 737. Top: the residuals of right ascension. Bottom: the residuals of declination. Crosses indicate Gaussian PSF-fitting centroid and circles indicate mean PSF fitting.

REFERENCES

Alard C., Lupton R. H., 1998, ApJ, 503, 325
Anderson J., 2016, Technical report, Empirical Models for the WFC3/IR PSF
Anderson J., King I. R., 2000, PASP, 112, 1360
Bailey S., 2012, PASP, 124, 1015
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Bradley L. et al., 2016, preprint (ascl:1609.011)
Cain S. C., Watts T., 2016, Opt. Eng., 55, 085104

Davies D. L., Bouldin D. W., 1979, IEEE Trans. Pattern Anal. Mach. Intell., PAMI-1, 224

Desiderà G., Carbillet M., 2009, A&A, 507, 1759

Diolaiti E., Bendinelli O., Bonaccini D., Close L. M., Currie D. G., Parmeggiani G., 2000, Proc. SPIE, 4007, 879

Gai M., Cancelliere R., 2007, MNRAS, 377, 1337

Gai M., Busonero D., Cancelliere R., 2013, PASP, 125, 444

Hotelling H., 1933, Journal of Educational Psychology, 24, 417

Høg E. et al., 2000, A&A, 355, L27

Janout P., Páta P., Skala P., Fliegel K., Vítek S., Bednář J., 2016, Proc. SPIE, 9971, 99710F

Jee M. J., Blakeslee J. P., Sirianni M., Martel A. R., White R. L., Ford H. C., 2007, PASP, 119, 1403

Joye W. A., Mandel E., 2003, in Payne H. E., Jedrzejewski R. I., Hook R. N., eds, ASP Conf. Ser. Vol. 295. Astron. Data Anal. Softw. Syst. XII. Astron. Soc. Pac., San Francisco, p. 489

Kohonen T., 1982, Biological Cybernetics, 43, 59

Konnik M. V., Welsh J., 2011, CoRR, abs/1412.4031

Krist J. E., Hook R. N., Stoehr F., 2011, Proc. SPIE, 8127, 81270J

Kuijken K., 2008, A&A, 482, 1053

La Camera A., Schreiber L., Diolaiti E., Boccacci P., Bertero M., Bellazzini M., Ciliegi P., 2015, A&A, 579, A1

Li B-S., Li G., Cheng J., Peterson J., Cui W., 2016, Res. Astron. Astrophys., 16, 007

Perrin M. D., Soummer R., Elliott E. M., Lallo M. D., Sivaramakrishnan A., 2012, Proc. SPIE, 8442, 84423D

Perrin M. D., Sivaramakrishnan A., Lajoie C.-P., Elliott E., Pueyo L., Ravindranath S., Albert L., 2014, Proc. SPIE, 9143, 91433X

Piotrowski L. W. et al., 2013, A&A, 551, A119

Schechter P. L., Mateo M., Saha A., 1993, PASP, 105, 1342

Stone R. C., 1989, AJ, 97, 1227

Stone H., Orchard M., Chang E-C., Martucci S., 2001, IEEE Trans. Geosci. Remote Sens., 39, 2235

Sun R., Jia P., 2017, PASP, 129, 44502

Sun R-Y., Zhao C-Y., 2013, Res. Astron. Astrophys., 13, 604

Sun R. Y., Zhao C. Y., Zhang Y. P., 2013, PASJ, 65, 110

Sun R., Zhao C., Zhang X., 2014, AJ, 147

Sun R-y., Zhan J-w., Zhao C-y., Zhang X-x., 2015, Acta Astronaut., 110, 9

Zackay B., Ofek E. O., Gal-Yam A., 2016, ApJ, 830, 27

This paper has been typeset from a TEX/LATEX file prepared by the author.