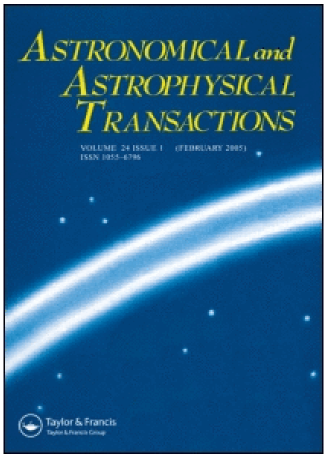

Astronomical & Astrophysical Transactions
The Journal of the Eurasian Astronomical Society
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713453505

The application of the Faint Object Classification and Analysis System (FOCAS) and Source Extractor (SExtractor) packages to the Digitized Sky Survey

M. Biernacka [a]; P. Flin [a]; J. Krywult [a]
[a] Institute of Physics, Pedagogical University, Kielce, Poland

Online Publication Date: 01 February 2006
To cite this Article: Biernacka, M., Flin, P. and Krywult, J. (2006) 'The application of the Faint Object Classification and Analysis System (FOCAS) and Source Extractor (SExtractor) packages to the Digitized Sky Survey', Astronomical & Astrophysical Transactions, 25:1, 75 - 94
To link to this article: DOI: 10.1080/10556790600849924
URL: http://dx.doi.org/10.1080/10556790600849924

Astronomical and Astrophysical Transactions
Vol. 25, No. 1, February 2006, 75–94

The application of the Faint Object Classification and Analysis System (FOCAS) and Source Extractor (SExtractor) packages to the Digitized Sky Survey

M. BIERNACKA*, P. FLIN and J. KRYWULT

Institute of Physics, Pedagogical University, ul. Swietokrzyska 15, 25-406 Kielce, Poland

(Received 5 June 2006)

In this paper the comparison between the Faint Object Classification and Analysis System (FOCAS) and Source Extractor (SExtractor) packages for automated star–galaxy classification is discussed. The analysis presented here uses a data set containing 43 Abell clusters of galaxies obtained from the Digitized Sky Survey. Comparison is based on four image parameters: the magnitude, area, ellipticity and position angle computed for each object. Statistical analysis shows that in the case of individual clusters, in general, there is good correlation between parameters obtained from both packages. The comparison of visual classification with automatic classification from FOCAS shows that the percentage of misclassified galaxies is about 20.

Keywords: Galaxy clusters; Catalogues; Faint Object Classification and Analysis System (FOCAS); Source Extractor (SExtractor); Digitized Sky Survey

1. Introduction

Many current astrophysical studies are carried out on catalogues of objects. This includes, for example, studies of the number counts of galaxies, clustering properties and colour–magnitude distribution. Since the late 1970s and the beginning of the 1980s, various computer programs have been developed to create galaxy catalogues automatically from astronomical images [1], e.g. the Faint Object Classification and Analysis System (FOCAS) [2], the Rome Observatory software [3], the Münster Red Sky Survey [4], the APM software [5], the COSMOS system [6], the PPP package [7] and Source Extractor (SExtractor) [8] which is based on a neural network.

Flin and Vavilova [9] showed that, for scans of old astronomical plates, obtained with the 48 in Palomar Schmidt telescope in the 1960s, where some cautions connected mainly with the necessity of visual checking are given, both the COSMOS and the Münster software can be successfully applied. Philip *et al.* [10] described the use of a new artificial neural network, the difference-boosting neural network (DBNN), for automated classification problems in astronomical data analysis. Comparison of the DBNN technique with the widely used

*Corresponding author. Email: bmonika@pu.kielce.pl

Astronomical and Astrophysical Transactions
ISSN 1055-6796 print/ISSN 1476-3540 online © 2006 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/10556790600849924

SExtractor package was described [10]. It was shown that, while the performance of the DBNN in star–galaxy classification is comparable with that of SExtractor, it has the advantage of a significantly higher speed and flexibility during training as well as classification. The original motivating purpose for FOCAS and SExtractor was the measurement of large numbers of objects to obtain a good sample for analysis. The FOCAS package contains over 100 programs which are used to make catalogues of objects from digitized photographic plates. The creation of the catalogue is accomplished by an automatic threshold detector where the threshold is measured relative to a simultaneously determined background. FOCAS was originally designed to optimize the detection and classification of faint objects. Nowadays the system can work with all kinds of digitized pictures and analyse many data. FOCAS calculates the catalogue parameters using the moments of pixel distribution in an object. There are three steps from the basic image to the object list in FOCAS: segmentation, area assembly and object evaluation. In the third step, the catalogue of objects is created. The various parameters characterizing the individual images in the segmented areas are calculated. In FOCAS the location of the object is defined by the centroids

$$\bar{x} = \frac{1}{M_{00}} \sum_{A} x_i [I(x, y) - I_{\rm s}], \tag{1}$$

$$\bar{y} = \frac{1}{M_{00}} \sum_{A} y_i [I(x, y) - I_{\rm s}], \tag{2}$$

where M_{00}, the zero moment, is

$$M_{00} = \sum_{A} [I(x, y) - I_{\rm s}]. \tag{3}$$

The summation over A means that the sum includes all pixels in the object-defining area A. $I(x, y)$ is the intensity corresponding to the density at the location (x, y) in the digital plate image. $I_{\rm s}$ is the intensity corresponding to the average plate density at the object location. Shape information about the object is obtained from the higher central moments

$$M_{ij} = \sum A(x - \bar{x})^i (y - \bar{y})^j [I(x, y) - I_{\rm s}], \quad i + j = 2, 3, \ldots. \tag{4}$$

FOCAS calculates the size-, rotation- and translation-invariant quantities from the central moments:

$$C_2 = \frac{M_{20} + M_{02}}{M_{00}}, \tag{5}$$

$$C_4 = \frac{M_{40} + 2M_{22} + M_{04}}{M_{00}}, \tag{6}$$

$$E = \frac{1}{M_{02} + M_{20}} [(M_{20} - M_{02})^2 + 4M_{11}^2]^{1/2}, \tag{7}$$

where C_2 and C_4 are the total second and fourth moments, respectively, while E is a parameter that measures the elongation of the object. The position angle of the object is calculate using the central moments [2]:

$$\tan(2\theta) = \frac{2M_{11}}{M_{20} - M_{02}}. \tag{8}$$

SExtractor is newer software especially designed to process, in batch mode, large digital images (up to 60 000 × 60 000 pixels), mainly analysis of charge-coupled device (CCD) frames.

It is a program based on a neural network. Special attention has been paid to speed and robustness in the extraction of objects in the image, regardless of their shape and size. The complete analysis of an image is achieved in six steps: estimation of the background, thresholding, deblending, filtering of the detected images, photometry and star–galaxy separation. We are interested in the measurement phase where the main parameters describing the objects were calculated. In SExtractor there are two categories of measurements. The first category made measurements from the isophotal object profiles (such as X_{image} and Y_{image}) and only pixels above the detection threshold are considered. The second-category measurements have access to all pixels of the image and are more sophisticated. In SExtractor the positional parameters are derived from the spatial distribution S of pixels detected above the extraction threshold and the pixel values I_i are taken from the filtered detection image. The position limits ($X_{\min}$, $Y_{\min}$, $X_{\max}$ and $Y_{\max}$) define two corners of a rectangle which encloses the detected object. The barycentre coordinates (X, Y) are computed as the first-order moments of the profile [11]:

$$M_{10} = \bar{x} = \frac{\sum_{i\in S} I_i x_i}{\sum_{i\in S} I_i}, \tag{9}$$

$$M_{01} = \bar{y} = \frac{\sum_{i\in S} I_i y_i}{\sum_{i\in S} I_i}. \tag{10}$$

The second-order moments analysis is convenient for measuring the spatial spread of a source profile. In SExtractor they are computed with

$$M_{20} = x^2 = \frac{\sum_{i\in S} I_i x_i^2}{\sum_{i\in S} I_i} - \bar{x}^2, \tag{11}$$

$$M_{02} = y^2 = \frac{\sum_{i\in S} I_i y_i^2}{\sum_{i\in S} I_i} - \bar{y}^2, \tag{12}$$

$$M_{22} = \overline{xy} = \frac{\sum_{i\in S} I_i x_i y_i}{\sum_{i\in S} I_i} - \overline{xy}. \tag{13}$$

From the second-order moments, SExtractor computes the basic shape parameters: A (the semimajor axis), B (the semiminor axis), θ (the position angle between the A axis and the x axis of the reference frame of the field counted in the counterclockwise direction). The semimajor axis is computed using

$$A^2 = \frac{x^2 + y^2}{2} + \left[\left(\frac{\overline{x^2} - \overline{y^2}}{2} \right)^2 + \overline{xy}^2 \right]^{1/2}, \tag{14}$$

the semiminor axis using

$$B^2 = \frac{x^2 + y^2}{2} - \left[\left(\frac{\overline{x^2} - \overline{y^2}}{2} \right)^2 + \overline{xy}^2 \right]^{1/2} \tag{15}$$

and the position angle using

$$\tan(2\theta) = 2\frac{\overline{xy}}{\overline{x^2} - \overline{y^2}}. \tag{16}$$

In our catalogues, we obtain the ellipticity parameter which is directly derived from A and B:

$$\text{ellipticity} = 1 - \frac{B}{A}. \tag{17}$$

Both packages use the moments of pixel distribution to compute the basic shape parameters such as the ellipticity and position angle. In both cases the procedures for calculating the shape parameters are the same. The estimations of the background are different for FOCAS and SExtractor. SExtractor constructs the 'background map' to calculate the precise estimation of the background level at any place of the image. SExtractor use the combination of the $\kappa\sigma$ clipping estimator and the mode estimator. The local background histogram is clipped iteratively until convergence at $\pm 3\sigma$ around its median. If σ is changed by less than 20% during that process, the field is uncrowded and the value for the background is the mean of the clipped histogram. Otherwise the mode is estimated (mode $=$ $2.5 \times$ median $- 1.5 \times$ mean) [8]. The method used to estimate the sky background in FOCAS tabulates a histogram of pixel values in the rectangular band surrounding the object. The mean of the distribution is computed, and a final mean is computed from sky values within ± 3 times the rms noise for object-free areas of the plate of the original computed mean. FOCAS also estimates the mode from mode $= 3 \times$ median $- 2 \times$ mean [12]. The parameters for subdivision of multiple objects in both pieces of software were optimized to separate galaxies correctly from stars. Both convolution filters are identical. If this conviction is correct, it is important to check whether the catalogues from the automatic procedure are error free. In our opinion, the best way to test the accuracy of parameters is to compare catalogues of galaxies obtained from applications of two different software packages to the same data.

2. Observational data and description of catalogue

The analysis presented here uses a data set consisting of 43 Abell clusters of galaxies obtained from the Digitized Sky Survey (DSS) [13]. All analysed clusters have a richness class $R < 4$ and a distance class $D < 4$. The area covering 4 Mpc $\times$ 4 Mpc on the sky for each cluster has been selected ($h = 0.75$; $q_0 = 0.5$).

Applying the FOCAS and SExtractor packages to the digitized scans of the Palomar Sky Survey, we have obtained two independent sets of catalogues. In each catalogue, we have the following data: right ascension, declination, position of the object centre, major semiaxis of the galaxy image and its position angle, object ellipticity, instrumental magnitude, object area and flags. Visual verification of the automated star–galaxy classification for FOCAS was carried out. From visual inspection we learned that most stars brighter than 17.5 instrumental magnitude were classified as galaxies. Moreover, we find that both programs recorded photographic plate errors as galaxies.

The FOCAS software uses the point spread function to perform star–galaxy separation and allocates a type to the object depending on its geometry. SExtractor is based on a neural network to assess the light distribution of the objects and assigns a 'stellar index' to each object ranging between 0 (galaxy) and 1 (star) [14]. All objects with a stellar index greater than 0.5 were considered as stars and have been excluded from the catalogues.

We found that small images drastically lowered the correlation coefficient. After several attempts we decided to rejected small objects with an area less than 150 arcsec^2 because they lie in the background noise.

We chose four parameters from both packages, namely the instrumental magnitude, ellipticity, position angle and area of object, as the observational basis of our study. The comparison was performed by plotting on one graph the parameters obtained from SExtractor on the y axis and from FOCAS on the x axis. For each plot the correlation between the investigated parameters was calculated.

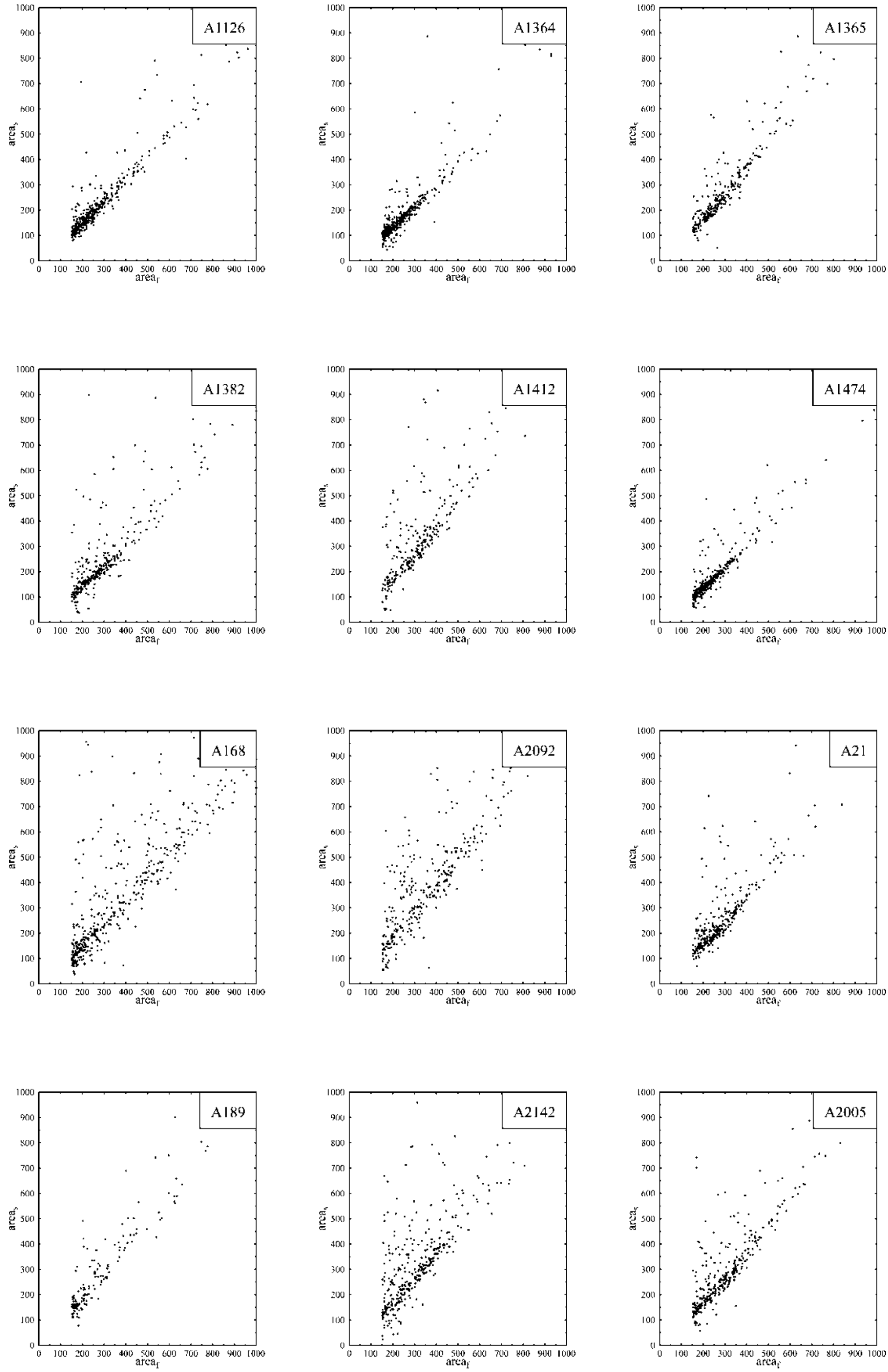

Figure 1. Correlation of the areas for all galaxies.

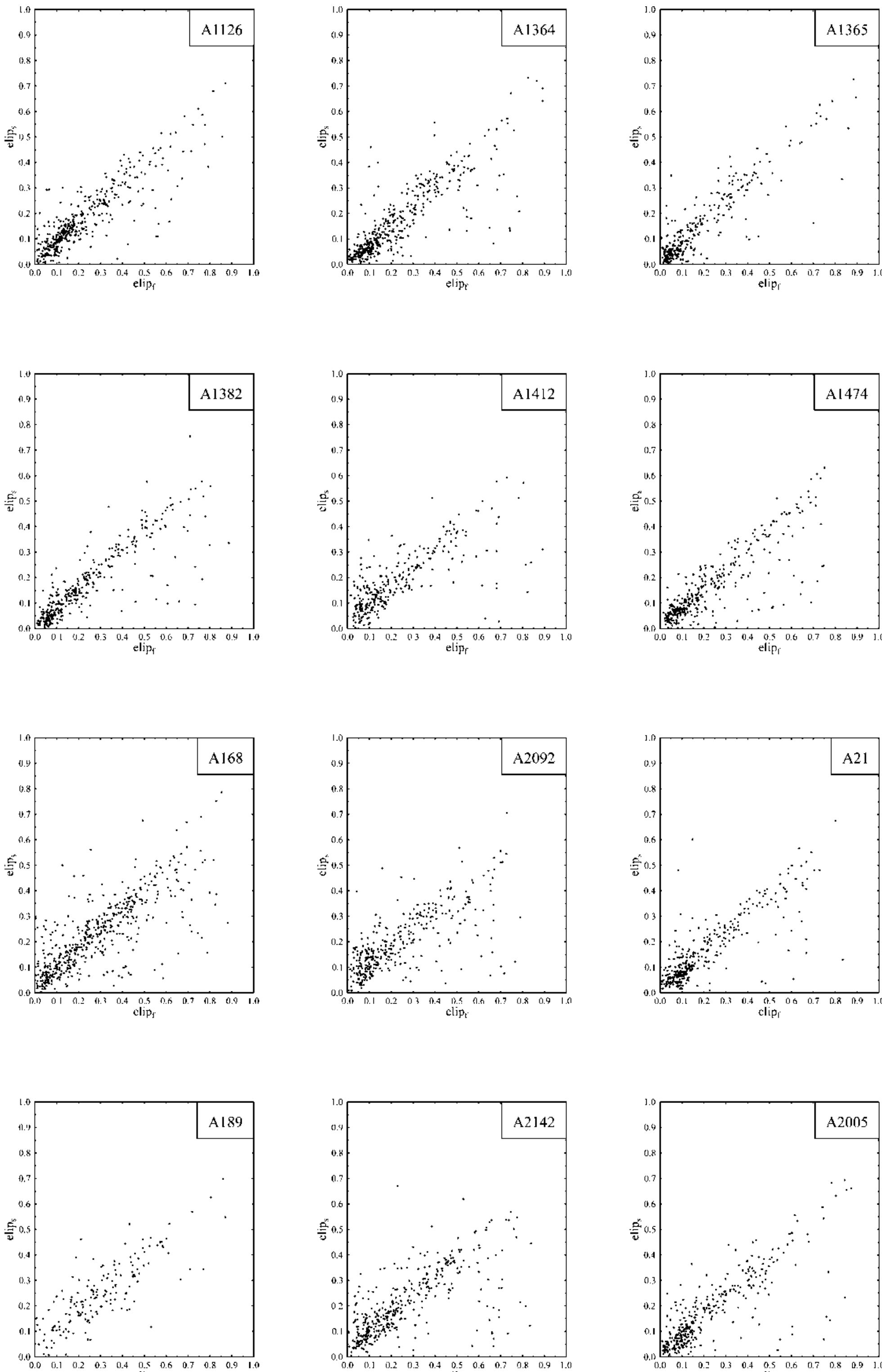

Figure 2. Correlation of the ellipticities for all galaxies.

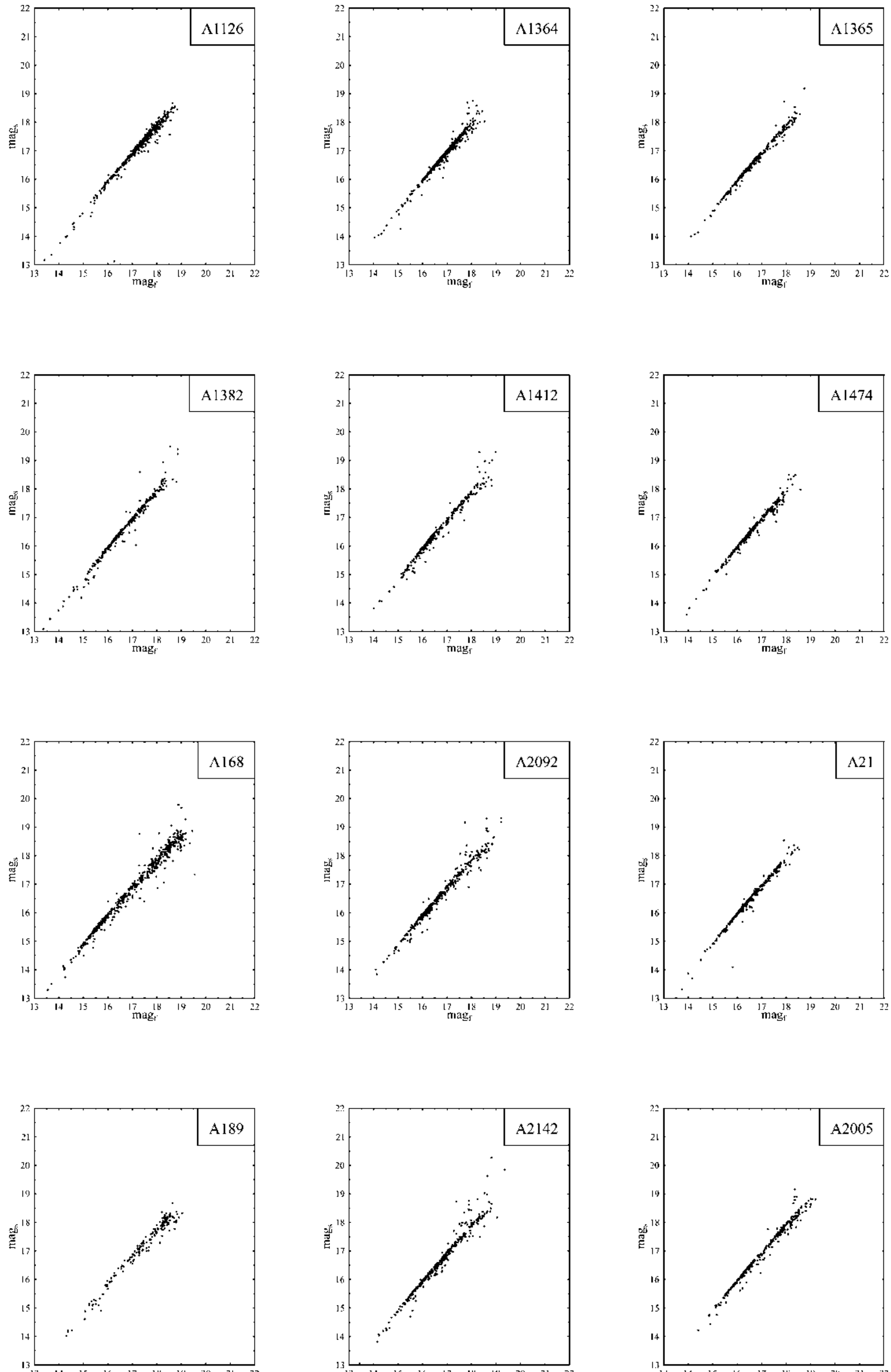

Figure 3. Correlation of the instrumental magnitudes for all galaxies.

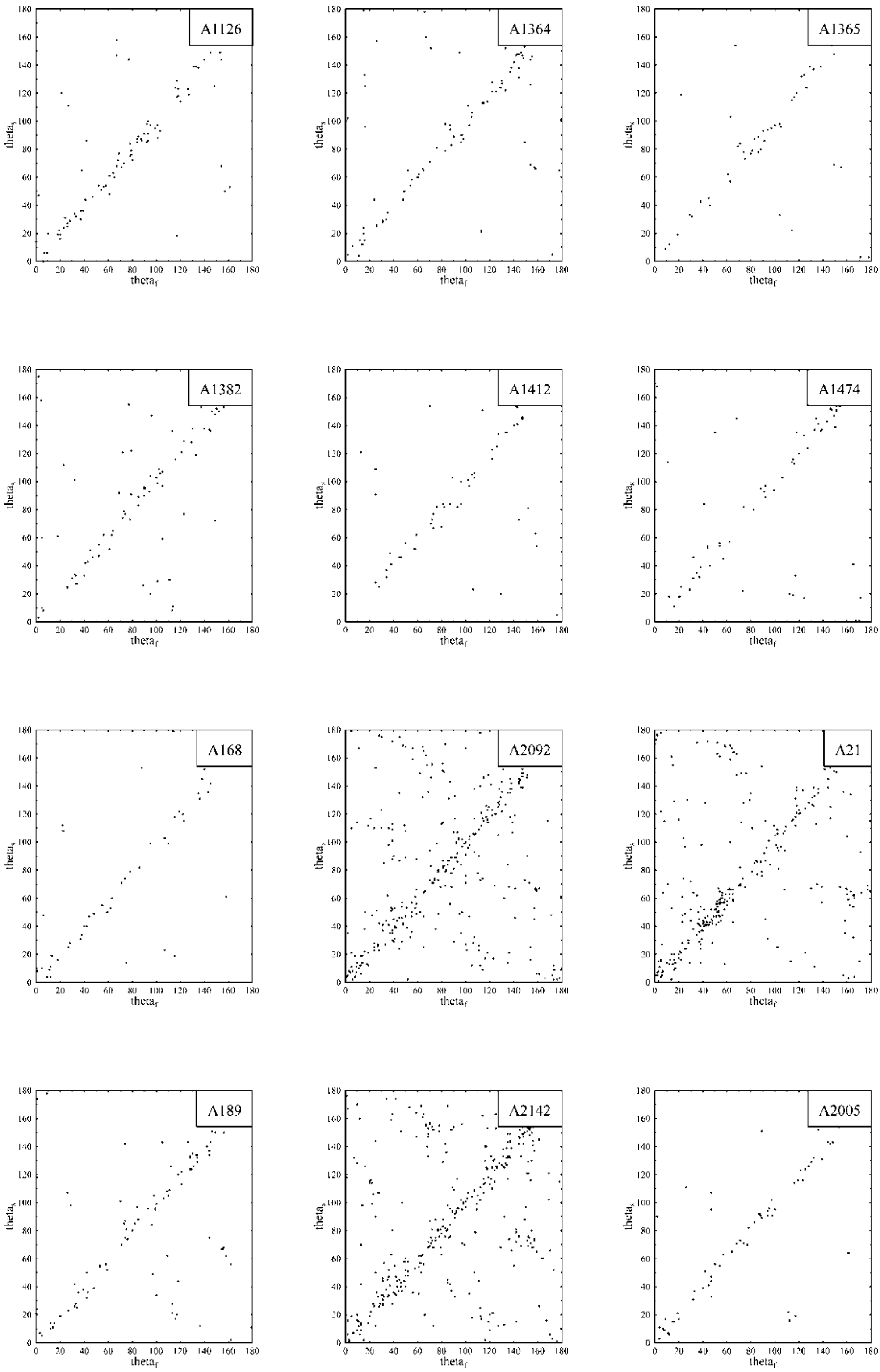

Figure 4. Correlation of the position angles for all galaxies.

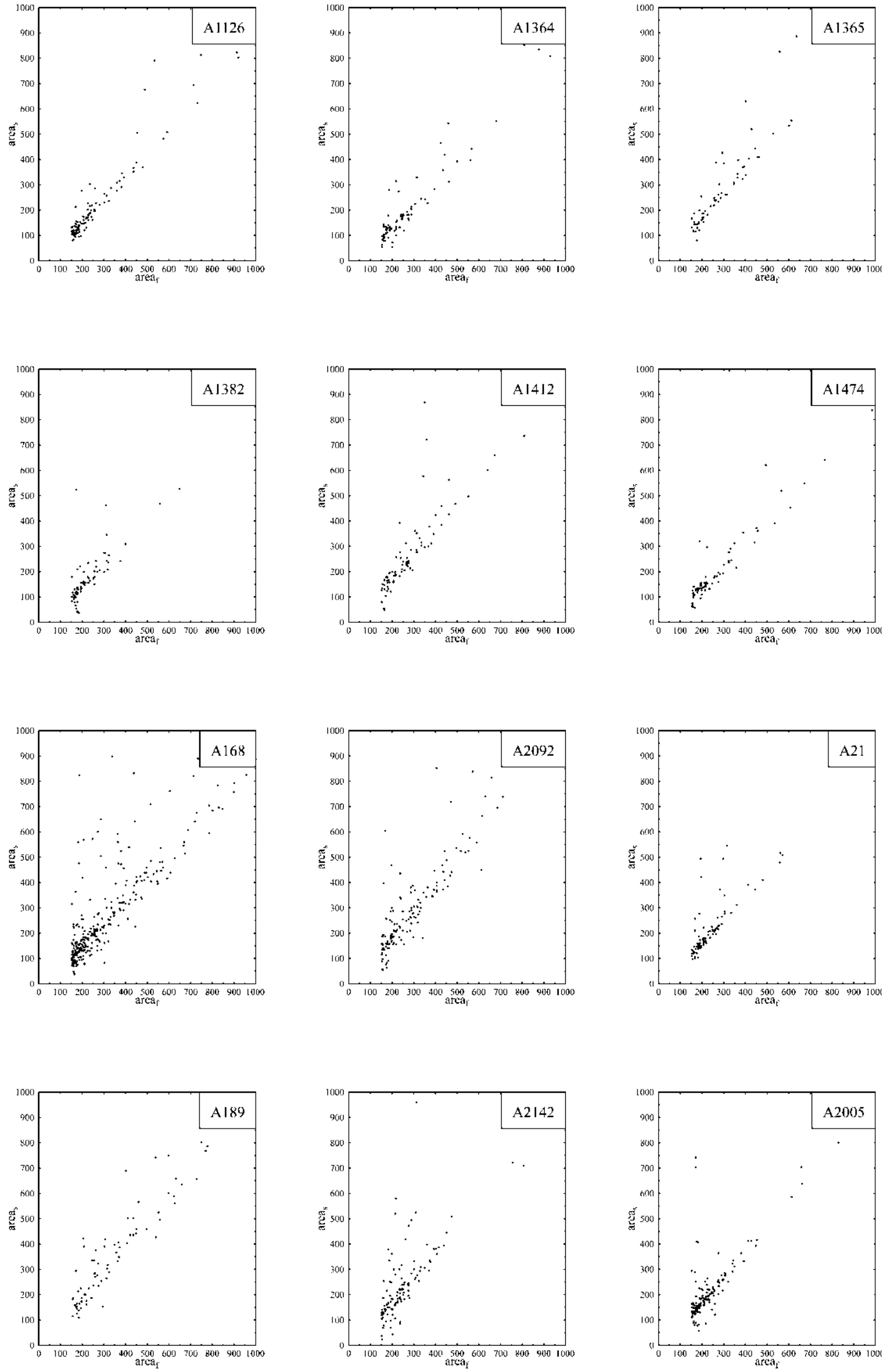

Figure 5. Correlation of the areas for all galaxies after visual classification.

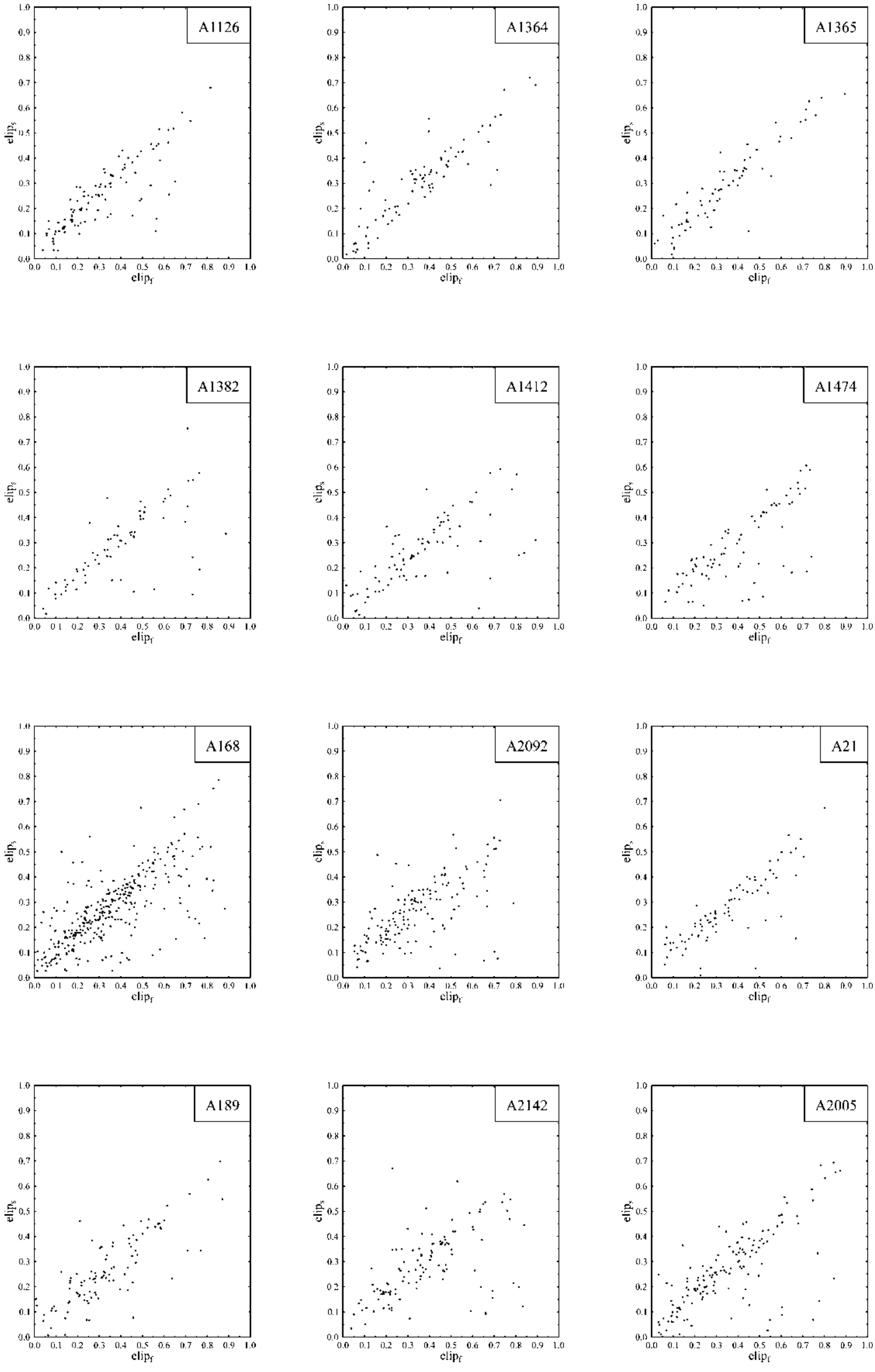

Figure 6. Correlation of the ellipticities for all galaxies after visual classification.

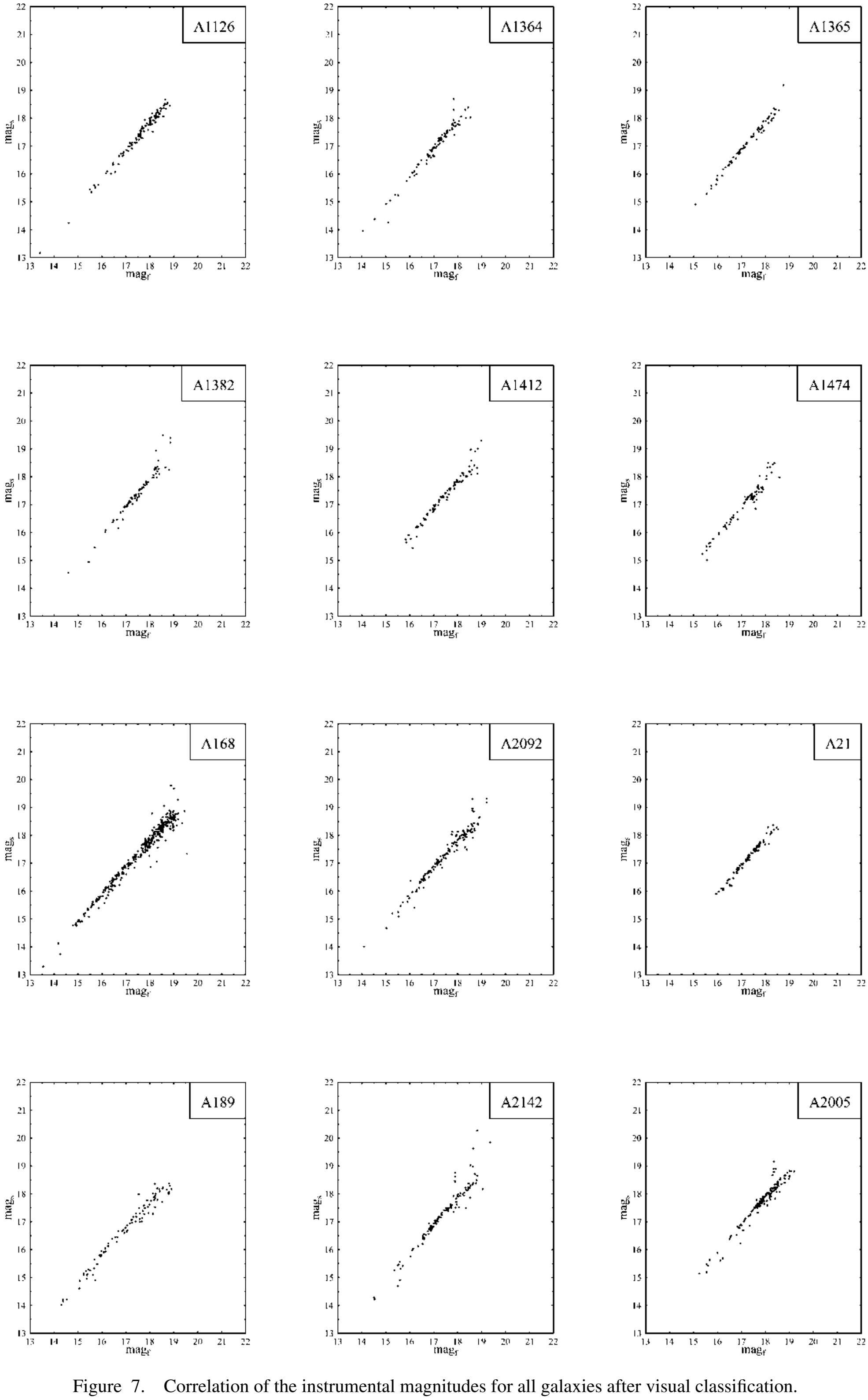

Figure 7. Correlation of the instrumental magnitudes for all galaxies after visual classification.

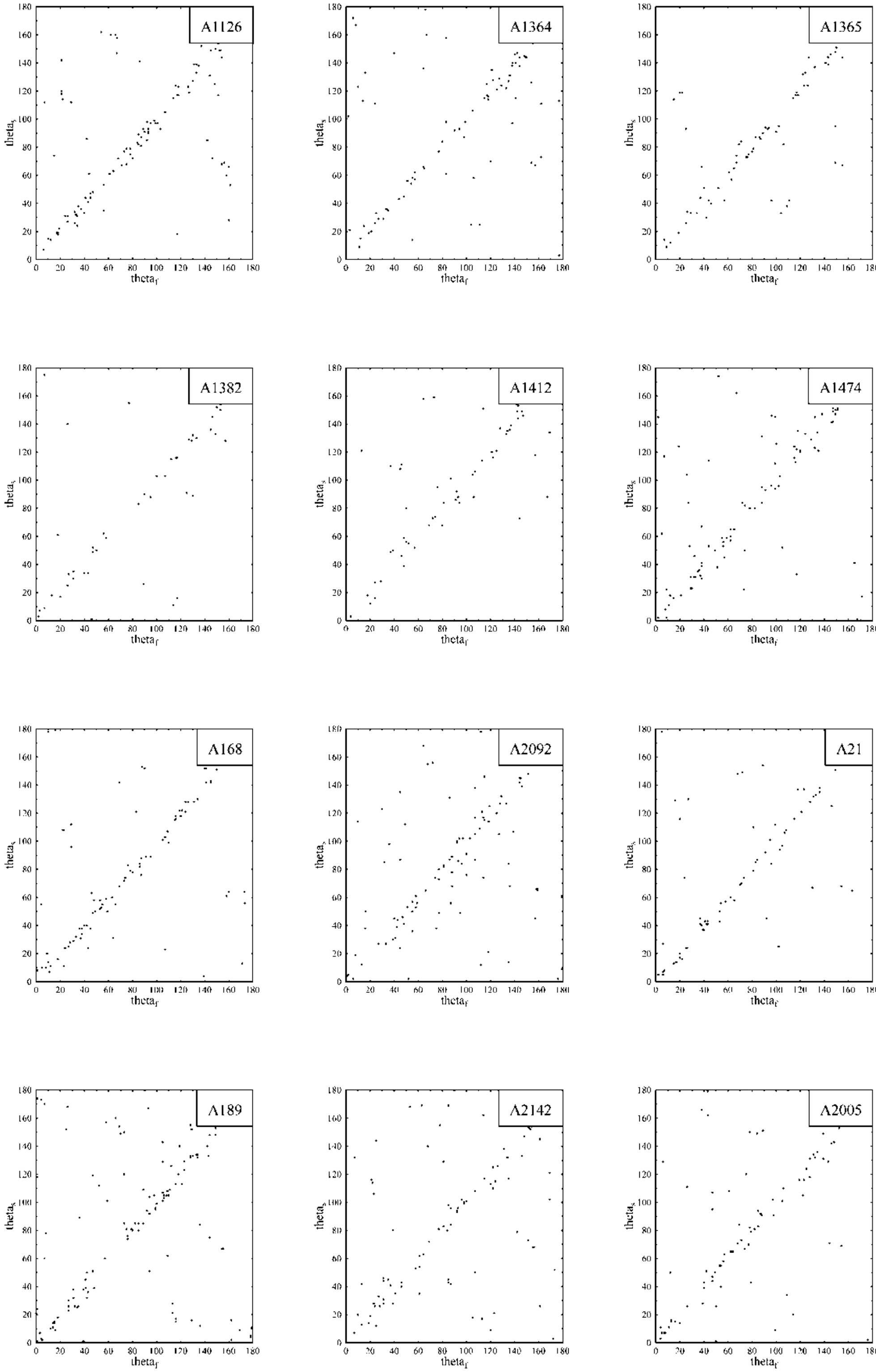

Figure 8. Correlation of the position angles for all galaxies after visual classification.

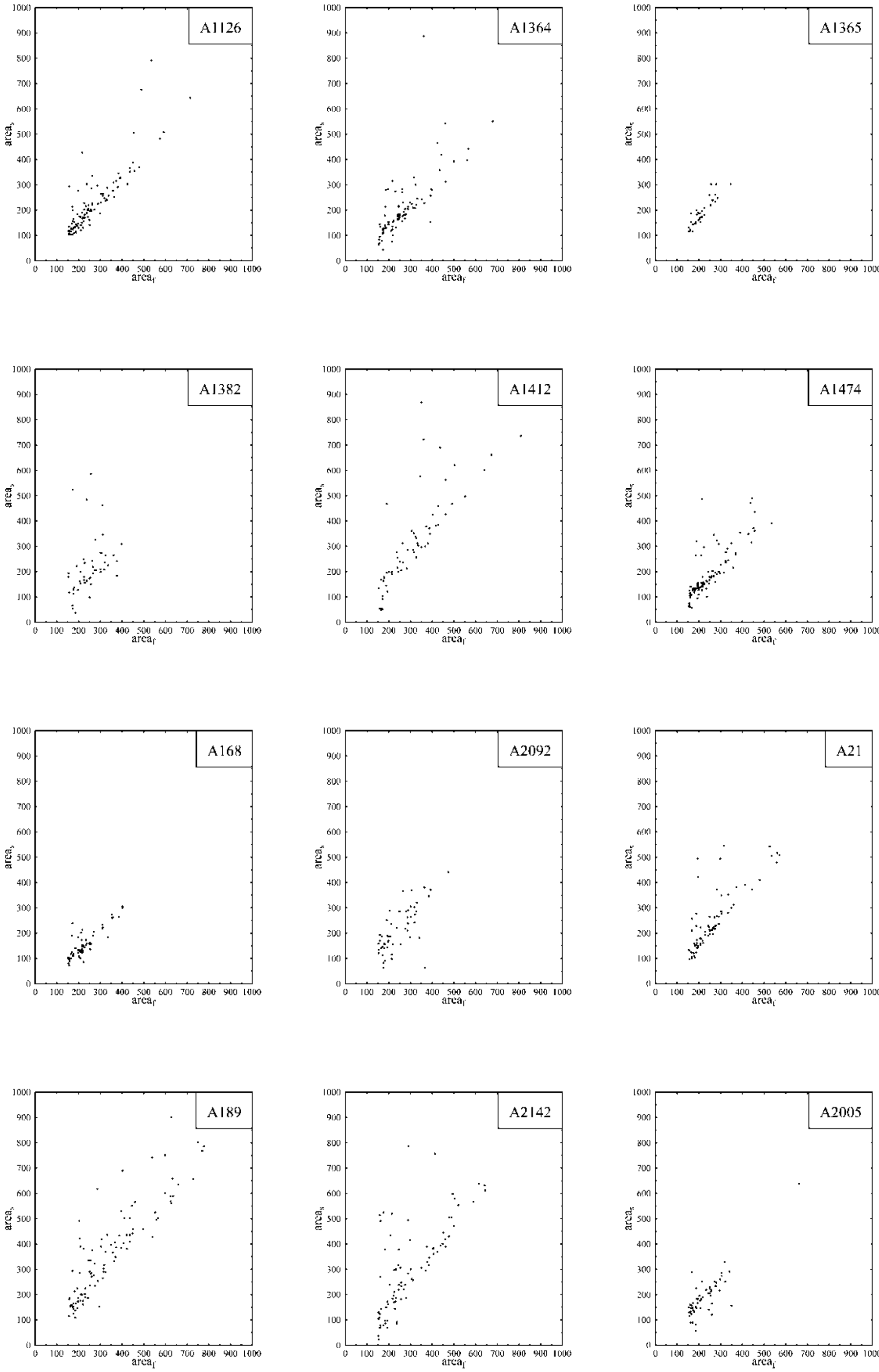

Figure 9. Correlation of the areas for galaxies after visual classification within the magnitude range $m_3 + 3$.

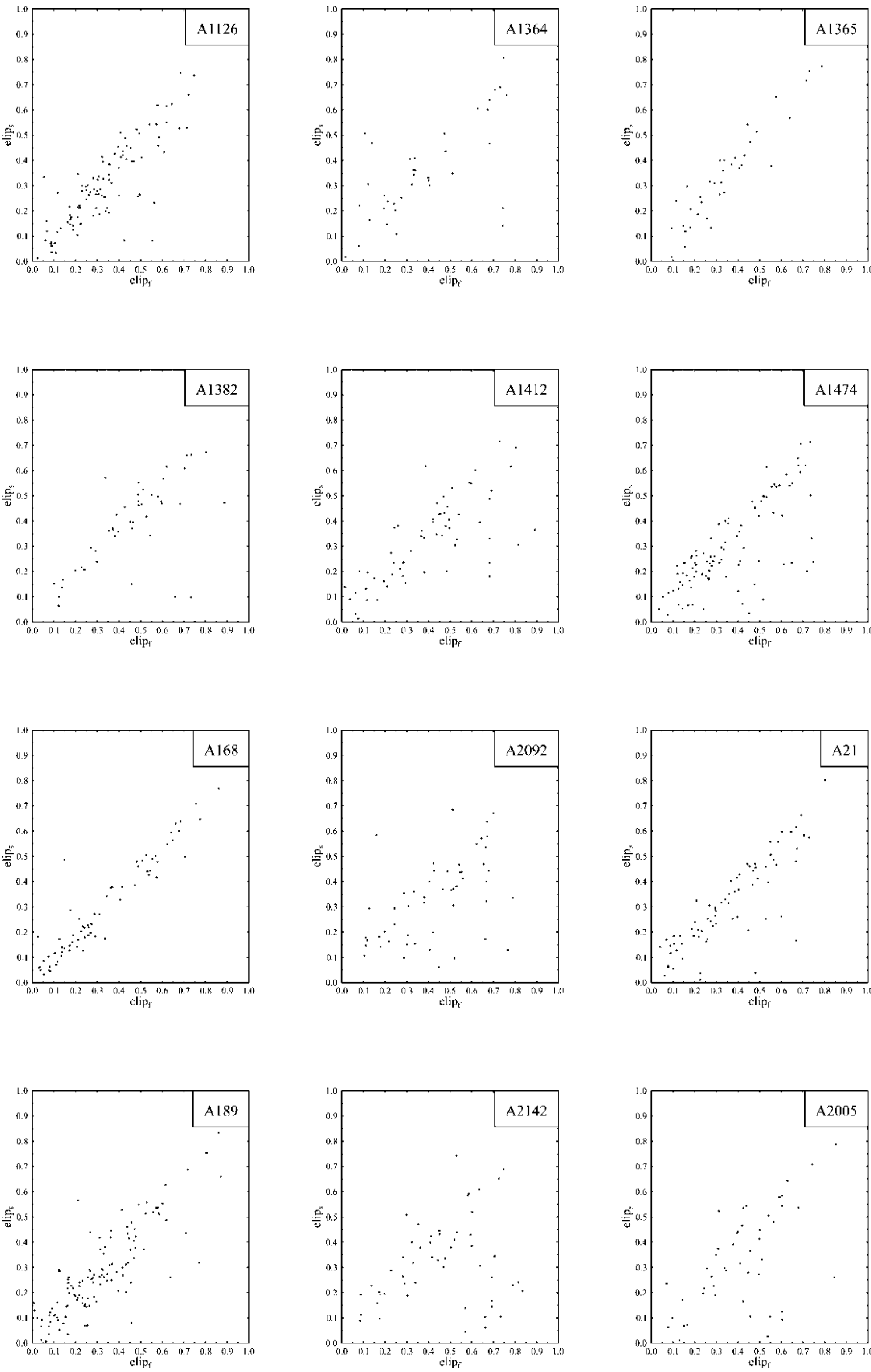

Figure 10. Correlation of the ellipticities for galaxies after visual classification within the magnitude range $m_3 + 3$.

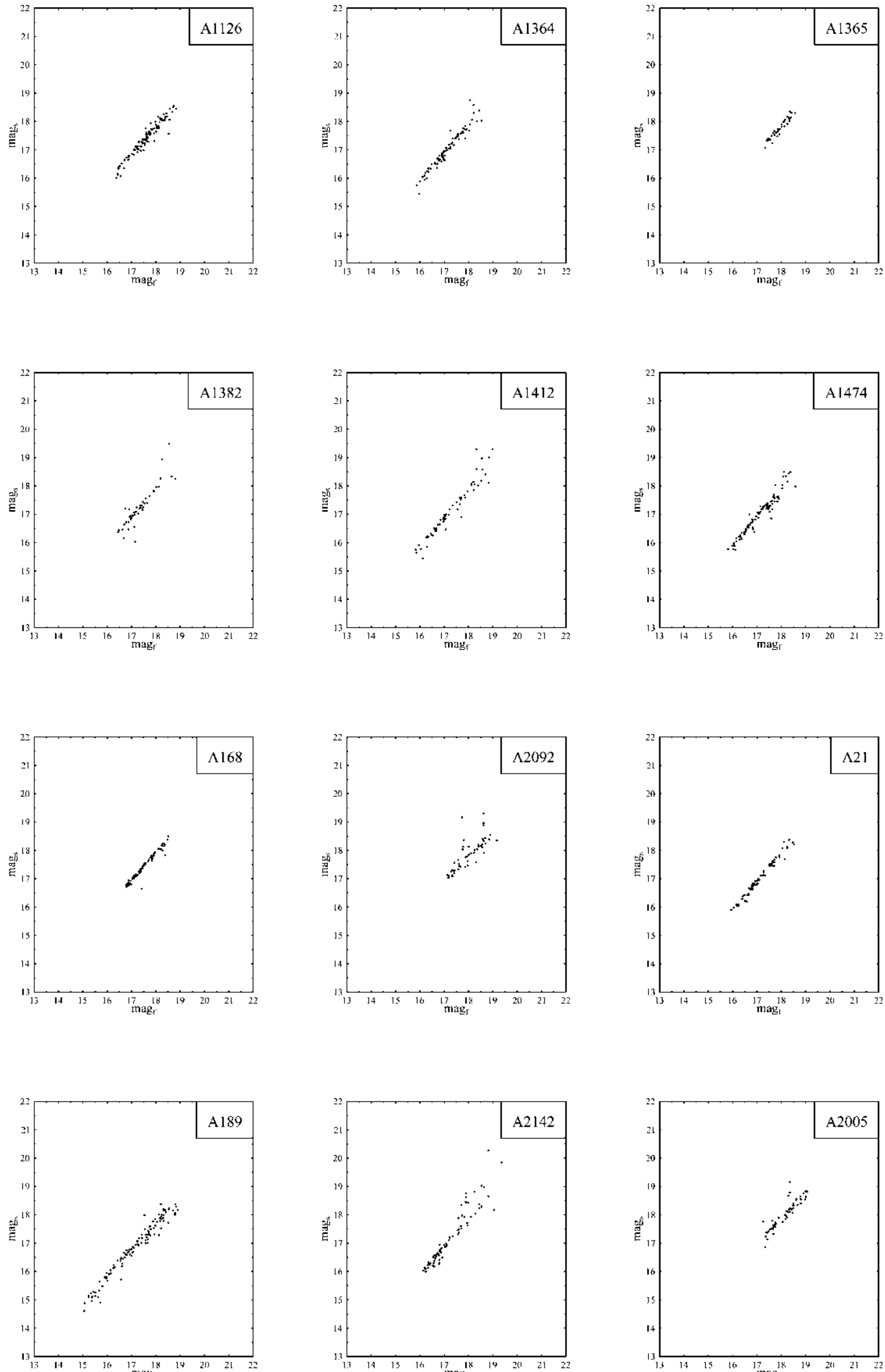

Figure 11. Correlation of the instrumental magnitudes for galaxies after visual classification within the magnitude range $m_3 + 3$.

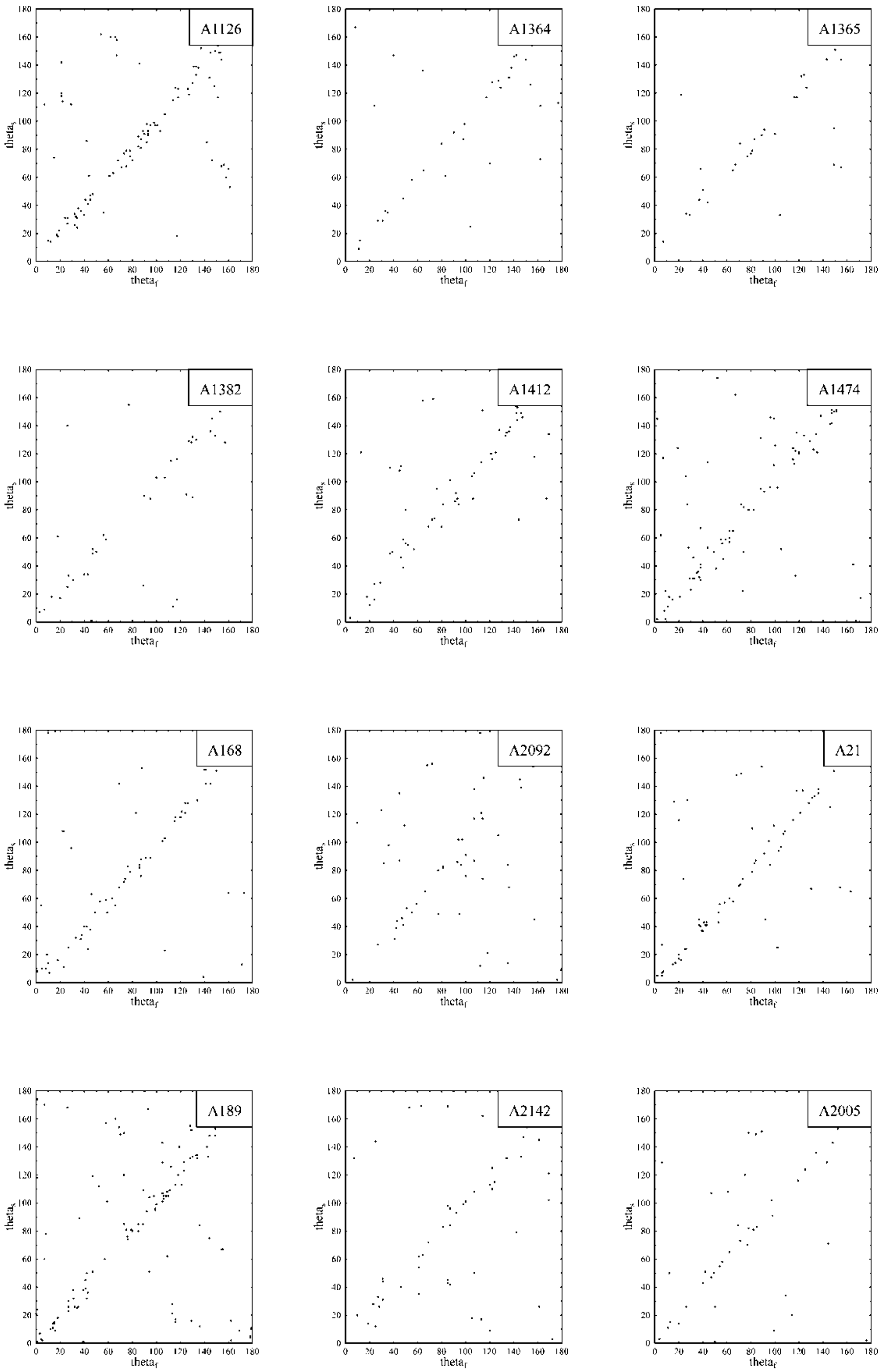

Figure 12. Correlation of the position angles for galaxies after visual classification within the magnitude range $m_3 + 3$.

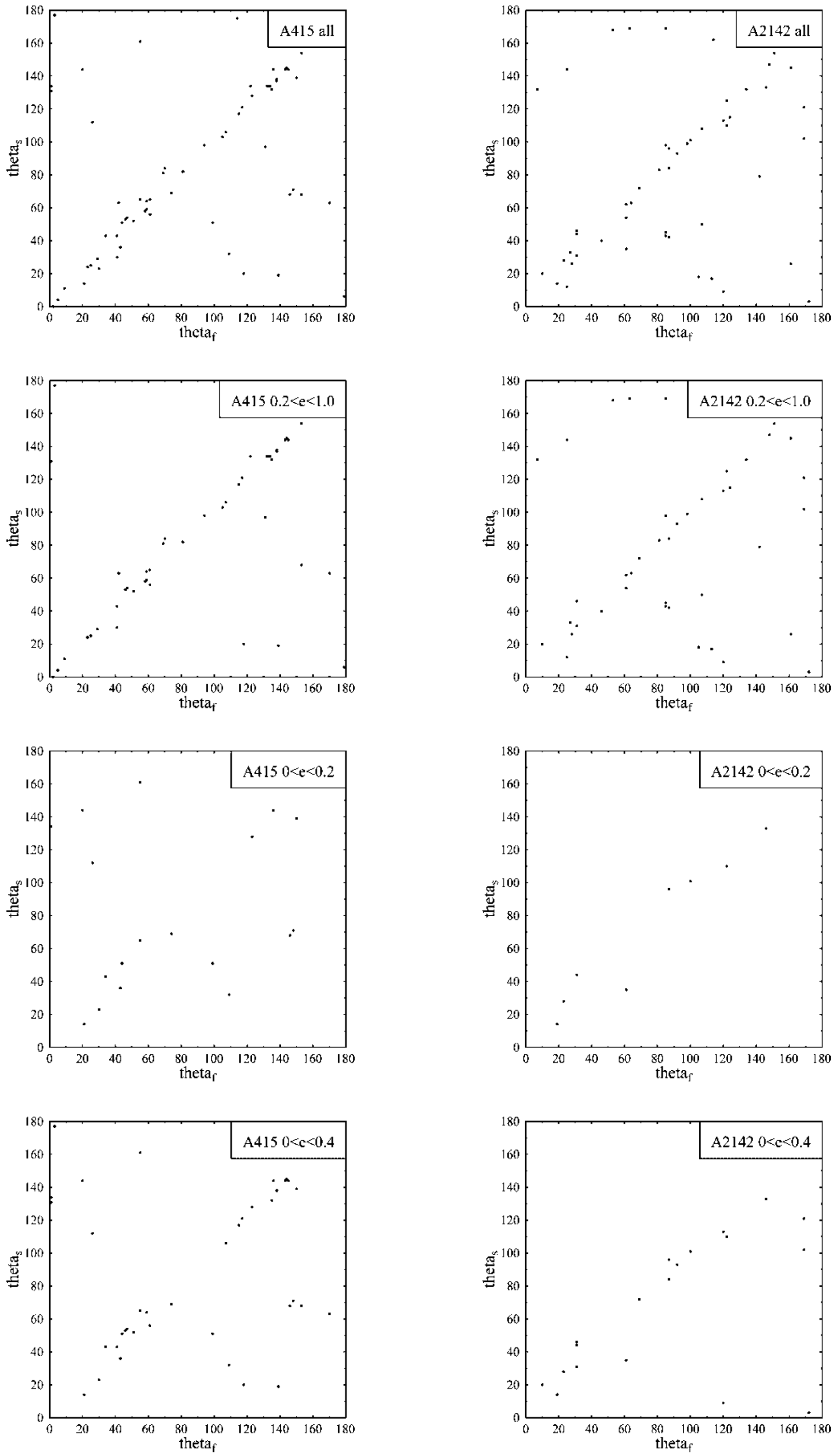

Figure 13. The dependence of the correlation between galaxy position angles on galaxy ellipticity.

3. Results and discussion

We compared four parameters important for astronomical investigations: the instrumental magnitude, area of object, ellipticity and position angle created by FOCAS and SExtractor. Moreover, we compared the values of the investigated parameters obtained from FOCAS after visual correction and SExtractor. Because of the large number of graphs, we present in this paper comparisons for 12, randomly chosen Abell clusters of our sample. In figures 1– 4, we present the correlations between investigated parameters for complete catalogues of galaxies. Figures 5–8 give the correlations after visual corrections of the FOCAS catalogues, while the next set of four figures (figures 9–12) deals with the correlations when the catalogues were restricted to the magnitude range $m_3 + 3$. The basic influence on the area of the object has an algorithm which specifies the level of the local background. The background estimators used in these two packages influence the number of pixels included in the object. This could be the reason why the areas of objects are different for FOCAS and SExtractor. The pixels lying above the background are included in the object and utilized to calculate the shape parameters. Because the areas calculated by the two packages were not identical, the shape parameters calculated by applying the same algorithms in both packages cannot be identical. As can be seen in the figures, which is supported by statistical analyses for individual clusters, the lines have different slopes; this could be due to the various sky background levels on each DSS plate. From figure 13, it follows that the galaxy position angle even in the ellipticity range $0 < e < 0.4$ is not as poorly determined as some researchers have claimed.

4. Conclusions

It is generally accepted that data obtained from automatic software are error free and correctly describe investigated parameters, which means that the results of any investigations using such parameters are assumed to be reliable. Therefore we decided to check the correctness of four parameters important in astronomy, namely the instrumental magnitude, area, ellipticity and position angle, when obtained from two frequently used packages. According to our best knowledge, this is the first attempt to compare these two packages. As the first step of our investigations we found that visual verification of objects selected for the catalogue is necessary. This is connected with the fact that brighter objects are usually classified as galaxies (brighter than 17.5 in DSS instrumental magnitude). This result is not new and some researchers as a remedy for this effect advise that such objects should be excluded from the analysis [4]. Ungruhe [4] found that galaxies suffer most from misclassification. Down to magnitudes of 13, the error is, independent of galactic latitude, at least 60%. Between magnitudes of 13 and 17, the percentage of misclassified galaxies for $b < 45°$ drops continuously to between 15% and 30% and is clearly dependent on galactic latitude. The classification of galaxies at low galactic latitudes is most strongly affected; in these regions, only half the galaxies are correctly classified. Errors found in this work thus lie, by a factor of 2–3, above the values quoted in the literature. Stars show classification errors of at most 10%, and this increases towards fainter magnitudes. For artefacts, noticeable errors occurs only for objects brighter than a magnitude of 15; this is mainly due to the saturation effects of the photographic emulsion. At magnitudes fainter than 15, the error is below 5%. No dependence on galactic latitude is seen. 1 300 000 galaxies, 815 000 stars and 647 000 perturbed objects previously classified automatically were reclassified by eye. We analysed 306 361 objects before and after visual verification. After automated star–galaxy classification we obtained 32 827 objects classified as galaxies and 223 257 classified as stars. We reclassified these objects by eye and we obtained 27 977 galaxies and 228 271 stars. The number of misclassified galaxies is about 15%. So we agree with Ungruhe that the number of misclassified galaxies is high, much greater that some researchers have claimed. We consider the visual verification as more correct because in this way some bright galaxies are not excluded from the analysis. Of course, visual verification is very laborious and is usually disregarded by astronomers. The star–galaxy separation programs works correctly for objects with areas greater than 150 arcsec^2. Smaller objects are a mixture of area objects and plate noise, which makes correct classification impossible. Both packages have difficulty at faint magnitudes. We agree with Haynes *et al.* [14] that FOCAS appears to be capable of detecting more faint objects than Sextractor is. This may be due to the different approaches to splitting multiple objects employed by the catalogue programs. The magnitudes at which this effect is most noticeable are close to the limit of the images. SExtractor suffers difficulties in separating faint objects from brighter companions and may therefore wrongly aggregate such objects together. This is most noticeable around the brightest stars in the image, where it fails to separate objects from the wings of the stars. So our comparison could be reliable for objects of area greater than 150 arcsec^2 and dimmer than a magnitude of 17.5 in the case of commonly used DSS data. The correlation in the case of ellipticity and area is weak and the best correlation is, as expected, in the case of instrumental magnitude. As was mentioned before, the shape parameters depending on the area of the object and the number of pixels are included in the object. As can be seen in the case of individual clusters, after visual verification the correlation between parameters in general is quite good. However, even in the case of individual clusters, when the correlation between values of the parameters obtained from the two packages is quite good, the slopes of correlation are different; this is caused by different sky background levels on the DSS plates. This could be the main reason why in general for the sample of 43 clusters the correlation is poor. Moreover, the value of correlation between the investigated parameters shows that after visual inspection the correlation is higher, which is good evidence for our claim for the necessity of visual inspection in order to perform correct star–galaxy separation. In our opinion, the FOCAS package is a very useful tool for identification of faint objects on the photographic plates. SExtractor was designed to identify objects on CCD arrays and uses algorithms based on a neural network. It is quite possible that, in the future, more developed software will be able to perform this separation correctly, making it unnecessary for astronomers to carry out these lengthy and laborious tasks.

Acknowledgement

We thank Dr Elena Panko for discussions.

References

[1] G. Pittella, Acta Cosmol. **15** 119 (1998).
[2] J.F. Jarvis and J.A. Tyson. Astron. J. **86** 476 (1981).
[3] D.C. Koo, R.G. Kron, D. Nanni *et al.* Astron. J. **91** 478 (1986).
[4] R. Ungruhe, PhD thesis, Universität Münster (1999).
[5] S.J. Maddox, W.J. Sutherland, G. Efstathiou *et al.* Mon Not. R. Astron. Soc. **243** 692 (1990).
[6] S.M. Beard, H.T. McGillivray and P.F. Thanish. Mon Not. R. Astron. Soc. **247** 311 (1990).
[7] H.K.C. Yee. Publs Astron. Soc. Pacif. **103** 396 (1991).
[8] E. Bertin and S. Arnouts. Astron. Astrophys. **117** 393 (1996).
[9] P. Flin and I.B. Vavilova, in *Proceedings of the 27th Meeting of the Polish Astronomy Society*, edited by J. Sarna and P.B. Marks (1996), p. 63.

[10] N.S. Philip, Y. Wadadekar, A. Kembhavi *et al.* Astron. Astrophys. **385** 1119 (2002).
[11] E. Bertin, SExtractor v. 2.0 User's manual, Institut d'Astrophysique de Paris, Sterrwacht Leiden, & European Southern Observatory (1997).
[12] M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Vol. 1, *Distribution Theory*, 4th edition (Griffin, Irvine, California, 1977).
[13] *The Digitized Sky Survey* (Space Telescope Science Institute, Association of Universities for Research in Astronomy, 1993, 1994).
[14] T. Haynes, G. Cotter, J.C. Haker *et al.* Mon Not. R. Astron. Soc. **334** 283 (2002).