FALSE DISCOVERY RATE IN OBJECTS DETECTION

František Mojžíš

Department of Computing and Control Engineering, ICT Prague, Technická 5, 166 28 Prague 6

frantisek.mojzis@vscht.cz

Abstract

This paper is devoted to the application of statistical methods in astronomical objects detection, which is one of the most fundamental topics in astronomical science. There is applied False Discovery Rate as a method of multiple hypotheses testing and compared on the hypothesis that input data are Poisson distributed. This method is than compared with commonly used method such like thresholding and application of two-dimensional filters.

1 Introduction

Objects detection [1] is one of the most fundamental topics in astronomical images processing [2, 3]. Analyzed data are usually acquired during the night when light conditions are poor. Thus it is necessary to use long exposure times.

For data acquisition, an astronomical CCD camera [4] is used. CCD sensor [4] is a source of several noises [5]. Suppose that it works as an photon counter, then it is logical that the images are contaminated by photon counting noise [2]. This is result of the fact, that the light is used as an information carrier and thus we must consider its behavior as a stream of photons.

The main goal of this paper is to describe method of exact astronomical objects localization. Main principle of this work is based on the objects detection via False Discovery Rate [6] (FDR) method, namely by evaluation of Sidak's correction The results of presented algorithms are Results section and further discussed in the Conclusion.

2 Objects detection

2.1 Noise Model

stronomical images can be expressed in mathematical way as follows

$$\mathrm{x}(k,l) = \mathrm{f}(k,l) + \mathrm{n}(k,l) \tag{1}$$

where $\mathrm{f}(k,l)$ are the data and $\mathrm{n}(k,l)$ represents noise called the dark current. This type of noise is caused by thermally generated charge, due to the long exposure times. Dark current should be simply removed by a dark frame, which maps mentioned thermally generated charge in CCD sensor. It can be considered that this type of noise is Poisson distributed [1] in the following way

$$\mathrm{n}(k,l) \sim \mathrm{Poisson}(\lambda(k,l)) \tag{2}$$

where λ(k,l) is expected number of occurrences in the CCD pixel cell (k,l) and $\lambda \in \mathbb{R}_0^+$. This claim can be verified on a sample of the dark images by a statistical test for the Poisson probability distribution, which can be found in [1, 7].

In the following text we will consider an average dark frame

$$\mathrm{d}(k,l) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{n}_i(k,l) \tag{3}$$

where m is the number of dark images, $\mathrm{n}_i(k,l)$ is a noise in i-th frame and futher we can assume that $\mathrm{d}(k,l) = \hat{\lambda}(k,l)$.

2.2 Thresholding

Thresholding is commonly used method for object segmentation in image processing. It is based on the idea that values lower then threshold T are set to zero and values greater or equal to value T are equal to one. This leads to the conclusion that result of thresholding is binary image. If $\mathrm{x}(k,l)$ is processed image function then image after threshold can be expressed as follows

$$\mathrm{x}_T(k,l) = \begin{cases} 1 & \mathrm{x}(k,l) \geq T \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

The key of this method is to select the threshold value T. In practice, there are used several methods such as histogram shape based methods, maximum entropy method, k-mean clustering, etc.

2.3 Edge Detection

Edge is detection is method based on identifying point in images at which the image intensity is changing sharply or has discontinuities. The aim of this method is to apply such image detector which result will lead to a set of connected curves indicating boundaries of objects. There exist many methods for object detection. Typical approach used to find discontinuities is application of a mask (operator) where the correlation between the mask and processed image regions is evaluated based on the following equation

$$\mathrm{x}_{\mathrm{E}}(k,l) = (\mathrm{x} * \mathrm{m})(k,l) \tag{5}$$

where $\mathrm{x}(k,l)$, $\mathrm{m}(k,l)$ and $\mathrm{x}_{\mathrm{E}}(k,l)$ are processed image, mask and result image with detected edges. Detection of sharp changes can be understood as a finding of function extreme, where the derivatives find it application. This leads to the conclusion that used masks represents first or second order derivatives. Commonly used masks are Sobel, Prewitt, Laplacian of Gaussian, Roberts, etc.

These methods are sometimes used in combination with thresholding, where the used sensitivity threshold T_{s} is used to ignore all edges, wich are not stronger then thresh.

2.4 False Discovery Rate

When multiple hypotheses are tested [6], it is necessary to control the portion of incorrectly rejected null hypotheses [6, 8] (type I errors). One of the procedures, used for this purpose, is FDR, based on relation

$$\mathrm{E}\left(\frac{V}{V+S}\right) = \mathrm{E}\left(\frac{V}{R}\right) \tag{6}$$

where V, S are the numbers of false positive (Type I error) and true positive [6] hypotheses and $R = V + S$. Application of FDR in object separation is based on the evaluation of critical p-values $p_{(k),\mathrm{crit.}}$ and confrontation with sorted p-values $p_{(k)}$, given as the result of multiple hypothesis testing.

Evaluation of FDR can be based on the Bonferroni correction, which presents multiple-comparison correction. That is when several dependent or independent statistical tests that are being performed simultaneously. FDR with Bonferroni correction is based on rejection rule

$$p_{(k)} \leq p_{(k),\text{crit.}} = \frac{k\alpha}{n} \tag{7}$$

where $k = 1, \ldots, n$, n is the number of tested hypotheses and α is the significance level [8], usually $\alpha = 0.05$.

A related correction, called the Sidak's correction [6], gives a weaker but valid bound than the Bonferroni correction and assumes that the individual tests are independent. This is given by

$$p_{(k)} \leq p_{(k),\text{crit.}} = 1 - (1 - \alpha)^{k/n}. \tag{8}$$

Critical p-values create a curve, that can or cannot cross original sorted p-values in ascending sequence. The number of p-values, that occur under this curve, presents the real number of hypotheses, which can be really rejected and are statistically significant. The portion between correctly rejected and previously rejected null hypotheses presents the FDR, which should be less than $\alpha/2$.

3 Results

Algorithms used in the previous section were then applied on model image, see Fig. 1 (a). In Fig. 1 can bee also seen the dark image Fig. 1 (b), sorted p-values after FDR evaluation and critical p-values, Fig. 1 (c) and the detected objects, Fig. 1 (d).

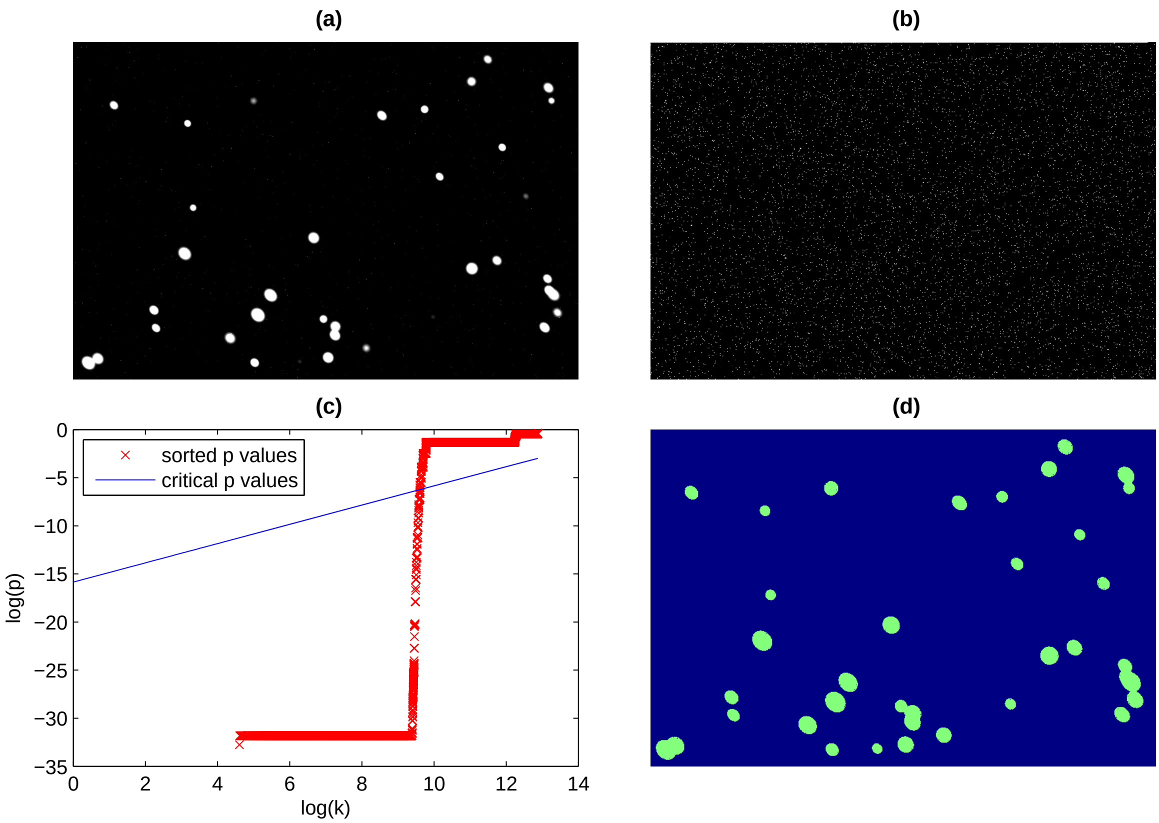

Figure 1: Model image object detection (a)model light image, (b)model dark image, (c) FDR evaluation, (d) detected objects in model image.

To show FDR advantage against thresholding and edge detection, presented algorithm were applied to the cut of mentioned model image, Fig. 2 (a). To highlight some invisible object there was used logarithm of of this image, Fig. 2 (b) where there can be seen and object invisible by eye whch is present in the red circle.

From the presented results in Figs. 3 - 5 can be found that FDR detection has no problem to detect this object while there are no false detected objects. Thresholding with different settings of threshold T did not identified this object. The Laplacian of Gaussian gives better results for T_{s}, see Fig. 4 (c) but there were also false detected objects from the background.

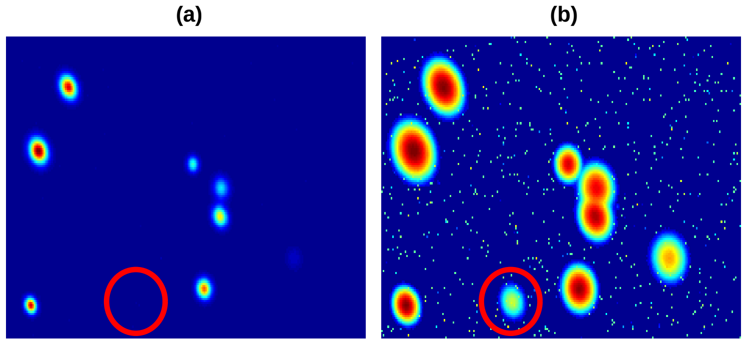

Figure 2: (a) cut of original image, (b) logarithm of light image model cut.

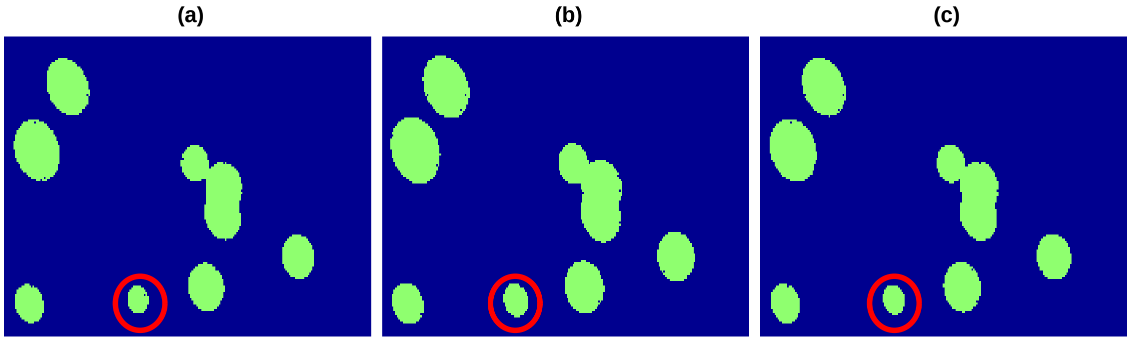

Figure 3: Significance level influence to object detection via. FDR (a) detected objects $\alpha = 0.01$, (b) detected objects $\alpha = 0.05$, (c) detected objects $\alpha = 0.10$.

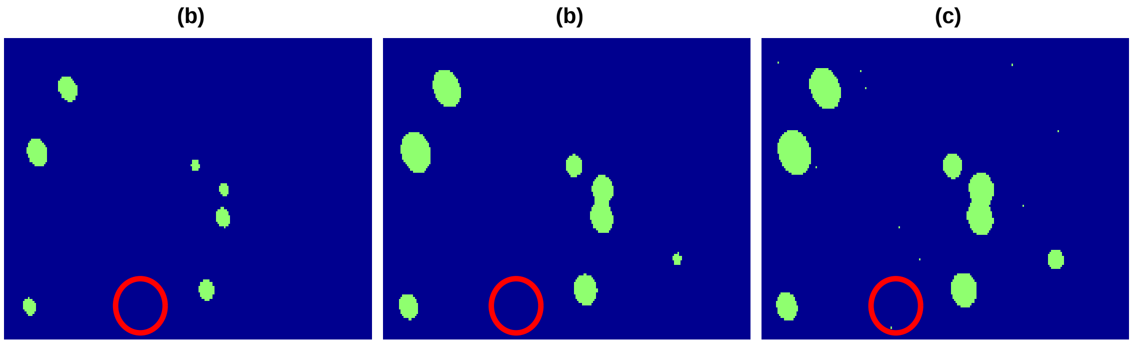

Figure 4: Threshold influence on object detection (a) detected objects $T = 0.050$, (b) detected objects $T = 0.010$, (c) detected objects $T = 0.001$.

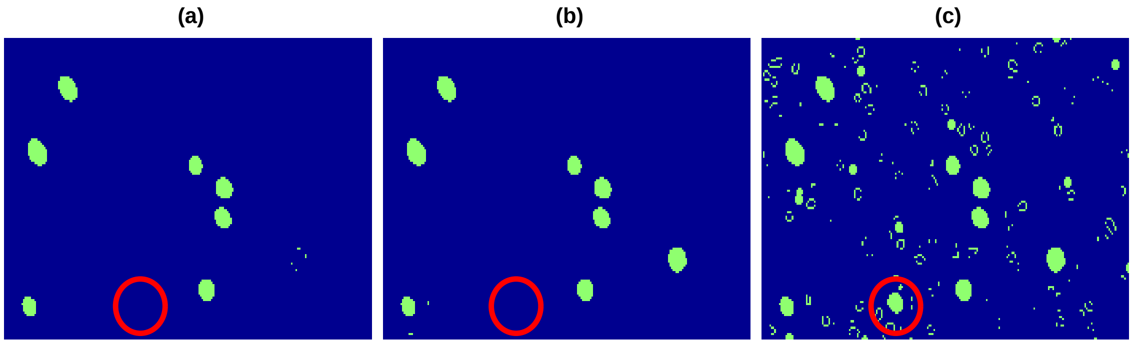

Figure 5: Object detection via. Laplacian of Gaussian (a) detected objects $T_s = 10$, (b) detected objects $T_s = 5$, (c) detected objects $T_s = 1$.

4 Conclusion

In this paper there was introduced a method of objects detection based on mathematical statistics, namely on hypothesis testing and further False discovery rate evaluation. As it can be seen from the results, FDR gives better results compared to the other two presented methods and has no problem with false detected objects. To mention its one disadvantage, it can be used only in combination with dark frame and where there is no dark frame the only option is to use commonly available methods.

References

[1] Mojzis, F., Kukal, J., Svihlik, J., Astronomical Systems Analysis and Object Detection, *In Proceedings of 22nd International Conference Radioelektronika*. Brno (Czech Republic), 2012, p. 201 - 204.

[2] Starck, J. L., Murtagh, F., Bijaoui, A., *Image processing and data analysis: The multiscale approach*, Cambridge University Press, 1998.

[3] Starck, J. L., Murtagh F., *Astronomical Image and Data Analysis*, Springer, 2006.

[4] Buil C., *CCD Astronomy: Construction and Use of an Astronomical CCD Camera*, Willmann-Bell, 1991.

[5] Gonzalez, C. R., Woods, E. R., *Digital Image Processing*, Prentice Hall, 2002.

[6] Efron, B., *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*, Cambridge University Press, 2010.

[7] Brown L. D., Zhao L. H., *A new test for the Poisson distribution*. 29 pages. [Online] Cited 2013-08-12. Available at: http://www-stat.wharton.upenn.edu/ lzhao/papers/newtest.pdf.

[8] Papoulis A., Pillai U. S. *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, 2002.

František Mojžíš was born in Prague, Czech Republic in 1986. He received his M.Sc. from the Institute of Chemical Technology Prague (ICT), in 2011. Now he is a Ph.D. student at ICT. His research interests include image processing and image denoising.