ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

Using Point Pattern Analysis Techniques to Describe Spatial Arrangement of Points in Images

Baraa Taha Yaseen[1], Zinah Mohsin Arkah[2] , Reem Ibrahim Hasan[3] , Sefer Kurnaz[4]

M.S. Student, Department of Information Technology, Altinbas University, Turkey[1]

University of Information Technology and Communications, Baghdad, Iraq[2]

Iraqi Commission for Computer and Informatics (ICCI), Baghdad, Iraq[3]

Department of Computer Engineering, Altinbas University, Turkey[4]

ABSTRACT: A large number of methods for the analysis of point pattern data have been developed in a wide range of scientific fields. Point pattern analysis (PPA) is the study of spatial arrangement of points in space. Describing the pattern of data can tell a lot of details about the data special pattern and can even help to diagnosis diseases. Many point pattern techniques have been applied to describe the pattern however not all of them give clear description of the data in the space. Some of them omit some points which consider a drawback. In this paper, Different clustering algorithms (G, F, K spatial point pattern analysis techniques) are applied to describe spatial arrangement of points. We also give an explanation of spatial point pattern analysis techniques. We compare results of each technique to others to have better understanding of spatial point pattern of Points then compare to R package. Based on results of spatial point pattern analysis techniques it can be decided that the cells are evenly spaced or clustered.

KEYWORDS: point pattern; K-Function; clustering; R package

I. INTRODUCTION

Point pattern analysis is the investigation of spatial arrangement of points in space. It has been used to describe the locations of disease, crimes and plants of different species [1]. In this work we give an explanation of spatial point pattern analysis techniques. We apply these techniques to describe spatial arrangement of points. Three major spatial point patterns are complete spatial randomness, regularity, and clustering [2], illustrated in Figure 1.

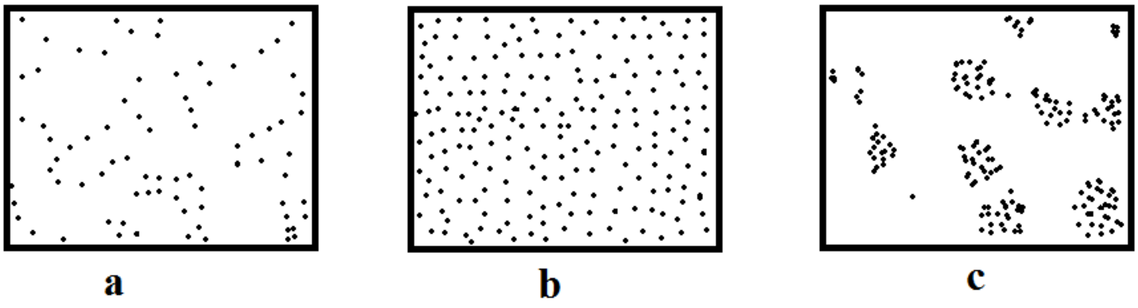

Fig1: Spatial point pattern types (a) Complete Spatial Randomness (CSR)(b) regularity (c) clustering

Point patterns can be described either by first order e□ects that analyze number of events per unit area (point density) or second order e□ects that describe interaction between points (point separation) . In this work we explore use of second order distance-based methods to describe distribution pattern of points. Specifically the measures F-

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

function, G-function, and K-function described below have been used to describe and di erentiate nuclei distribution patterns. In order to have a single complete processing pipeline, we have developed Matlab programs to compute these measures. To validate our results, we compared the output of our programs to the output of R package [3,4,11]. We have applied G,F and K functions to describe the pattern. We found out that G and F omit a lot of points while K gives a clear description of points in space (evenly or clustered).

II. RELATED WORK

Clustering is considered one of the most critical unsupervised learning problems. It endeavors to find an accurate structure in a collection of unlabeled data. Then organizing objects into groups whose members have similar general properties. A cluster is therefore a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other clusters. Ayala, Guillermo, et al. have used pattern analysis techniques to describe the pattern of medical images and compare them to other images [5]. Huang, Xiaohu, et al. have presented K-Function to monitor unformed data and show good results of K-function [6]. Similarly, Yunta, Mariña López, et al. have presented K-Function to analyze the spatial distribution of proteins [7]. Although the spectral clustering is a widely used technique but it is still facing some challenges such as the dimensionality of the data. For high dimensional data, it is hard to cluster similar objects by only depending on the distance among them. One of the famous algorithms to calculate the distance is called Euclidian Distance method. This method depends on the precision of building the similarity matrix. [8, 9, 10] have mentioned that the eigenvectors can be driven from similarity matrix. A large extent of data dimensions is one of vectors that the accuracy of clustering matrix depends on. For high dimensional data, it is hard to use Gaussian Mixture Model (GMM) to estimate the density as well. To overcome these difficulties such as noise and redundant attributes that is related with high dimensional data.

III. PROPOSED ALGORITHM

A. SPATIAL POINT PATTERN ANALYSIS USING DISTANCE-BASED MEASURES

Distance-based point pattern analysis methods rely on distances between points (often nearest-point distance). In this work, points correspond to centroids of nuclei in Points or corneal fibrils in transmission electron microscopy. Let Ci and Cj be two points (centroids). The distance between Ci and Cj is computed as

$$d(C_i, C_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (1)$$

Once the distances between each pair of points in an image are computed, the nearest distance for each point can be found. Let dmin(Ci) denote the distance of point Ci to its nearest point. Mean nearest-point distance is then computed as

$$\overline{d}_{min} = \frac{\sum_{i=1}^{n} d_{min(C_i)}}{n} \quad (2)$$

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

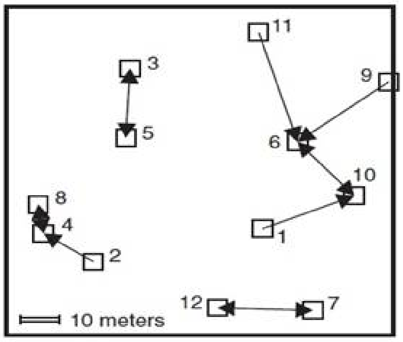

Fig 2: Distances between points and their nearest neighbours [1]

Points	X	Y	Nearest point	d_{min}
1	66.33	32.54	10	12.59
2	22.52	22.39	4	15.64
3	31.01	81.21	5	21.11
4	9.47	30.02	8	9.00
5	30.78	60.10	3	21.14

Table 1: Sample calculations for the nearest-point distances for the point pattern shown in Figure 2 [1].

.

dmin summarizes all the nearest-point distances by a single mean value but omits many details about the point pattern. More detailed measurements are needed to capture the di□erences between point distribution patterns. The following subsections will describe some well-established point pattern measures using the simple synthetic case illustrated in Figure 6.2 [1] where for each point, its nearest point is marked with an arrow pointing from the original point to its nearest point. Table 1 list for each point in Figure 2, id of the nearest point and associated distance. Note that the nearest-point relationship is not symmetric.

B. G-Function

G-function, also known as refined nearest point, uses information in Table 1. In-stead of calculating the mean minimum distance, computes the cumulative frequency distribution of the nearest-point distances. It is described by Eq.3.

$$G(d) = \frac{\#(d_{min}(C_i) < d)}{n} \qquad (3)$$

G (d) gives the fraction of all the nearest-point distances in the pattern that is lower than d. Figure 3 shows the G-function corresponding to points in Figure 2. The shape of the G-function gives information on how points are spaced in a pattern. If points are closely clustered together, then G-function increases rapidly at short distances. If points tend to be evenly spaced, then G increases slowly up to the range of distances at which most points are spaced, and only then increases rapidly [1].

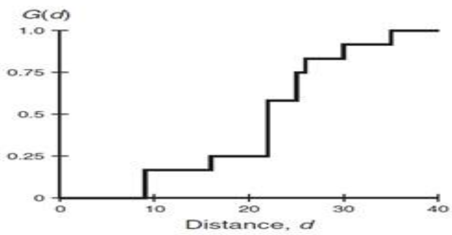

Fig 3: G-function for the points shown in Figure 2 [1].

C. F-Function

F-function is closely related to G-function but rather than accumulating the fraction of nearest-point distances between points in a pattern, selects random point locations within study region, computes minimum distance from these locations to any point in the original point set. F-function is computed as follows:

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

$$F(d) = \frac{\#(d_{min}(P_i, C) < d)}{m} \quad (4)$$

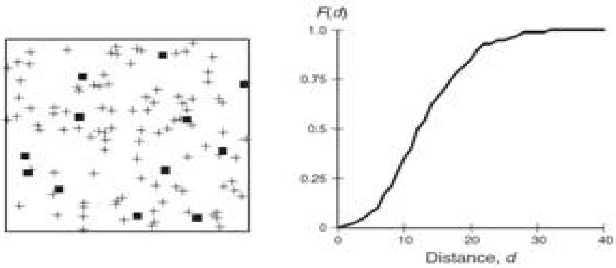

Fig 4: F-function for the points shown in Figure2 [6].

Where Pi is the randomly selected point locations, d.min(Pi, C) is the minimum distance between the randomly selected location and the points in the original point set. The advantage of F-function is its ability to boost sample size m to get a smoother cumulative frequency curve (Figure 4) [1].

D. K-Function

K-function, also known as Ripley's K-function, is a more descriptive point pattern measure. G-function and F-function rely only on the nearest point distances to de-scribe a point pattern. K-function uses all of the distances between point pairs in the pattern. K(d) corresponds to area normalized mean of number of points within distance d of another point. K-function is computed as follows:

$$\begin{aligned} K(d) &= \frac{\sum_{i=1}^{n} \#(P \in \mathcal{C}(C_i, d))}{n\lambda} \quad (5) \\ &= \frac{a}{n}\frac{1}{n}\sum_{i=1}^{n} \#(P \in \mathcal{C}(C_i, d) \end{aligned}$$

where P is the set of points and C(Ci, d)) is a circle with radius d centered at point Ci. Figure 5 illustrates the K-functions for two point patterns.

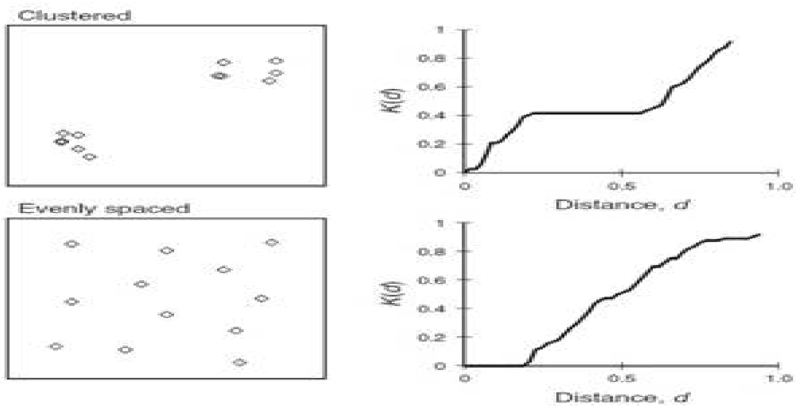

Fig 5: K-function for clustered and evenly spaced patterns [1].

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

V. SIMULATION RESULTS

A-Dataset

Random points (synthetic points) with evenly space and clustered points have been used to check the algorithms. Then we have used breast cancer dataset [12] consists of 537 H&E stained histopathological images. Size of the images is 2200 × 2200 pixels. Each image contains approximately 1500 nuclei. Ground-truth nuclei centers are available for a 400×400 sub region of each image. We have used the Ground-truth to describe the pattern.

B- Experimental Results

We implemented the point pattern description functions F-, G-, and K- described in Matlab and used on random points(synthetic points) Below we show sample results. As can be seen particularly in histopathology images K-function nicely captures point distribution pattern, differentiates between regularly spaced points versus clustered points distributions. Spatial distribution of points can reveal useful information about pattern of points. We will use point pattern analysis and the described measures to quantitatively study spatial distributions of points. Then compared to R package[3,4] which has same functions.

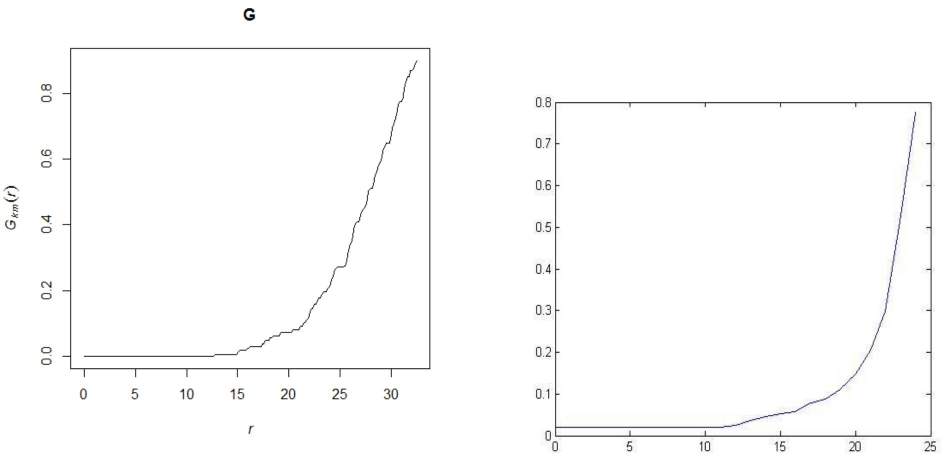

Fig7: G-function, first with R on the left and Second one with matlab which is ours on the right.

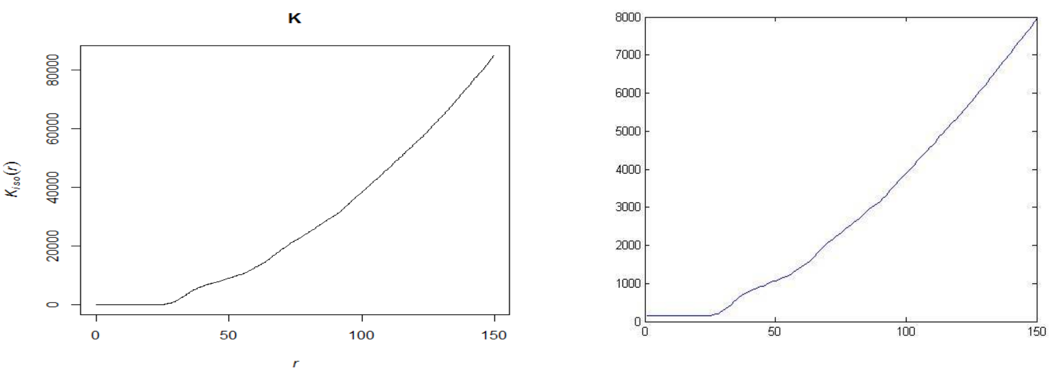

Fig7: K-function, first with R on the left and Second one with matlab which is ours on the right.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

F function has almost same results of G function.It is very clear that K-function describes the input image in evenly spaced better then G, F functions. In order to emphases on K- function we describe it with different input images in the next section.

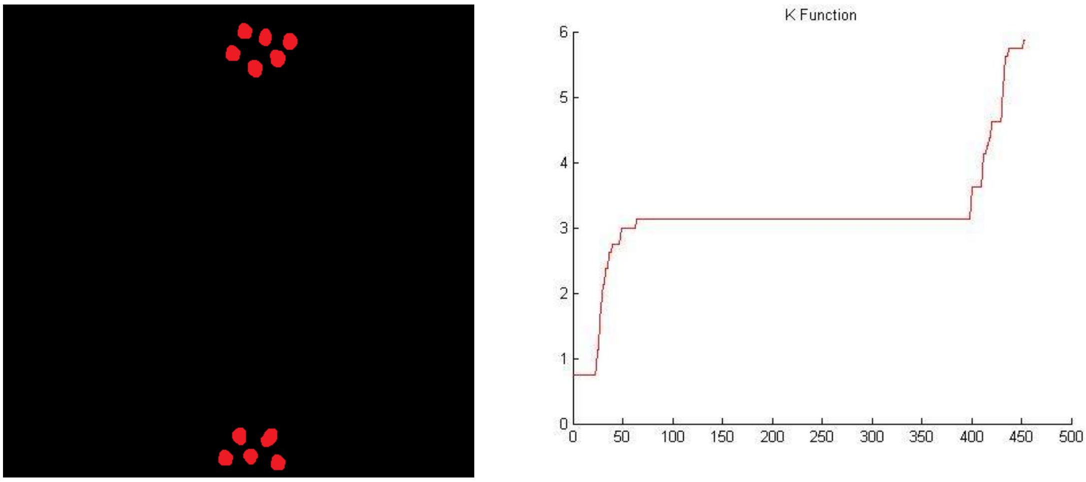

Fig 8: K-function with clustered image

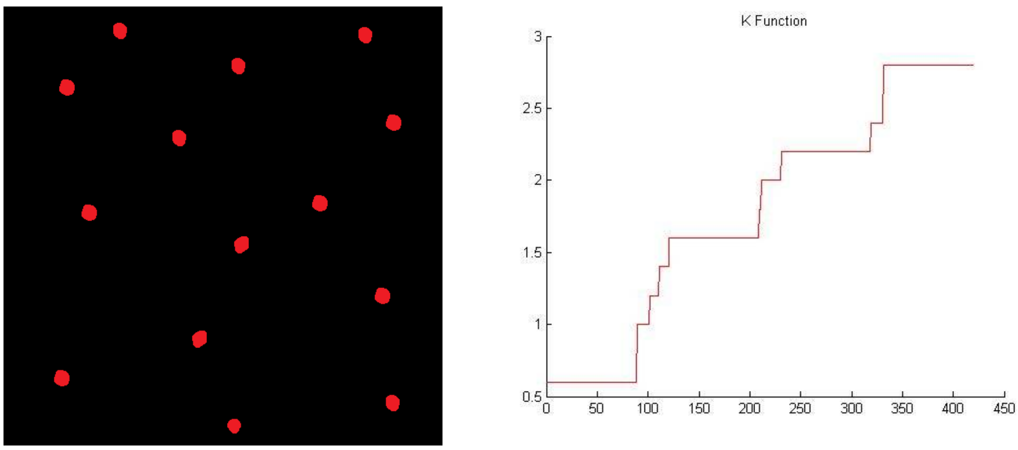

Fig 9: K-function with evenly spaced image

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

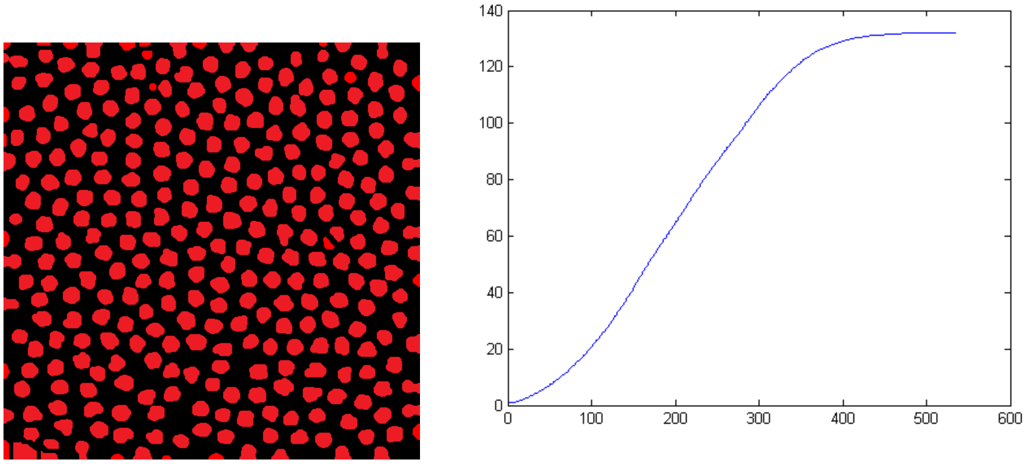

Fig 10: K-function with evenly spaced image and it show how the curve is smooth

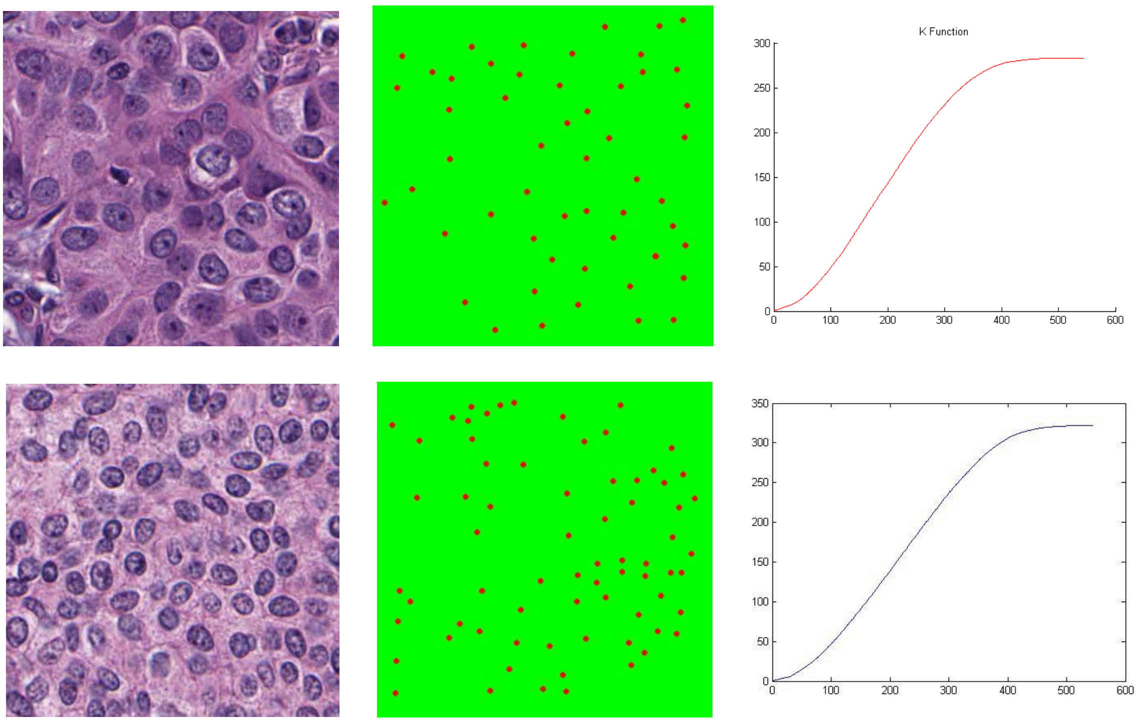

Fig 11: output of images from dataset [12]. First column is original image, second column is ground truth of the original image, and third column is the results of K-function on dataset [12].

VI. CONCLUSION

In this work, we have been concerned with developing a clearer idea of the concept of pattern and how it can be related to process. In principle, any pattern can be described using a variety of measures. We have applied G, F, K functions which are spatial point pattern analysis techniques to describe spatial arrangement of points. We also presented a brief

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

explanation of spatial point pattern analysis techniques. We compare results of each technique to others to have better understanding of spatial point pattern of Points then compare to R package. Based on results of spatial point pattern analysis techniques it can be decided that the cells are evenly spaced or clustered.

We applied these functions using Matlab and we found that K function has the best results compared to other functions.

REFERENCES

1. David O'Sullivan and David Unwin. "Geographic information analysis". John Wiley & Sons, pp.132-190, 2014.
2. Loosmore, N. Bert, and E. David Ford. "Statistical inference using the G or K point pattern spatial statistics". Ecology 87.8: 1925-1931:2006.
3. Baddeley, Adrian. "Analysing spatial point patterns in R." Technical report, CSIRO, 2010. Version 4. URL https://research. csiro. au/software/r-workshop-notes, 2008.
4. Baddeley, Adrian, and Rolf Turner. "Spatstat: an R package for analyzing spatial point patterns". Journal of statistical software 12.6 : 1-42, 2005.
5. Ayala, Guillermo, et al. "Clustering of spatial point patterns". Computational statistics & data analysis 50.4: 1016-1032:2006.
6. Huang, Xiaohu, et al. "Monitoring spatial uniformity of particle distributions in manufacturing processes using the K function". IEEE Transactions on Automation Science and Engineering14.2: 1031-1041:2017.
7. Yunta, Mariña López, et al. "A statistical analysis of spatial clustering along cell filaments using Ripley's K function". Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on. IEEE, 2014.
8. Stella X. Yu, Jianbo Shi, "Multiclass Spectral Clustering", Computer Vision, 2003. Proceedings. Ninth IEEE International Publication, Page(s): 313 - 319 vol.1 Year: 2003.
9. Wen-Yen Chen, Yangqiu Song, Hongjie Bai, Chih-Jen Lin, and Edward Y. Chang, "Parallel Spectral Clustering in Distributed Systems" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 33, NO. 3, MARCH 2011.
10. Jianbo Shi and Jitendra Malik, Member, IEEE, "Normalized Cuts and Image Segmentation" IEEE TRANSCTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 22, NO. 8, AUGUST 2000.
11. van Lieshout, M. N. M. "ADRIAN BADDELEY, EGE RUBAK, AND ROLF TURNER, Spatial Point Patterns: Methodology and Applications with R. Boca Raton, FL: CRC Press." Biometrics 72, no. 2 : 660-661:2016.
12. Jun Xu, Lei Xiang, Qingshan Liu, Hannah Gilmore, Jianzhong Wu, Jinghai Tang, and Anant Madabhushi. "Stacked sparse auto encoder (SSAE) for nuclei detection on breast cancer histopathology images". IEEE Transactions on Medical Imaging, 35(1):119–130, 2016.