

Crowd Detection in Airborne Images using Spatial Point Statistics

¹Abdullah H. Özcan, ²Cem Ünsalan, ³Peter Reinartz

¹TÜBİTAK BİLGEM, Turkey

²Yeditepe University, Turkey

³German Aerospace Center (DLR), Germany

(e-mail: unsalan@yeditepe.edu.tr)

Abstract—The crowd density in public places increases in social events. If an emergency occurs during such events, authorities should take urgent measures to prevent casualties. Therefore, crowd detection and analysis is a critical research area. Even though there are several studies on person detection from street or indoor cameras, these may not be directly used to detect or analyze the crowd formed from people. In this study, we approach the problem using aerial images. We propose two novel methods to detect the crowd using spatial statistics. The first novel method is based on the first-order statistics. It uses the nearest neighbor relations for each person in the image. The second novel method is based on the second-order statistics. Here, the spatial position of persons are checked whether they are clustered or randomly distributed. We test these two methods on a sample test image and provide performance measures.

I. INTRODUCTION

During social events like concerts, sports matches, and festivals the crowd density increases in specific regions. If an unexpected problem occurs during such events, authorities must take urgent measures to eliminate or limit casualties. Therefore, detecting and analyzing crowds in images is a critical research area.

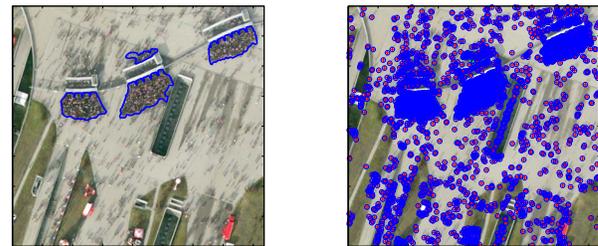
There have been several studies on crowd detection. The first group of these are based on street cameras [1], [4]. The second group of the studies is based on aerial images. Since the topic of this study is also crowd detection using aerial images, let's review them in detail. Arandjelovic and Zisserman [2] used SIFT features to detect crowds. However, this method is based on SVM classification which needs a proper training set. Perko *et al.* [7] presented a method for people counting and crowd monitoring from airborne imagery. This method is optimized for oblique views. Hence it is not suitable for nadir images. Meynberg and Kuschik [6] proposed a FAST feature extraction and SVM classification based method. Sirmacek and Reinartz [9] also used FAST features on airborne images for possible people detection. They approached the problem as kernel density estimation.

In this study, we propose two novel methods for crowd detection in aerial images. These are based on spatial point pattern statistics [3]. Specifically, we benefit from the first and second order statistics. The first-order statistics are about extracting the intensity of points (objects of interest) in a study region. The second-order statistics are about the distribution

of distance of point pairs. We take the location of a person as object of interest. Based on the first and second order statistics, we aim to detect whether persons are clustered (forming a crowd) or randomly distributed. We test our method on an aerial image and provide quantitative test results.

II. SPATIAL POINT STATISTICS

Spatial point statistics are based on the location of objects in spatial space. As for the object, we take the person location in the image detected by FAST [8]. We use these to extract the first and second order spatial statistics for crowd detection. In Fig. 1, we provide the FAST feature locations for the test image.



(a) Test image with ground truth

(b) FAST feature locations

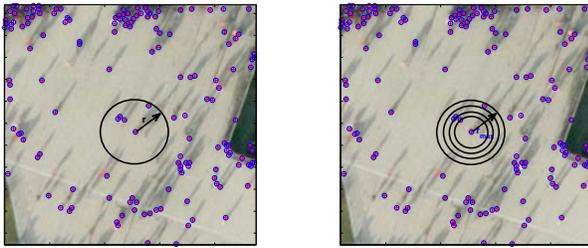
Fig. 1. Test image and possible detected people locations.

A. First-Order Statistics

The first-order statistics are based on the density of observations (here points representing people). In order to extract this statistic, we apply the following procedure. Let's assume that FAST features (each representing a person) extracted from the image are represented by their locations as (x_i, y_i) for $i = 1, \dots, K$. For each feature point, we extract the intensity value defined as

$$\lambda_i(x_i, y_i) = \sum_{d_l \leq r} \frac{3}{\pi r^2} \left(1 - \frac{d_l^2}{r^2}\right)^2 \quad (1)$$

where d_l is the distance between point (x_i, y_i) and its neighbors within the disk of radius r [5]. In Fig. 2(a), we provide an example on the intensity calculation. In this example, for each person the people that are in r neighborhood to that person is used to calculate λ .



(a) First-order statistics.

(b) Second-order statistics.

Fig. 2. Illustration of the calculation of the first and second order statistics.

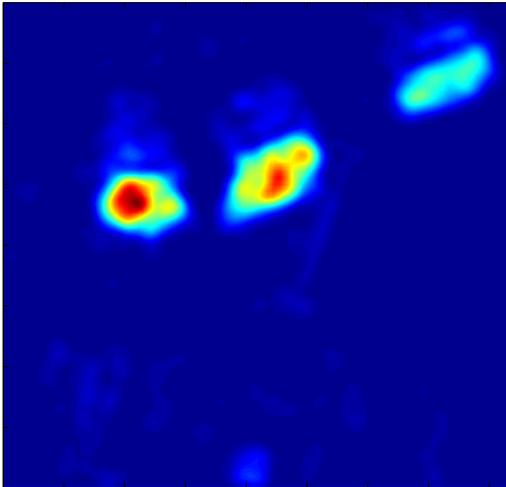
For FAST features within sparse people groups, λ_i will be small. For FAST features within crowded regions, this value will be large. We will use this information to detect crowd regions in the image. To do so, we merge all $\lambda_i(x_i, y_i)$ values for $i = 1 \dots K$ and obtain

$$\lambda(x, y) = \sum_{i=1}^K \lambda_i(x_i, y_i) \quad (2)$$

To estimate the crowd density map, we convolve $\lambda(x, y)$ with a symmetric Gaussian function as

$$D(x, y) = \lambda(x, y) * \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (3)$$

where $\sigma = r/5$. In this way, we obtain a smooth pdf. The crowd regions in $D(x, y)$ will have higher density than sparse people groups. To extract crowd regions, we use Otsu's automatic thresholding method on this density map. We provide the extracted crowd density using the first-order statistics in Fig. 3. We also provide the detected crowd boundaries in Fig. 4.


 Fig. 3. Calculated crowd density using the first-order statistics with $r = 60$.

B. Second-Order Statistics

The first-order statistics are based on the cumulative statistics of an observation point. The second-order statistics are



Fig. 4. Detected crowd boundaries using the first-order statistics. The detected crowd boundaries are represented in black. The ground truth for crowd boundaries is represented in blue.

based on the statistics of two neighboring points within radius r . There are several measures to calculate this statistics as K-function, L-function, nearest neighbor distribution, kth nearest neighbor distribution, and pair correlation function [11]. In this study, we use the pair correlation function to obtain the crowd density in the image. Pair correlation function, $g(r)$, is the distance derivative of Ripley's K-function as

$$g(r) = \frac{1}{2\pi r} \frac{dK(r)}{dr} \quad (4)$$

where

$$K(r) = \frac{A}{n^2} \sum_{i \neq j} I_r(r_{ij}) \quad (5)$$

$K(r)$ measures the expected number of points that are in the distance r of the observation point (obtained by FAST) [5]. Here A is the area, n is the number of points, r_{ij} is the distance between the i^{th} and j^{th} point, I_r is the indicator that is one if $r_{ij} < r$ and zero otherwise. We can say that pair correlation function measures the intensity for a ring with width r . This value is normalized to obtain complete spatial randomness. After normalization, if the points are completely randomly distributed $g(r) = 1$. If the points are clustered, we expect $g(r) > 1$. If the points are strongly distributed, we expect $g(r) < 1$.

$g(r)$ gives statistics for all points in the image. Therefore, for each point we calculate a local $g(r)$ from its neighboring points in r as illustrated in Fig. 2(b). For a point (person) that is in a cluster (crowd), we expect $g(r)$ to give similar results at each ring or each scale (r). It should not change much for different scales unless the largest scale is outside the range of the crowd.

In order to obtain a statistical feature to detect crowd regions, we propose the following method. We calculate $g(r)$ for each feature point for different r values. We expect $g(r)$ to

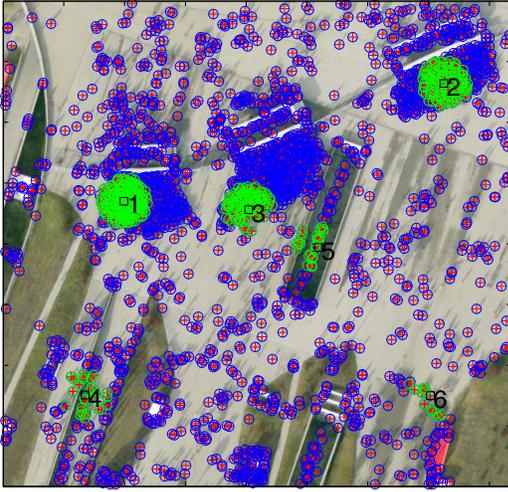


Fig. 5. Six examples of crowd and sparse people regions. Labels one to three on the image are chosen from crowd regions and four to six are chosen from sparse people groups.

oscillate less in feature points inside the crowd region since the spatial distribution of points is homogeneous through scales. For sparse people regions, $g(r)$ will be different for each scale. Therefore, there will be an oscillation through scales. These observations indicate that we can use the oscillation information to separate crowd and sparse people locations. We formalize this by using the variance of $g(r)$ as

$$\lambda_i(x_i, y_i) = 1/\text{var}[g(r)] \quad (6)$$

where (x_i, y_i) is the FAST feature location. Let's explain this method using the test image. We pick six samples from crowd regions and sparse people groups as in Fig. 5. In this figure, the black square marker indicates that we will calculate $g(r)$ for that person and it will be calculated from the neighboring people locations in r distance in which they are plotted in green markers. We provide the graphs for $g(r)$ for these in Fig. 6. Here, the top row is for crowd regions one to three. The bottom row is for sparse people regions from four to six. As expected, for crowd regions $g(r)$ oscillates less at all scales since the spatial distribution of points is homogeneous through scales. For sparse people regions, we find high differences in $g(r)$ for neighboring scales which look like oscillations. These results indicate that, we can use the oscillation information to separate crowd and sparse people locations.

As in the first-order statistics, we merge $\lambda(x_i, y_i)$ values using Eqn. 2. Afterwards, we obtain the density map using Eqn. 3. Here, again we choose $\sigma = r/5$. In this method, we provide the density map for $r=60$ in Fig. 7. Here, we calculate $g(r)$ in Eqn. 6 for $r = 1, \dots, 60$. We will apply the same strategy in Section III also. The obtained density map is different from the previous method's result. At sparse people locations, the density is almost zero due to the high variance of $g(r)$. We use Otsu's automatic thresholding method to detect the crowd regions. We provide the extracted crowd boundaries in Fig. 8.

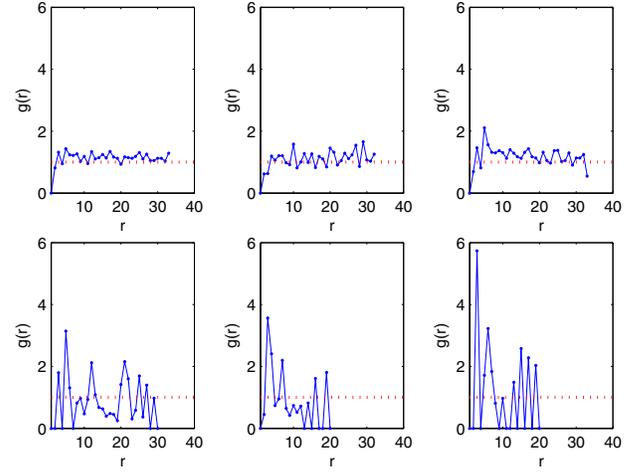


Fig. 6. Calculated $g(r)$ values wrt r for crowd and sparse people regions.

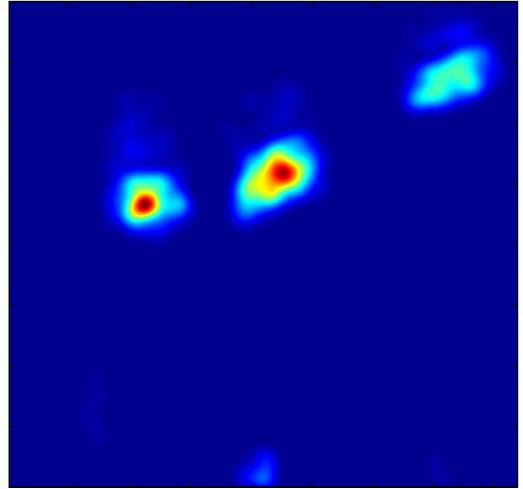


Fig. 7. Calculated crowd density using second-order statistics.



Fig. 8. Detected crowd boundaries from pair correlation function (in black) with ground truth (in blue) for $r = 60$.

III. EXPERIMENTS

We test the proposed crowd detection methods on an airborne image acquired from a stadium entrance. The German Aerospace Center (DLR) provided this image for test purposes. It is acquired by the 3K-Camera-System mounted on a Cessna aircraft from 1000 m flight altitude and ground sampling distance (GSD) of 15 cm [10]. The test image has been already shown in Fig. 1(a). There are sparse and crowd people groups in this image. The crowds are gathered at the gate entrances. Also, in the image the true crowd boundaries are plotted in blue.

Since the proposed methods only use the single parameter r , we obtained quantitative results for $r=[40\ 60\ 80\ 100]$ in terms of true positive (TP) and false alarm (FA). In Table I, we provide the results for both methods in percentages. As can be seen in this table, increasing r increases TP but FA is also increased for both methods. The reason for this is that, as r is increased the density map becomes smoother.

TABLE I
CROWD DETECTION PERFORMANCES IN PERCENTAGES FOR THE FIRST AND SECOND-ORDER STATISTICS BASED METHODS.

r	First-order statistics		Second-order statistics	
	TP	FA	TP	FA
40	80.27	0.51	82.51	0.72
60	81.96	0.67	81.40	1.60
80	83.97	1.27	86.34	5.37
100	85.89	2.39	94.27	13.06

We also tested the processing times of both methods. In tests, we used a laptop with Intel Core i7 quad core processor and 8 GB RAM. We implemented the methods in MATLAB 2014a on the Windows 7 operating system. We used MATLAB's built in FAST corner detection algorithm for possible people detection. Also for r -neighbor searches, we used MATLAB's built in function. In the test image, 3124 FAST features are detected where the size of the image is 799×836 pixels. For these feature points, the first method takes less than a second for all r values. In the second method the processing time increases with r as follows. For $r=40$, the time needed is 8 seconds. For $r=60$, this time becomes 17 seconds. For $r=80$, it becomes 26 seconds. Finally, for $r=100$, the time needed becomes 38 seconds.

IV. CONCLUSION

In this study, we propose two novel methods to detect crowd regions on airborne images. These methods are based on spatial point statistics. Both methods depend on a single parameter r , which is the distance measure of people locations. Initial test results indicate that these methods can be used for crowd detection in aerial images. However, for a better performance evaluation they should be tested on a larger data set. This will be the topic of our future work.

ACKNOWLEDGMENT

This work is supported by TÜBİTAK through project number 114E199.

REFERENCES

- [1] S. Ali and M. Shah, "Floor fields for tracking in high density crowd scenes," in *ECCV 2008*, 2008, pp. 1–14.
- [2] O. Arandjelovic and A. Zisserman, "Crowd detection from still images," in *BMVC*, 2008, pp. 1–10.
- [3] S. Baykut, A. H. Ozcan, D. S. A. Sahinkaya, and I. K. Yalcin, "Analysis of spatial point process characteristics of radar detections in sea clutter region," in *SIU 2013*, April 2013, pp. 1–4.
- [4] W. Ge and R. T. Collins, "Marked point processes for crowd counting," in *CVPR 2009*, 2009, pp. 2913–2920.
- [5] W. L. Martinez, "Computational statistics in matlab®," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 3, no. 1, pp. 69–74, 2011.
- [6] O. Meynberg and G. Kusch, "Airborne crowd density estimation," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 1, no. 3, pp. 49–54, 2013.
- [7] R. Perko, T. Schnabel, G. Fritz, A. Almer, and L. Paletta, "Airborne based high performance crowd monitoring for security applications," in *Image Analysis*, 2013, pp. 664–674.
- [8] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 105–119, 2010.
- [9] B. Sirmacek and P. Reinartz, "Automatic crowd density and motion analysis in airborne image sequences based on a probabilistic framework," in *ICCV 2011*, 2011, pp. 898–905.
- [10] —, "Feature analysis for detecting people from remotely sensed images," *Journal of Applied Remote Sensing*, vol. 7, no. 1, pp. 073 594–073 594, 2013.
- [11] T. Wiegand and K. A. Moloney, *Handbook of Spatial Point-pattern Analysis in Ecology*. CRC Press, 2013.