

Sparse People Group and Crowd Detection using Spatial Point Statistics in Airborne Images

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—Crowd monitoring is an important task of security forces. If an emergency occurs during large events, authorities should take urgent measures to prevent casualties. Also understanding crowd dynamics such as tracking crowds or sparse people groups before an emergency occurs is a need. Therefore, crowd detection and analysis is a critical research area. There are several studies for crowd monitoring that use street or indoor cameras which may not be directly used for analyzing large crowds. In this study, we approach the problem using aerial images. We propose two novel methods. In the first method, we use first-order spatial point statistics. It uses the nearest neighbor relations for each person in the image to detect crowd regions. Our second method also uses the first order statistics with an additional sparse people group detection flexibility. We test the proposed methods on two aerial images and provide quantitative test results.

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. In the first group, street cameras are used [1], [4]. In the second group, airborne images are used.

We will review the studies that are based on aerial images since our method also uses aerial images. Arandjelovic and Zisserman [2] used SIFT features to detect crowds. However, this method is based on SVM classification which needs a proper training set which may not be available. Perko *et al.* [8] 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. Also, oblique views limit the image observation coverage which may be a serious drawback for observing large events. Meynberg and Kuschik [6] proposed a FAST feature extraction and SVM classification based method. They used Gabor filter banks with variety of different orientations and scales. Sirmacek and Reinartz [10] also used FAST features on airborne images for possible people detection. They approached the problem as kernel density estimation. They used Gaussian kernel for voting each FAST feature location and extracted a total crowd density map.

In this study, we propose two novel methods for crowd and sparse people group detection in aerial images. Proposed

methods are based on spatial point statistics. Spatial point statistics are extensively used in ecology. Spatial points connotate any object that wanted to be observed. In particular, ecologists mostly use them for observing patterns of plants, animals, or crater centers of volcanoes. In these studies, the basic usage of these statistics is in detecting if the points are clustered, randomly or regularly distributed. The interested reader may find a review of point pattern analysis in ecological studies [12]. Point pattern statistics has also been used in radar applications [3]. In this study, we take the location of a person that is detected from airborne image as object of interest. In the first method, we use first-order statistics and aim to detect whether persons are clustered (forming a crowd) or distributed. In the second method, we define the crowd in a quantitative manner. Therefore, the user will have a control on detecting any crowd density level. This leads to sparse people group detection besides crowds. We test these methods on two aerial images 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 [9]. We use these to extract the first order spatial statistics for crowd detection. We provide the test images to be used throughout this study in Figs. 1(b) and 1(d). We also provide the FAST feature locations for the test images in Figs. 1(b) and 1(d).

A. Crowd Detection with First-Order Statistics

The first-order statistics are based on the density of observations. In order to extract this statistic, we use people locations as observation points. 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 other saying, for each person the people that are in r neighborhood to that person is used to calculate λ .

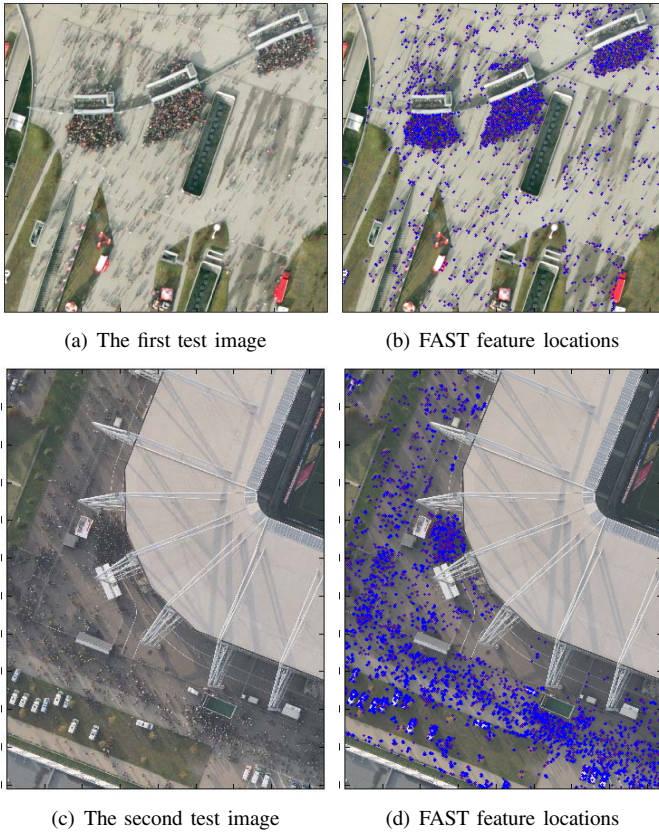


Fig. 1. Test images and possible detected people locations.

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 density map. The crowd regions in $D(x, y)$ will have higher density than sparse people groups.

B. Sparse People Group Detection

The definition of the crowded region may be relative by its nature. In basic terms, we can call a region as crowded if there are large number of people in unit area. Actually, this definition is related to the average of nearest neighbor distances between people. For example, if this distance is five meters in a region, then it is not crowded. However, if the distance is below one meter, then it can be called as crowded.

In the previous section, we used first order statistics to detect crowd regions using Eqn. 1. Let's call the average of the nearest people distances as d_p . Then, in $\pi \times d_p^2$ area we expect one person in average. Using this information, we can guess the expected number of people in $\pi \times r^2$ area as $N = \pi r^2 / \pi d_p^2 = r^2 / d_p^2$. Therefore, if the user specifies a value for d_p , then any crowd density level can be detected. This gives flexibility to detect very dense crowds or sparse people groups.

In order to extract the crowd density level, we apply the following method. For each person, number of people that are in the r neighborhood to that person is counted. We call this count as N_r . Setting a lower and upper threshold for N_r , any crowd density level is extracted. Here, the threshold will be $T = r^2 / d_p^2$. For example, for $N_r > T$ where $d_p = 1$ meter, we expect to detect dense crowds. As in the first method, we estimate the crowd density map for people that satisfy $N_r > T$.

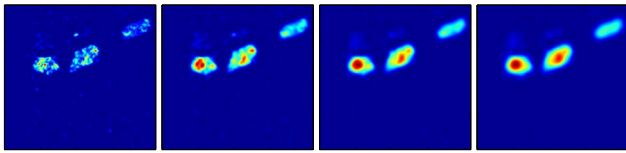
III. EXPERIMENTS

We test the proposed crowd and sparse people group detection methods on two airborne images acquired from stadium entrances. The German Aerospace Center (DLR) provided these images for test purposes. They are acquired by the 3K-Camera-System mounted on a Cessna aircraft from 1000 m flight altitude and ground sampling distance (GSD) of 15 cm [11]. The test images have been given in Fig. 1(a) and Fig. 1(c). There are sparse and crowd people groups in the images. In these images, people are gathered at the stadium gate entrances and at some other locations.

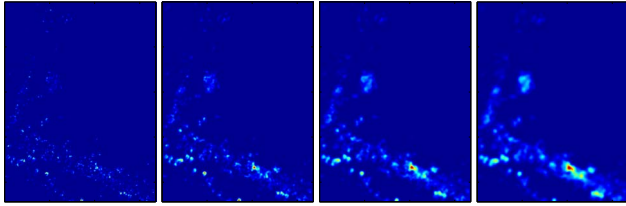
A. Visual Results

For different r values, estimated crowd density maps are given in Figs. 2(a) and 2(b) for the first and second test images respectively. Here, r values are taken as [40,60,80,100]. For lower r values, the calculated density maps have higher variance due to σ . For larger r values, the density map is smoothed as expected. To extract crowd regions, we use Otsu's thresholding method on the density map [7]. We provide the extracted crowd density boundaries for $r = 60$ in Figs. 3(a) and 3(b) for the first and second test images respectively. Here, we manually extracted the ground truth for crowd regions (given in blue color). In these figures, the detected crowd region boundaries are given in black color.

For the first test image, we provide the people satisfying the condition $N_r > r^2 / d_p^2$ for $d_p = 1$ and the corresponding crowd density in the first column of Fig. 4. To extract sparse people groups, we need upper threshold T_u and lower threshold T_l so that any crowd density can be detected. Instead, we find people for a lower threshold only such as we choose $d_p = 2$. However, this time the people that were detected in dense crowd regions are not included for this calculation. We provide the result for this case in the second column of Fig. 4. Similarly for $d_p = 3$, the people that were detected in the first and second steps are discarded. The result is given in the third

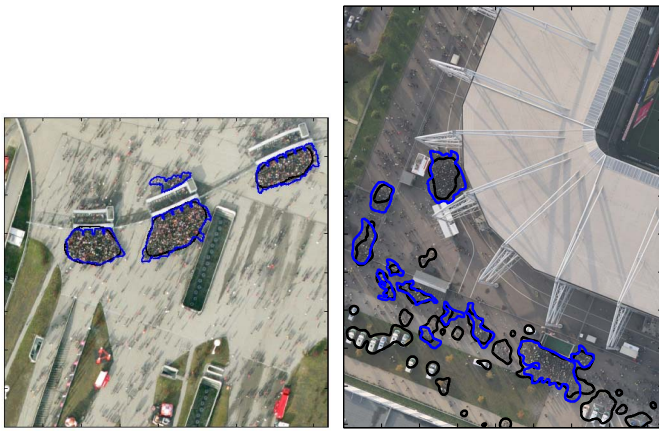


(a) Crowd density maps for the first test image



(b) Crowd density maps for the second test image

Fig. 2. Estimated crowd density maps for $r=[40\ 60\ 80\ 100]$.



(a) Detected dense crowds for the first test image (b) Detected dense crowds for the second test image

Fig. 3. Crowd detection results for $r=60$ where ground truth crowd boundaries are given in blue and detection results are given in black.

column of Fig. 4. Finally, for $d_p = 4$ the result is given in the last column of Fig. 4.

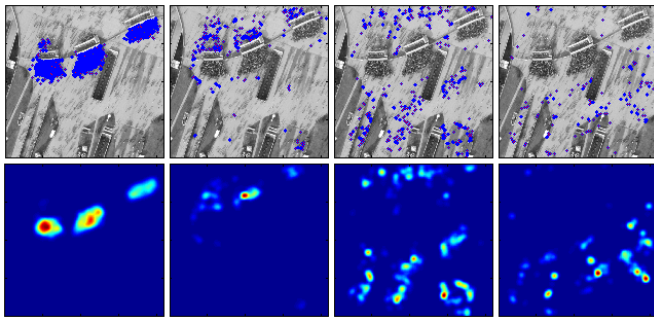


Fig. 4. The first test image, crowd and sparse people detection results. Here different crowd density levels are iteratively detected.

The first image is acquired over a stadium with entrances. Here, there are dense crowds at the gate entrances. These are detected in the first level by the proposed method. Behind the

gates, towards the upper side of the image, the crowd density is lower. Also, from the lower side of the image up to the gate entrances people are walking in sparse groups. These are detected at the second level by the proposed method. In the third and the fourth levels, the detected people are in sparser groups.

We also provide the sparse people detection results for the second test image in Fig. 5. In this image there are a few very crowded regions and some less crowded regions at different locations. As in the first test image, we detect the first level density, dense crowd of people for $d_p = 1$. The result is given in the first column of Fig. 5. In the other columns of Fig. 5, we provide the sparse people detection results for $d_p = 2$, $d_p = 3$, and $d_p = 4$ respectively.

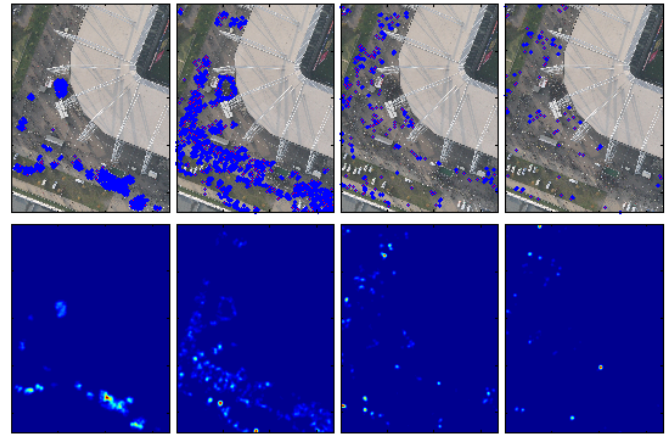


Fig. 5. The second test image, crowd and sparse people detection results. Here different crowd density levels are iteratively detected.

B. Qualitative Results

As we mentioned previously, deciding on a crowd region may be relative. We have provided sparse people group detection results in Figs. 4 and 5. We do not provide any quantitative performance results for these. We will only focus on the performance of the first method only where dense crowds are detected. Since the first proposed method only uses the single parameter r , we obtained quantitative results for $r=[40\ 60\ 80\ 100]$ in terms of true positive (TP) and false alarm (FA) of crowd region areas. In Table I, we provide the results for both test images in percentages. As can be seen in this table, increasing r increases TP but FA is also increased. The reason for this is that, as r is increased the density map becomes smoother. In the second test image, we detected FAST features on the top of the trees and cars. These produced false alarms. There are other blob detection algorithms in the literature which may decrease these false alarms. Since our approaches focus on using spatial statistics for crowd detection, we leave eliminating false detections in a future work.

TABLE I
CROWD DETECTION PERFORMANCES IN PERCENTAGES.

r	Test Image 1		Test Image 2	
	TP	FA	TP	FA
40	80.27	0.51	53.12	5.72
60	81.96	0.67	71.01	7.51
80	83.97	1.27	78.06	9.29
100	85.89	2.39	83.50	11.87

C. Computation Time

We also checked the processing time of our method. In tests, we used a PC 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 first test image, 3124 FAST features are detected where the size of the image is 799×836 pixels. In the second test image, 4797 FAST features are detected where the size of the image is 2218×1677 pixels. For these feature points, the proposed method takes less than a second for all r values and for both images.

IV. CONCLUSIONS

In this study, we propose a novel approach to detect crowd and sparse people groups in airborne images. The proposed methods are based on spatial point statistics. The first method for crowd detection depends on a single parameter r , which is the distance measure of people locations. For the sparse people detection, an additional parameter (nearest distance of people) should be entered which gives control on the crowd density levels. Initial test results indicate that, the proposed methods can be used for crowd and sparse people detection in aerial images.

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. Kuschik, "Airborne crowd density estimation," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 1, no. 3, pp. 49–54, 2013.
- [7] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on System, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [8] 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.

- [9] 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.
- [10] 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.
- [11] —, "Feature analysis for detecting people from remotely sensed images," *Journal of Applied Remote Sensing*, vol. 7, no. 1, pp. 073 594–073 594, 2013.
- [12] T. Wiegand and K. A. Moloney, "Rings, circles, and null-models for point pattern analysis in ecology," *Oikos*, vol. 104, no. 2, pp. 209–229, 2004.