International Journal of Engineering Research and Advanced Technology (IJERAT)

DOI: 10.31695/IJERAT.2021.3678

E-ISSN : 2454-6135

Volume.7, Issue 1
January -2021

The Use of Input Words Vectors to Detect Crowd Region in Images

Sally Ali Abdulateef, Lamyaa Mohammed Kadhim, Ekhlas Wattan Ghandawi

Department of Computer Science

College of Dentistry

University of Al-Mustansiriyah

Bhagdad, Iraq

ABSTRACT

Human crowd analysis has common utilizations from the urban engineering and traffic management to law enforcement. They all need a crowd for first being detected, and is the issue that has been dealt with in the present study. Considering an image, the algorithm that has been proposed in this paper performs a segmentation of that image to crowd and non-crowd areas. The fundamental concept is capturing two main characteristics of the crowd: (a) on a narrower scale, its main elements have to appear like humans (only weakly so, as a result of the low resolution, dressing variations, occlusion, and so on), whereas (b) on the wider scale, the crowd intrinsically includes elements of the redundant appearance. The proposed approach makes use of that through the utilization of underlying statistical framework which has been based on the quantized features of the SURF. The two previously mentioned characteristics of the crowds have been obtained through the resultant statistical model responses' feature vector, which describe the level of crowd-like appearances around the location of an image with the increase of the spatial level around it.

Keywords: SUPER Descriptor, Crowed detection.

1. INTRODUCTION

During the past years, the researchers made attempts towards solving the crowd detection in the images with the use of various methods. The early researches have been focused upon the detection approaches for recognizing certain parts of the body or the entire body with the use of the hand-crafted characteristics [1-4].

Whereas the detection based approaches have been hard to deal with the dense crowds due to the occlusion, some of the researches have studied learning a function of the mapping between the characteristics to the number of the people [5,6]. In addition to that, Lempitsky et al. [4-7] have suggested local characteristics for density map to make use of the spatial information. None-the-less, hand-crafted characteristics aren't adequate enough in the case of facing clutter and low image resolutions. The analysis of the crowd can be taken under consideration as one of the most interesting areas of research in a variety of the areas like the sociology, psychology, computer vision, and engineering [8]. For example, [9] sociologists and psychologists are focused on the exploration of the individual behaviors for the sake of the enhancement of the safety services for the individuals, especially in the over-crowded regions. This may be accomplished via the understanding of the people interactions in a certain region.

A wide range of the research agencies and universities like the British Engineering and Physical Sciences Research Council (EPSRC), the Defense Advances Research Projects Agency (DARPA), and the EU funded projects ADVISOR and PRISMATICA have been focusing on the detection, counting, estimation and tracking of the moving individuals with the use of video recordings or images. Initially, the monitoring and the management of the crowd has required the construction of a model providing crowd control and surveillance for the sake of responding to the situation of the event [10-13]. In the present paper the interest has been focused on the detection and the segmentation of human crowds in the static images. Considering an image, the aim is determining whether or not it includes a crowd, and in the case where it does, it has the task of determining the parts of image that it occupies. Being capable of inferring the existence of the crowd in the image is in fact a beneficial task: crowd formations might result in causing the delays in the shopping centers, underground passages, and streets, or it might be a civil unrest indication [9,10]. In automotive industries, crowds are interesting as possible road hazards. In addition to that, it is necessary to segment a crowd before the higher level tasks, like counting (or, in a more general sense, estimation) the number of

the people in the crowd, or the analysis of their interaction and behavioral dynamics. Therefore, the field on which the crowd detection may be implemented, differs from the psychological researches and the macro-engineering, to crime detection and prevention [14-18].

A variety of approaches were presented for the density estimation of a crowd. One of which is the background removal in [19] which has been utilized on a reference image with only background for segmenting the image pixels into background or pedestrians, i.e., it has been based upon the subtraction of foreground pixels from the background ones with the use of the statistical pixel information. Despite the fact that using this approach has led to promising outcomes, it operates optimally for crowds of low density only. In the majority of the earlier researches, the issue of the robust detection of the crowd is almost entirely evaded through the use of a simple background subtraction form.

For instance, Roqueiro and Petrushin [20], have utilized a static camera and the data which has been gathered over a long period for the estimation of the appearance of the background. Brostow and Cipolla [21] have applied independent motion detection to the crowds, which has efficiently performed the segmentation of the motion on a small group of the interest points. This method has suffered from the issue of the lack of a model of appearance and, as a result, unable of finding the still persons (which has increased the false negative errors) it also was unable of recognizing the cases where the objects in motion aren't people (which has increased the false positive errors). In addition to that, using the independent motions to count the people in the crowd is not certain, due to the fact that the crowds (or parts of them) usually show a level of behavioral coherences.

Rabaud and Belongie [22] have proposed an equivalent method that has virtually suffered from similar types of limitation. The key distinction is utilizing a weak geometrical framework, bounding box, constraining the spatial degree of every one of the independently moving bodies (in other words, a set of the interest points) at the same time as permitting the articulations in it (see [23] as well). Which is a result of the dependence of the high scale, due to the fact that authors didn't propose ways for the automatic selection of the size of the bounding box. Another group of hypotheses has been made by "Reisman et al." [24] where they have designed a system specifically for the detection of pedestrian crowds. They have been dependent upon detecting zebra crossings and right-to-left (or vice versa) movement of certain pedestrians, in relation to forward-facing camera on the moving vehicle. Unlike the abovementioned approaches, the suggested method in this study performed no background subtractions. Moreover, they have not used video or motion, and rather they have entirely depended on the cues of the appearance and the separate images. Ultimately, this method has not been based on the detection of the individuals and as a result, may be implemented on large as well as small crowds.

2. PROBLEM DIFINITION

The crowd can be defined as a set of the spatially close objects of one class. In the present paper, human crowds have been specifically considered, as type which is typically of the maximum interests in the practice. There is a number of reasons to why the detection of the crowds can be considered as a challenge. Initially, limited image resolutions, meaning the fact that evidences for a specific individual are typically quite scarce. Considering the fact that there is quite an abundance of the partial occlusions in the crowds, and variations in the clothes, poses and lighting, detecting persons as the key element cannot be considered as a potential method [24].

However, a method which searches for a number of individuals in a direct way, suffers from issues of modelling a considerably increased variability range in the combined appearances, in addition to factors that are crowd specific like the distances amongst the person, which is referred to as the density of that crowd.

3. THE PROPOSED METHOD OUTLINES

The concept of the crowd is intrinsically involved with the redundant occurrences, necessitating a specific spatial degree over which that redundancy is shown. Therefore, for the sake of propagating the local data, on top level segmentation is approached as an issue of the minimum graph cut. In a sense which the vertices of the graph are corresponding to true pixels of the image (and the closeness of those vertices to that in image), the suggested approach is of rather a high density in nature. None-the-less, the proposed method has sparse features in this appearance has been defined in relation to sparse group of the local characteristics. Extracting and modelling them is the initial process of the proposed approach and Figure 1 illustrate the flowchart of the proposed method.

DOI : 10.31695/IJERAT.2021.3678

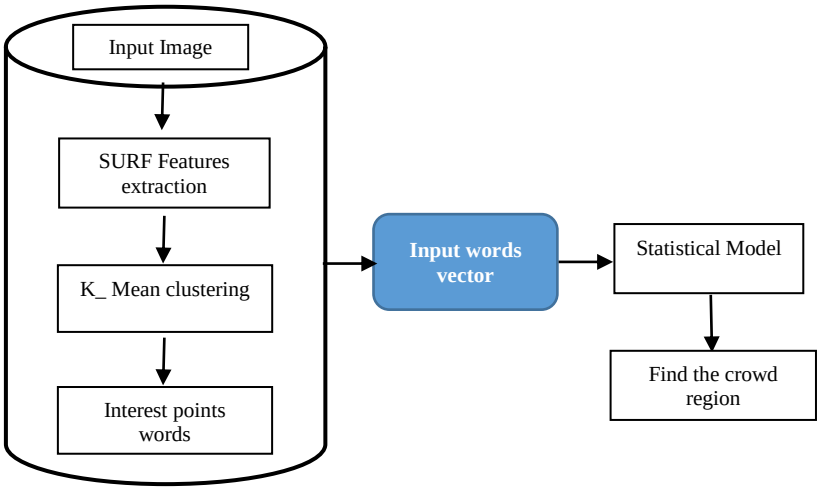

Figure 1. The Proposed System Flowchart

3.1. The Fundamental Features

On lowest level the local characteristics are utilized for characterizing the content of the image. Which are corresponding to a sparse interest point set, characterized as the scale-space extrema in a DoG pyramid which has been produced from the initial image. Naturally, crowds in general provide numerous interest points. Like Bay [25], SURF descriptor has been utilized for describing the neighborhood of every one of the points, at a scale where the interest point has been found.

Every one of the descriptors has been quantized through assigning that descriptor to the closest one of SURF words, or K clusters, which has been calculated by the use of the K-means clustering descriptors which have been obtained from the image. The key contribution concerns of this study are the way by which the calculated group of the SURF words has been utilized. As a first step, extract the SURF features from image, to create a wide range of features. To find the interest point from the big set of SURF features, second step, grouping the SURF features using *K*- mean clustering method. The size of *K* has high important in detect the number of interest point and the size of interest point words vector. Final step is putting the centroid of each cluster in a vector, called the interest point words vector.

3.2. Statistical Model

Assuming to be dealing with the crowds at predefined scales. The aim is deciding whether or not a specific part in the image (does not necessarily have to be the interest point's location) is corresponding to a crowd area or non-crowd area. For doing that, a K-cluster is considered around it and has the attempt of quantifying how much "crowd-like" it is. Due to the fact that the number of the interest points which have been obtained (and the obtained SURF words) are in general, insignificant compared with the number of the pixels of the image, in other words, For the purpose of finding the crowd area we first extract the important attributes and then we compose those important points to be the words and then we put those words extracted in one vector we use to calculate the size of the crowd area. Our work suggests discovering and sensing the crowd area by linking the number of SURFwords extracted in the image and the size of the cluster and its relationship to the size of the image points as follows:

The crowd region is the number of SURF words multiplying by the size of cluster.

$$C_R = S_W * K \quad (1)$$

Where S_W is the number of SURF words , and K is the cluster size . And the C_R is the crowd region if and only if :

$$\{C_R >= x*y/2\}\&\{ dis(S_W{}^{i}) - dis(S_W{}^{i+1}) >= 0.001\} \quad (2)$$

4. EXPERMENTAL RESULTS

For the sake of evaluation of the efficiency of the suggested approach, a data-base of 100 images has been collected, 50% of which include a crowd, as seen in Figure 2.

DOI : 10.31695/IJERAT.2021.3678

Figure 2. Example Images from Database.

15 images have been arbitrarily chosen from the data-base and after that, hand segmented to crowd areas and non-crowd areas and have been utilized for training the proposed algorithm. The training included (a) the detection of the interest points, (b) clustering related descriptors to 1,000 SURF words, (c) estimating the statistics model as has been explained in Sec3.2. the rest of the 85 images have been utilized for testing the proposed method's efficiency.

The images showing the crowd distribution were obtained from the on-line video and utilized in the present study after the processes of the preprocessing, which includes the trimming and resizing. The images of the crowd have been treated afterwards as a camera position and orientation function. Which is why, the horizontal and vertical (close and long range) images have been utilized. Figure 3 illustrates the image types that are used in the present research. Those image specifications have been shown as well in Table1. After that, the crowd characteristics were obtained with the use of a suggested algorithm according to SURF approach, have been utilized for the detection of the interest points that fundamentally represent the characteristics of the crowd. After that, the clustering processes have been utilized for the elimination of feature points that don't belong to the characteristics of the crowd.

Figure 3. Image Orientations used in Proposed Method

DOI : 10.31695/IJERAT.2021.3678

Table1. Image Specifications

Image orientation	Camera angle	Image size in pixel		
Vertical	90°	691	1359	3 unit8
Horizontal (close range)	45°	683	471	3unit8
Horizontal (long range)	45°	689	1366	3unit8

The obtained features of the crowd were illustrated in Figure 4 with the use of the suggested approach. As seen in the figure, the suggested method has been successful in the discerning of the features of the crowd from other classes of features. when comparing between the horizontal and the vertical images, it can be noted that in the case of the use of the vertical images, results representing the actual features of the crowd have been of a rather higher accuracy compared to the features that have been presented in both the horizontal images.

Which is due to the fact that the camera has been focused upon a region which is covered with crowds. None-the-less, in the cases of other images, the surrounding environment (i.e. streets, buildings, and trees) took a massive portion of image. As seen in Figure 4, showing the detected points of the SURF in the horizontal images, some detected characteristics have been marked as features of the crowd, whereas they are not in fact. Which results from the values of the intensity of these pixels (the features from the environment around, i.e. streets, trees, and building) have comparable values of intensity to the ones of the crowd features.

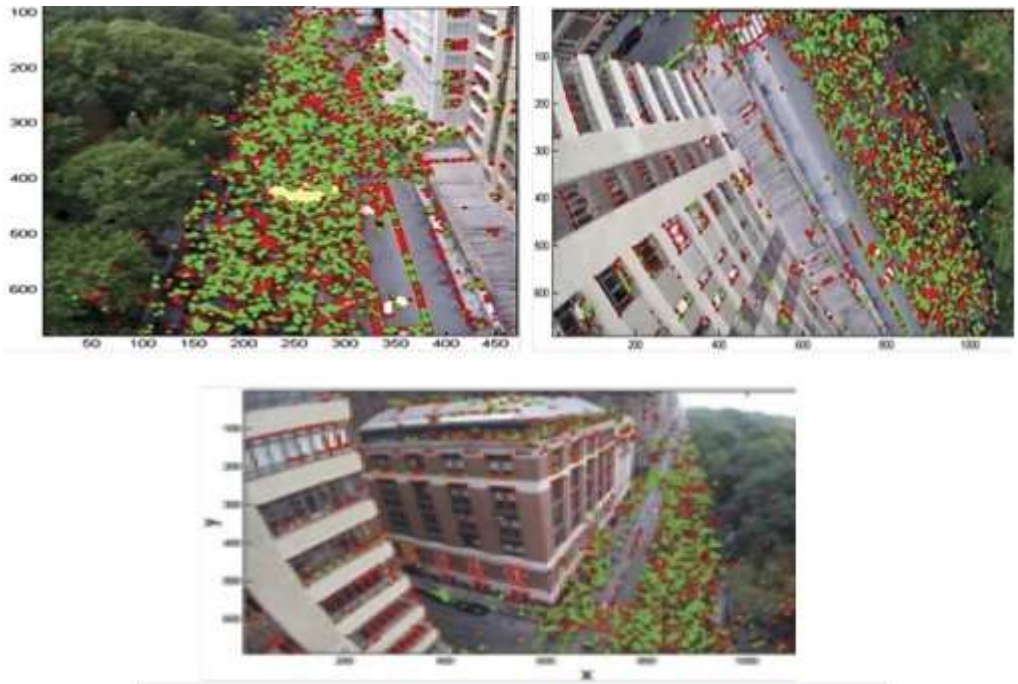

Figure 4. Crowd Feature Extraction

DOI : 10.31695/IJERAT.2021.3678

The characteristic group of the results has been illustrated in Figure 5, for the data which includes crowd, the detection of the crowds over various scales, view-points, and types of scenes which range from street crowds to political rallies and concerts.

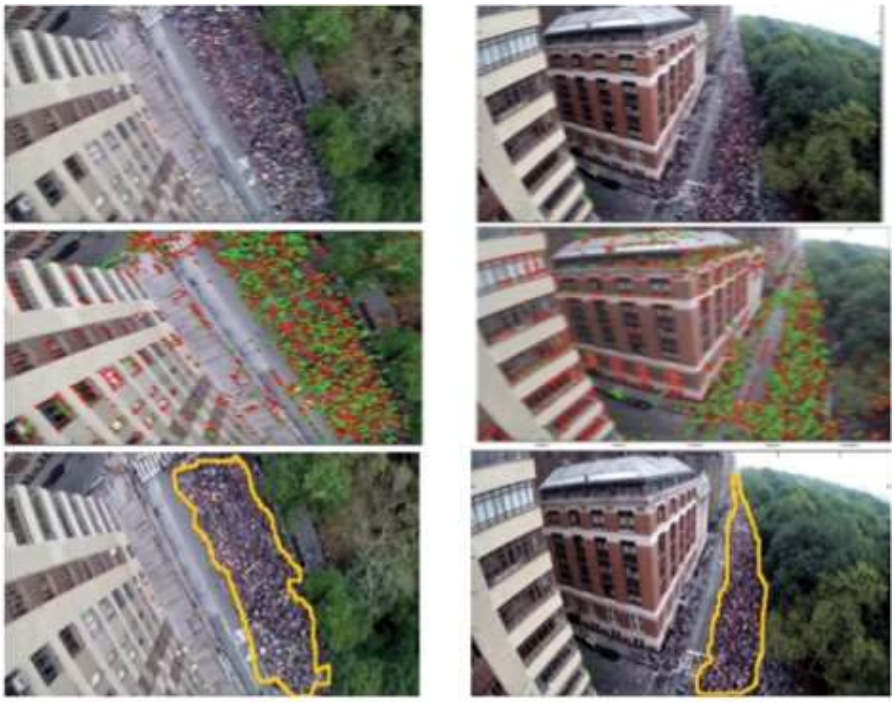

Figure 5. Results on Data Containing Crowds

5. SUMMARY AND CONCLUSION

The key contribution in the present research is a new approach which has been designed to detect crowds in the static images. This approach has been designed as appearance-based, utilizing a statistical, occurrences model of the quantized SURF words over the image. The suggested approach has shown potential outputs on the data-set that has considerable variations in the viewpoints, appearances of people in a crowd, the scale and density of those people, and the type of the background scene. In the present study, a new algorithm which is based upon the detection of the SURF features and clustering those features with the use of the K-mean approach was presented and implemented for the detection of the features of the crowd from a variety of the images. The vertical and horizontal images of the UAV have been utilized for the detection of the crowd features and for the mapping of crowd density levels. In the complicated types of environment, in which there are several classes of features in the image, the detection of the crowd features can be considered as a complicated task. Results have shown that the suggested approach has been capable of the detection of the crowd features amongst other classes of feature. Which was observed in all of the cases of the images. Future researches need to be focused on the detection of the crowd features with the use of the geo-referenced images for the generation of real crowd density maps.

7. REFERENCES

[1] H. Fradi, and J. L. Dugelay, “Crowd Density Map Estimation Based on Feature Tracks.” Paper Presented at MMSP 15th International Workshop on Multimedia, 2013.

[2] M. Jiang, J. Huang, X. Wang, and J. Tang, “An Approach for Crowd Density and Crowd Size Estimation.” Journal of Software 9 (3): 757−762. doi:10.4304/jsw.9.3.757-762, 2014.

[3] Y. Biadgie, and K. A. Sohn, “Feature Detector Using Adaptive Accelerated Segment Test.” Paper Presented at 5th International Conference on Information Science and Applications, Seoul, May 6−9, 2014.

[4] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, “Single-image crowd counting via multi-column convolutional neural network,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 589–597.

[5] D. Onoro Rubio and R. J. L ˜ opez-Sastre, “Towards perspective-free ob- ´ ject counting with deep learning,” in European Conference on Computer Vision (ECCV), 2016, pp. 615–629

[6] V. A. Sindagi and V. M. Patel, “Generating high-quality crowd density maps using contextual pyramid cnns,” in International Conference on Computer Vision (ICCV), 2017, pp. 1879–1888.

DOI : 10.31695/IJERAT.2021.3678

[7] Y. Li, X. Zhang, and D. Chen, "CSRNet: Dilated convolutional neural networks for understanding the highly congested scenes," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 1091–1100.
[8] Z. Shi, L. Zhang, Y. Liu, X. Cao, Y. Ye, M. Cheng, and G. Zheng, "Crowd counting with deep negative correlation learning," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 5382–5390.
[9] X. Liu, J. van de W., and A. D. Bagdanov, "Leveraging unlabeled data for crowd counting by learning to rank," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7661– 7669.
[10] D. Sam and R. V. Babu, "Top-down feedback for crowd counting convolutional neural network," in AAAI Conference on Artificial Intelligence (AAAI), 2018, pp. 7323–7330.
[11] F. Burkert, and Butenuth, M, "Event Detection Based on a Pedestrian Interaction Graph Using Hidden Markov Models". *Paper Presented at 2011 ISPRS conference on Photogrammetric image analysis, Munich, Germany*, October 5−7 , 2011.
[12] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005, pp. 886–893.
[13]] M. Li, Z. Zhang, K. Huang, and T. Tan, "Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection," in International Conference on Pattern Recognition (ICPR), 2008, pp. 1–4.
[14] A. B. Chan, Z. J. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–7.
[15] V. Lempitsky and A. Zisserman, "Learning to count objects in images,"in Advances in Neural Information Processing Systems (NeurIPS), 2010, pp. 1324–1332.
[16] H. Idrees, I. Saleemi, C. Seibert, and M. Shah, "Multi-source multiscale counting in extremely dense crowd images," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 2547–
2554.
[17 M. Nagai, T. Chen, R. Shibasaki, H. Kumugai, and A. Ahmed, "UAV-borne 3-D Mapping System by Multi-Sensory Integration." IEEE Transactions on Geoscience and Remote Sensing,2009, 47 (3): 701–708. doi:10.1109/TGRS.2008.2010314.

[18] M. Patterson, and A. Brescia, "Integrated Sensor Systems for UAS." Paper Presented at 23rd Bristol International Unmanned Air Vehicle Systems (UAVS) Conference, Bristol, United Kingdom, April 7–9,2008.
[19] J. Yin, S. Velastin, and A. Davies ," Image Processing Techniques for Crowd Density Estimation Using a Reference Image",.*Paper presented at 2nd Asia−Pacific Conference on Computer Vision*, Singapore, December 5−8 ,1995.
[20] D. Roqueiro, and V. A. Petrushin ," Counting people using video cameras", *International Journal of Parallel, Emergent and Distributed Systems (IJPEDS)*, 2007,22(3):193– 209.
[21] G. J. Brostow, and R. Cipolla ," Unsupervised bayesian detection of independent motion in crowds", *In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*,2006, 1:594–601.
[22] V. Rabaud, and S. Belongie ," Counting crowded moving objects", *In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*,2006.
[23] D. Kong, D. Gray, and H. Tao ," A viewpoint invariant approach for crowd counting",*In Proc. IEEE International Conference on Pattern Recognition (ICPR)*, 2006,3:1187–1190.
[24] P. Reisman, O. Mano, S. Avidan, and A. Shashua ," Crowd detection in video sequences",*International Symposium on Intelligent Vehicles*, 2004 ,pages 66–71.
[25] H. Bay, T. Tuytelaars, L. VanGool ," SURF: Speeded UpRobust Features,‖In ECCV (1),2006, pp. 404 –417.
[26] ,"Human face detection and recognition using contour generation and matching algorithm",

DOI : 10.31695/IJERAT.2021.3678