An Approach to Crowd Segmentation at Macroscopic Level

Darko Mihalić*, Darijan Marčetić**, Slobodan Ribarić**
* University of Zagreb, Faculty of Organization and Informatics, Zagreb, Croatia
** University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia
darkox3@gmail.com, {darijan.marcetic, slobodan.ribaric }@fer.hr

***Abstract* - In this paper, an approach to crowd segmentation at macroscopic level is proposed. The method is based on salient points within the region of dense optical flow that are tracked through N frames. Salient points were tracked by using the pyramidal Lukas-Kanade tracker. Around each salient point, an area is formed by using combination of a dense optical flow, a morphological operation dilation and Voronoi diagram. To each subarea, i.e. Voronoi cell of an optical flow (VcOF), the direction and magnitude of displacement is assigned based on the tracklet of the salient point. The VcOFs are grouped based on their directions and distances by hierarchical clustering. The method considers the temporal and spatial dynamics of the crowd motion. It has been tested on the video clips of real-world video sequences and has produced the results that are close to human perception and segmentation of a crowd.**

Keywords - crowd segmentation; optical flow; Voronoi diagram; tracklet

I. Introduction

Crowd behaviour analysis is becoming increasingly popular topic in the field of computer vision. The progress made in this research area has many practical applications [1, 2] such as: i) visual surveillance of crowded public places, ii) crowd management and predicting security threats, iii) counting people and estimation of crowd density, iv) designing public spaces where a large number of people are expected, and v) simulation of crowd movement and behaviour.

Generally, the first step in analysing and predicting crowd behaviour is crowd segmentation. Crowd segmentation is challenging task, especially for the high-level crowded scene, because conventional computer vision methods (object detection, tracking, individual behaviour recognition) are not appropriate for this level due to extreme clutters, a small number of pixels representing individuals, occlusions and ambiguities. Also, crowd of people may contain several groups of people with diverse directions of movement, velocities, and trajectories, so crowd movement can exhibit complex, even chaotic movement patterns and behaviours, which are hard to analyse and/or predict.

To solve crowd segmentation task at macroscopic level, we propose a new method of crowd segmentation based on the tracking of salient points in a set of short video clips. The salient points are located in an area predetermined by a dense optical flow calculated on the first two frames of each video clip. By using of the pyramidal Lukas-Kanade (L-K) tracker, the set of tracklets of salient points is obtained. Each tracklet after N frames, where N is a total number of frames in a short video clip, is represented with starting point (x, y), a magnitude of its displacement and direction.

978-1-7281-5973-7/20/$31.00 ©2020 IEEE

Note that some tracklets which have displacement of the starting points less than predetermined threshold value are neglected. The set of starting points of tracklets are used for Voronoi diagram generation. By overlapping the area obtained by combination of a dense optical flow, morphological operation dilation and Voronoi diagram, the Voronoi cells of an optical flow (VcOFs) are obtained. Each VcOF is characterized by starting position, magnitude and direction of the corresponding tracklet. The method considers the temporal and spatial dynamics of the crowd motion. It has been tested on the video clips of 50 real-world video sequences and has produced the results close to human perception, segmentation and interpretation of a crowd. The paper is organized into five sections: after Introduction, Section II. gives a brief overview of the related work. Section III. provides a detailed description of the proposed method. Section IV. presents preliminary experimental results, and section V. gives the conclusion and provides insights for future work.

II. Related works

Overview papers [1, 2] describe major research milestones that have been achieved in the area of the analysis of crowd behaviour within last 15 years. Solmaz et al. [3] analyse the advection of optical flow particles using linearization of a continuous dynamic system around critical points. From the determinant and trace values of the Jacobian matrix, and the eigenvalues of the matrix, inferences can be made about the trajectories of the crowd, such as bottlenecks, fountainheads, lanes, arches and blocking. Ali and Shah [4] treat the moving crowd as a dynamic system and, using an optical flow, calculate Lagrange trajectories of the flow particles. The FTLE field (Finite Time Lyapunov Exponent) is calculated from the motion of particles along the x and y axes, and the boundaries of different flows within the crowd are reflected as LCSs (Lagrangian Coherent Structures) in the FTLE field. The final flow segmentation is performed using the normalized cuts algorithm. Sharma and Guha [5] divide the frame into blocks and follow each block through a series of frames. The resulting paths are grouped using direction, location, and path density information, and based on the resulting groups, they define crowd flow segments. In [6], the authors observe feature points of optical flow and calculate LTDS (local-translation domain segmentation) for them and based on this they derive a conclusion about the shape and motion of the crowd. Papers [7, 10] use motion vectors extracted from video to segment the flow of the crowd, and focuses on pedestrian monitoring, considering pedestrian social responses as well as their orientation toward the destination. The authors Li et al. [9] use a dense optical flow that they display as a grey image and then create a histogram of that image. From the maximum and minimum of the histogram curve, they obtain the data needed to segment the crowd flow. Optical flow for crowd segmentation is also used by Wang et al. [11]. They select a set of points within a dense optical flow, and for these points the density and distance attributes are calculated. These attributes are used for crowd segmentation. Xu and Anjulan [12] isolate the corresponding peaks of optical flow and connect them into tracklets. The positions of the tracklet represent the vertices of the Delaunay triangulation graph. Over the time they monitor the spatial relations between the vertices and create visual descriptors based on which they draw conclusions about the behavior of the crowd. J. Cezar S. J. et al. [13] used the position of each detected individual as a site for the Voronoi diagram at each frame to determine and temporal evaluate some sociological and psychological parameters, such as distance to neighbours and personal spaces.

III. Proposed Method For Crowd Segmentation

The proposed method has the following consecutive stages: A) Feature extraction, B) Tracklets generation, C) Generation of Voronoi cells of an optical flow (VcOFs), D) Hierarchical clustering of VcOFs by direction and distance. The flowchart of the method is showed in Fig. 1.

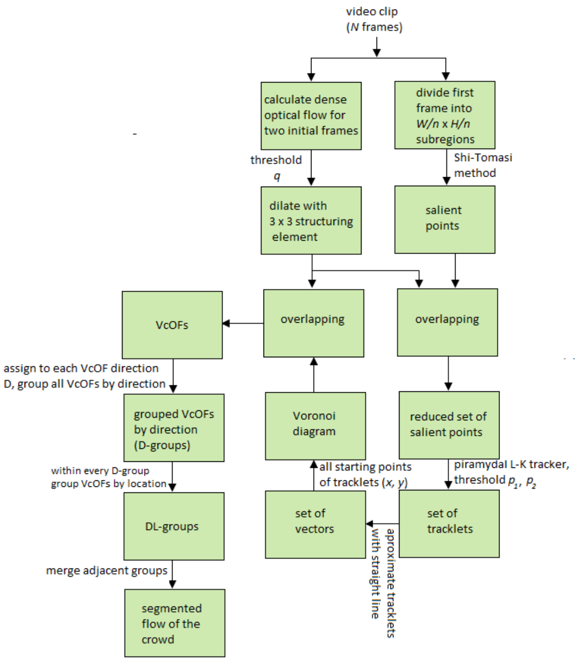


Figure 1. The flowchart of the method.

A. *Feature extraction*

N successive frames from the video sequence are extracted to obtain a video clip. The first frame of a video clip is divided into $n \times n$ rectangles (regions) where each rectangle is width W/n and height H/n, and W and H are width and height of a frame (in pixels) respectively (Fig. 2.). Within each rectangle, by using the Shi-Tomasi method, salient points, i.e. corners are extracted (Fig. 2.). A dense optical flow based on the two initial frames in the video clip is obtained by using the Horn-Schunk method. The pixels of the dense optical flow whose magnitude is below the threshold q are rejected. The dense optical flow area is dilated with 3×3 structuring element (Fig. 3.). All salient points within the region of dilated dense optical flow are used for further processing (Fig 4.).

Figure 2. Initial frame divided on rectangles and extracted salient points.

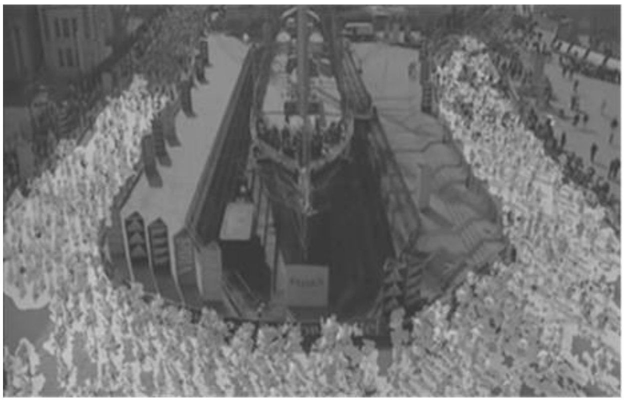


Figure 3. The dense optical flow area.

Figure 4. Salient points for further processing.

B. *Tracklets generation*

The salient points from frame *1* to frame *N* are tracked by using the pyramidal L-K tracker. After tracking, all salient points whose displacement through *N* frames is less than threshold p_1 are discarded. All salient points whose displacement along the x or y axis, between two consecutive frames, is greater than threshold p_2 are also discarded. For each remaining salient point, its trajectory through *N* frames (tracklet of salient point) is created. Each tracklet is approximated with a straight line (Fig. 5.). For each tracklet its location of a starting point (x, y), direction D and magnitude M are recorded.

C. Generation of Voronoi cells of an optical flow (VcOFs)

A Voronoi diagram of filtered salient points is generated (Fig. 6. a). The Voronoi diagram is overlapped with the area determined by a dense optical flow (Fig. 6. b) and the Voronoi cells of optical flow (VcOFs) are obtained (Fig. 6. c). Each VcOF is characterized by a starting point, magnitude and direction of the corresponding tracklet. Each pixel in the VcOF assumes the direction and magnitude of its salient point.

D. Hierarchical clustering VcOFs based on direction and distance

By using the hierarchical clustering algorithm [14], VcOFs are grouped based on direction and distance. First, VcOFs are grouped into D-groups based on direction. The number of groups depends on the distribution of directions of the cells and selected threshold direction d_1. Each D-group is further clustered by the DBSCAN algorithm based on distances of VcOFs - DL-groups are obtained (Fig. 7.). The number of DL-groups depends on the threshold l_1 of distance between VcOFs. Spatial adjacent DL-groups are merged if directions of the VcOFs along the common boundary of the two groups do not differ more than threshold d_1. Figure 8. depicts the final result of crowd segmentation - only one crowd group is detected.

IV. Preliminary Experiments

The method described in Section III. has been tested on video clips of 50 video sequences of different lengths (from 300 to 2500 frames) and resolutions (from 320 × 240 to 720 × 400 pixels). A clip was composed of 10, 20 or 30 successive frames. All video sequences depict real-life situations and were recorded with a stationary camera. The preliminary experiments have shown that results are close to the human perception and segmentation of a crowd. The method has demonstrated robustness to handle different scenes and different types of dense crowds. Figures 8. and 9. illustrate the examples of acceptable results of crowd segmentation.

Fig. 10. depicts a deficient crowd segmentation result. The problem arouse due to fact that the overlapping of the salient points and a dense optical flow results with areas corresponding to both moving people and moving objects (escalator; green and yellow areas).

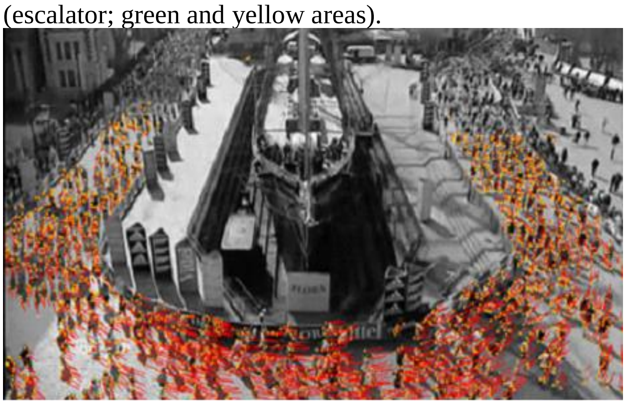
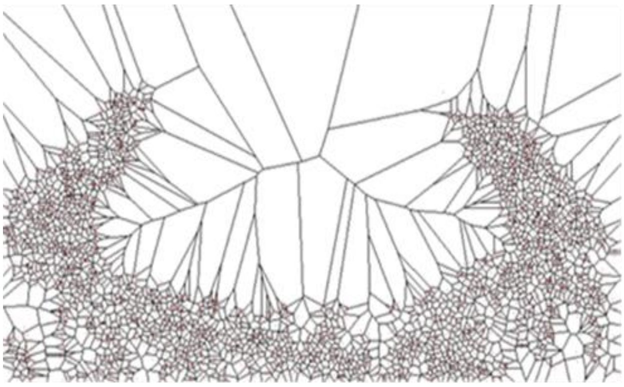
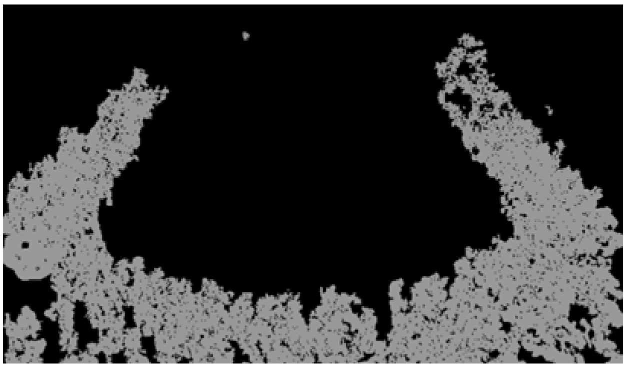


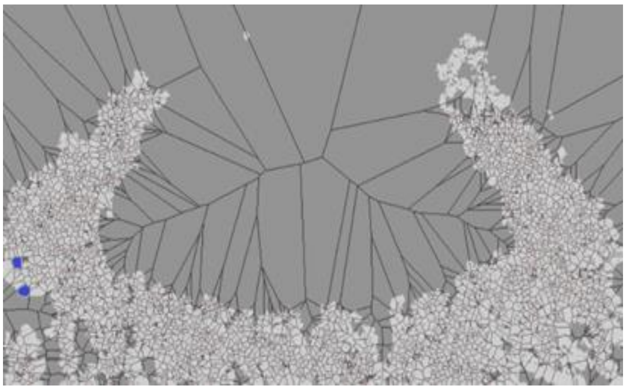
Figure 5. Result of tracklets generation.



a)



b)



c)

Figure 6. Voronoi cells of the optical flow (VcOFs): a) Voronoi diagram with salient points, b) dense optical flow, c) VcOFs (two different VcOFs are marked in blue).

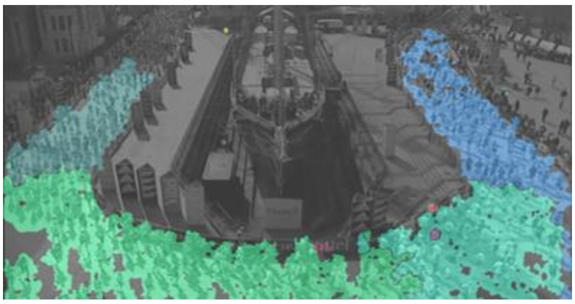


Figure 7. DL-groups of VcOFs (N = 10, n = 6, W = 720, H = 400, q = 0,5, p_1 = 2, p_2 = 10, d_1 = 60º, l_1 = 20).

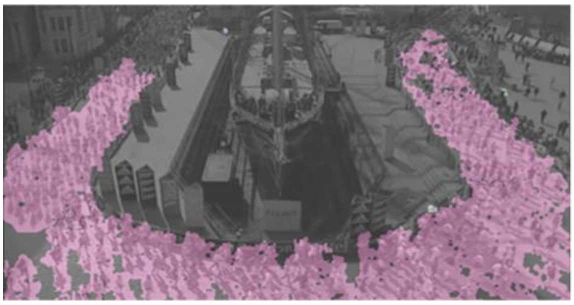


Figure 8. An example of the final result of crowd segmentation.

Figure 9. An example of acceptable result of crowd segmentation.

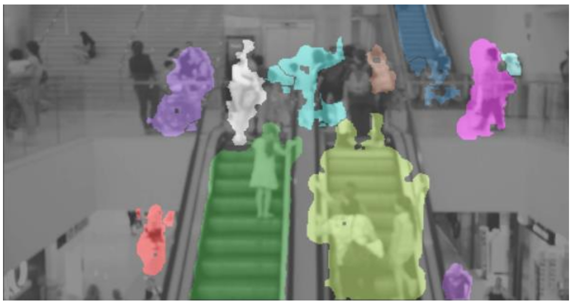


Figure 10. An example of a poor result of crowd segmentation.

V. CONLUSION AND FUTURE WORK

In this paper, an approach to segmentation of a crowd is proposed. The method consists of the following consecutive stages: i) Feature extraction, ii) Tracklets generation, iii) Generation of Voronoi cells of an optical flow (VcOFs), iv) Hierarchical clustering of VcOFs by direction and distance. The preliminary experiments based on video clips from 50 video sequences have shown that results are close to the human perception and segmentation of a crowd. The method has demonstrated enough robustness to handle different crowded scenes and density of crowds. For future work, the experiments based on the more discriminative salient points for crowd detection and segmentation will be performed. The idea is to aggregate results of crowd segmentation (group, direction and displacement) of successive video clips to obtain the elements of a meta-trajectory suitable for crowd behaviour analysis.

ACKNOWLEDGMENTS

This work has been supported by the Croatian Science Foundation under projects IP-2018-01-7619 "A Knowledge-based Approach to Crowd Analysis in Video Surveillance" (KACAVIS) and IP-2016-06-8345 "Automatic Recognition of Actions and Activities in Multimedia Content from the Sports Domain" (RAASS).

REFERENCES

[1] T. Li, H. Chang, M. Wang, B. Ni, R. Hong and S. Yan, "Crowded Scene Analysis: A Survey," IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 3, 2015, pp. 367–386.

[2] J. Cezar S. J. Junior, S. R. Musse, C. R. Jung, "Crowd Analysis Using Computer Vision Techniques," 2010, pp. 66–77.

[3] B. Solmaz, B. E. Moore, M. Shah, "Identifying Behaviors in Crowd Scenes Using Stability Analysis for Dynamical Systems", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 10, 2012, pp. 2064–2070.

[4] S. Ali and M. Shah, "A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis," IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, USA, 17-22 June 2007, IEEE, 2007, pp. 1–6.

[5] R. Sharma and T. Guha, "A trajectory clustering approach to crowd flow segmentation in videos," IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25-28 Sept. 2016, IEEE, 2016, pp. 1200–1204.

[6] S. Wu and H.S. Wong, "Crowd motion partitioning in a scattered motion field," IEEE Trans. Systems, Man, and Cybernetics, Part B, vol. 42, no. 5, pp. 1443–1454.

[7] S. Biswas, R.G. Praveen, and R.V. Babu, "Super-pixel based crowd flow segmentation in H.264 compressed videos," IEEE International Conference on Image Processing (ICIP), Paris, France, 27-30 Oct. 2014, pp. 2319–2323.

[8] S. Pellegrini, A. Ess, K. Schindler, and L. Van Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 2009, pp. 261–268.

[9] W. Li, J-H. Ruan, and H-A. Zhao, "Crowd movement segmentation using velocity field histogram curve," Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition, Xian, China, 15-17 July 2012, pp. 191–195.

[10] S.S. S. Kruthiventi and R.V. Babu, "Crowd flow segmentation in compressed domain using CRF," IEEE International Conference on Image Processing (ICIP), Quebec City, Canada, 27-30 Sept. 2015, pp. 3417–3421.

[11] Z. Wang, c. Cheng and X. Wang, "A Fast Crowd Segmentation Method," IEEE International Conference on Audio, Language and Image Processing (ICALIP), Shanghai, China, 16-17 July, 2018, pp. 242–245.

[12] H. Fradi, B. Luvison, Q. C. Pham, "Crowd Behavior Analysis Using Local Mid-Level Visual Descriptors," IEEE Transactions on Circuits and Systems for Video Technology, vol. 27, no. 3, 2017, pp. 589–602.

[13] J. Cezar S. J. Junior, A. Braun, J. Soldera, S. R. Musse and C. R. Jung, "Understanding people motion in video sequences using Voronoi diagrams: Detecting and classifying groups," Pattern Analysis and Applications, vol. 10, no. 4, 2007, pp. 321–332.

[14] S. Theodoridis, K. Koutroumbas, Pattern Recognition, Sea Harbor Drive Orlando, FL, Elsevier, 2008.