

Survey on Holistic approach Representation Methods of Crowd Analysis

Jose Mary A¹P.G. Student, Department of Computer Science and Engineering, College of Engineering and Management, Punnappra,
Kerala, India¹

ABSTRACT: Dynamic crowd flow analysis has been a great interested area in computer vision. The abnormal behaviour detection is one of its most important research area because of the increasing rate of peoples. However, the analysis of crowd scene is a challenging task because of its density and high occlusions in some areas. There are different kinds of segmentation, object, and particles based methods for the analysis of a dense crowd, for this analysis holistic approaches uses exact feature extraction methods, which are computationally affordable and it is more suitable for the analysis purposes. There are different representation schemes are used to represent the extracted features. This paper presents a survey on holistic crowd activity representation methods, also considered different data sets available for the analysis and detection purposes

KEYWORDS: Crowd analysis, Holistic approach, Features, Abnormal Behaviour, Representation Methods, Spatio temporal.

I. INTRODUCTION

Video based crowd flow analysis is most complicated and challenging task in computer vision. Crowd analysis purposes consists of object based approach and a holistic approach. The dense crowd make it impractical for the object identification and individual tracking. Feature generation methods are one of the solution for analysing dense crowd. There are mainly particle based approaches such as optical flow, force flow etc. There are wide range of applications in crowd flow analysis. Fig 1 represents the taxonomy of crowd analysis.

Automatic crowd analysis application includes density estimation or counting of individuals in the crowd scene, here these models are divided into three level based methods. The pixel level analysis methods have low level features for counting the peoples in the scene, in other cases uses background subtraction techniques and edge detection techniques for the estimation of crowd density [1]. The method [2] proposed by Regazzoni, which extract features based on some edge detection algorithms and find the vertical edges. Training phases have some dynamic programming and this model is used to identify the number of peoples presented in the monitored scene. A neural network based method to estimate the crowd density in substations was proposed by Cho and his collaborators [3]. This have approximately 90% of correctness overall tested methods to detect high density crowd. Multicamera and segment based approach was introduced by Yang et al. [4], in this work the segmentation of foregrounded objects was done by group of sensors.

The texture analysis consists of higher level features as compared to pixel level analysis. It estimates the number of peoples, instead, of identifying the individuals. The work [5] by Wu and colleagues is make a perspective projection model, which generate a series of multi resolution cells. These texture information from the cells are extracted using a gray-level dependence matrix method. The trained SVM taken this input and find the velocity to dense crowd. A comparison of this matrix and translation invariant orthonormal Chebyshev moments are made by Rahmalan in his work [6]. He finds out that translation invariant orthonormal Chebyshev moments provides overall best results.

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 4, April 2018

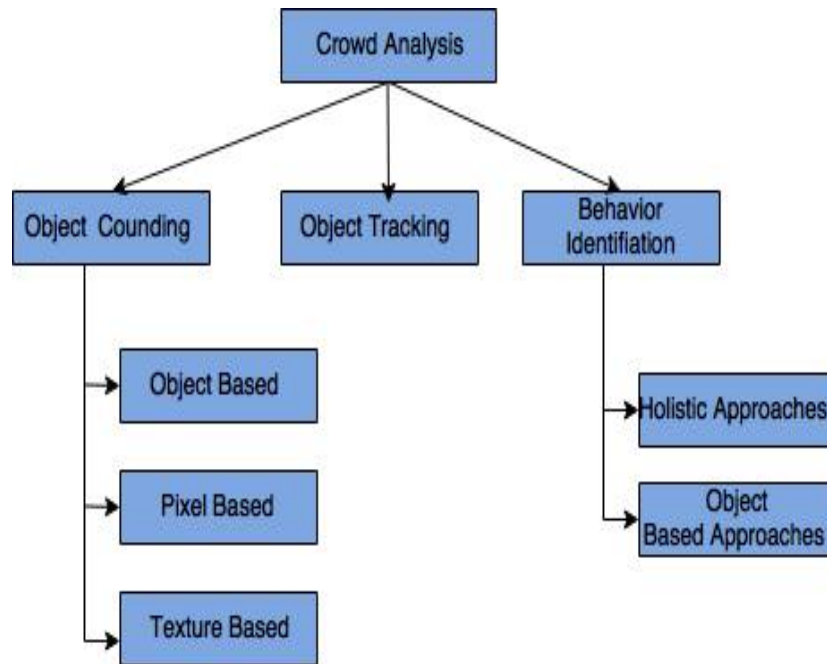


Fig 1: Taxonomy of crowd analysis

A segment based texture analysis method is developed by Chan et al. [7], consists of a mixture of dynamic textures. Segment the crowd into variable motion directions; then, extract some segment features like area, orientation, edge features, texture features etc... . Finally Gaussian process is used to identify the number of people in each segment.

The object tracking methods are used to identify the position of person in different sense. Ali and Shah [12] introduce an approach for people tracking. It uses a particle oriented approach in which a person consists of different particle of cells. The unstructured environment can also be used for people tracking [13]. The crowd behaviour identification models specifies temporal information which contains directions [14], velocity [1] and unusual motions [15]. Models can be classified as object based approaches and holistic approaches. An object based approaches the behaviour of a group is identified by the segmentation or direction of individuals. These are most suitable only for low or moderately crowded scenes. [16] Specifies an individual tracking, but it is difficult about a dense scene. The work [17] uses an unsupervised learning frameworks, it builds a hierarchical Bayesian model to connect low level virtual features, activities, and interactions.

A different concept is used in holistic approach, which consider the crowd as a single entity. The work [1] Propose a pixel level motion features extraction and Andrade et al. [18] identifies the behaviour of a crowd scene based on optical flow using Hidden Markov Model. The holistic approach can also use the segmentation methods on crowd flow [15]. The Lagrangian particle dynamics are used to find the instabilities of crowd. Merhan and colleagues [19] develop a social force model. This explore the abnormality in row flow by using social force in the particle over the image. Moreover, the abnormal video event can be classified into two classes.

1. Local Event
2. Global Event

The local event implies an abnormality of individual from a group of people. But global event means that the group behaves abnormally. Consider the example in fig 2 a) Specifies a local event in which one object shows an abnormality form the normal flow. b) In global event group of objects shows the abnormal behaviour.

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 4, April 2018

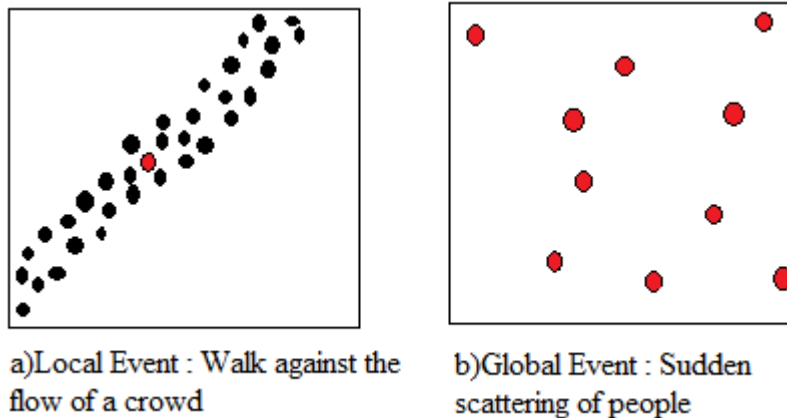


Fig 2. Local and global abnormal crowd event

II. CROWD ACTIVITY REPRESENTATION METHODS

The motion representation is the basis of crowd scene analysis. There are different kinds of motion features which includes flow based features, local spatiotemporal features and trajectory. These features can be used for various crowd behaviour identification applications. To use this features, it is needed to represent in any of the following representing methods

a) Multiresolution histogram

The frame scenes at various levels of scale can be used for the framework in multiresolution histogram generation and which is built with pyramid algorithms [20]. This will reduce the complexity in image processing operations. The image segmentation is based on the distribution of intensity levels and selection of threshold from a histogram of an image. Multiresolution histogram represents several levels of resolution from lower level to the finer levels [21]. The main problem with crowd analysis is to identify the activity and understanding of scene. These activities can be recognized by a spatiotemporal texture. Lu et al. [22] propose a method that recognize the specified dynamic textures by a multiresolution histogram representation of velocity and acceleration field. But in [23] Chetverikov and Peteri uses the normal flow features like divergence, orientation, periodicity for representation. [24] Also specifies a histogram based feature analysis method to identify the unusual activities in large video set. The image feature used by this method is a spatial histogram of the detected objects, so these are sufficient to detect complex activities. The work by Dmitry Chetverikov [25] introduces a combined approach of multiresolution histogram with optical flow. Spatial image information like shape and textures an also be represented using multiresolution histograms.

b) Spatiotemporal cuboids

The spatiotemporal cuboids are most commonly used representation for crowd flow analysis. For each video sequence, the spatiotemporal cuboids are extracted randomly. One approach proposed by Bo Wang et al. [26] uses size adapted spatio temporal features for identify the abnormal crowd behaviour. The advantage of using this method is that size and number of cuboids for the representation related automatically. The research [27] Kratz and Nishino propose a method for abnormality detection by a spatiotemporal cuboids and [28] consists of a 3-D multivariate Gaussian distribution of the spatiotemporal gradients for crowd behaviour detection. A variation of normal cuboids, which represent the aggressive behaviour of the people in the crowd is recognized using a spatiotemporal elastic cuboid trajectories. The crowd unusual event analysis and novel framework for motion patterns of crowd scene using a cuboid method was implemented by Louis Kratz [29]. These representation models are computationally efficient.

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 4, April 2018

c) Motion Descriptors

The motion descriptors are represented the motion features in the video frames. The motion is related to both motion object and camera. There are different kinds of motion descriptors based on the way of motion feature selection. The motion trajectories are extracted via object tracking and create a signature description of motion trajectories [30]. The collective motion patterns are represented using a curl and Divergence of motion trajectories (CDT) descriptors [31]. These are helps to identify the lane, clockwise arch, bottleneck crowd behaviours. The camera motion descriptors [32] represents the global motion of camera in a normal 3D scene. Another one, the object motion trajectory descriptor describes the movement of objects in 3D or 2D plane. The work [33] uses a feature descriptor called Lagrangian Scale Invariant Feature Transform Descriptor (LaSIFT), which specifies the direction of motion features.

d) Spatiotemporal Volumes

The spatiotemporal volume can be used to identify the static crowd behaviour. The local spatio temporal motion patterns of the video sequence are captured and its variations can be represented as video volume [29]. These variations in the neighbourhood events to the detected events of a space time corners of image data can also be represented using a spatiotemporal volume [34]. The approach [35] uses a shape based event detection and it is extended to operate with over segmented spatiotemporal volumes. This volume is more useful for the textures of optical flow based anomaly detection. The initial video 3D volumes or spatiotemporal patches is used to extract features such as motion textures, spatial information's.

These feature based representations are used with different approaches of crowd flow analysis works. Selection of features determines which representation models are more useful for the flow analysis process such as people counting, tracking, and behaviour identification. Following table 1 summarizes the various representation models

| Representation Methods | Crowd flow analysis models | Efficiency of method |
|---------------------------|--|---|
| Multiresolution Histogram | Motion segmentation and activity representation[Spatiotemporal Textures, Dynamic Textures] | Computational efficiency Robust to noise Spatial efficiency |
| Spatiotemporal Cuboids | Spatiotemporal features[Motion features] | Computationally efficient Easier to represent |
| Motion Descriptors | Trajectory based approaches | Independent to image postures and it is scalable |
| Spatiotemporal Volumes | Spatiotemporal motion patterns[Shapes],Textures of optical flow | Better for segmentation methods and provides temporal continuity |

Table 1: Summary of representation methods and corresponding analysis models with Its efficiency

III. CONCLUSION

This paper is present a survey of crowd activity representation methods. The analysis and behaviour identification is based on some features of crowd. The holistic approach of crowd analysis consists of different motion features. This features are needed to be represented in an efficient way. There are multiple feature representation models. This survey shows that which crowd analysis models can use appropriate representation of features. The efficiency of each approves are also specified in this survey.

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 4, April 2018

REFERENCES

- [1] A. C. Davies, J. H. Yin, and S. A. Velastin, "Crowd monitoring using image processing," *IEEE Electron. Commun. Eng. J.*, vol. 7, no. 1, pp. 37–47, 1995.
- [2] C. S. Regazzoni, A. Tessei, and V. Murino, "A real-time vision system for crowding monitoring," in *Proc. Int. Conf. Industrial Electronics, Control, and Instrumentation*, 1993, vol. 3, pp. 1860–1864.
- [3] S.-Y. Cho, T. W. S. Chow, and C.-T. Leung, "A neural-based crowd estimation by hybrid global learning algorithm," *IEEE Trans. Syst., Man, Cybern.*, 1999, vol. 29, no. 4, pp. 535–541.
- [4] D. B. Yang, N. Héctor, H. González-Ba, and L. J. Guibas, "Counting people in crowds with a real-time network of simple image sensors," in *Proc. IEEE Int. Conf. Computer Vision*, Washington, DC, USA, 2003, p. 122.
- [5] X. Wu, G. Liang, K. K. Lee, and Y. Xu, "Crowd density estimation using texture analysis and learning," in *Proc. IEEE Int. Conf. Robotics and Biomimetics*, 2006, pp. 214–219.
- [6] H. Rahmalan, M. Nixon, and J. Carter, "On crowd density estimation for surveillance," in *Proc. Institution of Engineering and Technology Conf. Crime and Security*, 2006, pp. 540–545.
- [7] A. Chan, Z. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008, pp. 1–7.
- [8] E. Leibe, B. Seemann, and B. Schiele, "Pedestrian detection in crowded scenes," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Washington, DC, 2005, pp. 878–885.
- [9] M. J. Jones and D. Snow, "Pedestrian detection using boosted features over many frames," in *Proc. Int. Conf. Pattern Recognition*, 2008, pp. 1–4.
- [10] V. Rabaud and S. Belongie, "Counting crowded moving objects," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2006, pp. 705–711.
- [11] G. J. Brostow and R. Cipolla, "Unsupervised Bayesian detection of independent motion in crowds," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Washington, DC, 2006, pp. 594–601.
- [12] S. Ali and M. Shah, "Floor fields for tracking in high density crowd scenes," in *Proc. European Conf. Computer Vision*, 2008, pp. II:1–14.
- [13] M. Rodriguez, S. Ali, and T. Kanade, "Tracking in unstructured crowded scenes," in *Proc. IEEE Int. Conf. Computer Vision*, Kyoto, Japan, 2009, pp. 1389–1396.
- [14] A. M. Cheriyyadath and R. Radke, "Detecting dominant motions in dense crowds," *IEEE J. Select. Topics Signal Process.*, vol. 2, no. 4, pp. 568–581, Aug. 2008.
- [15] S. Ali and M. Shah, "A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2007, pp. 1–6.
- [16] J. C. S. Jacques, Jr., A. Braun, J. Soldera, S. R. Musse, and C. R. Jung, "Understanding people motion in video sequences using voronoi diagrams," *Pattern Anal. Applicat.*, vol. 10, no. 4, pp. 321–332, 2007.
- [17] X. Wang, X. Ma, and W. E. L. Grimson, "Unsupervised activity perception in crowded and complicated scenes using hierarchical Bayesian models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 31, no. 3, pp. 539–555, 2009.
- [18] E. L. Andrade, S. Blunsden, and R. B. Fisher, "Modelling crowd scenes for event detection," in *Proc. Int. Conf. Pattern Recognition*, Washington, DC, 2006, pp. 175–178.
- [19] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2009, pp. 935–942.
- [20] G. Bongiovanni, L. Cinque, S. Levialdi, and A. Rosenfeld, "Image segmentation by a multiresolution approach," *Pattern recognition* vol 26, No 12, pp. 1845–1854, 1993.
- [21] Z. Lu, W. Xie, and J. Huang, "Dynamic texture recognition by spatiotemporal multiresolution histogram," *IEEE Workshop on Motion and Video Computing*, 2005.
- [22] D. Chetverikov and R. Peteri, "A brief survey of dynamic texture description and recognition," in *4th Int Conf Comput Recognit Syst*, 2005.
- [23] Hua Zhong, Jianbo Shi, Mirko Visontai, "Detecting Unusual Activity in Video", Departmental Papers (CIS), Department of Computer & Information Science, 2004.
- [24] Dmitry Chetverikov and Renaud Peteri "A brief survey of dynamic texture description and recognition", 2005
- [25] E. Hadjemetriou, M. D. Grossberg and S. K. Nayar, "Spatial information in multiresolution Histograms", computer science, Columbia university, New York, N Y 10027.
- [26] Bo Wang, Mao Ye, Xue Li, Fenguan Zhao, "Abnormal crowd behavior detection using size adapted spatiotemporal features", *International journal of control, automation and systems*, 2011.
- [27] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2009, pp. 1446–1453.
- [28] Ismael Serrano, Oscar Deniz, Gloria Bueno, Guillermo Garcia-Hernando, Tae-Kyun Kim, "Spatio-temporal elastic cuboid trajectories for efficient fight recognition using Hough forests", *Machine Vision and Applications*, 2018, Volume 29, Issue 2, pp 207–217.
- [29] Louis Kratz, Ko Nishino, "Anomaly Detection in Extremely Crowded Scenes Using Spatio-Temporal Motion Pattern Models", Department of Computer Science, Drexel University.
- [30] Shandong Wu, Y. F. Li, Jianwei Zhang, "Motion descriptor: A motion trajectory signature", *Information and Automation*, 2009. ICIA '09, June 2009.
- [31] Shuang Wu, Hua Yang, Shibao Zheng, Hang Su, Yawen Fan, Ming-Hsuan Yang, "Crowd Behavior Analysis via Curl and Divergence of Motion Trajectories", *International journal of computer vision*, 2017.
- [32] Sylvie Jeannin, Radu Jasinschi, Alfred She, Thumpudi Naveen, Benoit Mory, Ali Tabatabai, "Motion descriptors for content-based video representation", *Signal Processing: Image Communication*, September 2000.
- [33] Tobias Senst, Volker Eiselein, Alexander Kuhn, and Thomas Sikora, "Crowd Violence Detection Using Global Motion-Compensated Lagrangian Features and Scale-Sensitive Video-Level Representation", 2017 IEEE.
- [34] I. Laptev, "On space-time interest points," *Int. J. Comput. Vision*, vol. 64, no. 2-3, pp. 107–123, 2005.
- [35] Ramin Mehran, Brian E. Moore, Mubarak Shah, "A Streakline Representation of Flow in Crowded Scenes", Computer Vision Lab, Department of Mathematics, University of Central Florida.