Machine learning methods from group to crowd behaviour analysis

Luis Felipe Borja-Borja[1], Marcelo Saval-Calvo[2], and Jorge Azorin-Lopez[2]

[1] Universidad Central del Ecuador,
Ciudadela Universitaria Av. América, Quito, Ecuador
[2] Computer Technology Department, University of Alicante,
Carretera San Vicente s/n, 03690, San Vicente del Raspeig (Spain)

Abstract. The human behaviour analysis has been a subject of study in various fields of science (e.g. sociology, psychology, computer science). Specifically, the automated understanding of the behaviour of both individuals and groups remains a very challenging problem from the sensor systems to artificial intelligence techniques. Being aware of the extent of the topic, the objective of this paper is to review the state of the art focusing on machine learning techniques and computer vision as sensor system to the artificial intelligence techniques. Moreover, a lack of review comparing the level of abstraction in terms of activities duration is found in the literature. In this paper, a review of the methods and techniques based on machine learning to classify group behaviour in sequence of images is presented. The review take into account the different levels of understanding and the number of people in the group.

Keywords: Human Behavior Analysis, Motion analysis, Trajectory Analysis, Machine Learning, Crowd Automated Analysis, Computer Vision.

1 Introduction

Nowadays, video surveillance of people is a widely used tool because there are many cameras that facilitate the capture and storage of video. Most of these products are dependant on an operator to analyze the content of stored information. Knowing this limitation it is necessary to provide systems of video surveillance that make possible the automatic identification of behavior. These types of system can be carried out using computer vision techniques, since they allow the identification of patterns of people behavior in an unsupervise manner as gestures, movements and activities among others. In general terms, machine learning, it is possible to model the behavior of people in open or closed spaces such as universities, shopping malls, parks or streets, and then analyze them using automatic learning methods.

There are currently many researches on Human Behavior Analysis such as, [1] that have resulted in the identification of various types of people's behavior in video sequences. These behaviors have been classified from the simplest to the most complex taking into account their duration, from seconds to hours. For these behaviors a classification has been proposed in [2].

The objetive of this paper is to provide a classification of human behavior analysis proposals taking into account the size of the group or crowd, identifying the number of people that comprises it, the type of behavior detected, the level of abstraction(from simple actions to complex behaviors) and the techniques used for its treatment and analysis. The most important public datasets are also reviewed which are used to test algorithms there exist several studies on the identification of human behaviors such as [2], [3], [4], [5]. In [6] a taxonomy of groups with fewer and more members is established, in addition the methods to analyze them are specified. There are works such as [7], where it is proposed to analyze the behavior of crowds by classifying them into two levels, macro and micro. Despite research efforts to analyze behavior in groups and crowds, we still have many fronts on this subject for researchers.

According to the above the objectives of this paper are: to propose a classification of group and crowd behaviour analysis proposals according to the number of members and the level of abstraction regarding the duration of behavior detected.

2 Aspects of Human Behavior Analysis

In this section the main aspects of the human behavior analysis are explained. First we will present the different levels of understanding and later the main datasets available for experimentation.

2.1 Description of human behavior types and semantics (gesture, motions, activities, behavior)

In order to identify human behavior according to the level of abstraction and understanding the data has to be classified depending on the meaning, duration and complexity of tasks performed by humans.

Classifications of activities has taking as its main reference the level of complexity of them, from the easiest to the most complex. The complexity factor is directly related to the time duration of the activity, generally, an activity is considered complex if it has a longer duration. In [8] four levels related to their semantics:

- **Level 1 (Gestures):** Basic movements of parts of the body that last a time. Examples of gestures can be movements of the hand, arm, foot or head among others.
- **Level 2 (Actions):** Also called atomic, consists of actions performed by a single person, their duration is larger than a gesture. An example of actions could be walking, runing, jumping.
- **Level 3 (Interaction):** In this category human-human or human-object interaction activities are performed. Examples of these interactions can be two people dancing, kissing, running one behind another, children playing, people cycling.

– **Level 4 (Group Activity):** At this level of description it conforms to two or more groups of people, one or more objects can intervene in the scene. An athletic race, basketball team forwarding, pedestrians crossing a street, a football game, a fight in a stadium can be examples of group activities.

Another taxonomy of human behavior that classifies it according to the complexity and duration time is proposed in [2]. In this approach, the analysis is classified on the degree of semantics in four levels:

– **Level 1 (Motion):** Detection in seconds or frames.
– **Level 2 (Action):** Detection of simple tasks in terms of seconds. The human can interact with objects, or be sitting, standing, walking.
– **Level 3 (Activity):** These are tasks from of minutes to hours. They constitute the sequence of actions, such as cleaning a room, washing a vehicle.
– **Level 4 (Behavior):** This is the higher level of understanding since its duration time can be hours and days. Example behavior can be daily routines of a person, personal habits, mix of two activities in logical sequence.

Both taxonomies described above are based on the daily activities of people, taking into account important factors such as the level of semantics, the duration and the activities composed of other simpler parts such as movements and actions. They described the levels/orders of behavior from the simple movements lasting seconds, to complex activities performed by people for several minutes, hours and even days. The aim of the researchers has been to propose a general classification human behaviour. There are other classification, however, in this work we are going to base our proposal on these focused on group and crowd behavior classification

2.2 Specialized Datasets

In [9] Blusden and Fisher presented a set of datasets which include sequences for individual and group behavior which are part of the BEHAVE project and include some form of ground truth. Since this paper is focused on group and crowd analysis, the individual datasets are not studied, but authors refer to the original paper for further details.

In group analysis, there are three datasets belonging to BEHAVE project: CAVIAR, CVBASE, ETISEO. Examples of behavior detected in these datasets are: InGroup(The people are in a group and not moving very much), Approach(Two people or groups with one (or both) approaching the other), WalkTogether(People walking together), Meet(Two or more people meeting one another), Split(Two or more people splitting from one another), Ignore(Ignoring of one another), Chase(One group chasing another), Fight(Two or more groups fighting), RunTogether(The group is running together), Following(Being followed).

In this paper we analyze the behavior of groups and crowds such as pedestrians, crowds in public places such as stadiums or squares, interactions of large and small groups, sport actions such as soccer and basketball, and others. The datasets used by the researchers are numerous, being the main ones the following: BEHAVE, BIWI, VSPETS, ETH, DGPI, UHD, HMDB, SportsVU, PETS, UNM, ViF, Bus STATIONS, Subway STATIONS, others. Also in some cases the researchers use their own datasets or videos obtained on YouTube. In [10] it is proposed a study and dataset classification taking into account the behaviors, number of people involved, techniques used to recognize behaviors, types of scene, year of publication, among other characteristics. From this study, an absense of RGB-D (Color and depth) datasets is shown.

With the objective of studying human behavior, in the last years several public datasets have been created. In these dataset, video sequences with contents of several activities in different scenarios and situations are stored. There are also sites dedicated to study particular activities such as a movement or action of a sport, identification of abandoned objects, or daily activities (ADL) such as having a cup of coffee, detection of falls of human, gait study, gesture analysis.

These studies are directly related to public datasets, where tests of the algorithms and techniques used in each case are performed, in certain studies more than one dataset is used to check the accuracy of the recognition systems developed, in other cases it is used custom datasets or the researcher's own, video sequences obtained in public places like bus stations or trains, also of people who carry out activities in squares, streets and commercial centers of a city, are also used. There are very few studies that use YouTube as a source for video footage for research.

Video analysis to perform such a study requires effort and time for researchers, thousands of man-hours are used for the labeling of the different situations that need to be identified in a video. Currently, in cities, it is common to find camcorders capturing video that are later stored. However, all this large amount of information is not available for public access and experimentation.

3 Classification of the level of understanding of groups

To analyze human behavior by using video surveillance cameras, a system based on computer vision requires following a series of ordered steps as suggested in [11]. This paper aims to organize a classification of human behavior according to the number of people that make up a group or crowd, and the techniques, algorithms or frameworks used for analysis.

Human behaviour analysis (HBA) investigations have different applications: improving the quality of life of human beings, in aspects such as support in the health area to detect unusual behaviors, for example falls of elderly people in assisted living environments (AAL) [12],[11], [3]; surveillance of pedestrians, fights, people running, assaults, ingesting liquor in public places, for example.

The classification of tasks performed by humans described in the previous section are analyzed in [3] according to the level of semantics (in ascending order according to the duration time of this is): Movement (seconds), actions (seconds, minutes), activity (minutes, hours), behavior (hours, days). Each of these tasks must be recognized and modeled, using different techniques, algorithms and other tools suitable for this task.

Turaga et al. [4] proposed a scale of recognition of human activities from simple (actions) to complex (activities), for actions called simple uses (Non-Parametric, Volumetric, and Parametric), for activities called complex uses (Graphical Models, Syntactic , Knowledge Based). Another organization proposal for recognition of activities is set out in [11], where it proposes the Chain of Activity Recognition. This approach divides the recognition process into different procedures, which are: Data Acquisition, Preprocessing, Segmentation, Characteristic extraction, Classification, Decision. Most current research focuses on the last two procedures of this proposal and is often referred to as the learning and decision phases.

In the studies about human behavior of groups and crowds analyzed, it was found that there are few works dealing with RGBD cameras and analysis of human behavior using 3D information. It is important to highlight the work of Wu et al. [13]. They proposed the MoSIF method is combined with HMM [13] to analyze video sequences obtained from a Microsoft Kinect RGBD device. The accuracy obtained is 60% for 3600 video sequences. However, according to the authors, a better result could be obtained if they used more videos to improve learning.

The methods of classification can be supervised and not supervised, and can be used individually or combined using boosting techniques.

On the subject of behavior and trajectories of groups of people there are also some approaches that are based on (HBA) study individually, for example: to recognize activities of groups of people we use the Group Activity Descriptor Vector (GADV) Proposed in [14]. This method has as its predecessor the Activity Description Vector (AVD) revised in [14], [15], and aims to recognize human behavior in advance.

3.1 Features of a groups and crowds

For example, Andrade et al. [16] detected behaviour of a crowd in different scenarios considered unusual or an emergency, usually provoked by a minority of people in the crowd. These behaviors are coded in Hidden Markov Models (HMM) with mixture of Gaussians output(MOGHMMs), detecting within the different scenes according to their density of people that conform it. It should be considered that the system must be previously trained to detect a type of behavior considered normal that usually have the majority of members of a crowd analyzed. Analyzing specifically the modeling of dense crowds is still an open problem of researchers.

In a public space, where there are a lot of people, the behaviour could be analysed by two variables: actions and duration. Its behavior and its duration. A general trend could be noticed and described as the actions considered normal ones have an extended duration,... a general trend that would be described as that the behaviors considered normal ones have an extended duration, in which most people make up the crowd, while the behaviors considered abnormal are caused by few people in the crowd and in short times of duration. For the study of these types of behaviors, Hu et al.[17] proposed to use a statistical exploration method analyzing the video in a separate way as sliding windows in which the behaviors considered anomalous are detected, taking into account that the algorithm used in this technique requires monitoring.

As we have previously described in order to understand the behavior of crowd, we must take into account the social behavior of the masses, since in this one can observe patterns of behavior that can be modeled by computer studying their structure and special characteristics as proposed in [18]. This study analyzes the human activity of medium level in the granulity, that is to say in the number of people that conform it based on algorithms for the detection of pedestrians and tracking of several moving objects. A particular fact is that the study considers small groups of people traveling together considering the hierarchy of smaller to larger size of the group. It takes into account the proximity of pairs of people and their speed when walking in a particular scene. According to [18], a group is formed from two people, in addition it must feet other parameters such as: if they are within 2,13 meters of each other and not separated by another in the middle, have the same speed up to within 0,15 meters per second, and is traveling in the same direction within 3 degrees. When a member of the group stops fulfilling these characteristics or complies with them, it can be said that he or she is inside or outside the group.

The datasets can be chosen by the researchers according to their criteria, taking into account the suitability of their objective. The data are grouped into two categories the heterogeneous, referring to the general activities and the specific when these actions have a special treatment. A third category is included in [10], which specify techniques for motion capture such as the use of infrared, thermal and motion capture (MOCAP).

3.2 Behavior of groups and crowds

This paper shows in Table 1 and Table 2 a classification of the group size according to the number of members and the activities that each type of group performs, besides specifying the methods, algorithms and forms of recognition that can be used for their study. We can see the following analyzed fields: Ref=Reference to the article, CL=Classification (number of people if exist), TE = Technique, D = Dataset, LA = Level Abstraction. In the column LA = Level Abstraction we show three levels of abstraction: Mot = Motion, Act = Action, Actv = Activity, also two automatic tasks, CP = Count-People and Tra=Tracking.

The classification according to the number of people is in two main sets GROUP and CROWD. Group is defined as the compound of two or more people in a given site and performing an action or activity. Crowd is a composition of people larger that a group that performs simultaneous activities.

The types of behaviors analyzed using video surveillance are limited and specific. The most frequently studied behaviors are the following: Tracking, trajectories, bicyclist, pedestrian, skateboarders, count people in a group or crowd,

Table 1. Classification of proposals reviewed

AR	CL	TECHNIQUE	D	LA
[15]	G	Self-Organizing Map (SOM) Supervised Self-Organizing Map (SSOM) Neural GAS (NGAS) Linear Discriminant Analysis (LDA) k-Nearest Neighbour (kNN) Multiclassifier (MC)	CAVIAR	Actv
[19]	G	Convolutional Neural Net- works (CNN) Long Short-Term Memory (LSTM)	UAV	Actv
[20]	G	Multiple Object Tracking Accuracy (MOTA) K-Shortest Pats Optimization(KSP) Markov Decision Process(MDP) Recurrent Neural Networks(RNN)	TOWN ETH HOTEL STATION	
[21]	C	Collective Transition priors (CT) Mixture of dynamic texture (DTM) Hierarchical clustering (HC) Coherent filtering (CF)	CUHK	Mot
[22]	C	Pedestrian Simulation(PS) Person re-identificatio(PT) Pedestrian tracking(MPF)	NY Station Shanghai-Expo	Mot
[23]	C	Motion Pattern Features(MDA)	N	Mot
[24]	G	Stability Features(HDP)	BEHAVE	Actv
[25]	G	Hidden Markov Models(HMM) Dynamic Probabilistic Networks(DPN)	HMDB BEHAVE	Mot
[26]	G(50)	Inter-Relation Pattern Matrix(IRPM) Game-Theoric Conversational Groups(GTCG) Spectral Clustering (R-GTCG SC)	DGPI	Actv
[27]	C	Model Dynamic Textures Temporal(MDT-temp) Local Motion Histogram(LMH) Spatail(MDT-spat)	UNM UCSD	Mot
[28]	G(25)	Markov Chain Monte Carlo (MCMC) Gaussian Mixture Model (GMM)	FIFA WC 2006	Tra
[29]	G	Category Feature Vectors (CFVs) Gaussian Mixture Models(GMM) Recognizing algorithm (CFR)	N	Actv
[30]	G	Convolutional Neural Network (CNN) Feed Forward Network (FFN)	SportsVU	Actv
[31]	G	Multiple Human Tracking (MHT) Correct Detected Tracks (CDT) False Alarm Tracks (FAT) Track Detection Failure (TDF)	ETH UHD	Tra
[32]	G	Neural Network(NN)	N	Act
[33]	C	Histogram of Oriented Gradients (HOG) Histogram of Optical Flow (HOF) Motion Boundary Histogram (MBH)	UMN UCSD CUHK PETS2009 ViF Rodriguezs UCF Own Dataset	Act
[34]	C	Hidden Markov Models (HMM) Support Vector Machine (SVM) Robust Local Optical Flow (RLOF)	PETS UMN	Act

Table 2. Classification of proposals reviewed

AR	CL	TECHNIQUE	D	LA
[35]	G	Cumulative Match Characteristic (CMC) Synthetic Disambiguation Rate (SDR) Center Rectangular Ring Ratio-Occurrence(CRRO) Block based Ratio-Occurrence (BRO)	2008 i-LIDS MCTS	Act
[36]	C	Accumulated Mosaic Image Difference (AMID) OpticalFlow+BackgroundModel (OFBM) Markov Random Fields (MRF) Support Vector Machine (SVM)	Subway Station Bus Station Plaza	Mot
[37]	C	Support Vector Machine (SVM) Library for Support Vector Machines(LIBSVM) Basis Radial Function(BRF) Block Matching Algorithm (BMA)	UMN	Act
[38]	C	Fast Corner Detect(FAST) Support Vector Machine (SVM)	BEHAVE	Act
[39]	G	Evolving Networks(EN) Monte Carlo(MC)	N	Mot
[40]	G	Linear Trajectory Avoidance (LTA)	N	Mot
[7]	G(20)	Bag of words modelling (BoW)	Novel Dataset	Mot
[41]	G	Gaussian Mixture Model(GMM) EM algorithm	N	Actv
[42]	G	Minimum Description Length (MDL)	COLLECTIVE ACTIVITY BEHAVE	Actv
[43]	G	Hidden Markov Models(HMM) Dynamic Bayes Networks(DBN)	BIWI	Tra
[44]	G(20)	Multi-model MHT	Own	Tra
[45]	G	Voronoi Diagrams Model(VDM)	N	Mot
[5]	G	Dynamic Probabilistic Networks (DPNs) Dynamically Multi-Linked (DML) Hidden Markov Model(HMM)	PETS 2004 YouTube	Mot
[46]	G(25)	Support Vector Machines (SVM)		Act
[47]	C	Hidden Markov Model (HMM)	N	Mot
[43]	G	Sampling Importance Resampling (SIR) Discrete Choice Model (DCM) Multi Hypothesis Tracking (MHT) Statistical Shape Modeling(SSM)	BIWI	Tra
[48]	G(90)	Heuristic learned(HL)	N	CP
[49]	C	Bag of Words (BoW) Locality-constrained Linear Coding (LLC) Vector Quantization (VQ)	BEHAVE	Mot
[50]	C	Unsupervised Bayesian Clustering Framework(UBCF)	N	Mot
[51]	C	Bayesian Marked Point Process (MPP)	CAVIAR VSPETS SOCCER	CP
[52]	C	Social Force Model(SFM) Pure Optical Flow(POF)	UNM	
[53]	C	Detection of moving regions	METRO	Tra
[54]	C	Linear Fitting(LF) Unpervised Neural Network(UNN)	N	CP

street fights, interaction objects-people, motions or actions in sports, human actions (walking, jogging, running, boxing, hand waving and hand clapping).

The datset frequently used for experimentation to analyze the behavior of groups and crowds are the following: BEHAVE, BIWI, CAVIAR, VSPETS, ETH, DGPI, UHD, HMDB, SportsVU, PETS, UNM, ViF, Bus STATIONS, Subway STATIONS, others. Also in some cases the researchers use their own dataset or videos obtained on YouTube.

Based on the information analyzed in the papers, it is possible to propose a classification according to the level of abstraction of the analyzed human behavior of groups and crowds according to the case, in order of shortest to longest duration of behavior we propose three levels of abstraction: Motion, Action, Activity, also two automatic tasks, Count-People and Tracking.

The techniques or methods frequently used to analyze human behavior of groups and crowds using video surveillance are the following: Bag of Words, Deep Neural Networks, Hidden Markov Models, Monte Carlo, Gaussian Mixture Model, Multiple Human Tracking, Support Vector Machines. Many authors use these methods including specific tunings and mask the name with a slight modification.

4 Conclusions and future directions

In this work the human behavior of groups and crowds has been approached taking into account the degree of semantics and especially the size of people that integrate the group or crowd, in addition has been considered behaviors like; Sports teams of soccer and basketball, pedestrians, groups of people in metro and bus stations, people grouped in parks and squares. We propose a classification of behavior of groups and crowds according to degree of semantics has been carried out in three types; Motion, Action, Activity, also two automatic tasks, Count-People and Tracking. It has included techniques and algorithms that researchers use for analysis, and has included the datasets used, which in most of the investigations are traditional and in a few cases custom datasets or YouTube videos are used.

As future work, it is important to address the issue of video sequences using RGBD cameras, as this type of technology is currently in increasing use.

References

1. J. Azorin-Lopez, M. Saval-Calvo, A. Fuster-Guillo, J. Garcia-Rodriguez, and S. Orts-Escolano, “Self-Organizing Activity Description Map to Represent and Classify Human Behaviour,” *IJCNN 2015*, 2015.
2. A. A. Chaaraoui, P. Climent-Pérez, and F. Flórez-Revuelta, “A review on vision techniques applied to Human Behaviour Analysis for Ambient-Assisted Living,” *Expert Systems with Applications*, vol. 39, no. 12, pp. 10873–10888, 2012.
3. F. Cardinaux, B. Deepayan, A. Charith, M. S. Hawley, S. Mark, D. Bhowmik, and C. Abhayaratne, “Video Based Technology for Ambient Assisted Living : A

review of the literature," *Journal of Ambient Intelligence and Smart Environments (JAISE)*, vol. 1364, no. 3, pp. 253–269, 2011.
4. P. Turaga, R. Chellappa, V. Subrahmanian, and O. Udrea, "Machine Recognition of Human Activities: A Survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, pp. 1473–1488, nov 2008.
5. M. S. Ryoo and J. K. Aggarwal, "Recognition of high-level group activities based on activities of individual members," *2008 IEEE Workshop on Motion and Video Computing, WMVC*, vol. 2008, no. January, 2008.
6. L. Mihaylova, A. Y. Carmi, F. Septier, A. Gning, S. K. Pang, and S. Godsill, "Overview of Bayesian sequential Monte Carlo methods for group and extended object tracking," *Digital Signal Processing: A Review Journal*, vol. 25, no. 1, pp. 1–16, 2014.
7. P. Climent-Pérez, A. Mauduit, D. N. Monekosso, and P. Remagnino, "Detecting events in crowded scenes using tracklet plots," *Proceedings of the International Conference on Computer Vision Theory and Applications*, 2014.
8. S. Vishwakarma and A. Agrawal, "A survey on activity recognition and behavior understanding in video surveillance," *The Visual Computer*, vol. 29, no. 10, pp. 983–1009, 2013.
9. S. Blunsden and R. B. Fisher, "The BEHAVE video dataset: ground truthed video for multi-person behavior classification," *Annals of the BMVA*, vol. 2010, no. 4, pp. 1–11, 2010.
10. J. M. Chaquet, E. J. Carmona, and A. Fernández-Caballero, "A survey of video datasets for human action and activity recognition," *Computer Vision and Image Understanding*, vol. 117, no. 6, pp. 633–659, 2013.
11. O. Banos, M. Damas, H. Pomares, F. Rojas, B. Delgado-Marquez, and O. Valenzuela, "Human activity recognition based on a sensor weighting hierarchical classifier," *Soft Computing*, vol. 17, no. 2, pp. 333–343, 2013.
12. D. Bruckner, G. Q. Yin, and a. Faltinger, "Relieved commissioning and human behavior detection in Ambient Assisted Living Systems," *Elektrotechnik und Informationstechnik*, vol. 129, no. 4, pp. 293–298, 2012.
13. Y. Wu, Z. Jia, Y. Ming, J. Sun, and L. Cao, "Human behavior recognition based on 3D features and hidden markov models," *Signal, Image, Video Processing*, 2015.
14. J. Azorin-Lopez, M. Saval-Calvo, A. Fuster-Guillo, J. Garcia-Rodriguez, M. Cazorla, and M. T. Signes-Pont, "Group activity description and recognition based on trajectory analysis and neural networks," pp. 1585–1592, 2016.
15. J. Azorin-Lopez, M. Saval-Calvo, A. Fuster-Guillo, and A. Oliver-Albert, "A predictive model for recognizing human behaviour based on trajectory representation," *2014 I.J.C on Neural Networks (IJCNN)*, pp. 1494–1501, 2014.
16. E. Andrade, S. Blunsden, and R. Fisher, "Hidden Markov Models for Optical Flow Analysis in Crowds," *18th International Conference on Pattern Recognition*, no. January, pp. 460–463, 2006.
17. Y. Hu, Y. Zhang, and L. S. Davis, "Unsupervised abnormal crowd activity detection using semiparametric scan statistic," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 1, pp. 767–774, 2013.
18. W. Ge, R. T. Collins, and B. Ruback, "Automatically detecting the small group structure of a crowd," in *2009 Workshop on Applications of Computer Vision, WACV 2009*, 2009.
19. K. Goel and A. Robicquet, "Learning Causalities behind Human Trajectories," 2015.
20. A. Maksai, X. Wang, and P. Fua, "Globally Consistent Multi-People Tracking using Motion Patterns," *arXiv preprint arXiv:1612.00604*, vol. 1, 2016.

21. J. Shao, C. C. Loy, and X. Wang, "Scene-Independent Group Profiling in Crowd," pp. 2219–2226, 2014.
22. S. Yi, H. Li, and X. Wang, "Pedestrian Travel Time Estimation in Crowded Scenes Shenzhen Institutes of Advanced Technology , Chinese Academy of Sciences," pp. 3137–3145, 2015.
23. S. Yi, H. Li, and X. Wang, "Understanding pedestrian behaviors from stationary crowd groups," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June-2015, pp. 3488–3496, 2015.
24. A. Al-Raziqi and J. Denzler, "Unsupervised Framework for Interactions Modeling between Multiple Objects," *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, vol. 4, no. Visigrapp, pp. 509–516, 2016.
25. S. Jiao, "Small Group People Behavior Analysis Based on Temporal Recursive Trajectory Identification Cong Shen , Rong Xie , Liang Zhang , Li Song Inst. of Image Communication and Network Eng. Cooperative Medianet Innovation Center,"
26. S. Vascon, E. Z. Mequanint, M. Cristani, H. Hung, M. Pelillo, and V. Murino, "Detecting conversational groups in images and sequences: A robust game-theoretic approach," *Computer Vision and Image Understanding*, vol. 143, no. May 2016, pp. 11–24, 2016.
27. W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly detection and localization in crowded scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 18–32, 2014.
28. J. Liu, X. Tong, W. Li, T. Wang, Y. Zhang, H. Wang, B. Yang, L. Sun, and S. Yang, "Automatic Player Detection, Labeling and Tracking in Broadcast Soccer Video," *Procedings of the British Machine Vision Conference 2007*, pp. 3.1–3.10, 2007.
29. W. Lin, M.-T. Sun, R. Poovandran, and Z. Zhang, "Human activity recognition for video surveillance," *IEEE International Symposium on Circuits and Systems*, no. June, pp. 2737–2740, 2008.
30. M. Harmon, P. Lucey, and D. Klabjan, "Predicting Shot Making in Basketball using Convolutional Neural Networks Learnt from Adversarial Multiagent Trajectories," *arXiv preprint arXiv:1609.04849*, 2016.
31. M. Camplani, A. Paiement, M. Mirmehdi, D. Damen, S. Hannuna, T. Burghardt, and L. Tao, "Multiple Human Tracking in RGB-D Data: A Survey," jun 2016.
32. K. Wickramaratna, M. Chen, S. C. Chen, and M. L. Shyu, "Neural network based framework for goal event detection in soccer videos," *Proceedings - Seventh IEEE International Symposium on Multimedia, ISM 2005*, vol. 2005, pp. 21–28, 2005.
33. H. M. Hamidreza Rabiee, Javad Haddadnia, "Emotion-Based Crowd Representation for Abnormality Detection Hamidreza," *International Journal on Artificial Intelligence Tools*, 2016.
34. H. Fradi and J. L. Dugelay, "Spatial and temporal variations of feature tracks for crowd behavior analysis," *Journal on Multimodal User Interfaces*, vol. 10, no. 4, pp. 307–317, 2016.
35. S. Gong, M. Cristani, S. Yan, and C. C. Loy, "Person Re-Identification," *Acvpr*, 2014.
36. L. Cao and K. Huang, "Video-based crowd density estimation and prediction system for wide-area surveillance," *China Communications*, vol. 10, no. 5, pp. 79–88, 2013.
37. H. Liao, J. Xiang, W. Sun, Q. Feng, and J. Dai, "An abnormal event recognition in crowd scene," *Proceedings - 6th International Conference on Image and Graphics, ICIG 2011*, no. September 2011, pp. 731–736, 2011.

38. M. C. Chang, N. Krahnstoever, S. Lim, and T. Yu, "Group level activity recognition in crowded environments across multiple cameras," *Proceedings - IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2010*, no. February, pp. 56–63, 2010.
39. A. Gning, L. Mihaylova, S. Maskell, S. K. Pang, and S. Godsill, "Group object structure and state estimation with evolving networks and Monte Carlo methods," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1383–1396, 2011.
40. S. Pellegrini, A. Ess, and M. Tanaskovic, "Wrong turn-no dead end: a stochastic pedestrian motion model," *Computer Vision and*, 2010.
41. M. Perše, M. Kristan, S. Kovačič, G. Vučkovič, and J. Perš, "A trajectory-based analysis of coordinated team activity in a basketball game," *Computer Vision and Image Understanding*, vol. 113, no. 5, pp. 612–621, 2009.
42. Y. Yin, G. Yang, and H. Man, "Small human group detection and event representation based on cognitive semantics," *Proceedings - 2013 IEEE 7th International Conference on Semantic Computing, ICSC 2013*, no. September 2013, pp. 64–69, 2013.
43. W. Ge, R. T. Collins, and R. B. Ruback, "Vision-based analysis of small groups in pedestrian crowds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 5, pp. 1003–1016, 2012.
44. B. Lau, K. O. Arras, and W. Burgard, "Multi-model hypothesis group tracking and group size estimation," *International Journal of Social Robotics*, vol. 2, no. 1, pp. 19–30, 2010.
45. J. C. S. Jacques, A. Braun, J. Soldera, S. R. Musse, and C. R. Jung, "Understanding people motion in video sequences using Voronoi diagrams," *Pattern Analysis and Applications*, vol. 10, no. 4, pp. 321–332, 2007.
46. C. Schuldt, L. Barbara, and S. Stockholm, "Recognizing Human Actions : A Local SVM Approach Dept . of Numerical Analysis and Computer Science," *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 3, pp. 32–36, 2004.
47. E. L. Andrade, S. Blunsden, and R. B. Fisher, "Modelling Crowd Scenes for Event Detection," 2006.
48. P. Kilambi, E. Ribnick, A. J. Joshi, O. Masoud, and N. Papanikolopoulos, "Estimating pedestrian counts in groups," *Computer Vision and Image Understanding*, vol. 110, no. 1, pp. 43–59, 2008.
49. C. Zhang, X. Yang, W. Lin, and J. Zhu, "Recognizing human group behaviors with multi-group causalities," *Proceedings of the 2012 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops, WI-IAT 2012*, pp. 44–48, 2012.
50. G. J. Brostow and R. Cipolla, "Brostow: Unsupervised Bayesian Detection of Independent Motion in Crowds,"
51. W. Ge and R. T. Collins, "Marked Point Processes for Crowd Counting," *IEEE Computer Vision and Pattern Recognition, 2009*, pp. 2913–2920, 2009.
52. R. Mehran, A. Oyama, and M. Shah, "Abnormal Crowd Behaviour Detection using Social Force Model," *IEEE Conference on Computer Vision and Pattern Recognition*, no. 2, pp. 935–942, 2009.
53. F. Cupillard, F. Brémond, and M. Thonnat, "Tracking groups of people for video surveillance," *Video-Based Surveillance Systems*, 2002.
54. D. Kong, D. Gray, and H. Tao, "A viewpoint invariant approach for crowd counting," *Proceedings - International Conference on Pattern Recognition*, vol. 3, pp. 1187–1190, 2006.