Abnormal behaviour detection on queue analysis from stereo cameras

Luis Patino and James Ferryman
University of Reading, Computational Vision Group
Whiteknights, Reading RG6 6AY, United Kingdom
{j.l.patinovilchis, j.m.ferryman}@reading.ac.uk

Csaba Beleznai
AIT - Austrian Institute of Technology
Donaucitystr. 1, 1220 Vienna, Austria
csaba.beleznai@ait.ac.at

Abstract

In this paper we perform an analysis of human behaviour for people standing in a queue with the aim to discover, in an unsupervised way, ongoing unusual or suspicious activities. The main activity types we focus on are detecting people loitering around the queue and people going against the flow of the queue or undertaking a suspicious path. The proposed approach works by first detecting and tracking moving individuals from a stereo depth map in real time. Activity zones (including queue zones) are then automatically learnt employing a soft computing-based algorithm which takes as input the trajectory of detected mobile objects. Statistical properties on zone occupancy and transition between zones makes it possible to discover abnormalities without the need to learn abnormal models beforehand. The approach has been tested on a dataset realistically representing a border crossing and its environment. The current results suggest that the proposed approach constitutes a robust knowledge discovery tool able to extract queue abnormalities.

1. Introduction

The systematic analysis of visual streams acquired from surveillance cameras, is an essential aspect of safety and security systems. Depending on the application domain, such systems may specialise in the detection of some targeted events. Sensitive infrastructures are interested in detecting a person in a forbidden area [15, 16]. For systems deployed to monitor disabled/elderly persons, a person falling is an important event [20]. Counting people is an important feature in space/ environment planning in ambient intelligence applications [12]. Border control areas, for instance at airports, require a dedicated analysis for people queuing up to present documentation and have access granted. Detecting events related to queues is thus an important feature in ensuring the safety and security of border control areas. Queues have been the subject of previous studies where the main aim is to calculate some important statistics such as the length of the queue in order to measure the service quality and/or give customers an idea of the waiting time [10, 11, 19].

In this work we address the task of suspicious detection from video recordings acquired on networks of cameras deployed in queuing areas. Thus in this work we are focusing on identifying those activities which are abnormal (or unexpected) in the sense they differ from the frequent patterns observed in the monitored area. We focus on detecting people loitering around the queue and people undertaking a suspicious (unfrequent) path. The proposed approach works by first detecting and tracking the mobiles (moving individuals); a challenging task in itself given the high human density (occluded visibilities) and typical occasional motion (stop-and-go). Activity zones (including queue zones) are then automatically learnt employing a soft computing-based algorithm which takes as input the trajectory of detected mobile objects. Statistical properties on zone occupancy and transition makes it possible to discover abnormalities without the need to learn abnormal models beforehand. Although recognising these type of activities is essential on border control areas, no comparable study has been undertaken in queue analysis. The present paper constitutes the first work in this domain. The remainder of the paper is organised as follows. The next section gives a short overview of the related work. The general system description is presented in Section 3. Section 4 presents how object detection and tracking is achieved in crowded scenarios such as queuing areas. In Section 5 it is explained how trajectories are analysed and abnormal behaviour recognition is performed. Section 6 gives the main results and evaluation. Finally, Section 7 draws the main conclusions and describes possible future work.

2. Related work

Queue analysis has been the subject of recent interest mainly from a managerial and client-service point of view.

978-1-4673-7632-7/15/$31.00 ©2015 IEEE

Research trends are addressed towards monitoring statistics like Queue length, Throughput rate, Growth rate, Wait time to help businesses enhance service efficiency, improve customer satisfaction and increase revenue. Estimating queue parameters is an important part of many business operations, including retail shops, public transport hubs and airports. Diverse works have addressed this problematic either employing video content analytics on imagery acquired by surveillance cameras [11] or other devices and signal types such as WiFi and smartphones [10, 19]. However, sensitive areas, such as border control areas at airports, require more sophisticated monitoring to detect anomalous behaviour. Although behaviour detection has not been expicitly adressed in queue analysis, it constitutes a very active research domain in computer vision. Behaviour extraction corresponds mainly to matching information coming from sensors observing the scene with predefined event models which humans are using to understand the scene. Such event models, in some cases, can be set manually with the domain-expert knowledge. Such video interpretation systems have been built in the past. PRISMATICA [18] was a video surveillance system tested on-site in Paris and London undergrounds and able to detect overcrowding/congestion, intrusion, and stationarity of people. Similarly, ADVISOR [4] was tested in Brussels and Barcelona metro stations and was able to detect fighting between persons and vandalism. Poppa et al. [14] designed a system to recognise predefined events in a shopping centre based on the combination of trajectory classification, action recognition and zone information (i.e. time spent in a given zone). The trajectory classification has three categories (disoriented, looking around and goal oriented). The number of behaviours that can be recognised is limited to 6 manual models (looking for support, disoriented buying, disoriented, goal oriented buying, looking around buying). These kind of systems generally rely on manually-set thresholds; however, the adequate values might be difficult to find or might work only on very specific situations. Furthermore the manual definition of zones of interest is often required as a spatial constraint in the event recognition. This makes the system very application/context dependent. The research challenge is to attempt to learn the behaviour models. The most popular techniques include Hidden Markov Models (HMM) [1, 7], attribute multiset grammars/Bayesian Networks [5] and graph mining [17]. However, in these approaches the complexity of models increasees considerably to cope the many different varieties at which the event may occur and still achieve the event recognition.

Our contribution to the state of the art is thus a behaviour analysis-based approach for queue analysis able to discover abnormalities without the need to learn specific behaviour models beforehand. The proposed approach is based on automatically learnt activity zones, which include queuing areas by taking as input the trajectory of detected mobile objects. We have developed specific algorithms to detect and track people in particularly crowded areas. Zone characteristics such as occupancy and transition allow detecting people loitering around the queue and people going against the flow of the queue or undertaking a suspicious path.

3. General system description

Behaviour characterisation is based on the analysis of trajectories from detected mobile objects. Object detection and tracking is thus the first processing step working on the streams of video acquired on the monitoring area. Behaviour recognition is achieved by learning the activity zones where mobiles objects move along a certain path. As observed from Figure 1, the input to the zone learning procedure are the extracted trajectory points of interest, which correspond to mobile changes in speed or direction. In a third step the mobile activity is characterised as a series of visited activity zones (activity extraction module). Such characterisation allows delivering behaviour events indicating the mobile activity. As a final step a check is made to extract those abnormal activity patterns which deviate from the usual behaviours.

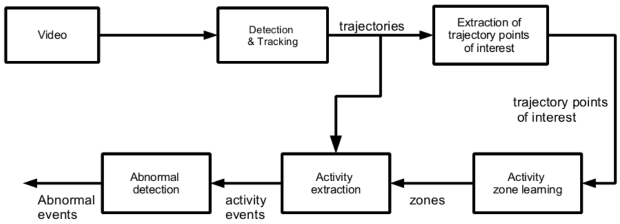

Figure 1. Processing chain for the proposed approach

4. Detection and Tracking

We use an in-house developed sensor (Figure 2) to extract intensity information, employing a canonical stereo setup (three monochrome cameras mounted in parallel), with a baseline of $0.4m$ between the two cameras located at the ends of the rig. The board-level industrial cameras have a USB2 interface and the resolution of the sensor is 1280×1024 pixels, resampled to 1150×920 with 8 bit quantization. This trinocular camera setup is calibrated offline. The stereo matching process outputs depth data alongside with rectified intensity images, congruent to the depth image. Depth information is computed via a pyramidal implementation [6] of a Census-based stereo matching algorithm, which is an explicit adaption and optimization of the well-known Census transform in respect to embedded real-time systems in software. Depth is computed for all three available baselines thus improving the quality of obtained depth map at the different spatial ranges. At the given resolution, the sensor delivers approximately 10 fps, when stereo computation is performed on a modern PC.

Figure 2. Our customized trinocular stereo camera setup with a baseline of 40cm (between the left and right cameras).

4.1. Human detection

We estimate the 3D pose of a dominant ground plane in the scene using vertical disparity statistics [8]. Given the recovered plane parameters, a perpendicular projection of the depth data onto the 3D plane can be computed, forming an ortho-map. Given the relation to the ground plane, depth data is filtered according to its height above the ground, retaining only data points between 5 cm (approximately corresponding to the variance of data noise) and 2 m. Humans standing and moving on the ground plane appear as persistent peaks in the ortho-map.

Our objective is to delineate multiple stable structures in the ortho-map in a stable manner. We introduce a stability criterion which controls a scale-adaptive mean shift [3] search for clusters which remain stable across varying scales. We apply spatial grouping of stable clusters generating compact convex hull-based contour representations. We exploit the integral image concept [2] to efficiently compute intensity sums and higher order statistical moments within windows of arbitrary scales to obtain fast estimates for the local covariance matrix. A stability criterion is defined as a stopping criterion by finding local minima in the rate of change of a given cluster area over iterations. In such a way multiple cluster hypotheses are generated, which are combined in a simple fusion step: first, elliptic contours are approximated by resampling, creating a closed polygonal contour representation. Next, the convex hull of strongly overlapping polygonal cluster boundaries is generated. The compact convex hull-based representation of the outline has the advantage that nearby, truly distinct, oriented objects remain spatially separated after the spatial grouping step, thus well outlining the multi-modal distribution of nearby humans in the ortho-map. For tracking,we employ a standard Bayesian filter based multi-target tracking scheme where data association is performed by the Hungarian algorithm and states are estimated by a Kalman filter. The target states are represented by (x, y, z) coordinates which are metric coordinates on the ground plane, with z being height of the centroid of the 3d object standing on the ground plane.

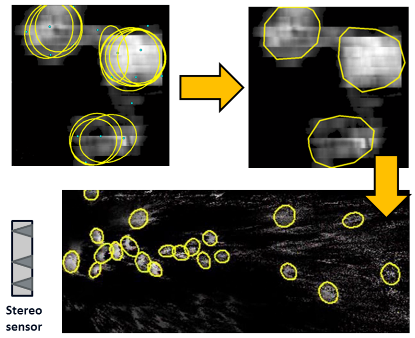

Figure 3. Top-left: Stable clusters estimated for a fast scale-adaptive CamShift method. Top-right: Computed convex hulls delineating the final objects. Bottom: Sample detection results in a queue seen from the top.

5. Behaviour analysis

We achieve complex activity recognition by first learning activity zones where mobiles evolve in the scene. We interpret activity zones as those important areas on the observed scene where mobiles interact with other mobiles or perform behavioural changes. Such behavioural changes include: stop walking, speed up walking or simply stand waiting. Note that this information can be extracted from the analysis of the mobile speed profile. The first task is thus to analyse the mobile speed profile and obtain those speed changing points. The second task is to cluster speed changing points to build the final activity zones. Finally we perform the activity extraction by representing each track as a sequence of transitions through zones. By analysing zone occupancy and transition we identify abnormal behaviours.

5.1. Speed changing points extraction

Each trajectory is defined as the set of points $[x_j(t), y_j(t)]$ corresponding to their position on the ground on the t-th frame. The instantaneous speed for that mobile at point $[x_j(t), y_j(t)]$ is then $v\left(t\right) = \left(\dot{x}\left(t\right)^2 + \dot{y}\left(t\right)^2\right)^{\frac{1}{2}}$, and the direction θ that the mobile takes at that point is $\theta\left(t\right) = arctan\left(\dot{y}\left(t\right)/\dot{x}\left(t\right)\right)$.

Each of these two time series is analysed in the frame of a multiresolution analysis [9] with a Daubechies Haar smoothing function, $\rho_{2^s}\left(t\right) = \rho\left(2^s t\right)$, to be dilated at different scales s. The analysis is performed through six dyadic scales. The effect at performing a broader approximation is to smooth out signal variations at each scale. We select as speed changing points and direction changing points those points seen as sharp discontinuities which remain present across scales despite the smoothing procedure.

5.2. Zone computation

Activity zones are computed having as input the track speed (or direction) changing points calculated as explained in last section. These points are first clustered by a fast partitioning algorithm. In a second step the partition is corrected leading to the final activity zones.

Computing Initial Activity Zones. The algorithm works on-line without needing to specify the number of clusters in advance. The first point is assigned as Leader representative of a new cluster. Then the next point is assigned to an existing cluster or defines a new cluster depending on the distance between the point and the cluster leading representative. The process is repeated until all input points are assigned to clusters. In our application, the cluster influential zone, Z_n, is defined by a radial basis function (RBF) centered at the position of the point designed as cluster leader (or leading representative), L; and the membership of a new point $p(u, v)$ to that zone is given by:

$$Z_n(L,p) = \phi(L,p) = exp(-\frac{\|p - L\|^2}{T^2}) \quad (1)$$

The RBF function has a maximum of 1 when its input is $p = L$ and thus acts as a similarity detector. An object element will be included into a cluster Z_n if $Z_n(L,p) \geq 0.5$; the cluster receptive field (hyper-sphere) is controlled by parameter T.

Final Activity Zone calculation. For the refinement of this initial zone partition we replicate the soft computing-based methodology given by [13] as this technique has shown to be effective for activity zone learning. We merge together initial zones which fullfill the following relationships. $R_{1_{ij}}$: Zone Z_{n_i} overlaps Zone Z_{n_j}. $R_{2_{ij}}$: zone Z_{n_i} and zone Z_{n_j} are destination zones for mobiles departing from any same activity zone Z_{n_k}. $R_{3_{ij}}$: zone Z_{n_i} and zone Z_{n_j} are origin zones for mobiles arriving to the the same activity zone Z_{n_k}. $R_{4_{ij}}$: zone Z_{n_i} and zone Z_{n_j} have about the same number of detected mobiles stopping at the zone. $R_{5_{ij}}$: zone Z_{n_i} and zone Z_{n_j} have about the same mobile interaction time. All relations can be aggregated employing a typical bounded product T-norm soft computing operator $R = \max(0, R_1 + R_2 + R_3 + R_4 + R_5 - 4)$.

5.3. Behaviour extraction

Having discovered in total $k = 1, ..., K$ zones; and AZn_k^α is one zone resulting from the zone learning procedure, we understand then a mobile behaviour as the sequence of transitions between learned zones in its trajectory. Two different transitions can be defined:

- Mobile from Zone AZn_k^α to Zone $AZn_{k'}^\alpha$
- Mobile stays at Zone AZn_k^α

The complete behaviour is then characterised as the ordered sequence of transitions generated as the mobile moves between zones.

5.4. Abnormal activity finding

In this work two kinds of abnormal events are essentially targeted: people undertaking unusual paths and people staying unusually long in an activity area. As a rule of thumb, it is interpreted that the person is loitering if she/he undertakes an unusual path and at some point has also stayed long in any activity zone. In all cases the decision on abnormality is based on the calculation of statistical confidence levels. This Statistics-based methodology is based on the idea that 'normal' data objects follow a generating mechanism, e.g. some given statistical process (distribution model). 'Abnormal' objects deviate from this generating mechanism's statistics. Such a statistical measure is mathematically defined with the following equation:

$$CL = 1.96 * \left(\frac{\sigma}{\sqrt{n}}\right) \quad (2)$$

where σ is the standard deviation of a given measured parameter, μ, and n is the number of observations of such parameter. The confidence level can be interpreted as having 95% confidence that the true value of μ lies between $[\overline{\mu} - CL, \overline{\mu} + CL]$. Such statistical bounds are employed as a reference to decide whether a measurement could be considered suspicious by comparing whether it lies inside or outside such statistical bounds. The two parameters we measure for abnormality are thus the time a mobile object spends in the given activity zone, and the path frequency for a given trajectory.

6. Experimental results and evaluation

We applied the proposed approach on a staged queue scenario recorded in an isolated room, which could well form the hall area in an airport or at another border checkpoint. The experimental recording lasted about 25 minutes where the involved actors were not given any other particular instruction than forming the queue. Despite the crowded environment, we successfully detected and tracked the actors on the monitored area. There are 437 mobiles detected in the queue. Actors spontaneously performed some abnormalities such as joining or leaving the queue in the middle and not at the ends; joining or exiting the queue by undertaking 'irregular' paths. Some actors were as well spontaneously off the queue, standing and looking at the queue advancing. This kind of behaviour is what we call loitering. The proposed approach succeded at detecting all of these abnormalities, working thus as a knowledge discovery tool indicating which behaviours were adrift from normal patterns. An example of a loitering behaviour is pictured in Figure 4.

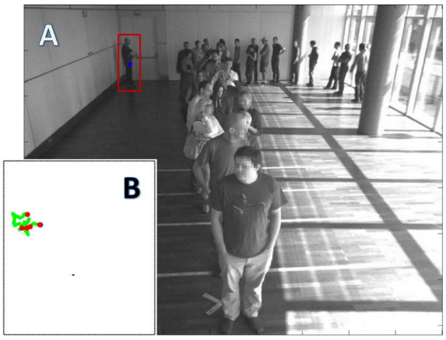

Figure 4. Loitering detection. A) An individual is behaving with a suspicious path and spending a significant amount of time off the queue. B) The associated trajectory, which led to the abnormal detection. Dots in the trajectory represent stop or change direction points.

MobileID	Abnormal Event	Visual Interpretation	Visual validation
6926	Loitering	Loitering off queue	TP
6985	Loitering	Loitering off queue	TP
7433	suspicious path	Joins queue then comes back	TP
9036	suspicious path	Abandons queue in the middle	TP
9066	suspicious path	Joins queue in the middle	TP
9321	Loitering	Loitering near queue	TP
9595	suspicious path	joins queue in the middle	TP
9653	suspicious path	exits queue wrong direction	TP

Table 1. Abnormal behaviours detected with the proposed approach.

An example of abnormal behaviour by someone exiting the queue in the wrong direction can be observed in Figure 5. Table 1 lists all abnormal behaviours correctly retrieved.

Although from visual inspection of the recorded video we detected all abnormalities, we still have a number of false positive detections; which are detailed in Table 2. There are two main difficulties we are facing in our system, which account for most of the wrong detections. Mobiles are detected with enough accuracy in most of the monitored area. However, we face a typical approach/leave tracking challenge, which consists in accurately following the mobile as it approaches or goes away from the camera. The challenge namely being able to handle the change in size of the mobile as seen from the camera. Mobiles far from the camera are harder to identify and for the tracker it is easier to wrongly associate targets ('ID change'). This is emphasised by the fact we are working with a crowded environment. The second problem we face is the typical occlusion tracking challenge. In the recording, the queue passes behind a column and this occlusion induces some erroneous positions before recovering the tracked mobile. In both cases the delivered trajectory is interpreted as an abnormal pattern by the behaviour module and the mobile is thus flagged as having abnormal behaviour. It is to be remarked that the number of False Positives is relatively small given the number of mobiles detected and particularly the fact that the area is crowded. We still have a couple of detections which we flagged as False Positive because although the undertaken path from the mobiles was unfrequent; the behaviour itself cannot be semantically considered abnormal, i.e. joining the end of the queue slightly by the side compared to other people. Overall we can state the proposed approach is very promising as in general we obtain a sensitivity of 100% and a specificity of over 90%.

Abnormal event	Number of FP Instances	False detection reason
suspicious path	15	ID change
suspicious path	9	Occlusion
suspicious path	2	Unfrequent path

Table 2. False alarms on abnormal behaviour detection.

7. Conclusions and future work

In this work we have addressed the challenging problem of detecting abnormal behaviour in queues formed by people. We have succeeded to detect and track individuals moving in the queue with sufficient accuracy to understand their behaviours. To overcome the difficulty to perform with high human density (occluded visibilities) and typical occasional motion (stop-and-go), we have worked with a canonical stereo set-up (three monochrome cameras mounted in parallel) and employed a Bayesian filter-based multi-target tracking scheme. We employ the trajectories from detected individuals to extract behaviour and detect abnormalities. We base our analysis on automatically learning the main activity areas of the scene (including queueing zones); mobile objects are then characterised in relation to the learned activity areas: either as 'staying in a given activity zone' or 'transferring from an activity zone to another'. Statistical properties on zone occupancy and transition between zones make it possible to discover abnormalities. The approach has been useful to discover two types of abnormalities: 'people loitering' and 'suspicious path'. We analysed 25 minutes of staged data and succeeded to discover such abnormalities appearing in the video. Overall the results we obtain in sensitivity and specificity are very encouraging. We still produced a number of false positives meaning we need to render our system more robust to deal with potential errors due to 'occlusion' and/or 'change of ID' in object detection and tracking. We will address this aspect in our future work as well as analysing the possibility to extract some other behaviours also in multiple queues, i.e. 'Jockeying' (a person may move from one queue to another).

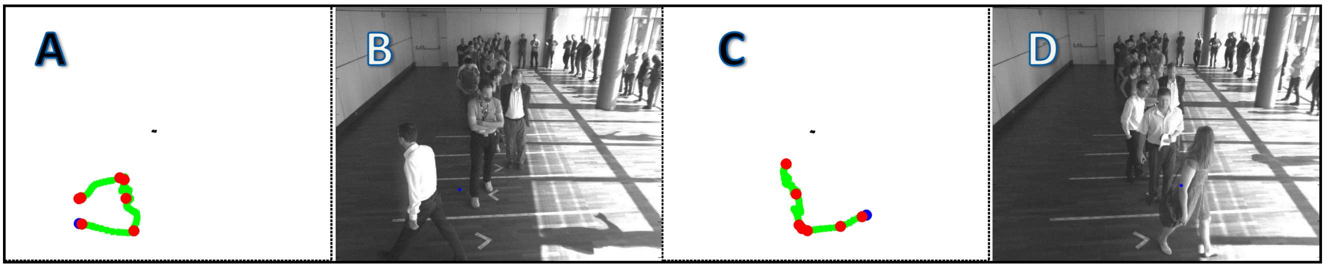

Figure 5. Wrong path detection. A) Mobile trajectory detected as abnormal. B) Individual undertaking wrong queue exit. C) Mobile trajectory detected as normal. D) Individual undertaking normal queue exit.

Acknowledgement

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 312583.

References

[1] F. Bashir, A. Khokhar, and D. Schonfeld. Object Trajectory-Based Activity Classification and Recognition Using Hidden Markov Models. *IEEE Transactions on Image Processing*, 16:1912–1919, 2007.

[2] C. Beleznai, P. Sommer, and H. Bischof. Scale-adaptive clustering for object detection and counting. In *Proc. IEEE Workshop on Performance Evaluation in Surveillance and Tracking*, 2008.

[3] G. R. Bradski. Computer vision face tracking for use in a perceptual user interface. *Intel Technology Journal*, (Q2):15, 1998.

[4] F. Cupillard, A. Avanzi, F. Bremond, and M. Thonnat. Video understanding for metro surveillance. In *Networking, Sensing and Control, 2004 IEEE International Conference on*, volume 1, pages 186–191 Vol.1, March 2004.

[5] D. Damen and D. Hogg. Explaining activities as consistent groups of events. *International Journal of Computer Vision*, 98(1):83–102, 2012.

[6] M. Humenberger, C. Zinner, M. Weber, W. Kubinger, and M. Vincze. A fast stereo matching algorithm suitable for embedded real-time systems. *Comput. Vis. Image Underst.*, 114(11):1180–1202, Nov. 2010.

[7] M. G. Kim, S. H. Lee, and I. H. Suh. Learning of social skills for human-robot interaction by hierarchical hmm and interaction dynamics. In *Electronics, Information and Communications (ICEIC), 2014 International Conference on*, pages 1–2, Jan 2014.

[8] Y. Li and Y. Ruichek. Occupancy grid mapping in urban environments from a moving on-board stereo-vision system. *Sensors*, 14(6):10454, 2014.

[9] S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(7):674–693, 1989.

[10] T. Okoshi, Y. Lu, C. Vig, Y. Lee, R. K. Balan, and A. Misra. Queuevadis: Queuing analytics using smartphones. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, IPSN '15, pages 214–225, New York, NY, USA, 2015. ACM.

[11] V. Parameswaran, V. Shet, and V. Ramesh. Design and validation of a system for people queue statistics estimation. In C. Shan, F. Porikli, T. Xiang, and S. Gong, editors, *Video Analytics for Business Intelligence*, volume 409 of *Studies in Computational Intelligence*, pages 355–373. Springer Berlin Heidelberg, 2012.

[12] L. Patino, H. Benhadda, E. Corvee, F. Bremond, and M. Thonnat. Extraction of activity patterns on large video recordings. *Computer Vision, IET*, 2(2):108–128, June 2008.

[13] L. Patino, F. Bremond, and M. Thonnat. Online learning of activities from video. In *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on*, pages 234–239, 2012.

[14] M. C. Popa, L. J. M. Rothkrantz, C. Shan, T. Gritti, and P. Wiggers. Semantic assessment of shopping behavior using trajectories, shopping related actions, and context information. *Pattern Recogn. Lett.*, 34(7):809–819, may 2013.

[15] G. Sanrom, L. Patino, G. Burghouts, K. Schutte, and J. Ferryman. A unified approach to the recognition of complex actions from sequences of zone-crossings. *Image and Vision Computing*, 32(5):363 – 378, 2014.

[16] K. Smith, P. Quelhas, and D. Gatica-perez. Detecting abandoned luggage items in a public space. In *in PETS, 2006*, pages 75–82, 2006.

[17] M. Sridhar, A. G. Cohn, and D. C. Hogg. Relational graph mining for learning events from video. In *Proceedings of the 2010 conference on STAIRS 2010: Proceedings of the Fifth Starting AI Researchers' Symposium*, pages 315–327, Amsterdam, The Netherlands, The Netherlands, 2010. IOS Press.

[18] S. Velastin, B. Boghossian, B. Lo, J. Sun, and M. Vicencio-Silva. Prismatica: toward ambient intelligence in public transport environments. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 35(1):164–182, Jan 2005.

[19] Y. Wang, Y. Chen, and R. Martin. Leveraging wi-fi signals to monitor human queues. *Pervasive Computing, IEEE*, 13(2):14–17, Apr 2014.

[20] N. Zouba, F. Bremond, and M. Thonnat. An activity monitoring system for real elderly at home: Validation study. In *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, pages 278–285, Aug 2010.