Meeting detection in video through semantic analysis

Meeting detection in video through semantic analysis

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

Abstract

In this paper we present a novel approach to detect people meeting. The proposed approach works by translating people behaviour from trajectory information into semantic terms. Having available a semantic model of the meeting behaviour, the event detection is performed in the semantic domain. The model is learnt employing a soft-computing clustering algorithm that combines trajectory information and motion semantic terms. A stable representation can be obtained from a series of examples. Results obtained on a series of videos with different types of meeting situations show that the proposed approach can learn a generic model that can effectively be applied on the behaviour recognition of meeting situations.

1. Introduction

Behaviour analysis is an essential task in modern video surveillance systems. Traditionally, their main purpose has been to raise an alarm on detection of specific threats such as abandoned luggage, loitering or intrusion in forbidden zones [12, 13]. However, extraction of behaviour (activity) patterns has also proved valuable on fields other than security, such as daily living monitoring [17] or space management to learn the main flows of people and/or dense areas of the observed scene [9]. Recently special attention has been placed on social analysis or interaction detection between people in the monitored space. Firstly, when detecting that people are meeting into groups, detection and tracking systems could potentially better adapt to occlusion and other related problems when being aware of the meeting situation; secondly, because at a higher level of behaviour analysis it is important to know when a group is forming and what kind of interaction its members may have.

Although it would seem easy to think that people meeting detection can be achieved by detecting people being close to one each other, this simplistic approach does not show the intention or voluntariness of the members to form the group [14] leading to the detection of short-lived false alarms.

Current vision-based systems attempt to model the meeting situation either manually or with machine learning. The first type of approaches are difficult to create because these models generally rely on manually-set thresholds and the adequate values might be difficult to find or might work only on very specific situations. Machine-based learned models have started to show significant results on detecting accurate voluntary meeting situations [4] but at the same time the complexity of models has incremented while trying to build an enough generic model capable of recognising the many different varieties at which the event may occur. Analysing long temporal storylines, setting causal relationships between mobiles, including pixel-based analysis for action recognition, are just some examples.

In this work we claim that the trajectory information has yet not been fully exploited. We propose that the visual behaviour can be translated from trajectory into semantic terms and a generic model can be obtained to recognise a meeting situation from a semantic analysis. We show the model can be learned employing a soft-computing clustering algorithm. We have applied our approach to the public database CAVIAR and the results are encouraging when compared with other state of the art approaches. 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. In Section 4, it is explained how trajectories are analysed, then how we employ the soft-computing clustering algorithm is presented in Section 5. Section 6 explains how the semantic model is learnt. Section 7 gives the main results and evaluation. Finally, Section 8 draws the main conclusions and describes possible future work.

2. Related work

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 domain-expert knowledge. The re-

search challenge is evidently to attempt to learn the behaviour models.

Setting the model to recognise people meeting and group forming has been researched in both categories. In the first case, for instance, Artikis et al. [1] employ an Event Calculus (introduced by Kowalski and Sergot [6]) consisting in manually setting short-term events with temporal constraints that, if satisfied, lead to the recognition of the behaviour of interest. The approach is however complex in setting thresholds and parameters. In the second case, statistical methods have been mostly employed to enhance tracking while groups are forming. Bazzani et al. [2], for instance, employ particle filtering and a joint tracking of individuals and groups, which feed each other. Pellegrini et al. [11] perform trajectory prediction with Conditional Random Fields having modelled group appearance. Although HMMs and Bayesian Networks are very common in computer-vision modelling, they are less popular on detecting people meeting due to the combinatorial complexity that increases with the number of individuals and possible states. Olivier et al. [8] study HMM and CHMM for modeling the group interaction; Bayesian networks have been employed for instance by Intille et al. [5]. The current trend and the most Interesting results have been recently reported by employing Trajectory clustering. Zaidenberg et al. [16] employ mean-shift to cluster trajectories of individuals between T frames to find similar trajectories, representative of groups. The approach is more suited to regroup people close to each other rather than willingling to meet. Other works have researched to define social features to better represent the voluntariness to meet, for instance, attraction and repulsion forces [14, 15]; or taking into account social theory such as the proxemic space between mobiles [3]. Choi et al. [4] attempt to enhance the recognition of interaction between mobiles by adding supplementary information to trajectory position such as 'pose'. Sanroma et al. [12] also look to extract information from action recognition.

Our contribution to the state of the art is a novel semantic approach to model people meeting situations. We translate trajectory motion into semantic terms. The event detection is then transformed in the semantic domain and achieved when the coordinated semantic terms between the involved mobiles appear. The semantic model is first learnt from a series of videos containing the meeting situation. In a first step we employ a soft-computing clustering algorithm where trajectory information and motion semantic terms are employed to identify the meeting situation; in a second step a stable representation with only semantic terms is extracted from the different meeting situations. The semantic model can then be applied directly to any other video.

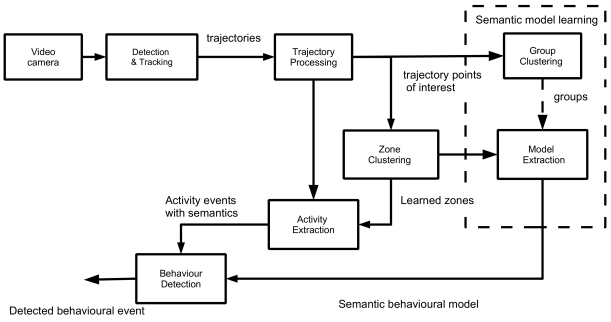

Figure 1. Processing chain for the proposed approach

3. General system description

The proposed approach consists in performing the event recognition in the semantic domain. That is, trajectory motion is translated into semantic terms, then having available a semantic model of the meeting situation, those semantic terms contained in the model are identified from the semantic characterisation of those mobiles involved in the meeting. The complete processing chain of the proposed approach is shown in Figure 1. The proposed system would then start by the analysis of detected mobile trajectories (Trajectory Processing), which consists in extracting trajectory points of interest indicating mobile change of speed or direction. Moreover, speed changing points are given semantic labels allowing to better understand the mobile behaviour (i.e. 'stopping' or 'increasing speed').

Trajectory points of interest are the input to several following modules. They are first employed to learn Activity Zones of the scene (those zones where mobiles enter/exit the scene or have social interactions). Having activity zones calculated, the mobile trajectory and its points of interest are employed together with the learned zones to individually characterise mobile activity as a series of visited activity zones (activity extraction module). Such characterisation allows delivering behaviour events which already contain semantic terms from the trajectory speed analysis. In these semantic terms the meeting situation between mobiles can be recognised (Behaviour Detection) if the semantic model is available. Such semantic model is obtained during a learning phase. Trajectory points of interest are also input to a clustering algorithm aiming to discover meeting situations (Group Clustering). Mobiles involved in the meeting situation have their individual activity characterised, as mentioned before, employing learned activity zones. Semantic terms leading to the meeting situation are identified; if several meeting examples are available, a stable representation will be built from them (Model Extraction).

4. Trajectory Processing

Behavioural indicators for meeting situations are for instance 'people stopping walking to meet ' or people who 'change direction' to approach someone else. Information of this type can be extracted from both the analysis of the mobile trajectory speed and direction profile. There are thus two parallel processes: The first is to analyse the mobile speed profile and obtain those speed changing points. The second is to analyse the mobile direction profile and obtain those direction changing points.

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 $v(t) = \left(\dot{x}(t)^2 + \dot{y}(t)^2\right)^{\frac{1}{2}}$, and the direction θ that the mobile takes at that point is $\theta(t) = arctan(\dot{y}(t)/\dot{x}(t))$.

Each of these two time series is analysed in the frame of a multiresolution analysis [7] with a Daubechies Haar smoothing function, $\rho_{2^s}(t) = \rho(2^s t)$, to be dilated at different scales s.

In this frame, the approximation A of $v(t)$ by ρ; where b is a translation parameter spanning the time domain of $v(t)$, is such that $A_{s-1}(v) = \int v(t)\,\rho\left(2^{s-1}t - b\right)dt$ is a broader approximation of $A_s v$ and correspondingly for $A_{s-1}(\theta)$ and $A_s\theta$. 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 strong variations in the signal; such points remain present in all scales despite the smoothing procedure.

Speed changing points are then labelled according to the direction of the speed change : 'with decreasing speed', 'with increasing speed', 'with normal speed', 'stopping'. Change direction points are labelled on a single category: 'Change direction'.

5. Clustering

We employ clustering at two different moments in the processing chain. First to allow the system to automatically learn activity zones where mobiles show behavioural change (possibly for interaction with other mobiles). Then to detect those mobiles which are effectively grouping. In both cases, we employ a soft computing clustering algorithm. The motivation is that soft computing provides uncertain information processing capability and set a framework to work with symbolic/linguistic terms; a key feature for our approach based on the mobile semantic term analysis. It is to be noted that soft computing clustering has already shown to be effective for activity zone learning [10]. We reproduce thus the automatic zone generation algorithm proposed in [10]. For the detection of mobiles meeting and thus forming a group, we propose in this work a new set of soft computing relationships, which we describe here below.

5.1. Soft computing relation clustering

Any relation between two sets *X* and *Y* is known as a binary relation *R*:

$$R = \{((x,y), \mu_R(x,y)) \mid (x,y)\, \epsilon X \times Y\}$$

and the strength of the relation is given by $\mu_R(x,y)$

Let's consider now two different binary relations, *R1* and *R2*, linking three different fuzzy sets *X*, *Y*, and *Z* :

$R1$ = x is relevant to y; $R2$ = y is relevant to z

It is then possible to find to which extent *x* is relevant to *z* by employing the extention principle (noted $R = R1oR2$):

$$\mu_{R=R1\circ R2}(x,z) = \max_y \min\left[\mu_{R1}(x,y), \mu_{R2}(y,z)\right]$$

R can be made furthermore closure transitive following the next steps

Step 1. $R' = R \cup (R \circ R)$
Step 2. If $R' \neq R$, make$R = R'$ and go to step1
Step 3. $R = R'$ Stop.

(1)

R is the transitive closure where

$$R \circ R(x,y) = \max_z \min\left(R(x,z), R(z,y)\right) \qquad (2)$$

If we define a discrimination level α in the closed interval [0,1], an $\alpha - cut$ can be defined such that

$$R^\alpha(x,y) = 1 \Leftrightarrow R(x,y) \geqslant \alpha \qquad (3)$$

From the classification point of view, R^α induces a new partition π^α with a new set of clusters $\pi^\alpha = \left\{CL_1^\alpha, \cdots, CL_k^\alpha, \cdots, CL_{|\pi^\alpha|}^\alpha\right\}$ such that cluster ${CL_k}^\alpha$ is made of all initial elements x, y, z which up to the alpha level fullfill the final similarity relation in Equation 2.

5.2. Relation setup for group detection

Here we set out to establish the appropriate spatio-temporal relationships identifying the meeting between mobiles in the observed scene. The relation definition is based on the natural assumption that the meeting situation happens when mobiles are coming spatially and temporally closer one to each other. The first pair of relations to be included in the clustering algorithm are thus:

$R1_{ij}$: mobile object *O(i)* spatial position is close to mobile object *O(j)* spatial position

$R2_{ij}$: mobile object *O(i)* temporal position is close to mobile object *O(j)* temporal position

We strengthen this natural definition by adding behavioural cues meaningful for the meeting situation and which are obtained from the mobiles speed and direction analysis. Namely, we establish

$R3_{ij} = 1$ IF (mobile object *O(i)* speed label is 'Stopping' AND mobile object *O(j)* speed or direction label is 'Stopping' or 'with decreasing speed' or 'Change direction') OR IF (mobile object *O(i)* speed label is 'with decreasing speed' AND mobile object *O(j)* speed or direction label is 'Stopping' or 'with decreasing speed' or 'Change direction'); $R3_{ij} = 0$ otherwise.

In the above, 'close to' is a linearly decreasing fuzzy triangular membership function outputting 1 for a null distance between mobiles. All relations can be aggregated employing a soft computing aggregation operator such as

$R = R = R1 \cap R2 \cap R3 = \max(0, R1 + R2 + R3 - 2)$ and made transitive with Equation 1. Clusters of activity are obtained after applying an $\alpha - cut$ discrimination level as indicated in Equation 3.

6. Meeting behaviour model learning

Trajectory information can be translated into semantic terms with the help of discovered zones and speed and direction labels.

Let us assume, we have in total $k = 1, ..., K$ learned zones; and AZn_k^α is one learned zone. The different kinds of behaviours that can now be identified with learned zones, and taking into account the set of speed and direction semantic labels, are thus:

• Mobile with $speed - direction - label$ from Zone AZn_k^α to Zone $AZn_{k'}^\alpha$
• Mobile at Zone AZn_k^α with $speed - direction - label$

Having at hand a representative number of meeting situations, it is possible to mine the stream of behaviours corresponding to the mobiles involved in the meeting. Practically, this can be achieved by extracting the behaviours related to the common learned zone, corresponding to the meeting situation, and identifying those which are common to all meeting situations.

7. Experimental results and evaluation

We have evaluated our approach on the publicly available CAVIAR dataset. The dataset is representative of the challenge addressed as it contains, among others, eight

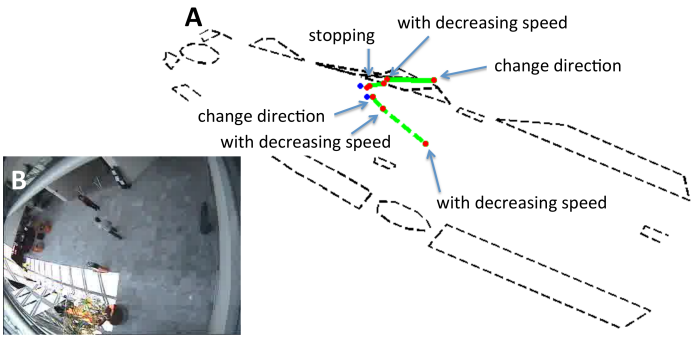

Figure 2. Group meeting detected in one of the CAVIAR sequences (mwt2gt). **Panel (A)** shows tracks (in green), speed and direction changing points (in red) and corresponding semantic labels generated by our approach. Remark that for each of the two mobiles meeting, not all of their trajectory is shown, but only that portion leading to the meeting situation as outputted by the Group clustering algorithm. **Panel (B)** shows the mobiles meeting in a corresponding frame.

acted sequences of people meeting and in some cases having a fight. The dataset is challenging because the meeting situations are varied. In some cases the meeting involves a significant amount of movements and short displacements from the actors (particularly when a fight is involved). Often the meeting situation is short-lived and in several scenarios other people walk near the mobiles meeting, or some people simply walk near each other, which can create confusion on what are the true meetings. The dataset contains annotated detection and tracking of the mobiles appearing in the scene. Mobiles forming groups in the scene are identified and their behaviour annotated as 'joining' or 'interacting'. We have employed the available detection/tracking as input to our system, and the Group annotations to evaluate our approach. We followed the standard machine learning leave-one-out testing methodology. All video sequences except one are used for training and the remaining one is used as test case. This process is iterated until each video sequence is used as test case exactly once. Figure 2 shows, as an example, one of the video sequences processed during the learning phase (mwt2gt). The result shows the mobile tracks, while meeting, as outputted by the Group clustering algorithm. It can be observed that the stream of semantic terms contained in the group cluster are indeed compatible with a possible natural description of a meeting situation.

As previously mentioned, we achieve behaviour recognition by first learning activity zones where mobiles change position in the scene; then the mobile movement is characterised as a pattern of visited activity zones. Table 1 shows the mobile characterisation taking activity zones into account for one pass in the Leave-one out evaluation. Sequence 'mwt2gt' shown graphically in Figure 2 corre-

Sequence	Stream
fcgt	**mobile1** Chg direction, at Zone6 THEN Chg direction, with normal speed at Zone6 **before mobile2** with normal speed at Zone8 THEN with decreasing speed at Zone6 THEN with decreasing speed at Zone6 THEN Chg direction, at Zone6
fomdgt	**mobile1** with normal speed at Zone19 THEN Chg direction, with decreasing speed at Zone4 THEN with decreasing speed at Zone19 **before mobile2** with decreasing speed at Zone4 THEN Chg direction, at Zone21 THEN Chg direction, with normal speed at Zone21 THEN with decreasing speed at Zone21
fra1gt	**mobile1** with increasing speed at Zone28 THEN Chg direction, at Zone2 THEN with normal speed at Zone2 THEN with increasing speed at Zone2 **before mobile2** Stopping at Zone29 THEN with decreasing speed at Zone29 THEN with normal speed at Zone29 THEN Chg direction, at Zone29 THEN with normal speed at Zone29 THEN with normal speed at Zone29 THEN with increasing speed at Zone29 THEN with decreasing speed at Zone2 THEN Chg direction, with decreasing speed at Zone30 THEN Chg direction, with normal speed at Zone30
mws1gt	**mobile2** Chg direction, with increasing speed at Zone8 THEN with increasing speed at Zone10 THEN with decreasing speed at Zone10 THEN with decreasing speed at Zone3 THEN with normal speed at Zone3 THEN with decreasing speed at Zone3 THEN with normal speed at Zone3 THEN with decreasing speed at Zone3 THEN with normal speed at Zone3 THEN with decreasing speed at Zone3 THEN with decreasing speed at Zone3 **before mobile1** with normal speed at Zone4 THEN with decreasing speed at Zone3 THEN Chg direction, with normal speed at Zone3 THEN Chg direction, at Zone3 THEN Chg direction, at Zone3 THEN with decreasing speed at Zone3 THEN Chg direction, at Zone3
mwt1gt	**mobile1** with normal speed at Zone4 THEN with decreasing speed at Zone3 THEN Chg direction, with decreasing speed at Zone3 THEN with normal speed at Zone3 THEN Chg direction, at Zone3 **before mobile2** Chg direction, at Zone21 THEN with decreasing speed at Zone3
mwt2gt	**mobile2** with decreasing speed at Zone10 THEN with decreasing speed at Zone23 THEN Chg direction, at Zone17 THEN Chg direction, with normal speed at Zone17 **before mobile1** Chg direction, at Zone16 THEN Chg direction, at Zone17 THEN Chg direction, at Zone17 THEN with decreasing speed at Zone17 THEN Stopping; at Zone17 THEN Stopping; at Zone17 THEN Stopping; at Zone17 THEN with normal speed at Zone17
	Extracted Pattern
	mobileA (Chg direction AND with normal speed) OR with decreasing speed before mobileB with decreasing speed

Table 1. Semantic characterisation of the meeting situation for a set of six sequences and the meeting pattern extracted from them employing the learned activity zones.

sponds to the last processed sequence in the table. The common representation for those sequences taken into account is highlighted in the last row of the table. This is the result given by the 'Model Extraction' module in our system architecture. Note that the learned rule represents indeed a generic semantic model of the meeting situation. The test sequence for this particular iteration of the leave-one-out process is sequence 'fra2gt' where thanks to the semantic model learned, the meeting situation is correctly recognised.

It is to be noted that scenarios, such as 'mwt2gt', employed during the learning process, and 'fra2gt', employed for test, are completely different in their meeting situation. In the former scenario one of the mobiles stands and waits for the other mobile to meet. In the latter scenario the meeting is rather abrupt as it involves the two mobiles suddenly involved in a fight. The spatial spread and dynamics (such as individual speeds and approaching directions) are very different; yet, a spatial closeness between the mobiles can

Test Sequence	Semantic Model	TP	FP	FN
mwt2gt	mobileA (Chg direction AND with normal speed) OR with decreasing speed before mobileB with decreasing speed	2	0	0
mwt1gt	mobileA (Chg direction AND with normal speed) OR with decreasing speed before mobileB with decreasing speed	1	0	0
mws1gt	mobileA (Chg direction AND with normal speed) OR (Chg direction AND with decreasing speed) before mobileB with decreasing speed	1	1	0
fra2gt	mobileA (Chg direction AND with normal speed) OR with decreasing speed before mobileB with decreasing speed	1	0	0
fra1gt	mobileA (Chg direction AND with normal speed) OR with decreasing speed before mobileB with decreasing speed	1	0	0
fomdgt1	mobileA (with normal speed AND Chg direction) OR (with normal speed AND with decreasing speed) before mobileB with decreasing speed	1	0	0
fcgt	mobileA (Chg direction AND with normal speed) OR with decreasing speed before mobileB with decreasing speed	1	0	0
mw3ggt	mobileA (Chg direction AND with normal speed) OR with decreasing speed OR before mobileB with decreasing speed	0	0	1

Table 2. Recognition results for each iteration of the leave-one-out validation process.

be asserted when a common activity zone is learned, and particularly because among the stream of semantic events leading to the meeting situation, those semantic terms in the learned model can be recognised.

The complete set of results of our experiments are summarised in Table 2. It can be observed that a generic model for the behaviour of interest can be learned. The model appears to be very constant and can successfully be applied on a series of scenarios in which the way the mobiles meet varies considerably.

In our results from Table 2, only one False Positive was detected which means the approach does rather well at differentiating between mobiles simply walking near to each other and actually meeting. We still obtained 1 False Negative in sequence 'ms3g'. The behaviour detection is challenging because the meeting situation appears to start when both mobiles are further away from each other. The semantic terms characterising the meeting semantic model occur but do not temporally overlap when mobiles share a common activity zone.

We compared our approach with two methods from the state of the art having different foundations. The first approach [16] is a clustering-based algorithm. The second approach [1] employs event calculus and it manually sets the event components to achieve the behaviour recognition. Both are evaluated on the same CAVIAR dataset. The first approach [16] approach reports best performance evaluation but only targeting three meeting situations in the dataset. Artikis et al [1] works with the same video se-

Method	Instances	TP	FP	FN	Precision	Recall
[16]	3	3	0	0	100%	100%
[1]	9	6	1	3	86%	67%
Proposed	9	8	1	1	89%	89%

Table 3. Comparison of different results.

quences as us. Our approach has a better evaluation in terms of TP, FN. Although the results could still be improved, the current performance is encouraging.

8. Conclusions

This work addresses the problem of automatically detecting people meeting. The key-novelty in the proposed approach is translating people behaviour from trajectory information into semantic terms. A generic model of a meeting situation can be learned from a series of examples (training set). For this stage, in a first step, a soft-computing clustering algorithm is employed to identify meeting situations combining trajectory information and motion semantic terms; in a second step a model containing only semantic terms is extracted from the different examples. The recognition of an unseen meeting situation can be then performed in the semantic domain.

The approach is evaluated in the publicly available CAVIAR dataset. Our current results are encouraging as we obtain high values of Precision and Recall. The performance of the proposed approach can concurrence other state of the art techniques. We have the advantage that no thresholds on distances or time must be set to recognise the meeting behaviour. In order to perfect the performance of the proposed approach we consider to add supplementary semantic terms and potentially, apply the semantic analysis at different spatial and temporal resolutions to better capture meeting situations with different spatio-temporal spread.

Acknowledgement

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 607567.

References

[1] A. Artikis, M. Sergot, and G. Paliouras. A logic programming approach to activity recognition. In *Proceedings of the 2Nd ACM International Workshop on Events in Multimedia*, EiMM '10, pages 3–8, New York, NY, USA, 2010. ACM.

[2] L. Bazzani, M. Zanotto, M. Cristani, and V. Murino. Joint individual-group modeling for tracking. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 37(4):746–759, April 2015.

[3] S. Calderara and R. Cucchiara. Understanding dyadic interactions applying proxemic theory on videosurveillance trajectories. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 20–27, June 2012.

[4] W. Choi and S. Savarese. Understanding collective activitiesof people from videos. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(6):1242–1257, June 2014.

[5] S. S. Intille and A. F. Bobick. Recognizing planned, multiperson action. *Computer Vision and Image Understanding*, 81(3):414 – 445, 2001.

[6] R. Kowalski and M. Sergot. A logic-based calculus of events. *New Gen. Comput.*, 4(1):67–95, Jan. 1986.

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

[8] N. Oliver, B. Rosario, and A. Pentland. A bayesian computer vision system for modeling human interactions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):831–843, Aug 2000.

[9] 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.

[10] 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.

[11] S. Pellegrini, A. Ess, and L. Van Gool. Improving data association by joint modeling of pedestrian trajectories and groupings. In K. Daniilidis, P. Maragos, and N. Paragios, editors, *Computer Vision ECCV 2010*, volume 6311 of *Lecture Notes in Computer Science*, pages 452–465. Springer Berlin Heidelberg, 2010.

[12] 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.

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

[14] J. Sochman and D. Hogg. Who knows who - inverting the social force model for finding groups. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 830–837, Nov 2011.

[15] K. Yamaguchi, A. Berg, L. Ortiz, and T. Berg. Who are you with and where are you going? In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1345–1352, June 2011.

[16] S. Zaidenberg, B. Boulay, and F. Bremond. A generic framework for video understanding applied to group behavior recognition. In *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on*, pages 136–142, Sept 2012.

[17] 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.