Multicamera trajectory analysis for semantic behaviour characterisation

Conference or Workshop Item

Accepted Version

www.reading.ac.uk/centaur

CentAUR

Multicamera trajectory analysis for semantic behaviour characterisation

Luis Patino and James Ferryman
University of Reading, Computational Vision Group
P.O. Box 225, Whiteknights, Reading RG6 6AY, United Kingdom
{j.l.patinovilchis, j.m.ferryman}@reading.ac.uk

Abstract

In this paper we propose an innovative approach for behaviour recognition, from a multicamera environment, based on translating video activity into semantics. First, we fuse tracks from individual cameras through clustering employing soft computing techniques. Then, we introduce a higher-level module able to translate fused tracks into semantic information. With our proposed approach, we address the challenge set in PETS 2014 [1] on recognising behaviours of interest around a parked vehicle, namely the abnormal behaviour of someone walking around the vehicle.

1. Introduction

In recent years, behaviour understanding from multicamera analysis has acquired a significant importance due to the increasing use of cameras and particularly networks of cameras in our daily lives. Ambient intelligence cameras [2], traffic cameras and indoors or public space monitoring cameras [10] are just some examples. Understanding human activities is important for a series of reasons including 1) security reasons, where the essential task is to detect threats on the camera video streams and to send appropriate alarms; 2) management reasons, such as vehicle traffic control or people flow control in transport systems; 3) health reasons, such as building the personal daily living activity profile on people suffering from Alzheimers. Multi-camera analysis has the inherent advantage of larger and/or more-detailed coverage of the monitored space but at the same time brings the challenge to fuse the multi-camera information to build an unified view of the human activities across the network of cameras. Furthermore, once a unified activity representation is obtained, it still remains to characterise it, or extract the interesting behaviour patterns such that the vision system can indeed identify the ongoing human activities in the monitored space. In this paper we have addressed these two challenges. We propose to fuse individual people detections from a network of cameras to obtain a unified trajectory of

the mobile inside the monitored space. In a second step, we analyse the trajectory to understand the mobile behaviour and express it in semantic terms. Important behaviours of interest can then be detected and extracted from a semantical analysis. We have applied our approach on the PETS 2014 [1] dataset and addressed the recognition of the abnormal behaviour of someone walking around the vehicle. Our current results are encouraging as all instances in the dataset are detected with a very few number of false positives. The remainder of the paper is organised as follows. Section 2 presents the related work from the literature. The overview of our system is given in Section 3. The fusion of tracks from individual cameras is explained in Section 4; and Section 5 describes how unified trajectories are processed. The semantic activity extraction and analysis is explained in Section 6. Our results on the PETS2014 dataset are given in Section 7. Our conclusions and future work are summarised in Section 8.

2. Related work

Learning behaviour based on trajectory analysis is now a well established area in video understanding. This type of analysis is most popular mainly due to its effectiveness in detecting normal/abnormal behaviours; for instance, on abnormal trajectory detection on roads [18, 17] or pedestrian trajectory characterisation [3]. Hidden Markov Models (HMM) have also been employed to detect different states of pre-defined normal behaviour [4]. However, most of the work done in behaviour understanding is mainly done on a single camera view.

Behaviour analysis in a multicamera environment is of deep interest because of the increased global view of the space under monitoring as a single camera can only have a limited view. Multicamera analysis brings the advantage to better study the global activity and the longer 'history' of single activities but at the same time faces as the first challenge the correct multi-object correspondence across cameras.

There are two different approaches to analyse and understand multicamera information: These are either centralised

or distributed. The former is probably the most widespread approach consisting to infer the multicamera topology correspondance and achieve tracking on the ground plane. A variety of techniques have been considered on this aspect, for instance, Kim and Davis [13] proposed a multi-view multi-hypothesis approach. Fleuret et al. [8] proposed an algorithm combining a probabilistic occupancy map. Du and Piater [5] presented an approach to perform ground-plane single target tracking fusing multiple camera information with sequential belief propagation. Yao and Odobez [20] present an approach to automatically detect and track a variable number of people in a multi-camera environment with partial field of view overlap using Bayesian tracking. S. Khan and M. Shah [12] calculate the 'edges' of a camera view as seen in other cameras. These methods are however very dependent on the overlap between cameras to solve the multicamera correspondence problem, which actually still remains an open problem.

When the camera views are disjointed or there is almost no overlap, the second approach employed to analyse and understand multicamera information is then distributed, in which tracking is conducted independently at each camera, and then results from the different cameras are fused and combined at a higher level. In this case it is often the appearance model which is set in correspondance between camera views. Once the correspondence is established, trajectories from different camera views can be fused and analysed globally. Numerous works have adressed this problem [7, 11, 19].However, matching appearance models between cameras is also difficult because the appearance of objects may vary between cameras as objects may be observed with a different pose, with different resolutions, or are subject to different illumination or occlusion conditions.

In this paper we propose an innovative way for behaviour recognition employing distributed tracking. First, we fuse tracks from disjoint cameras through clustering employing soft computing techniques. Then, we have the addition of a higher-level module able to translate fused tracks into semantic information. With our proposed approach, we address the challenge set in PETS 2014 [1] on recognising behaviours of interest around a parked vehicle, namely the abnormal behaviour of someone walking around the vehicle.

3. System description

Behaviour characterisation is based on the analysis of trajectories from detected mobile objects; with Low-level tracking information transformed into high-level semantic descriptions. This paper is however not addressing the multicamera tracking challenge. Instead, we employ the tracking results available with the PETS2014 dataset [1]. The hypothesis is thus that an online module has already performed the tracking process. As observed in Figure 1, we

assume a distributed multicamera tracking system, which is the case for the data available in PETS 2014. The first step in our system is thus the fusion of trajectories from different camera views. In a second step fused trajectories are analysed globally to extract points of interest and then employed to achieve activity zone learning. As previously mentioned, we achieve complex activity/behaviour recognition by first learning activity zones where mobiles evolve in the scene. In a fourth step the mobile movement is characterised as a pattern of visited activity zones. Employing the inherent soft relation properties, the reported activities are labeled with understandable semantics. In the final step behaviour recognition is achieved by the analysis of the reported semantics.
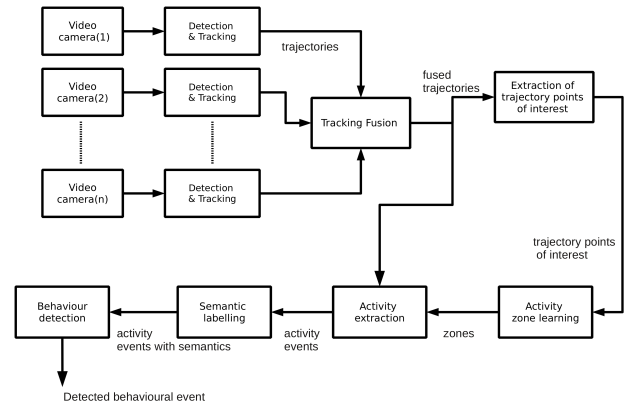

Figure 1. Processing chain for the proposed approach

4. Trajectory Fusion

In this work we understand behavioural activity as the coherent spatio-temporal trajectory interaction across cameras. We propose in this work to group trajectories from all cameras in order to form those coherent clusters of activity. We propose for this task a soft computing clustering algorithm. The motivation is that soft computing provides uncertain information processing capability; set a framework to work with symbolic/linguistic terms and thus allows producing natural language-like interpretable results. This is a key feature in our approach as behaviour recognition is based on the mobile semantic term analysis.

4.1. Fuzzy relation clustering

A fuzzy set is a set of ordered pairs such that $A = \{(x, \mu_A(x)) \mid x \epsilon X\}$. Given two fuzzy sets, A and B, it is possible to calculate their union or intersection with typical T-norm and T-conorm operators such that,

$$\mu_{A \bigcap B}(x) = T(\mu(A), \mu_B(x)) = min(\mu_A(x), \mu_B(x))$$

$$\mu_{A \bigcup B}(x) = S(\mu(_A), \mu_B(x)) = max(\mu_A(x), \mu_B(x))$$

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

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

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 $R = R1oR2$):

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

It is interesting to verify whether the resulting relation is symmetric, $R(x,y) = R(y,x)$, reflexive $R(x,x) = 1$, which make of R a compatibility relation and occurs in most cases when establishing a relationship between binary sets. Because R was calculated employing the extention principle, R is also a transitive relation. $R(x,y)$ is a transitive relation if $\exists z \in X, z \in Y / R(x,y) \geqslant \max_z \min[R(x,z), R(z,y)]$

R can be made furthermore closure transitive following the next steps

Step1. $R' = R \cup (R \circ R)$
Step2. If $R' \neq R$, make $R = R'$ and go to step1
Step3. $R = R'$ Stop.

$$(1)$$

R is the transitive closure where

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

R is now a transitive similarity relation with R indicating the strength of the similarity. 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 $\pi^\alpha = \{CL_1^\alpha, \cdots, CL_k^\alpha, \cdots, CL_{|\pi^\alpha|}^\alpha\}$ 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. It should be noted that this relation clustering was first proposed by Patino et al. [15] but employed solely for the learning of activity zones.

4.2. Relation setup for clustering

We now set out to establish the appropriate relations between detected mobiles across cameras reflecting spatio-temporal similarities in order to obtain coherent groups of trajectories reflecting a unified activity.

With this aim, we define the following relations:

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

$R2_{ij}$: mobile object *O(i)* end speed is close to mobile object *O(j)* start speed

$R3_{ij}$: mobile object *O(i)* temporal end is close to mobile object *O(j)* start

$R4_{ij}$: mobile object *O(i)* end direction is close to mobile object *O(j)* start direction

$R5_{ij}$: mobile object *O(i)* does not overlap temporally with mobile object *O(j)*

All relations can be aggregated employing a soft computing aggregation operator such as
$R = R = R1^t \cap R2^t \cap R3^t \cap R4^t \cap R5^t = \max(0, R1^t + R2^t + R3^t + R4^t + R5^t - 4)$ and made transitive with equation 1. Clusters of activity are obtained after applying an $\alpha - cut$ discrimination level as indicated in Equation 3.

5. Trajectory Processing

In our work, the mobile behaviour is characterised as a pattern of visited activity zones. We understand 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 to meet someone', 'change direction while walking', 'speed up walking' or simply 'stand waiting'. Also this information can be extracted from the analysis of the mobile speed profile. Once trajectories across cameras have been fused and unified, the first task is thus to analyse the mobile speed profile and obtain those speed changing points and those direction changing points (indicattors of behavioural changes). The second task is then to cluster these key points to build the final activity zones. 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.

Let us consider the dataset $O = \{tr_j\}, j = 1 \ldots N$ made up of *N* trajectories. Each trajectory is defined as the set of points $[x_j(t), y_j(t)]$ corresponding to their po-

sition 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 $\theta(t) = arctangent\left(\dot{y}(t)/\dot{x}(t)\right)$.

Each of these two time series is analysed in the frame of a multiresolution analysis [14] with a Daubichis 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 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 correspondly for $A_{s-1}(\theta)$ and $A_s\theta$. The analysis is performed through six dyadic scales. Speed changing points and direction changing points correspond to those important discontinuities which remain present across scales.

A clustering algorithm is then applied to the speed and direction key points, employing a fast partitioning algorithm such as the well-known Leader algorithm [9, 6]), allowing us to quickly create an initial set of zones Zn. Then, the partition is corrected employing again the same generic fuzzy relation clustering described in section 4.1 with an appropriate set of relationships: $R1_{ij}$: Zone Zn_i overlaps Zone Zn_j, $R2_{ij}$: zone Zn_i and zone Zn_j are destination zones for mobiles departing from any same activity zone Zn_k, $R3_{ij}$: zone Zn_i and zone Zn_j are origin zones for mobiles arriving to the the same activity zone Zn_k, $R4_{ij}$: zone Zn_i and zone Zn_j have about the same number of detected mobiles stopping at the zone, $R5_{ij}$: zone Zn_i and zone Zn_j have about the same mobile interaction time. Note that this is the same online procedure as employed by Patino et al. [16]. However, unlike [16], we also analyse the direction profile in order to extract as well the information on sudden direction changes or trajectories constantly changing direction.

6. Activity Extraction

Let us assume, we have in total $q = 1, ..., Q$ user-defined zones $Zctx_q$ (also named contextual zones), and $k = 1, ..., K$ learned zones. If $\pi^\alpha = \{AZn_1^\alpha, \cdots, AZn_k^\alpha, \cdots, AZn_K\}$ is the partition for the α level, and AZn_k^α is one learned zone from the induced partition; the different kinds of behaviours that can now be identified for the α level are thus:

- Mobile from Zone $Zctx_q$ to Zone $Zctx_{q'}$
- Mobile from Zone $Zctx_q$ to Zone AZ_k^α
- Mobile from Zone AZn_k^α to Zone $Zctx_q$
- Mobile from Zone AZn_k^α to Zone $AZn_{k'}^\alpha$
- Mobile at Zone $Zctx_q$
- Mobile at Zone AZn_k^α

We attribute discovered zones a semantic label, which allows description of the discovered zone in terms of user-defined zones when available. To this end we consider two new soft-computing relations: The comparison of areas between discovered and user-defined areas, and the distance relationships between discovered and user-defined areas.

The activity is then characterised as the ordered sequence of simple behaviours generated as the mobile visits different learned or user-defined zones. The recognition of displacements of interest (walking around the veicle in this case) comes down to the semantic interpretation of zone-based delivered events.

7. Experimental results

We have addressed the challenge set in PETS2014 [1] regarding the behaviour recognition of someone walking around a parked vehicle. We have thus processed all sequences marked in the dataset description as containing at least one instance of someone walking around the vehicle. The total number of sequences procesed from the PETS2014 dataset amounts then to seven. Each sequence is composed of four different recordings corresponding to four non-overlapping cameras, mounted on the vehicle itself, and covering the space around. Figure 2 shows an example (sequence 06_01) of the person walking around the vehicle and being detected and tracked separatly on each of the four onboard cameras.

Each dataset was processed independently meaning that no accumulation of evidence was necessary for the behaviour recognition of the person walking around the vehicle. Our fusion of tracks within the same camera across cameras works successfully. Figure 3 shows, as an example, how the proposed algorithm succeds at fusing the tracks from across cameras into a coherent single activity. Remark that although the clustered activity contains only a few tracks across cameras, our fusion algorithm has to extract the activity clusters from about 300 to 600 tracks contained in each sequence with all four cameras taken into account. To achieve the semantic activity extraction, we employ 4 user-defined areas: truck, pedestrian way-out, service and smoking areas. (seen in Figure 2). When a unified trajectory is analysed, the extracted activity is then reported in terms of these user-defined areas. Figure 3 shows as well the semantic being obtained for the fused trajectory from the person walking around vehicle shown in Figure 2.

Some other examples of the walking around vehicle behaviour as reported by our system in other sequences are:

* From just north of zone Truck to just east of zone Truck; Then to just north-east of zone Truck; staying in just north-east of zone Truck; Then to 14.4 metres away northeast of zone Bus Shelter

* From 14.2 meters away north of zone Coffee to just south-east of zone Truck; Then to just north of zone Truck; staying in just north of zone Truck; Then to just east of zone Truck; Then to just north-east of zone Truck

Some examples of other general activities detected by our system are for instance:

* From 15.5 meters away north-west of zone Coffee to 11.6 meters away east of zone Truck; Then to 9 meters away east of zone Truck; staying in 9 meters away east of zone Truck

The recognition of the behaviour of interest (walking around truck) comes down to the interpretation of semantic delivered events. As such, all learned zones tagged as just-east, just-west, just-north, just-south are interpreted as arriving at/exiting the vehicle/area or interest. For example, the event From just west of zone Truck; staying in just west of zone Truck; Then to just south-west of zone Truck; Then to just east of zone Truck; staying in just east of zone Truck gives a clear indication of the person walking around vehicle. The key words 'From just * Truck' and 'to just * Truck' provides us enough information to detect the targeted behaviour, where * is one of the directional tags attributed by our system.

Applying this semantic event recognition we obtain the results shown in Table 1. On the seven analysed sequences, we detected 23 events corresponding to a person walking around the vehicle. From them, our visual inspection on the video indicates that 19 of them correspond correctly to the behaviour (TP=19) and four of them are wrong (FP=4).

False positives that we obtained can be explained as follows. In two cases (sequence 06_04, merged MobileID=3 and MobileID=2), the detected mobiles walked close to the vehicle and seen from two cameras. The semantic given was [From just south-east of zone Truck to just south of zone Truck] but the visual inspection indicates the intention was not to walk two sides of the truck although in our engine we allow such interpretation for the case one person walks partially one side and then another one. The remaining false positives occurs because of a tracking error. Indeed a false object detection identifies the truck plastic cover at one side, occasionally as one person moving along the vehicle. Tracking errors of this type, or some others like mobiles associated as a group in one view, then tracked separately in another view, or broken tracks in a single view occur frequently in the dataset but our algorithm still copes with these problems particularly fusing broken tracks in a single view and establishing coherent associations across cameras. In general, it can be established from our results that the targeted behaviour is successfully recognised despite obtaining some false positives.
.

8. Conclusions

We have presented in this paper a new approach for behaviour recognition, from a multicamera environment. The novelty lies on the fact that we translate video activity into semantics and we analyse these to interpret and detect be-
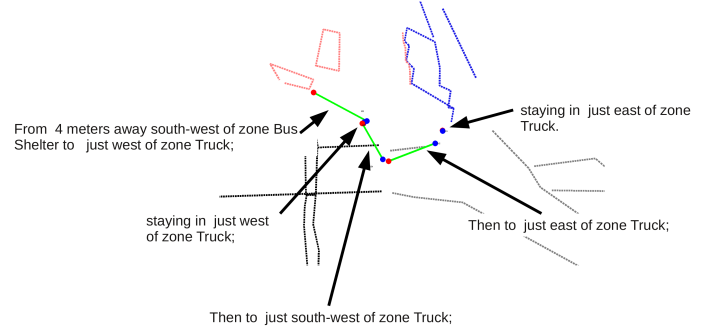

Figure 3. Unified trajectory containing those separate tracks from Figure 2. For simplicity, the trajectory is represented as straight lines between start and end points of the individual contained tracks. Corresponding to each mobile displacement it is shown the semantic description automatically generated by our algorithm.

Sequence	MobileID	Event start time	Event end time	Visual validation
10_03	1	14:53:20.836	14:53:34.307	TP
10_03	3	14:52:52.4333	14:53:10.625	TP
10_03	4	14:53:00.925	14:53:06.525	TP
06_01	14	13:43:46.779	13:45:14.039	TP
06_01	11	13:44:58.9263	13:45:39.439	TP
06_01	7	13:44:37.7263	13:44:49.279	TP
06_01	12	13:43:18.211	13:44:56.611	TP
06_01	13	13:44:59.0597	13:45:02.2597	TP
06_01	21	13:45:50.961	13:45:55.961	TP
10_04	5	14:54:44.056	14:54:47.675	TP
10_04	1	14:54:15.885	14:54:15.885	FP
10_05	5	14:56:23.223	14:56:24.823	TP
10_05	7	14:55:58.873	14:56:03.373	TP
10_05	4	14:55:52.333	14:56:07.354	FP
14_01	3	16:30:52.809	16:31:07.687	TP
14_07	12	16:52:53.967	16:52:59.517	TP
14_07	7	16:52:25.111	16:52:25.911	TP
06_04	7	14:03:00.34	14:03:15.9853	TP
06_04	3	14:03:29.64	14:03:35.5187	FP
06_04	10	14:02:38.17	14:03:33.52	TP
06_04	8	14:03:12.5187	14:03:37.02	TP
06_04	2	14:03:01.09	14:03:46.9853	FP
06_04	9	14:03:33.02	14:03:42.109	TP

Table 1. Walking around vehicle-recognised behaviour in the PETS 2014 dataset.

haviours of interest. We have set up a set of soft computing relationships and clustering algorithm to fuse tracks from non-overlapping cameras. Then, we have in addition a higher-level module able to translate fused tracks into semantic information perform the recognition of behavious of interest. Our results obtained in the PETS 2014 dataset for the targeted behaviour (walking around vehicle) are encouraging. We have succeeded to coherently fuse tracks within a single camera view, and also across cameras. We have detected all dataset instances and the number of false positives is low. However, to enhance the system we will work to set more strict constraints in the activity semantic interpretation and lower the number of false positives.
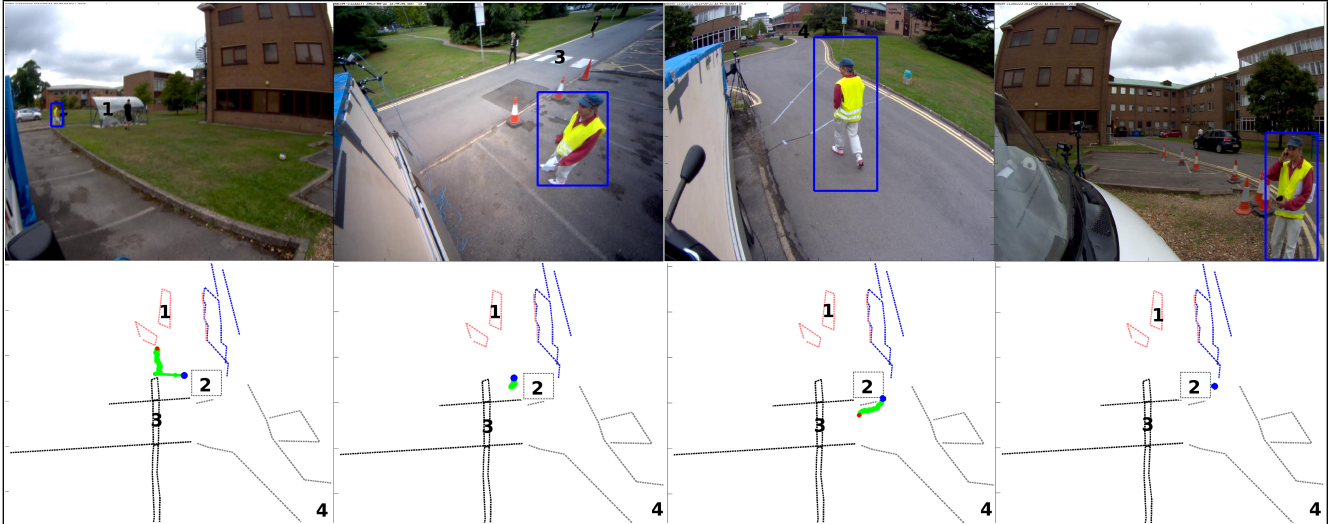
Figure 2. Walking around vehicle PETS2014 sequence 06_01. Top panel: Object detection on each individual on-board camera. Bottom panel: Distributed object tracking. Green lines represent the object trajectory on each camera. Red point indicates start of the trajectory; blue point indicates end of the trajectory. User zones of interest are indicated with numbers: 1) Bus shelter, 2) Truck, 3) Pedestrian way and 4) Coffee zone.

Acknowledgements

References

[1] PETS 2014 the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, www.pets2014.net.

[2] G. Acampora, D. Cook, P. Rashidi, and A. Vasilakos. A survey on ambient intelligence in healthcare. *Proceedings of the IEEE*, 101(12):2470–2494, Dec 2013.

[3] N. Anjum and A. Cavallaro. Single camera calibration for trajectory-based behavior analysis. In *AVSS 2007, IEEE Conference on Advanced Video and Signal Based Surveillance*, pages 147–152, 2007.

[4] 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.

[5] W. Du and J. Piater. Data fusion by belief propagation for multi-camera tracking. In *Information Fusion, 2006 9th International Conference on*, pages 1–8, July 2006.

[6] R. Duda, P. Hart, and D. Stork. *Pattern Classification and Scene Analysis*. John Wiley & Sons, Inc., New York, 1995.

[7] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2360–2367, June 2010.

[8] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua. Multicamera people tracking with a probabilistic occupancy map. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(2):267–282, Feb 2008.

[9] J. A. Hartigan. *Clustering algorithms*. John Wiley & Sons, Inc., New York, 1975.

[10] Y. Huang, J. Cui, F. Davoine, H. Zhao, and H. Zha. Head pose based intention prediction using discrete dynamic bayesian network. In *Distributed Smart Cameras (ICDSC), 2013 Seventh International Conference on*, pages 1–6, Oct 2013.

[11] O. Javed, K. Shafique, and M. Shah. Appearance modeling for tracking in multiple non-overlapping cameras. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 26–33 vol. 2, June 2005.

[12] S. Khan and M. Shah. Consistent labeling of tracked objects in multiple cameras with overlapping fields of view. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(10):1355–1360, Oct 2003.

[13] K. Kim and L. S. Davis. Multi-camera tracking and segmentation of occluded people on ground plane using search-guided particle filtering. In *Proceedings of the 9th European Conference on Computer Vision - Volume Part III*, ECCV'06, pages 98–109, Berlin, Heidelberg, 2006. Springer-Verlag.

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

[15] J. L. Patino, F. Brémond, and M. Thonnat. Activity discovery from video employing soft computing relations. In *IJCNN*, pages 1–8. IEEE, 2010.

[16] 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.

[17] F. Porikli. Learning object trajectory patterns by spectral clustering. In *2004 IEEE International Conference on Multimedia and Expo (ICME)*, volume 2, pages 1171–1174. IEEE, 2004.

[18] C. Stauffer and W. E. L. Grimson. Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):747–757, 2000.

[19] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu. Shape and appearance context modeling. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8, Oct 2007.

[20] J. Yao and J.-M. Odobez. Multi-Camera Multi-Person 3D Space Tracking with MCMC in Surveillance Scenarios. In *Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications - M2SFA2 2008*, Marseille, France, 2008. Andrea Cavallaro and Hamid Aghajan.