A HYBRID ARCHITECTURE FOR TRACKING PEOPLE IN REAL-TIME USING A VIDEO SURVEILLANCE CAMERA: APPLICATION FOR BEHAVIOURAL MARKETING

Kheireddine AZIZ[1], Djamal MERAD[2], Jean-Luc DAMOISEAUX[3] and Pierre DRAP[2]

[1]SeaTech Toulon, Toulon University, La Gardes, France
[2]LSIS Lab, Aix-Marseille University, Marseille, France
[3]IUT R&T, Aix-Marseille University, Marseille, France

ABSTRACT

This paper describes a novel method for tracking customers using images taken from video-surveillance cameras. This system analyzes the number of customers and their motions through the aisles of big-box stores (supermarkets) in real-time. The originality of our approach is based on the study of the blobs properties for managing the splitting/merging issues using a mathematical morphology operator. In the order hand, in order to manage a high number of customers in real-time, we combine the advantage of two tracking algorithms.

KEYWORDS

Deterministic tracking, splitting/merging blobs, behavioural marketing.

1. INTRODUCTION

The economic climate since 2008, as well as increasing competitive pressure has forced major retailers to diversify their sale offers (drive-thru, e-commerce, etc.) and also to improve the performance of their commercial outlets.

The work presented here falls within the scope of behavioural marketing analysis by video, and aims to better understand the purchasing behaviour of customers by analyzing their motions in a densely-populated sales area covered by a single camera. Our industrial partner required us to use the camera network that was already in place and that was initially designed for video-surveillance. Using overhead cameras, we should determine the trajectory of customers in the retail space and manage the occlusions and/or interaction issues between people.

In the retail space, understand and analyze the evolution of the trajectory and the consumption of a customer has several goals. First, these data provide a concrete vision on surfaces to sell a brand to a particular location in the store. They also provide a number of operations as a tool of direct marketing such as the presentation of samples of new products or the development of end displays.

DOI : 10.5121/sipij.2015.6601

Analyze the trajectory of the customer has the advantage of better manage and allocate human resources, according to the measures of the frequentation at the rays at specific times and waiting times for customers at checkout. These analyse further review and reorganize the plan of the store and the areas allocated by department promoting impulse purchases.

These methods provide an effective way to calibrate the surface of a point of sale depending on the extent of his attendance and sales. Finally, it will improve the personalization of offers. This is possible due to the extraction and analysis of gaze orientation clients and analyzing their trajectories.

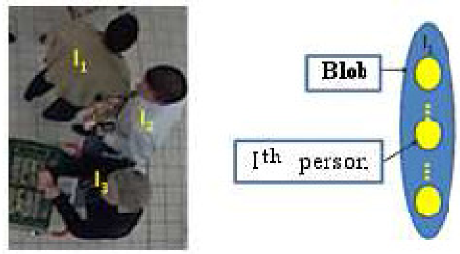


Figure 1. Notion of the blob.

Generally, in a sales area, a stream of customers is frequently seen on camera as a connected blob (Figure 1). This pixel-based entity can be further broken down inside analysis area and then generate other ones. Moreover, the interest objects that are initially separated may merge during motion and form then a new connected blob grouping them together. If a blob splits up inside the analysis zone, it would be necessary to save the information coming from these new blobs. In this case, it is essential to be able to predict the splitting of the blob at time ***t+1***. In our method, applying a mathematical morphology operator (erosion) on the blob at time ***t*** allows to make this prediction. In the event that the blob splits up, the set of new blobs will share the same spatio-temporal information from the original blob (path segment before separation, time and access point of the analysis zone, etc.). As for their merging, this is detected by calculating the overlapping of neighbouring blobs.

In the following sections, we discuss the latest developments in tracking field and present the architecture of our proposed solution with a description of each component. We conclude this study by analysis our results.

2. STATE-OF-THE-ART

The tracking of people is fundamental in all video-based behaviour analysis systems. In literature, we find several research studies that summarize the various methods for tracking objects. We can class them into two categories: deterministic and probabilistic approaches. With deterministic approaches, objects are tracked either by the association of observations, or by resolving an optimization problem. Gheissari [7]et al. split the silhouette into a set of homogenous regions characterized by points of interest. These points are then compared to determine if there are any matches between the two sets, the detection of a person is deduced based on their numbers. Thome et al.[11] propose a method more robust against occlusions that use a template to identify shapes, then by determining the maximum of the correlation function between the template of the person at time **t** and the person to locate in the image at time **t+1**. In the same category, we can mention the work of Aziz et al.[3]. Using a graphical model of the human body, they are able to identify people's heads and track them. Using this method, they are able to remedy the issue of occlusions. Urtasun and Fua[12] conducted a 3D tracking of the human body by shape recognition using volumetric primitives obtained from the principal component analysis (PCA). Using these primitives, the authors were able to describe the movements as being a linear combination of "distinctive motion vectors", and to estimate the posture of a person by minimizing a cost function of their 3D model and data provided by stereo cameras.

Approaches based on probabilistic methods have been used to compensate for problems with deterministic methods which there are confusion of subject and occlusions (partial, complete) issues for multiple objects tracking. Elgammal et al.[5] proposed a method for tracking people by modelling as random variables the values and location of feature element composing the interest region. The predicted location of the object is obtained by finding the geometric transformation between two consecutive images able to maximize a similarity function using the mean shift algorithm.

Perez et al.[6] and Stanley et al.[4] used a particle filter for tracking people. The accepted observation model was a histogram defined as the concatenation of the 2D – Hue-Saturation histogram and the brightness component V added when the chrominance component are too weak; while the latter used a spatiogram that included for each bin the average value and covariance of the pixel location contributing to each bin.

Ryoo et al.[8] put forward an approach for tracking people and objects in highly-occluded areas. They introduced a new paradigm for tracking of multiple hypotheses "observe and explain" as opposed to the old paradigm of "suppose and test". During tracking and at each multiple hypothesis, the system switches into "Observation" mode and waits until collecting enough data. This approach offers many possibilities for tracking by generating several probable "explanations" after having gathered a sufficient quantity of observations.

Works combining both tracking methods have also been proposed. Wang et al.[13] make this using the mean shift algorithm combined with a particle filter; the former is used to optimize the scale and position of each particle, while the latter due to its multi-hypothetical nature enables the mean shift to manage the redundancy of the particles. Andriluka et al.[2], for their part, used an articulated people detection model for tracking by combining a deterministic approach (tracking by detection) with a limb-detection method. This allowed them to build a robust and dynamic human model that can be extended for people detection in a more reliable way. The approximate posture of each person is detected according the local features that model the appearance of different human body parts. A previous knowledge of possible articulations and the temporal consistency within the walk cycle are modelled using a hierarchical Gaussian process with a latent variable model.

The methods for tracking people described above are designed to track a limited number of people, and moreover their effectiveness depends on the field of application. In real-life scenarios, difficulties may arise when a single object splits into two or more regions (or "blobs"). Likewise, the same applies when two moving objects merge into a single blob. The tracking algorithm must then be able to recognize this type of situation and react accordingly, with the ability to match a given number of objects with a given number of blobs. To address these issues, the probabilistic methods model this lack of precision as a noise. Unlike of the merging case, the various approaches used to track people can't predict a split-up situation. This case can be dealt by using a non-linear filter with a joint state space, but the temporal and computational complexity of these methods makes them less effective for real-time tracking applications.

In the scope of our project, we will develop a tracking application able to track objects from their entry up until their exit from a given analysis zone. In our case, the interest object is a blob that may include at least one people and the main limitation is to manage a variable number of these interest objects. Moreover, the algorithm must be able to store in memory the entry/exit points of each object notably when these blobs merge and/or split up (appearance and disappearance of objects of interest in the analysis zone).

3. MULTIPLE-TARGET TRACKING ARCHITECTURE

A large number of methods for tracking multiple targets have been proposed by the scientific community. Nonetheless, the problem of tracking a large number of people in heavily-populated environments remains difficult and has not been sufficiently addressed.

In order to resolve this problem, we have perfected an application for tracking trajectories that combines several algorithms. The main components are shown in Figure 2.

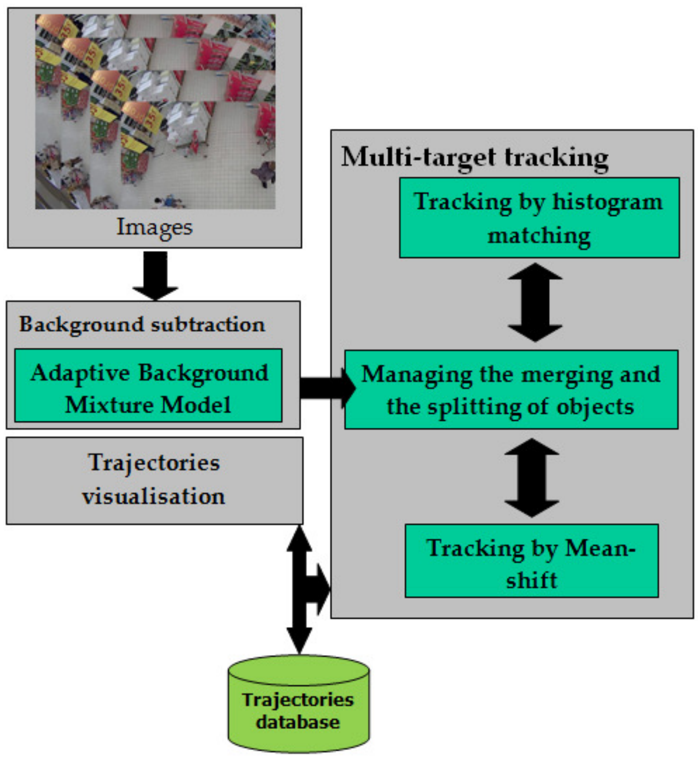


Figure 2. Multiple-target tracking architecture.

The aim of the background extraction module is detecting moving regions. These regions present the input into the multi-target tracking module that is the main module of our architecture. It consists for managing the merging/splitting of interest objects obtained from the background extraction component, and provides a hybrid tracking mode that combines two tracking methods.

The first tracking method consists to make matching between detected blobs in the simple cases (single-to-single) which is a quick and very efficient method. The second tracking method is based on the colour distribution to remedy the merging issue (single-to-multiple). In fact, once blobs have been merged, making just a simple matching of distributions is no longer conceivable.

3.1. Background extraction module

The background extraction module (Adaptive Background Mixture Model) identifies the moving regions in each image of the sequence. It computes a map of the distances between each of the pixels in the current image and a reference image that models the static parts of the image stream to be analyzed.

More specifically, we use the background subtraction technique proposed by Stauffer and Grimson[10] and modified by Shimada and Arita[9] where each pixel is modelled using a mixture of Gaussian distributions whose the number may vary throughout time.

3.2. Multi-target tracking methods

The used tracking method rests on the continuous updating of blob attributes during the tracking. It consists on the association of detected blobs until the latter be merged or splitted. In the case of splitting blob and in order to save a continuous trajectory for each new one, we simply have to identify the objects resulting from the splitted blob. When blobs are merged, we must also detect this occurs. This step is described in the next Section.

To be able to track objects in real-time, we combined two tracking algorithms. The first one is a histogram matching algorithm which is used for the isolated interest objects. It based on the measure of matching pairs. For this step, we use a Bhattacharya metric for comparing the HSV colour histograms calculated on N bins (see Eq.1)

$$\mathrm{d}(\mathrm{H}_{\mathrm{obj1}},\mathrm{H}_{\mathrm{obj2}}) = \sqrt{1 - \sum_{\mathrm{i}=0}^{\mathrm{N}-1} \sqrt{\mathrm{H}^{\mathrm{i}}_{\mathrm{obj1}} . \mathrm{H}^{\mathrm{i}}_{\mathrm{obj2}}}} \qquad 1$$

The second method tracks colour distributions that used for interest objects after their merging. Concretely, we apply the mean shift method. Then, the search for the target at the current time is based on colour distributions (histograms) in a simple geometric area. More specifically, starting from the image of the object at time **t**, we try to identify the area having the same colour as the initial object using a rectangular area that selected from an image at time ***t+1***. The estimation of the new distribution having the same size and position of the target in the image at time **t**, consists to use an analysis model. This model moved around the research area until we find the area that matches' the best. We perform each time the following steps:

1. Calculate the barycentre of each pixel ***(x,y)*** in the window, as well as their sum (zero-order moment).

$$M_{00} = \sum_{x,y} I(x,y), M_{10} = \sum_{x,y} xI(x,y), M_{01} = \sum_{x,y} yI(x,y) \Rightarrow x_c = \frac{M_{10}}{M_{00}}, y_c = \frac{M_{01}}{M_{00}}$$

2. Adjust the rectangle encompassing the target on the barycentre coordinates (x_c, y_c).

3.3. Trajectory database

The trajectory database contains the trajectories calculated during the tracking process. Each trajectory is defined by 4 data items: entry points, exit points, time of entry, and time of exit. Based on this data, the some measurements can be extracted such as how often customers move in each direction, density of people, people waiting time, etc.

4. MANAGEMENT OF CHANGING OBJECTS

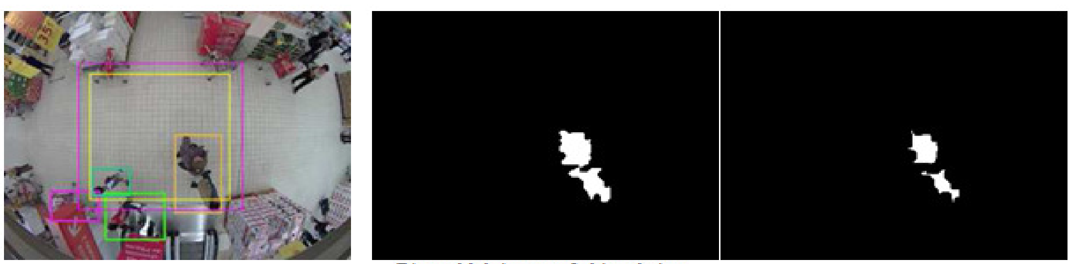
In order to resolve the splitting/merging issues, we use two different methods.

4.1. Splitting objects

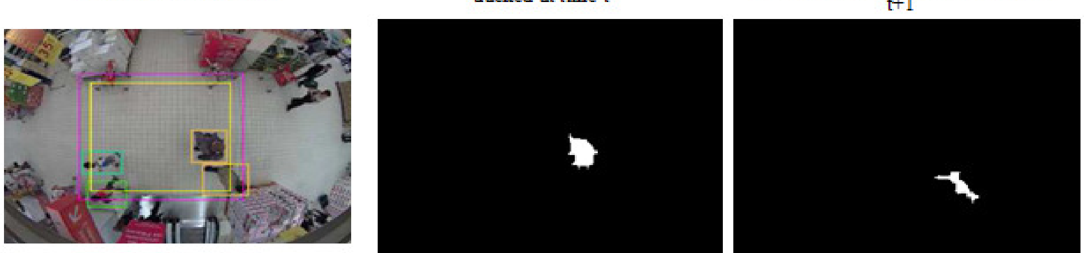
The separation of an object (or blob) at time ***t*** into a set of related objects (or blobs) at time ***t+1*** can be predicted by applying an erosion function to the object at time ***t***. Figure 3 illustrates the utility of erosion for separating touching objects. Figures 3-a and 3-b and illustrates how erosion could predict separation into blob. The iteration number of the erosion operator is depending on the rate frame the image stream acquisition. In our case, we set the number of iterations to 2 with a rate frame equal to 25fps. In this way, the eroded blob enables to distinguish the following cases:

1. If the set of child blobs, resulting from the application of the erosion operator on the parent blob, is included in the set of candidate blobs that detected in the image at time ***t+1***, then each blob in this set will inherit the same information from the parent blob at time ***t*** (access point, time of entry, path segment), and then the parent blob will be deleted. The matching between the predicted blobs and detected blobs is made by the histogram comparison algorithm.
2. Otherwise, the set of child blobs resulting from the erosion of the parent blob will be temporarily saved in the list of interest objects.

Figure 3. Illustration for separating touching objects in binary image using erosion operator.



Blobs tracked at time t | Binary blob image of object being tracked at time t | Prediction of blob separation at time t+1



Two new blobs after separation in the image $t+1$, they have the same colour rectangle as the one encompassing the blob at time t | Binary image of the first new blob at time t+1 | Binary image of the second new blob at time t+1

Figure 4. Prediction of the separation of objects at time t+1 using the “erosion” operator.

4.2. Merging objects

The merging situation is produced when there are junctions or overlapping the paths of the interest objects. Consequently, these last situations results a new connected blob covering those objects. Tracking these objects using just a simple histogram matching method is practically impossible. In this case, the objects are tracked using a colour distribution estimation method that applies a mean shift algorithm to resolve the merging issue. However, merging objects must first be detected before applying this type of tracking method. The merging of two interest objects is detected by calculating the overlap using the following formula:

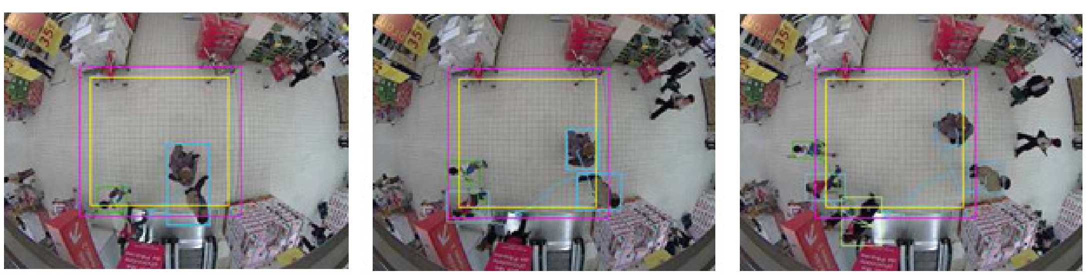
$$Overlap(Obj_i^t, Obj_j^t) = \begin{cases} 1, \dfrac{surface(Obj_i^t \cap Obj_j^t)}{\min(surface(Obj_i^t), surface(Obj_j^t))} > S_r \\ 0, sinon \end{cases}$$

where S_r is the threshold of the overlap.

5. RESULTS

Our method is based on two solutions: the first, by choosing an appropriate tracking algorithm able to reduce the computing time as much as possible. The second lets us efficiently process the tracking of blobs when they merge or split up especially in what concerns the storage of their trajectory histories.

The tracking process (Figure 5) is started as soon as the rectangle covering the detected blob intersects with the analysis area (yellow rectangle). In Figure 5-a, the detected blob (blue rectangle) at the entrance of the analysis zone is composed of two people. Then, these people are going into two different directions which will lead to their separation and creating of two new blobs containing each one a single people. After having predicted their separation, the two new targets will still keep the same colour before their separation. Figure 5-b illustrates how people can be tracked when they merge and how it is easy to track them in this type of situation with the mean shift algorithm. We note that the targets outside the purple square will be discarded by the system.



(a) Illustration of two people separating.

(b) Illustration of three people merging.

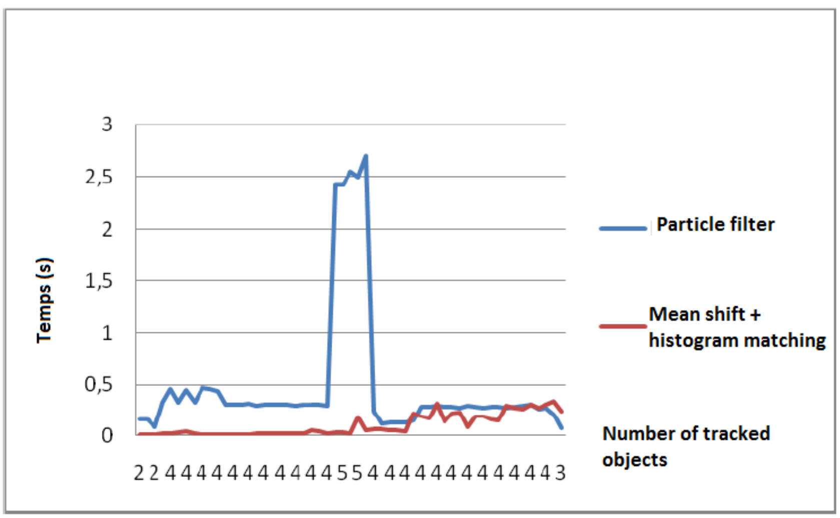
Figure. 5: Example of multi-target tracking with overhead camera.

After having compared our algorithm with non-linear methods, we can confirm that these methods are inefficient because they assume that a blob corresponds to only one state space (Figure 6).

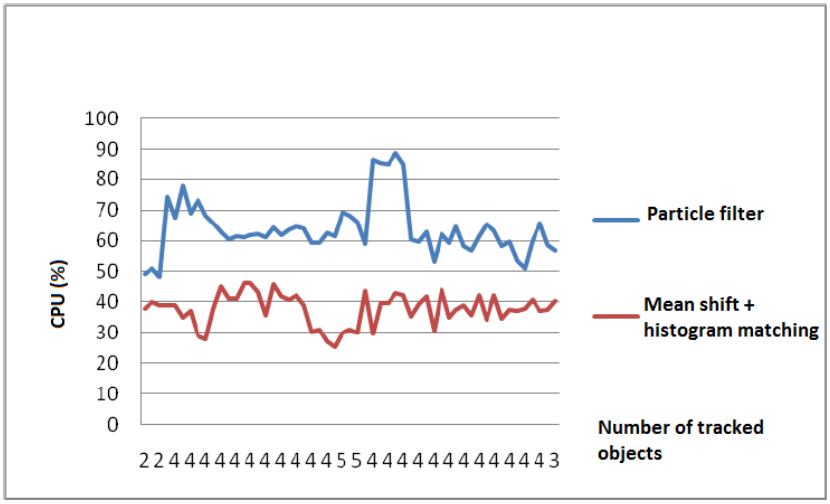
Figure. 6: Limitation of the non-linear tracker.

A blob is considered as a single object, thus the instability of the tracking in the event that it splits up. Another weakness of using these methods is that the presence of large number of objects in the analysis area may considerably slow down the algorithm's computation time. In fact, the function of the cost of object matching with another depends on the number of hypotheses, the geometry of the bounding boxes, their size and the number of the bins used for calculating their colour histograms.

In terms of performance, we compared our tracking method with many particle filters tracking launched together. Objects were tracked with the particle filter with histograms coded on 110 bins and using 100 particles.



(a) Graph of the CPU time spent by the particle filter and our proposed method, to track multiple targets. This time is proportional to the number of targets to track and their size.



(b) Graph of CPU resources used when tracking multiple-targets using the particle filter and our proposed method.

Figure 7. CPU resources used by various tracking methods.

Figures 7-a and 7-b compare the cost of using parallel particle filters with our method. For tracking using the mean shift algorithm and the comparison of histograms, we used 3D histograms coded on 10 bins. Tracking by particle filter slows down when the size of the objects to track is relatively large (blue peak). Increasing the number of objects also slows down the tracking process. This increases the CPU resources consumption due to the calculation of the comparisons of the histograms corresponding to the observations (particles).

Our method presents some limitations when using side-view cameras. In this configuration, the problem with tracking people becomes particularly difficult due to complete occlusions. As a whole, the objects are correctly tracked; although sometimes the model drops a target only to pick it up again later in the following sequences.

Table 1. People counting results from commercial deployments.

People counting results from commercial deployments.			
Camera ID	Ground truth count	Nb People	Accuracy
1	30	29	96.6
2	73	69	94.5
3	147	136	92.5
4	91	85	93.4
5	78	73	93.6
6	206	188	91.3
7	324	308	95.1
8	145	127	87.6
9	709	665	93.8

Table 1 shows people counting results from a number of customer. The camera views are illustrated in Figure 6. Our ground truth contents 24 hours of video from these scenarios and measured counting accuracy by comparing the manual ground truth count with that reported by our system. In most deployments the counting accuracy of the system was around **93%**. The primary reason for undercounting was occlusions due to large crowds. Over-counting was caused mostly by non-human targets like chariot detected as humans.

5. CONCLUSION

This system proposed for people tracking in a shopping centre provides an original solution to the problem thanks to its hybrid method that is able to correctly manage the merging and splitting up of targets. We tested our algorithm in real-life situations and the results that we obtained allowed us to evaluate the ability of our system to track a large number of targets in a small area. However, we also observed that when a target is partially hidden, our algorithm quickly diverges and can't insure a continues tracking.

REFERENCES

[1] Bhattacharyya, A. On a measure of divergence between two statistical populations defined by their probability distributions. In Bulletin of the Calcutta Mathematical Society. 1943

[2] Mykhaylo Andriluka, Stefan Roth, and Bernt Schiele. People-tracking-by-detection and people-detection-by-tracking. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR08), 2008.

[3] Kheir-Eddine Aziz, Djamel Merad, Nicolas Thome, and Bernard Fertil. Pedestrian head detection and tracking using graph skeleton for people counting in crowded environments. In Proceedings of the 12th IAPR Conference on Machine Vision Applications (MVA), pages 252–256. IAPR, June 2011.

[4] Stanley T. Birchfield and Sriram Rangarajan. Spatial histograms for region-based tracking. Electronics and Telecommunications Research Institute (ETRI) Journal, 29(5):697–699, october 2007.

[5] Ahmed Elgammal, Ramani Duraiswami, and Larry S. Davis. Probabilistic tracking in joint feature-spatial spaces. In Proceedings of the 2003 IEEE computer society conference on Computer vision and pattern recognition, CVPR'03, pages 781–788, Washington, DC, USA, 2003. IEEE Computer Society.

[6] Daniel Gatica-Perez, Jean-Marc Odobez, Sileye O. Ba, Kevin C. Smith, and Guillaume Lathoud. Tracking people in meetings with particles. In Proc. Int. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), invited paper, June 2005.

[7] Niloofar Gheissari, Thomas B. Sebastian, and Richard Hartley. Person reidentification using spatiotemporal appearance. In CVPR, CVPR'06, pages 1528–1535, Washington, DC, USA, 2006. IEEE Computer Society.

[8] Michael S. Ryoo and Jake K. Aggarwal. Observe-and-explain: A new approach for multiple hypotheses tracking of humans and objects. In CVPR, 2008.

[9] Atsushi Shimada, Daisaku Arita, and Rin-ichiro Taniguchi. Dynamic control of adaptive mixture-of-gaussians background model. In Proceedings of the IEEE International Conference on Video and Signal Based Surveillance, AVSS '06, pages 5–13, Washington, DC, USA, 2006. IEEE Computer Society.

[10] Chris Stauffer and W. Eric L. Grimson. Learning patterns of activity using real-time tracking. IEEE Trans. Pattern Anal. Mach. Intell., 22:747–757, August 2000.

[11] Nicolas Thome, Djamel Merad, and Serge Miguet. Learning articulated appearance models for tracking humans: A spectral graph matching approach. Image Commun., 23:769–787, November 2008.

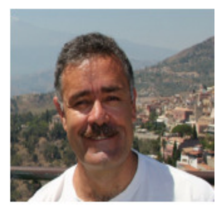
[12] Raquel Urtasun and Pascal Fua. 3d Human Body Tracking using Deterministic Temporal Motion Models. Technical report, 2004.

[13] Zhaowen Wang, Xiaokang Yang, Yi Xu, and Songyu Yu. Camshift guided particle filter for visual tracking. Pattern Recogn. Lett., 30:407–413, March 2009.

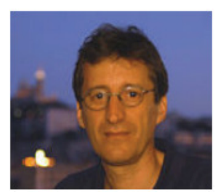
AUTHORS

Kheireddine AZIZ is a research engineer at SeaTech Toulon since 2014. He received his research master degree image processing from Caen Basse Normandie University in 2006. He received his professional master degree in database and image processing from Caen Basse Normandie University in 2007. After Ph.D degree in computer science from Aix-Marseille University and Cliris group company at 2012 on Multiperson tracking in constrained environment, he worked at IMRAEUROPE specialized in intelligent automotive application and Rn3d Lab.

Djamal MERAD After an initial training of engineer, supplemented by a DEA in Image and vision at the University of Nice, Djamal Merad concluded its academic formation by a doctorate in Robotics at the university from Evry. In September 2008, it integrates the team Images and Models of LSIS lab. as a university lecturer. Since, his research activity is focused on the pattern recognition field and he has participated in several research projects.



Jean-Luc Damoiseaux is associate professor Aix-Marseille University After a DEA in Artifical Intelligence at the University of the Mediterranean of Marseille, Jean-Luc Damoiseaux concluded its academic formation by a doctorate in pattern recognition at the University of Avignon and Pays de Vaucluse. Since 2002, he's working at the department of Networks and Telecommunications of the Aix-Marseille University Institute of Technology; currently is the director of this department.



Pierre Drap is researcher in CNRS, France, since 1997. In 2008 he joins the Image and Model team in the Since 2004, he was involved as partner or leader in more than 10 national or international projects dealing with these topics.