

State-Driven Particle Filter for Multi-Person Tracking

David Gerónimo¹, Frédéric Lerasle^{2,3}, and Antonio M. López¹

¹ Computer Vision Center and Department of Computer Science
Edifici O, 08193 Campus Universitat Autònoma de Barcelona, Bellaterra, Spain.

² CNRS-LAAS, 7 avenue du Colonel Roche, F-31077 Toulouse, France

³ Université de Toulouse (UPS), F-31077 Toulouse, France

{dgeronimo,antonio}@cvc.uab.es, lerasle@laas.fr

Abstract. Multi-person tracking can be exploited in applications such as driver assistance, surveillance, multimedia and human-robot interaction. With the help of human detectors, particle filters offer a robust method able to filter noisy detections and provide temporal coherence. However, some traditional problems such as occlusions with other targets or the scene, temporal drifting or even the lost targets detection are rarely considered, making the systems performance decrease. Some authors propose to overcome these problems using heuristics not explained and formalized in the papers, for instance by defining exceptions to the model updating depending on tracks overlapping. In this paper we propose to formalize these events by the use of a state-graph, defining the current state of the track (e.g., *potential*, *tracked*, *occluded* or *lost*) and the transitions between states in an explicit way. This approach has the advantage of linking track actions such as the online underlying models updating, which gives flexibility to the system. It provides an explicit representation to adapt the multiple parallel trackers depending on the context, i.e., each track can make use of a specific filtering strategy, dynamic model, number of particles, etc. depending on its state. We implement this technique in a single-camera multi-person tracker and test it in public video sequences.

1 Introduction

Human detection and tracking has been one of the most relevant research topics in computer vision for almost two decades. Nowadays, it is still an important subject of investigation due to the difficulty to develop techniques capable of reliably performing this task in many contexts. Humans are one of the most challenging classes in computer vision: they are dynamic, deformable, unpredictable, their variability in size and clothings is big and they can be affected by changing illumination. Furthermore, they can often be occluded, appear in groups or isolated. Some of the applications of human detection and tracking are surveillance [1], advanced driver assistance systems [2], human-robot interaction [3], etc. All these applications require multi-target detection and tracking of people from single camera, which is the focus of this paper.

The detection is defined as the process of providing information about objects in images in a frame-by-frame basis, hence the task of a detector is mainly restricted to just localizing the objects. The tracking identifies and follows these objects through a sequence of images, enlarging the tasks to the initialization of tracks, updating their state, managing the occlusions, reinitializing the tracks after occlusions, data association, etc. During the last decades, Particle Filters [4, 5] have become popular in the visual tracking literature thanks to their capability to provide a simple while robust framework to tackle some tracking process. Nowadays, the most successful approaches make use of the so-called tracking-by-detection approach, which takes advantage of the output of an object detector to (re)initialize and provide evidences through the frames [6, 7].

In spite of the plethora of tracking strategies, which are more or less efficient depending on the context, few papers address traditional difficulties such as occlusions with other targets or with the scene, temporal drifts of the tracks, lost tracks, etc., which can lead to typical tracking problems like target hijacking or track model drifting. These problems not only decrease the system performance but also result in misinterpretations of the scene. The traditional approach is to manage occlusions or lost targets by heuristics not detailed in the papers. For instance, in the well-known approach in [6], the authors just update the classifiers when tracks are not overlapped and terminate tracks with no associated detections during several frames. In this case, an optimization algorithm that makes use of an online classifier, a gait function and the particles position, is used to match the current detections and tracks at each frame without explicitly reasoning about the state of the targets. One of the approaches adapting the tracker strategies to occlusions is presented in [8], in which the authors propose different trackers for a multi-view object tracking system. Depending on whether the targets are viewed from different cameras or not and on whether they are isolated or not, the algorithms can be: Interactively Distributed Multi-Object Tracking, Bayesian Multiple Camera Tracking or Multiple Independent Particle Filter. However, apart from being addressed specifically to multi-camera systems, it does only exploit geometrical cues such as isolation and camera projection, not taking advantage of the possible cues that the target model can provide in a single camera. In [9], Zhou et al. propose an event analysis stage that identifies presence of targets in the scene and inter-object occlusion. In this case, contrary to our proposal, this reasoning is made as a posterior process to the filtering to provide high-level event information.

Based on a Particle Filter tracker, we propose a novel approach for multi-person tracking that defines a state for each track (*potential*, *tracked*, *occluded* or *lost*), represented in a graph. The nodes of a graph are used to depict the states (context), which define a specific tracking strategy, while the arcs represent the conditions that trigger the state transitions. As conditions we employ a traditional bounding box overlapping approach to detect non-isolated tracks, together with a Kalman Filter and Generalized Likelihood Ratio Test [10] to detect tracker failures. The benefits of this state-graph compared to the conventional implementation are clear. It provides an explicit representation to adapt

the multiple parallel trackers to the context, i.e., each track could make use of a specific filtering strategy, dynamic model, number of particles, sampling strategy, etc. depending on its state. For example, it allows to use different simultaneous filtering approaches like Condensation and ICondensation in the same scene, or to restrict the update of the track model to non-occluded targets, which amends the problem of model drifting but it is still more flexible than fixing a constant model. Furthermore, recent approaches dealing with occlusions [11] or tracker failures [12] can be embedded in the framework. We test the tracker combined with a HOG-SVM detector [13] in public sequences with multiple persons in real situations.

The structure of the manuscript is as follows. Section 2 recalls the conventional particle filtering algorithm. Section 3 presents our state-based approach. We first define the states of the graph, with the associated actions and adapted filtering strategy, and then describe the transition conditions. The experimental results are presented in Section 4. Finally, the conclusions and perspectives are summarized in Section 5.

2 Original Particle Filtering Formulation

In this section we first recall the main concepts and components of a conventional ICondensation particle filter [14]. Let us define a set of N particles representing hypotheses of the position and scale of a person (track). Each particle has an associated state vector $\mathbf{x}_k^{(i)} = \langle x, y, s, w \rangle$, where $i \in 1, \dots, N$ is the particle index, x, y, s are image coordinates and scale of the track bounding box at frame k , and w its weight. When a track is initialized at $k = 0$, we set $w_k^{(i)} = 1/N$ and draw particles according to a prior $p_0(\mathbf{x})$. Next, at each frame a process of prediction and correction is carried out following the next steps:

1. Prediction (Importance function), which propagates particles from $k-1$ to k according to dynamics from a predefined probability density function (*pdf*) p and/or measurements z_k (from current image), defined as:

$$q(\mathbf{x}_k^{(i)} | \mathbf{x}_{k-1}^{(i)}, z_k) = \alpha \pi(\mathbf{x}_k^{(i)} | z_k) + \beta p(\mathbf{x}_k^{(i)} | \mathbf{x}_{k-1}^{(i)}) + (1 - \alpha - \beta) p_0(\mathbf{x}_k^{(i)}) , \quad (1)$$

where $\pi(\mathbf{x}_k^{(i)} | z_k) \in [0..1]$ corresponds to an intermittent image cue (e.g., a human detector [13]), $p(\mathbf{x}_k^{(i)} | \mathbf{x}_{k-1}^{(i)})$ is the dynamics *pdf* and p_0 is the prior. $\alpha, \beta \in [0..1]$ represent the proportion of particles according to observations and dynamics, respectively.

2. Correction (Weighting), which assigns a weight $w_k^{(i)}$ to each particle according to:

$$w_k^{(i)} \propto w_{k-1}^{(i)} \frac{p(z_k | \mathbf{x}_k^{(i)}) p(\mathbf{x}_k^{(i)} | \mathbf{x}_{k-1}^{(i)})}{q(\mathbf{x}_k^{(i)} | \mathbf{x}_{k-1}^{(i)}, z_k)} , \quad (2)$$

where $p(z_k | \mathbf{x}_k^{(i)})$ is the likelihood of the measurement z_k with respect to the particle state $\mathbf{x}_k^{(i)}$.

3. Resampling, in which the particles are redrawn according to $w_k^{(i)}$.

3 State-Driven Particle Filtering

In any Particle Filtering approach, the propagation and weighting is made independently of the state of the track, e.g., the particles in an occluded track are propagated according to the same equation as in a non-occluded one. This can lead to an inefficient particles distribution and even a target loss if the target does not re-appear as the dynamics predict, which often derive in the aforementioned problems enumerated in Section 1. We propose a three-noded graph to maintain the state of the tracks, with an additional one for the detections that can potentially become tracks, that modulates the strategy and parameters of the filtering.

Figure 1 illustrates the model. Each node defines a track state (*potential*, *tracked*, *occluded* and *lost*) while each arc defines the conditions to move from one state to another (i.e., tracks overlapping or confidence).

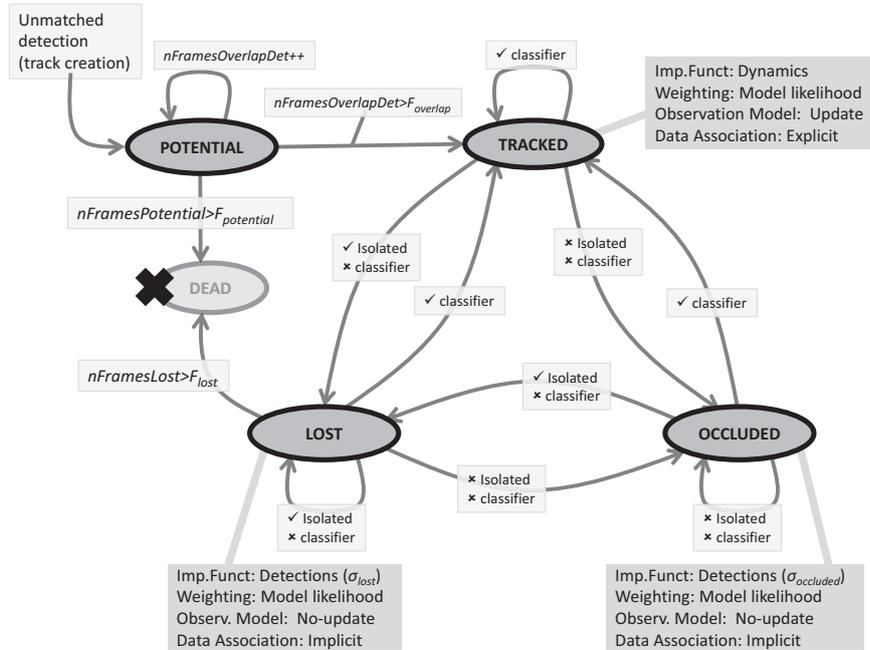


Fig. 1. Proposed state-graph.

In the following subsections we detail the state-actions and the state-transitions proposed in our graph.

3.1 State-actions (nodes)

The track state is represented as a node, which defines the corresponding actions on the tracking process of the specific target. For example, with this approach a *tracked* track can make use only of dynamics in the importance function without taking into account the detections (setting $\alpha = 0, \beta = 1$ in Equation (1) as in Condensation [4]), while an *occluded* track can depend both on detections and dynamics (i.e., $\alpha, \beta \neq 0$ in Equation (1) as in ICondensation [5]). In terms of implementation this can be represented by a weighted importance function that uses different sampling methods, which are adjusted for each state. As can be seen, the state-graph provides an explicit representation capable of working with two different trackers in the same image depending on the context. Figure 2 illustrates the different tracker strategies, which depend on the state of the track.

Potential Track. A *potential* track is created for each detection that does not match any of the current tracks. The matching is made if the overlapping between two detections in two consecutive frames exceeds a given threshold, and a counter is used to store the number of frames a track has been in this state. After $F_{overlap}$ frames in which the *potential* track matches a detection, it is upgraded to *tracked*. On the contrary, if the track stands for $F_{potential}$ frames as *potential* without enough matching detections, it is killed. Of course, this procedure is suited for the standard tracking test videos focused on surveillance or driver assistance. For other kind of applications, this procedure must be adapted to the specific video frame rate and the object detector performance of each case.

Tracked Track. Once a track has earned the state *tracked*, it can not return to the *potential* one, so from this frame a track model is used (initialized at this frame) and a complete particle filter process is carried out. The filtering process for this state is as follows:

- **Importance Function:** It is completely based on dynamics. In our proposal, we set $\alpha = 0$ and

$$p(\mathbf{x}_k^{(i)} | \mathbf{x}_{k-1}^{(i)}) = R(\mathbf{x}_{k-1}^{(i)}, \Sigma_{dyn}) \quad (3)$$

is used, where R is a Gaussian random walk function with covariance matrix Σ_{dyn} centered at $\mathbf{x}_{k-1}^{(i)}$.

- **Weighting Function:** The particles are weighted according to $p(z_k | \mathbf{x}_k^{(i)})$, i.e., the likelihood of the track model applied to their current position and scale in the image.
- **Data Association:** Explicit and carried out after the importance function and weighting. It matches the current detections with the *tracked* tracks, aimed at both discarding these detections to be used by *potential*, *occluded* or *lost* tracks and to compute the scale of the track as the average of the last

three detections. We construct a distance matrix with each tracked track t and each detection d using the equation:

$$Cost(t, d) = p(d|t) \cdot \mathcal{N} \left(\frac{1}{N} \sum_{i=0}^N dist(\mathbf{x}_k^{(i)}, d); 0, \sigma_{dist}^2 \right), \quad (4)$$

where $p(d|t)$ is the matching probability of the track t model applied to d (see Sect. 4); $dist(\mathbf{x}_k^{(i)}, d)$ is the Euclidean distance between a particle $\mathbf{x}_k^{(i)}$ from t and the center of the detection d ; and \mathcal{N} is a Gaussian function with mean 0 and variance σ_{dist}^2 . This cost function rewards detections that are similar to the track model and near the track particles. Finally, an optimization algorithm like Munkres or a greedy approach [6] is used to compute the best matching.

The data association is applied according to a state-based preference, i.e., first *tracked*, then *occluded* and finally *lost* tracks (as will be seen, their association is implicit). This approach is based on the assumption that *tracked* tracks are independent from detections, so the process is focused on explaining the scene status as better as possible with the last known data. Then, the algorithm tries to explain the unmatched detections with the *occluded* or *lost* tracks, whose confidence is lower than the *tracked* ones.

Occluded Track. An occluded track is a target which is significantly occluded by another one and whose track confidence decreases (these concepts will be detailed later in Sect. 3.2). In this case, the filtering is as follows:

- **Importance Function:** We draw particles in all the near unmatched detections weighted by their distance to the track ($\beta = 0$ in Equation (1)):

$$q(\mathbf{x}_k^{(i)} | z_k) = \pi(\mathbf{x}_k^{(i)} | z_k) \quad (5)$$

If there are not unmatched detections in the current frame, the particles are not sampled and the state is not changed.

- **Weighting Function:** The same equation as in *tracked* is used.
- **Data Association:** In this case, the association is implicit, i.e., no optimization is made. We assume that each *occluded* track provides the best possible matching with a detection, and we just make the matches track-detection according to the highest weighted particle. If a detection is matched by two occluded tracks, just the one with highest weight is taken while the other is left as unmatched.

Lost Track. A *lost* track is often the result of an occlusion with the scene or a sudden change in its appearance, which prevents the model to be updated on time, which make the maximum likelihood decrease suddenly. In this case, as explained later, the track is isolated from other tracks in the scene.

As importance function in this case we use the same as in *occluded*, however, the Gaussian used to draw the particles in the nearby detections is bigger since

we understand that an occlusion is local while a *lost* track can be found anywhere in the image if the sufficient number of frames has passed (e.g., could be occluded by a vehicle or street furniture) and appear in the other side of the image. In the case of weighting function and data association, we use the same as in the *occluded* state.

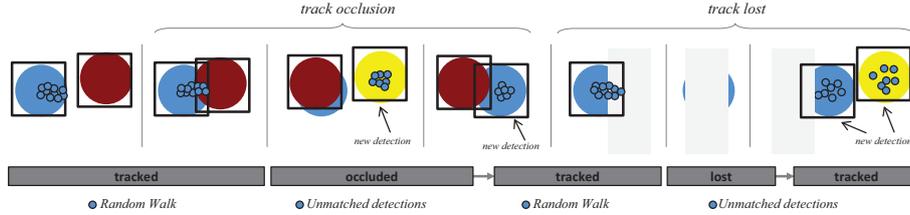


Fig. 2. Example of a track (blue) through the different states. Note: black boxes represent detections, blue, red and yellow circles represent objects. The figure can be seen as evolving from left to right.

3.2 State-transitions (arcs)

In order to go from one state to another, state-transitions depend on conditions of the tracks. In our proposal, the conditions are based on the classifier confidence output of each track and on the overlapping between tracks.

Classifier Matching. The classifier confidence of each track is computed as the maximum likelihood of its particles $C(\mathbf{x}_k | z_k) = \max(w_k^{(i)}) \forall i \in 0, \dots, N$. In order to detect track model failures, we propose to use a 1D Kalman Filter and the Generalized Likelihood Ratio Test (GLRT) to filter the confidence of an online classifier (Sect. 4). The GLRT is a statistical test used to compare two models that extends the Neyman-Pearson lemma [15] for hypotheses testing in a sub-optimal way. It involves the computation of the ratio $\frac{l_{H_1}}{l_{H_0}}$, where l_{H_0} is the log-likelihood of a null hypothesis and l_{H_1} is the log-likelihood of an alternative one. Whenever this ratio is greater than a threshold defined *a priori*, the test fails. In [10], Willsky and Jones proposed to filter the residual of a KF, i.e., the difference between actual measure and its prediction, with the GLRT to determine if a change in a signal has occurred.

In our proposal we set H_0 as no confidence change and H_1 as a confidence decrease. In Fig. 3 we illustrate the behaviour of the confidences in two tracks, one occluded and another non-occluded. As can be seen, depending on the state of the Kalman Filter, the GLRT is able to detect if a jump has been produced in the confidence signal.

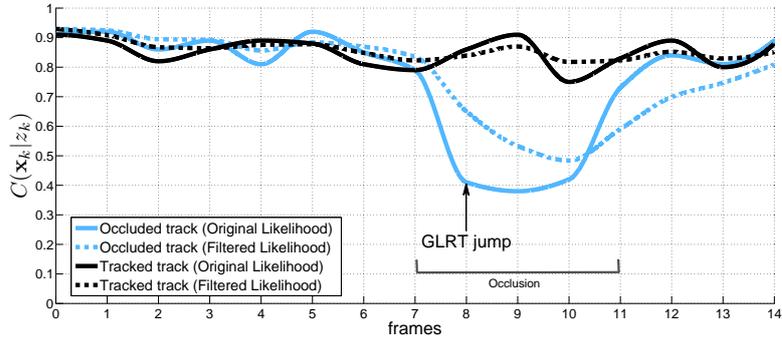


Fig. 3. In black, the confidence of a color model in a non-occluded person along 15 frames. In blue, the confidence of an occluded person. At frame 8, the GLRT successfully detects the decrease in the confidence.

Isolated Track. The isolated condition is computed by using the Jaccard overlapping criterion between two tracks:

$$J(t_0, t_1) = \frac{|t_0 \cap t_1|}{|t_0 \cup t_1|} > J_T \quad (6)$$

where $J(t_0, t_1) \in [0, 1]$, t_0 and t_1 are the areas of two given tracks and J_T is a fixed overlapping threshold (e.g., 0.5).

4 Experimental Results

In this section we evaluate the performance of the proposed approach in comparison with the original not state-based version. The public datasets used are TUD-Crossing, TUD-Campus⁴ [16], with 201 and 71 frames, respectively; and PETS2009-S2L1 View1⁵ with 795 frames. Video results can be found at <http://www.cvc.uab.es/~dgeronimo/projects/ACIVS12>.

The detector used as input is the well-known HOG/LinearSVM person detector by Dalal and Triggs [13], which is a current standard in the field. The particle filter uses 150 particles in the TUD sequences and 300 the PETS, all in a distributed filtering fashion. The online model of the track is based on a rgI color histogram of the torso region quantized into 16 bins per channel (in our experiments rgI works slightly better than RGB color space). The likelihood is based on the histograms difference, assumed to be normally distributed. As evaluation metrics, we make use of CLEARMOT [18], which defines an evalua-

⁴ <http://www.mis.tu-darmstadt.de/node/382>

⁵ <http://www.cvg.rdg.ac.uk/PETS2009/> (groundtruth annotations from [17])

tion protocol specific for multi-object tracking in terms of precision (MOTP⁶) and accuracy (MOTA⁷). The matching between detections and annotations is defined as in Eq. (6) with $J_T > 0.5$. For the sake of completeness, in addition to the CLEARMOT metrics, we also specify the number of total identity switches a long a full track (not only to the consecutive ones in the original formulation), the false negative rate FNR and the false positives per image $FPPI$.

Table 4 summarizes the performance of the system in the video sequences. As can be seen, the precision of both approaches is similar given that always a target is detected, it is based on the same detector and the likelihood model is the same in both cases. On the contrary, in the case of the accuracy, the state-driven approach improves the results AROUND A 7% in all cases. The FNR improvement is around 3 – 5% depending on the sequence; and the FPPI significantly improves (11 – 28%), this latter thanks to the reduction in the number of track drifts. In the case of ID switches, the improvement can be clearly appreciated in TUD-Crossing, in which there are many pedestrians and occlusions along the sequence. The improvement in FPPI is also significant in all the sequences.

Dataset	MOTP	MOTA	FNR	FPPI	IDS_w
TUD Crossing (Original)	77.8%	51.1%	36.7%	0.61	0 (12)
TUD Crossing (State-based)	77.7%	58.8%	33.7%	0.37	0 (5)
TUD Campus (Original)	78.6%	27.1%	64.9%	0.34	0 (1)
TUD Campus (State-based)	76.0%	34.8%	59.9%	0.23	0 (0)
PETS2009-S2L1 (Original)	73.9%	43.5%	48.0%	0.49	2 (40)
PETS2009-S2L1 (State-based)	75.0%	51.1%	45.2%	0.21	0 (27)

Table 1. Performance comparison between the traditional approach (Original) and our proposal (State-based). In parentheses, global ID switches for full tracks.

Fig. 4(a) illustrates the evolution of a track from *potential* to *tracked* (the brighter the particle, the higher its weight). Blue bounding boxes correspond to tracks (with their associated numeric id at the corner) and yellow boxes correspond to detections. Fig. 4(b) shows the algorithm behaviour after an inter-person occlusion, in which the foreground person is labeled as *tracked* while the person in the background is labeled as *occluded*. Fig. 4(c) shows an occlusion with an object of the scene.

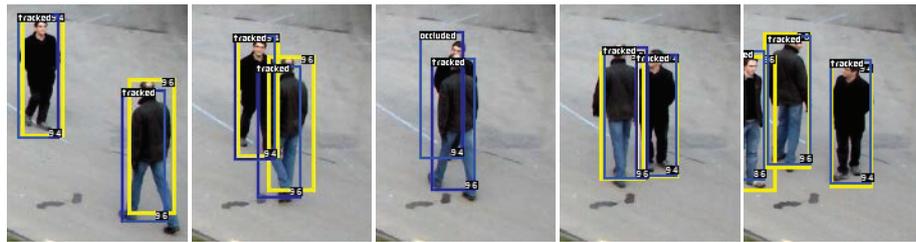
Fig. 5 shows a detailed reinitialization process after an occlusion with an undetected person. During the occlusion, the algorithm unsuccessfully tries to

⁶ $MOTP = \frac{\sum_{i,t} dst_t^i}{\sum_t c_t}$, where dst_t^i is the Euclidean distance between the annotation i and the corresponding detection and c_t is the number of matches, all in frame t .

⁷ $MOTA = 1 - \frac{\sum_t m_t - fp_t - mme_t}{\sum_t g}$, where m_t are object misses, fp_t are false positives, mme_t are mismatches and g_t are groundtruth objects at frame t .



(a) *potential* and *tracked* track



(b) *occluded* track



(b) *lost* track

Fig. 4. Examples of the tracker behaviour (a) in TUD-Crossing and (b,c) in PETS2009. Track particles in sequence (b) are not shown for visualization clarity. Note: detections in yellow, tracks in blue, potential track in dashed blue.

match the *lost* track with a near unmatched detection. After the occlusion, a new detection matches the online model and the track is reinitialized maintaining the same id. Without the proposed approach, a new track (with a new id) is created for the new detection resulting in an identity switch (counted in our full track switch criterion noted inside the parentheses in Table 4). These experiments demonstrate that even with simple components our proposal already provides improvement.

The errors in the tracking are mainly a result of the simple online track model used (based on the color histogram of the torso). For example, when persons are very similar in clothes and shapes, *lost* or *occluded* tracks can be misassigned after reinitialization. In addition, when the tracks change in scale, particles at different scales but centered in the torso region are equally weighted, leading

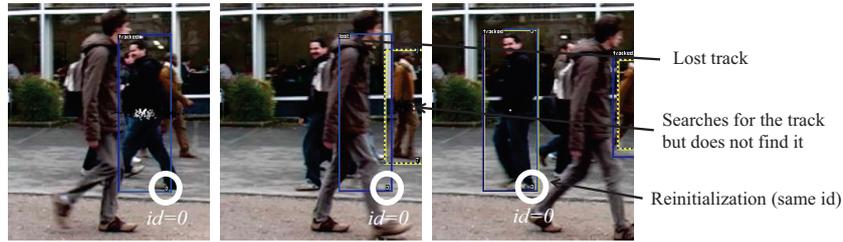


Fig. 5. Detailed reinitialization process of a track after an occlusion with an undetected person.

to a wrong track scale estimation. These problems could be amended by a more complex observation model exploiting shape in addition to color [6]. Other improvements such as a more elaborated occlusion manager [17, 11] or track failure detector [12] are likely improve the overall performance.

Finally, it is worth to mention that our proposal is not limited to distributed particle filters. In fact, it can be potentially applied in a centralized filtering approach by including the track state and the potential matched detections in the state vector, and in other filters such as Kalman or Mean-shift by adapting the actions and transitions of the state-graph.

5 Conclusion

In this paper we have presented a novel approach based on a state-graph able to deal with some problems of tracking like hijacking, occlusions or drifting. The main contribution is the formalization of the target state in a graph, with associated parameters and actions in the nodes and track-based conditions as node transitions. In our implementation, we demonstrate the benefits of our approach by switching between filtering strategies depending on the target status. The experimental results have proven that this representation is capable of overcoming the aforementioned problems in realistic complex video sequences.

As future work, in addition to improving the current components of the system as explained in the previous section, we plan to use state-dependent model cues (e.g., more complex appearance or color models if the track is lost), balance the number of particles according to the state, and perform graph interactions in order to provide a more robust state estimation.

Acknowledgements

This work was developed during D. Gerónimo stay at LAAS-CNRS, supported by the post-doctoral mobility grant AIRE-CTP 2010 CTP 00009 of the Generalitat de Catalunya, and the Spanish Ministry of Education and Science projects TRA2011-29454-C03-00 and Consolider Ingenio 2010: MIPRCV (CSD200700018).

References

1. Negre, A., Tran, H., Gourier, N., Hall, D., Lux, A., Crowley, J.L.: Comparative study of people detection in surveillance scenes. In: International Workshop on Structural and Syntactic Pattern Recognition, Kyoto, Japan (2006)
2. Gerónimo, D., López, A., Sappa, A., Graf, T.: Survey of pedestrian detection for advanced driver assistance systems. *TPAMI*, **32**(7) (2010) 1239–1258
3. Germa, T., Lerasle, F., Ouadah, N., Cadenat, V.: Vision and RFID data fusion for tracking people in crowds by a mobile robot. *Computer Vision and Image Understanding*, **114**(6) (2010) 641–651
4. Isard, M., Blake, A.: CONDENSATION – conditional density propagation for visual tracking. *IJCV*, **29**(1) (1998) 5–28
5. Isard, M., Blake, A.: ICONDENSATION: Unifying low-level and high-level tracking in a stochastic framework. In: ICCV, Freiburg, Germany (1998) 893–908
6. Breitenstein, M.D., Reichlin, F., B.Leibe, Koller-Meier, E., Van Gool, L.: Online multi-person tracking-by-detection from a single, uncalibrated camera. *TPAMI*, **33**(9) (2010) 1820–1833
7. Okuma, K., Taleghani, A., Freitas, N.D.: A boosted particle filter: Multitarget detection and tracking. In: European Conf. on Computer Vision, Prague, Czech Republic (2004)
8. Qu, W., Schonfeld, D., Mohammed, M.: Distributed bayesian multiple-target tracking in crowded environments using multiple collaborative cameras. *EURASIP Journal on Advances in Signal Processing* (2007)
9. Zhou, Y., Benois-Pineau, J., Nicolas, H.: Multi-object particle filter tracking with automatic event analysis. In: ACM International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Streams. (2010)
10. Willsky, A., Jones, H.: A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *IEEE Trans. on Automatic Control* **AC-21**(1) (1976) 108–112
11. Wang, F., Yu, S., Yang, J.: Robust and efficient fragments-based tracking using mean shift. *Int. Journal of Electronics and Communications* **64** (2010) 614–623
12. Di Caterina, G., Soraghan, J.: An improved mean shift tracker with fast failure recovery strategy after complete occlusion. In: 8th IEEE Int. Conf. on Advanced Video and Signal-Based Surveillance. (2011)
13. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR. Volume 1., San Diego, CA, USA (2005) 886–893
14. Pérez, P., Vermaak, J., Blake, A.: Data fusion for visual tracking with particles. *Proceedings of the IEEE* **92**(3) (2004) 495–513
15. Neyman, J., Pearson, E.: On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London, Series A, Containing Papers of a Mathematical or Physical Character* 231 (1933) 289–337
16. Andriluka, M., Roth, S., Schiele, B.: People-tracking-by-detection and people-detection-by-tracking. In: Int. Conf. on Computer Vision and Pattern Recognition, Anchorage, AK, USA (2008)
17. Andriyenko, A., Roth, S., Schindler, K.: An analytical formulation of global occlusion reasoning for multi-target tracking. In: 11th IEEE International Workshop on Visual Surveillance (at ICCV), Barcelona (2011)
18. Bernardin, K., Stiefelhagen, R.: Evaluating multiple object tracking performance: the clear mot metrics. *EURASIP Journal on Image and Video Processing* (2010)