Journal of Electronic Imaging

JElectronicImaging.org

Three hypothesis algorithm with occlusion reasoning for multiple people tracking

Carolina Reta
Leopoldo Altamirano
Jesus A. Gonzalez
Rafael Medina-Carnicer

SPIE. IS&T imaging.org

Journal of Electronic Imaging 24(1), 013015 (Jan/Feb 2015)

Three hypothesis algorithm with occlusion reasoning for multiple people tracking

Carolina Reta,[a,*] Leopoldo Altamirano,[a] Jesus A. Gonzalez,[a] and Rafael Medina-Carnicer[b]
[a]National Institute for Astrophysics, Optics, and Electronics, Department of Computer Science, Luis Enrique Erro No. 1, Puebla 72840, Mexico
[b]University of Cordoba, Department of Computing and Numerical Analysis, Campus de Rabanales, Edificio Einstein, 3[a] planta, Cordoba 14071, Spain

Abstract. This work proposes a detection-based tracking algorithm able to locate and keep the identity of multiple people, who may be occluded, in uncontrolled stationary environments. Our algorithm builds a tracking graph that models spatio-temporal relationships among attributes of interacting people to predict and resolve partial and total occlusions. When a total occlusion occurs, the algorithm generates various hypotheses about the location of the occluded person considering three cases: (a) the person keeps the same direction and speed, (b) the person follows the direction and speed of the occluder, and (c) the person remains motionless during occlusion. By analyzing the graph, our algorithm can detect trajectories produced by false alarms and estimate the location of missing or occluded people. Our algorithm performs acceptably under complex conditions, such as partial visibility of individuals getting inside or outside the scene, continuous interactions and occlusions among people, wrong or missing information on the detection of persons, as well as variation of the person's appearance due to illumination changes and background-clutter distracters. Our algorithm was evaluated on test sequences in the field of intelligent surveillance achieving an overall precision of 93%. Results show that our tracking algorithm outperforms even trajectory-based state-of-the-art algorithms. *© 2015 SPIE and IS&T* [DOI: 10.1117/1.JEI.24.1.013015]

Keywords: people tracking; occlusion; tracking graph; hypothesis management; spatio-temporal features; video surveillance.
Paper 14315 received May 29, 2014; accepted for publication Dec. 8, 2014; published online Jan. 13, 2015.

1 Introduction

Multiple people tracking is very important for the development of video surveillance technology. It provides useful information that can be used to classify pedestrian activities, such as walking, running, jumping, and waiting for something, among others. Multiple people tracking remains as an open problem when people move in real environments, such as parks, schools, shopping malls, lobbies, airports, borders, and so forth. This problem is challenging since in this domain the number of persons within the scene may vary over time and their dynamics are subject to sudden changes. Moreover, their clothing cannot be specified in advance and illumination changes and background-clutter distracters affect their perceived appearance. Another equally important issue is the occlusion mainly caused by the interaction among people.

In this work, we address the problem of multiple people tracking in uncontrolled sceneries using a single camera. Our solution deals with major problems, such as appearance changes of the individual's clothing, partial and total occlusions among persons, and confusion of identities of nearby or interacting targets. For this, we propose a detection-based tracking algorithm able to estimate the location and determine the trajectory of each person in a set of video frames. Our algorithm associates people trajectories with available detection responses and analyzes the interaction among targets to predict partial and total occlusions. Our algorithm keeps different hypotheses about the location of occluded people in order to avoid losing them during total occlusions. Our algorithm can track people when there are false positives (FPs) and false negatives (FNs) in the person's detection measurements. Results show that our tracking algorithm outperforms state-of-the-art algorithms.

The rest of the paper is organized as follows. Section 2 discusses the related work. Section 3 explains the detail of the proposed tracking algorithm. Section 4 shows experimental results of multiple people tracking in uncontrolled environments and it also presents a comparison of our results with other state-of-the-art approaches. Finally, Sec. 5 presents the conclusions of our work.

2 Related Work

The process of tracking multiple people consists of a target detection and representation stage and a temporal association stage. When people are separated and do not occlude each other, this process can be easily solved by running multiple independent trackers,[1] such as the bounding-box-based tracking,[2] the hybrid appearance-guided particle filter,[3] or the CamShift-guided particle filter.[4] However, in real sceneries, interactions and occlusions among people occur making both tracking stages difficult problems.

Some works in the literature have focused on improving the detection and representation stage to deal with partial occlusions. Senior et al.[2] use appearance models and probabilistic maps to localize people and vehicles blobs that were partially overlapped. Similarly, Vezzani et al.[5] proposed a probabilistic appearance-based approach that allows the estimation of the pixel-wise shape of each person during

*Address all correspondence to: Carolina Reta, E-mail: creta@inaoep.mx

occlusions. Both approaches are vulnerable to misclassification for similarly colored objects at object interactions. Wu and Nevatia[6] use body part detectors learned by boosting weak classifiers based on edgelet features. The combined detection responses and the part detection responses provide the observations used for tracking. Dalal and Triggs[7] propose a method for detecting people using the histogram of oriented gradient (HOG) descriptor and a linear support vector machine (SVM) classifier. Lin et al.[8] present a hierarchical part-template matching approach for human detection and segmentation. In their approach, human shapes are matched with an edge map using a Bayesian maximum a posteriori (MAP) framework approach that combines local part-based and global template-based features. Andriluka et al.[9] propose a dynamic limb-model based on a hierarchical Gaussian process latent variable model. This model is used to improve people detection performance by generating reliable people-tracklets in image sequences.

Since even a partial occlusion can mislead trackers, leading to fragments or total loss of tracks, some works have dealt with the occlusion problem by using a multiview configuration. Khan and Shah[10] develop a planar homographic occupancy constraint to localize people on the ground plane. Tracking is performed by minimizing an energy function that combines occupancy information and spatio-temporal proximity. Munoz-Salinas et al.[11] present an extension of particle filters to the Dempster-Shafer theory of evidence. To detect target occlusions, an occupancy map is calculated separately for each camera using a depth-ordering scheme. The evidence of visible people collected from all the cameras is fused to obtain the best estimation of target locations. Kaucic et al.[12] propose a method to link fragments of paths over sensor gaps by clustering spatially close pairs of fragments with a similar appearance and motion. Although a multiview configuration reduces the degree of occlusion, it does not solve the occlusion problem in settings in which there are plenty of occlusions caused by the interaction of multiple people. Ryu et al.[13] confirm this assertion and suggest a method to strategically place the cameras in order to minimize occlusions to tracking objects systems.

In this work, we address the tracking of multiple people task with a single-view configuration by strengthening the temporal association stage. The temporal association methods commonly used to track several objects are nearest neighbor (NN), global nearest neighbor (GNN), joint probability data association (JPDA), and multiple hypothesis tracking (MHT).[14–16] The NN and GNN algorithms[14,16] estimate the most likely assignment of the detection measurements with the existing trajectories in polynomial time. These algorithms are reliable when there is a moderate density of objects in the scenery and when the movement or change in the appearance of the object from frame to frame is low. However, these methods fail when the FP rate or the FN rate in the detection measurements increase, occlusions between objects occur, changes in the object appearance are significant, or when the maneuvers of the objects are complex. The JPDA algorithm[17] provides a suboptimal approximation of the Bayesian filter for a constant number of objects. Its main disadvantage is that it consumes a lot of computational cost (nondeterministic polynomial-time (NP)-hard). The JPDA algorithm is reliable in sceneries with a moderate density of objects. However, this algorithm presents the problem of interference from nearby objects. This influence may create collisions between objects that are moving in parallel, damage the recognition of the object, or damage the information used for its discrimination. The MHT[16] algorithm exhaustively enumerates all the possible hypotheses of the object estimations over a specific number of recent frames to choose the most likely estimation. Unlike the NN, GNN, and JPDA algorithms, in which a temporal association decision is irrevocable and is taken based on information from two consecutive frames, the MHT algorithm delays the temporal association decision until enough information of measurements from multiple frames is available to avoid wrong associations. Theoretically, the MHT algorithm finds the best solution to the problem but it is computationally expensive (NP-hard). This work proposes a solution to the problem in polynomial time.

Other works prefer to avoid tracking errors by temporal association algorithms that optimize the trajectories through the whole sequence analysis. Zhang et al.[18] define the temporal association problem as an MAP problem. The problem is mapped to a data flow network that does not allow overlapping between the trajectories. The optimal association of the trajectories is done using an algorithm that minimizes the network flow-cost. Wang et al.[19] also map the multitarget tracking problem to a network-flow. They solve the MAP problem using a global optimization framework that uses mixed integer programming. They show that the spatial location of the missing objects can be inferred from the estimation of the location of the other objects. Yang and Nevatia[20] formulate the tracking problem as an energy minimization problem, and propose an online learned condition random field approach for efficiently finding good tracking solutions with low energy costs. Collins[21] presents an iterative approximate algorithm to the multidimensional assignment problem under general cost functions. This algorithm uses a snake energy trajectory cost function to measure the quality of a proposed trajectory. Song et al.[22] analyze the statistical properties of segments of trajectories to develop assignments between them so that they can form larger trajectories. They propose a stochastic method based on the evolution of an association graph which has trajectories segments as nodes and affinity scores as weights. The association is done by estimating the MAP of connections between segments. Brendel et al.[23] formulate multitarget tracking as the maximum-weight independent set problem. They address the long gaps by iteratively linking smaller similar tracks into larger ones and splitting long unviable tracks until convergence. Some other works address the problem of temporal association by building up a graph where each node represents the observation of an object and the edges denote their path.[24–26] These works are based on the principle of adding measurements when an object is not detected and removing them when they correspond to false detections. They solve the temporal association task by adding edges to the tracking graph using algorithms that find the shortest path.

Taking into account the weaknesses of previous work, in Reta et al.,[27] we proposed a tracking strategy to track multiple people from the information of two consecutive frames. We introduced a description of a tracking graph approach based on human interaction rules able to maintain the tracking of people through occlusions. We presented an occlusion function that determines the depth ordering of

targets when an occlusion occurs and enables the successful update of people models. We tested our approach using controlled detection measurements obtained by adding noise to the ground truth. In the CAVIAR[28] benchmark sequences, we achieved an accuracy of 89.8% demonstrating that our approach yields robust tracking of people during partial or total occlusions.

In this work, we propose a novel tracking algorithm able to individually track multiple people in uncontrolled stationary environments. This work presents a significant extension of our previous approach. First, we propose a people detection scheme based on a human silhouette model that is able to find persons partially occluded (see Sec. 3.1). Second, we suggest a person's representation that is able to adjust to appearance changes due to variable illumination as well as deal with partial occlusions and changes in motion (see Sec. 3.2). Third, we present a temporal association algorithm that builds a graph structure whose main objective is to keep the tracking of people through occlusions (see Sec. 3.3, Algorithm 1). We also propose a similarity matching process that associates detection measurements with available people trajectories based on appearance and spatial motion similarity and location proximity (see Sec. 3.3.2). Our temporal association algorithm has a time complexity of $O(n^3)$, where n is the number of tracked targets, which is regulated by the execution time of the similarity association process. Finally, we extend experiments to test sequences in the field of intelligent surveillance so that we show our tracking algorithm is robust under different sceneries with various people interactions and occlusion events.

3 Proposed Work

3.1 *Detection of Persons*

This work proposes a detection scheme able to find multiple people who may be partially occluded in stationary sceneries. The detection scheme examines areas in the scenery where there is the presence of motion in order to locate targets that fit into a human silhouette model.

In our scheme, we first model the background of the scenery by building a mixture of Gaussian probability map that uses local texture features and invariant color features.[29] Because this map is adaptive to illumination changes as well as to the addition and removal of stationary objects, moving foreground regions are well-defined by a threshold filtering.

Second, we model the human silhouette by building a template in the form of an active basis, which is formed by a composition of Gabor wavelet elements that can be slightly perturbed so that the template is deformable. The active basis template is learned from images of persons in different poses and clothing[30] by the shared sketch algorithm.[31] After that, the fitting of the active basis template to a test image is achieved by the computational architecture of sum-max maps.[31] This architecture alternates between sum maps and max maps. The sum maps result from local filtering operations for detecting edge segments and shapes. The max maps result from local maximization operations that track shape deformations. In general, this architecture builds a log-likelihood map of the deformed active basis that can be interpreted as a shape filter map.

Finally, in order to detect people with different sizes, our scheme obtains shape filter maps at different scales to then find the best fits of the active basis template. We calculated the local maximum responses of all of the shape filter maps on the areas that correspond to moving foreground regions. To avoid over-fitting of the template through the multiples scales, once the global maximum response of the active basis template is obtained, the foreground region occupied by its silhouette is marked as occupied. Then, to find more persons in the scene, the search of subsequent maximum responses is repeated over the vacant areas of the foreground regions.

Figure 1 shows our people detection scheme in a test image.

3.2 *Persons' Representation*

Each person to be tracked is represented by his shape, appearance, and motion models.

The "human body shape" is modeled as an ellipse region with a parameter vector $S = (x_c, y_c, \phi, r_x, r_y)$, where (x_c, y_c) is the center of the ellipse, ϕ is its orientation angle according to the x axis, and (r_x, r_y) are the half-radii of the ellipse.

In this work, the parameters of the shape model of the object are computed directly from the bounding box location of the measurement provided by our people detection scheme.

The "appearance of the object" in the image is represented by histogram q, which describes the color distribution of the pixels inside the object's area S. To generate the q histogram, the color cube is divided into m equal-size bins. The function b: $S \subset \mathbb{R}^2 \rightarrow \{1, \ldots, m\}$ is defined to map the pixel at location p_j to the index $b(p_j)$ of the bin corresponding to its quantized color u. The color density distribution for each bin $\hat{q}(u)$ is then computed as

$$\hat{q}(u) = \frac{1}{|S|} \sum_{p_j \in S} \kappa[b(p_j) - u], \tag{1}$$

where κ is the Kronecker delta function and $|S|$ denotes the cardinality of S. The factor $1/|S|$ imposes the condition $\sum \hat{q}(u) = 1$ to normalize the resulting histogram $q = \{\hat{q}(u)\}_{u=1,\ldots,m}$.

In our experiments, each histogram is calculated in the RGB cube using bins of $16 \times 16 \times 16$ size on the elliptical region that models the shape of the object.

The "motion of the object" is represented by the list $M = [\hat{x}, \hat{P}]$, where $\hat{x}$ is the estimated state vector that includes parameters of the object's position and velocity, and $\hat{P}$ specifies the covariance of the state estimation error. Unlike our previous work[27] where the motion of the object was represented by parameters of the affine transformation from its optical flow estimation, in this work we use the Kalman filter theory[32] to predict and correct the dynamic model of the object from a series of incomplete and/or noisy measurements.

3.3 *Temporal Association Algorithm*

Let $G = \langle N, E \rangle$; be a tracking graph, where $N = \{N_o \cup N_h\}$ is the set of nodes of the graph that represents detected people measurements (N_o) and hypotheses of people that are not visible (N_h).

Each node $n \in N$ is identified by its index k at time t as n_k^t. Each node has a list of attributes $\text{Atr}_k^t = [\text{id}_k^t, S_k^t, q_k^t, M_k^t]$ that describes the represented person, where id is the identity

Algorithm 1 Temporal association algorithm.

Input: I: a sequence of images with people.

Onput: $G = \langle N_o, N_h, E \rangle$;: tracking graph.

Main variables

N: set of nodes of the graph.

E: set of edges of the graph.

N_o: represents visible objects.

N_h: represents hypotheses of non detected or occluded objects.

μ_ρ: threshold value for the appearance similarity metric.

μ_g: gate radius of the validation region.

m: required frames to prune hypotheses in the graph.

λ: percentage of overlapping among predictions of objects.

$\mathscr{V}$: set of detected objects in the image.

$\mathscr{O}$: set of objects in an occlusion event.

T: total number of frames.

t: current frame.

superscript t: it refers to the objects in frame t.

subscript k: it refers to the k'th object in the frame.

S: it refers to the shape area attributes.

$S^{t-1|t}$: it refers to the prediction area attributes.

$G = \langle \varnothing, \varnothing, \varnothing \rangle$;

Set values for μ_ρ, μ_g, m, and λ

for t: 1... T **do**

Assoc $= \varnothing$

$\mathscr{V} =$ `PeopleDetection` (I^t)

$\mathscr{O} =$ `Occlusion` (N^{t-1}, λ)

$\pi =$ `SortOcclusionTargets` $(\mathscr{O})$

foreach $v_k \in \mathscr{V}$ **do**

$N^t_{o_k} =$ `AddNoteToGraph` $(G, v_k, t,$ "N_o")

if $S^t_{o_k} \cap S^{t-1|t} \neq \varnothing$ **then**

$F_{\max} =$ `MaxSimilarity` $(N^t_{o_k}, N^{t-1}_j: S^t_{o_k} \cap S^{t-1|t}, \mu_\rho, \mu_g)$

if $F_{\max} \neq \varnothing$ **then**

`AddEdgeToGraph` $(G, E, (F_{\max}, N^t_{o_k}))$

Assoc $= \{$Assoc $\cup\ F_{\max} \cup N^t_{o_k}\}$

end

end

end

foreach $n_j \in N^t_o \setminus$ Assoc **do**

$F_{\max} =$ `MaxSimilarity` $(n_j, N^{t-1} \setminus$ Assoc$, \mu_\rho, \mu_g)$

if $F_{\max} \neq \varnothing$ **then**

`AddEdgeToGraph` $(G, E, (F_{\max}, n_j))$

Assoc $= \{$Assoc $\cup\ F_{\max} \cup n_j\}$

end

end

`UpdateAttributes` (N^t_o)

foreach $a_i \in N^{t-1} \setminus$ Assoc **do**

$N^t_{h_1} =$ `AddNoteToGraph` $(G, a_i, t,$ "N_h")

`AddEdgeToGraph` $(G, E, (a_i, N^t_{h_1}))$

`UpdateAttributes` $(N^t_{h_1}, \varnothing, 0,$ "the person retains his velocity and direction")

if `FirstOcclusion` $(a_i, \mathscr{O}, \pi)$ **then**

$N^t_{h_2} =$ `AddNoteToGraph` $(G, a_i, t,$ "N_h")

`AddEdgeToGraph` $(G, E, (a_i, N^t_{h_2}))$

`UpdateAttributes` $(N^t_{h_2}, \mathscr{O}, \pi,$ "the person follows the velocity and direction of the occluder")

$N^t_{h_3} =$ `AddNoteToGraph` $(G, a_i, t,$ "N_h")

`AddEdgeToGraph` $(G, E, (a_i, N^t_{h_3}))$

`UpdateAttributes` $(N^t_{h_3}, \varnothing, 0,$ "the person remains motionless during occlusion")

end

end

$G =$ `PruneHypothesis` (G, m)

end

of the tracked object, S is the shape model of the object, q is the color histogram that describes its appearance, and M is a list of parameters that describes its motion.

A directed edge $(n^{t-1}_i, n^t_j) \in E$ between two nodes in consecutive time steps is defined for two cases: (a) if $n^t_j \in N_o$, the similarity association process that matches tracked objects with detection measurements must be satisfied, i.e., $(n^{t-1}_i, n^t_j) \in \max \text{Similarity}(N^{t-1}_i, N^t)$, (b) otherwise,

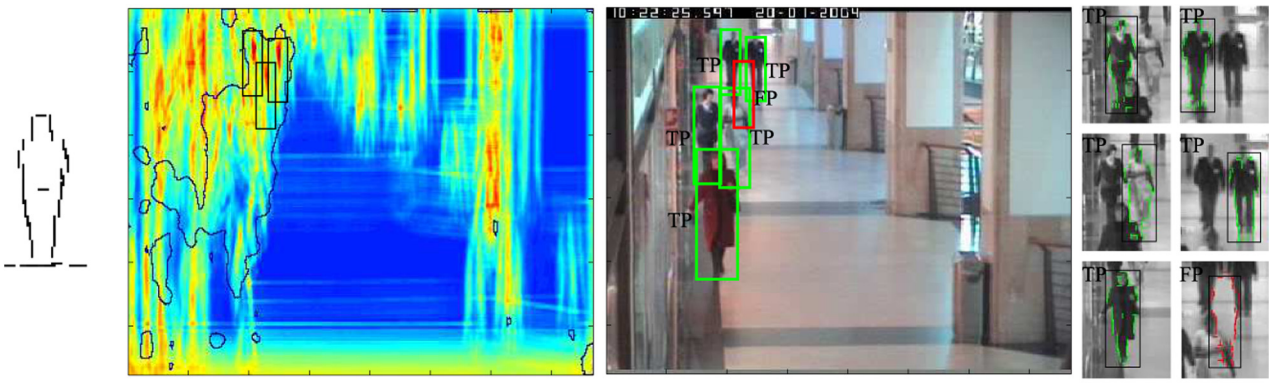


Fig. 1 People detection scheme. (a) The active basis template is learned from images of persons in different poses and clothing. The fitting of this template to a test image gives as result a log-likelihood map that can be interpreted as a shape filter map. (b) Shape filter maps at different scales are obtained to facilitate the detection of people with different sizes. The search of local maximum responses of all of the maps is reduced to the moving foreground regions (only a shape filter map is shown in the figure). (c) and (d) The fit of the template at different iterations produces as result the detection of multiple people. True positives and false positives (FP) in the detection measurements are also shown in these figures.

if $n_j^t \in N_h$ is a predicted successor of $n_i^{t-1} \in N$ then the edge is generated as a hypothesis trajectory to keep the identity of objects that are missing or occluded.

Algorithm 1 shows our novel tracking algorithm. Basically, our algorithm works as follows:

An iteration begins with a set of the object hypotheses trajectories built in the previous frame. For each hypothesis, a prediction is performed to estimate the location of the object in the current frame. Additionally, individual's measurements are obtained by our detection scheme to choose who should be tracked in every frame. Occlusion relationships between trajectory hypotheses are also obtained to determine which of those hypotheses correspond to objects that are occluded, as well as who occludes who in each occlusion set.

The process of linking tracked objects with candidate targets is performed in two stages. In the first stage, predictions for which area intersects with measurement areas are found. Then, the matching pairs are estimated by maximizing the total distance of the similarity of their appearance and spatial motion. In a second stage, predictions and measurements that were not matched in the first stage, but are inside the validation gate, are found. After that, the matching pairs are estimated by maximizing the total distance of the similarity of their appearance and position.

For both stages, current measurements are linked with predictions of tracked objects if their matching pairs satisfy an appearance similarity threshold. Otherwise, a hypothesis tracking for the new measurement is generated. This hypothesis will have to be confirmed in subsequent frames to determine if it is a measurement that corresponds to an object entering to the field of vision or if it is an FP measurement generated during the detection stage.

For each prediction that is not supported by a measurement, the algorithm assesses whether the object got out from the field of vision of the scene, or the object is being occluded, or it was simply not detected due to background noise. For the last two cases, a tracking hypothesis is generated to predict the location of the object. Of course, this hypothesis keeps the direction and speed of the tracked person. As a special case, when the total occlusion event occurs for the first time, two additional hypotheses are generated considering situations in which the person follows the direction and speed of the occluder, and the person remains motionless during occlusion.

To complete the iteration, contradictory occlusion hypotheses are pruned, hypotheses generated by FPs in the detection measurements are eliminated, and hypotheses corresponding to objects getting out of the scene are finalized.

The subsequent sections describe in detail three main components of our algorithm. Section 3.3.1 describes how our temporal association algorithm detects occlusions among people and sorts the objects involved in them (functions: Occlusion and SortOcclusionTargets, lines: 6 and 7). Section 3.3.2 describes the similarity association process between candidate targets and tracked associations carried out by our algorithm (function: MaxSimilarity, lines: 11 and 16). Finally, Sec. 3.3.3 describes how to update the attributes of the objects that are being tracked (function: UpdateAttributes, lines: 20, 24, 28, and 31).

3.3.1 *Occlusion relationships*

To represent possible occlusions among people, we verify which areas of the predicted regions of nodes overlap. Each prediction area is identified by its index k at time t as $S_k^{t|t-1}$ and is estimated by using the prediction step of the Kalman filter.[32]

We then build a set of binary relationships $\mathcal{O}$ between these overlapped regions as follows:

$$\mathcal{O} = \left\{ (n_i^{t-1}, n_j^{t-1}) | id_i \neq id_j, \frac{|S_i^{t|t-1} \cap S_j^{t|t-1}|}{|S_i^{t|t-1}| + |S_j^{t|t-1}|} > \lambda \right\}, \quad (2)$$

where the restriction $\mathrm{id}_i \neq \mathrm{id}_j$ prevents occlusion relationships among hypotheses generated for the same object, and

the threshold λ is established as the overlapping-percentage indicator to identify a possible occlusion.

To sort the elements in each occlusion pair of the set $\mathcal{O}$, we defined the function $\pi_{ij} \in \{+1, -1\}$ between objects i and j, where $\pi_{ij} = +1$ means i occludes j, and $\pi_{ij} = -1$ means j occludes i. To determine who occludes who, we first verify whether the tracked nodes in the occlusion set are visible or not in the current frame. A node $n_k^{t-1} \in \mathcal{O}$ is visible at time t if its appearance model matches against the model of a measurement detected on the overlapped region. In this case, we represent the corresponding successor of n_k^{t-1} that is visible at time t as $n_{k'}^t \in \mathcal{V}^t$.

Then, occlusion function π is evaluated in the current frame as follows:

$$\pi_{i'j'}^t = \begin{cases} +1 & \text{if } n_{i'}^t \in \mathcal{V}^t, n_{j'}^t \notin \mathcal{V}^t \text{ or } n_{i'}^t \in \mathcal{V}^t, n_{j'}^t \in \mathcal{V}^t, \text{up}_{y_c}(n_{i'}^t, \text{S}_j^{t|t-1}) > \text{down}_{y_c}(n_{i'}^t, \text{S}_j^{t|t-1}) \\ -1 & \text{if } n_{i'}^t \notin \mathcal{V}^t, n_{j'}^t \in \mathcal{V}^t \text{ or } n_{i'}^t \in \mathcal{V}^t, n_{j'}^t \in \mathcal{V}^t, \text{up}_{y_c}(n_{i'}^t, \text{S}_j^{t|t-1}) < \text{down}_{y_c}(n_{i'}^t, \text{S}_j^{t|t-1}) \\ \pi_{ij}^{t-1} & \text{otherwise} \end{cases}, \tag{3}$$

where up_{y_c} and down_{y_c} are the areas of the given shape region up and down the center of the reference object, respectively. The comparison of these area functions allows estimating the spatial ordering of the objects by considering the global alignment of their shape regions. This feature stands out against our previous work[27] in which a local alignment of objects was achieved by a comparison between the largest value of the y-coordinate of the regions.

As we can see in Eq. (3), occlusion relationships are deduced from the spatial visibility of the objects in the current frame. For cases where objects are not visible, their order before the occlusion is kept by inheriting the previous value of the occlusion function.

3.3.2 *Similarity association process*

According to the person models described in Sec. 3.2, we define similarity metrics that allow us to evaluate if a tracked person can be spatially and temporally linked with a candidate target.

The similarity of appearance between the color q of the tracked object and the color p of a candidate target is defined by the Bhattacharyya coefficient ρ as

$$\rho(p, q) = \Sigma_u \sqrt{\hat{p}(u)\hat{q}(u)}, \tag{4}$$

where $\hat{p}(u)$ and $\hat{q}(u)$ are the normalized color densities of the histogram bin u. The coefficient ρ is in the range [0,1], where $\rho = 1$ means that the two histograms are identical, and ρ decreases as the histograms differ.

The spatial motion between the predicted region S_q of the tracked object and the region S_p of a candidate target is measured by the Hamming distance δ as

$$\delta(S_p, S_q) = 1 - \frac{|S_p \cap \bar{S}_q| + |S_q \cap \bar{S}_p|}{|S_p| + |S_q|}. \tag{5}$$

Metric δ takes values in the range [0,1], where $\delta = 1$ means the two regions are identical, and δ decreases as the regions differ.

The proximity between the central points c_p and c_q of the object regions S_q and S_p is computed by the well-known Euclidian distance $d(c_p, c_q)$.

We also define a validation gate to delimit the area where temporal correspondences may occur. The validation gate is approximated by a circular region with the center at the prediction area and an established gate radius μ_g.

Detection measurements that lie inside the prediction gates are compared with their predictions in order to link tracked objects with candidate targets. The following similarity association process is proposed to set up these matching pairs.

1. Find from the predictions which area intersects with the measurement areas. If any prediction is found, then skip to step 5.
2. Compute the ρ and δ distances between the measurements and predictions obtained in step 1.
3. Choose the predictions that maximize the ρ and δ coefficients by using the Hungarian algorithm.[33]
4. Accept the predictions as matching pairs of their respective measurements if their ρ coefficients satisfy the appearance similarity threshold μ_ρ.
5. Find from the predictions which measurement areas are inside their validation gates μ_g, but where neither is in the matching pairs. If any prediction is found, then the process finishes.
6. Compute the ρ and d distances between the measurements and predictions obtained in step 5.
7. Choose the predictions that minimize the $1 - \rho$ and d coefficients by using the Hungarian algorithm.[33]
8. Accept the predictions as matching pairs of their respective measurements if their ρ coefficients satisfy the appearance similarity threshold μ_ρ.

Step 1 to step 4 of the process allow linking detection measurements with available people trajectories based on similarity metrics of appearance and spatial motion. Step 5 to step 8 of the process allow the matching based on similarity and location proximity metrics of the detection measurements that are inside the validation gate with the remainder of people trajectories that were not previously allocated.

3.3.3 *Attributes update*

To update the attributes of the objects, it is necessary to check if they appear in an occlusion relationship. This is done in order to know which object is occluded by which other object in the relationship. Knowing this information, the attributes update is carried out as follows:

- The "appearance model" of a tracked person is updated with the appearance model of his current measurement

when the person is absent from any occlusion. If we detect that the person is being occluded, the appearance model of the tracked person is kept constant.

- The "shape model" of a tracked person is updated with the shape model of his current measurement when the person is not being occluded, otherwise the shape model of the tracked person is updated with the model of his prediction area.
- The "motion model" is updated with the motion model calculated from his current measurement given that the person is not being occluded. If we detect that the person is being occluded, the motion model of the tracked person may be updated in three different cases. Case 1 states that the tracked person evolves independently during the occlusion event, keeping his velocity and direction. As this assumption can be violated during target interactions, case 2 states that the tracked person acquires the velocity and direction of his occluder. Case 3 states that the tracked person remains motionless during the occlusion.

4 Experimental Results

4.1 *Data Sets*

We evaluated our tracking algorithm with respect to reference sequences focusing on intelligence surveillance. We conducted tests in diverse real-world environments with complex situations of targets' interactions and occlusions.

We used CAVIAR[28] sequences to assess the performance of the algorithm in indoor environments. These sequences were captured in a corridor of a shopping mall in which a variable number of people in the scene perform activities, such as walking, talking, getting in and getting out of shops, waiting for someone else, and so on. The evaluation of our algorithm is carried out in the seven most challenging video sequences: *TwoEnterShop3*, *TwoEnterShop2*, *ThreePastShop2*, *ThreePastShop1*, *TwoEnterShop1*, *OneStopOneWait1*, and *OneStopMoveEnter1*.

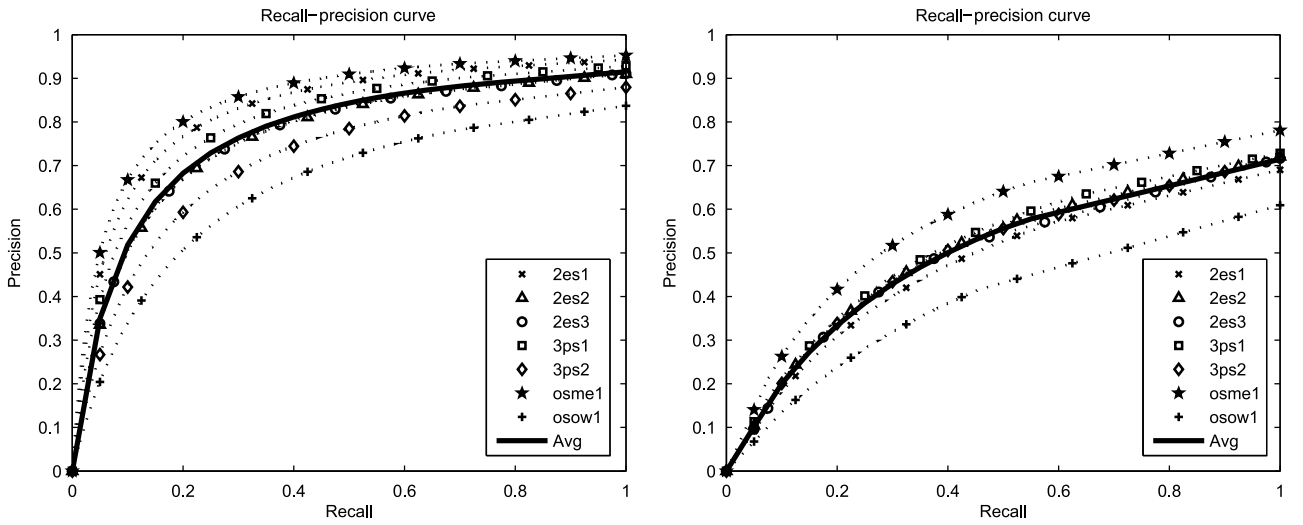
PETS[34] sequences allow us to assess the performance of our algorithms in outdoor environments. The sceneries of these sequences are filmed by multiple cameras focused on the crosswalk of a university. The sceneries involve actors who are walking either alone or in pairs. Actors also perform other activities, such as meeting with people, waiting for someone else, changing directions, as well as getting in and getting out of the scene. The evaluation of our algorithm is performed independently in four video sequences of the Dataset *S2: People Tracking, Scenery: L1* from *view 005* to *view 008*.

UCO[35] sequences are recorded in a laboratory room of a university from six cameras' points-of-view. The evaluation of our algorithm is performed independently in 24 video sequences of the Datasets *p2v1*, *p2v2*, *p3v1*, *p3v2* from *view 1* to *view 6*. The number of people in these sequences varied from 2 to 3, due to the limited field-of-view of the scenery. Although people were instructed to move around freely in the environment, they generated complex situations of frequent interactions and occlusions.

4.2 *Evaluation of Our Detection Scheme*

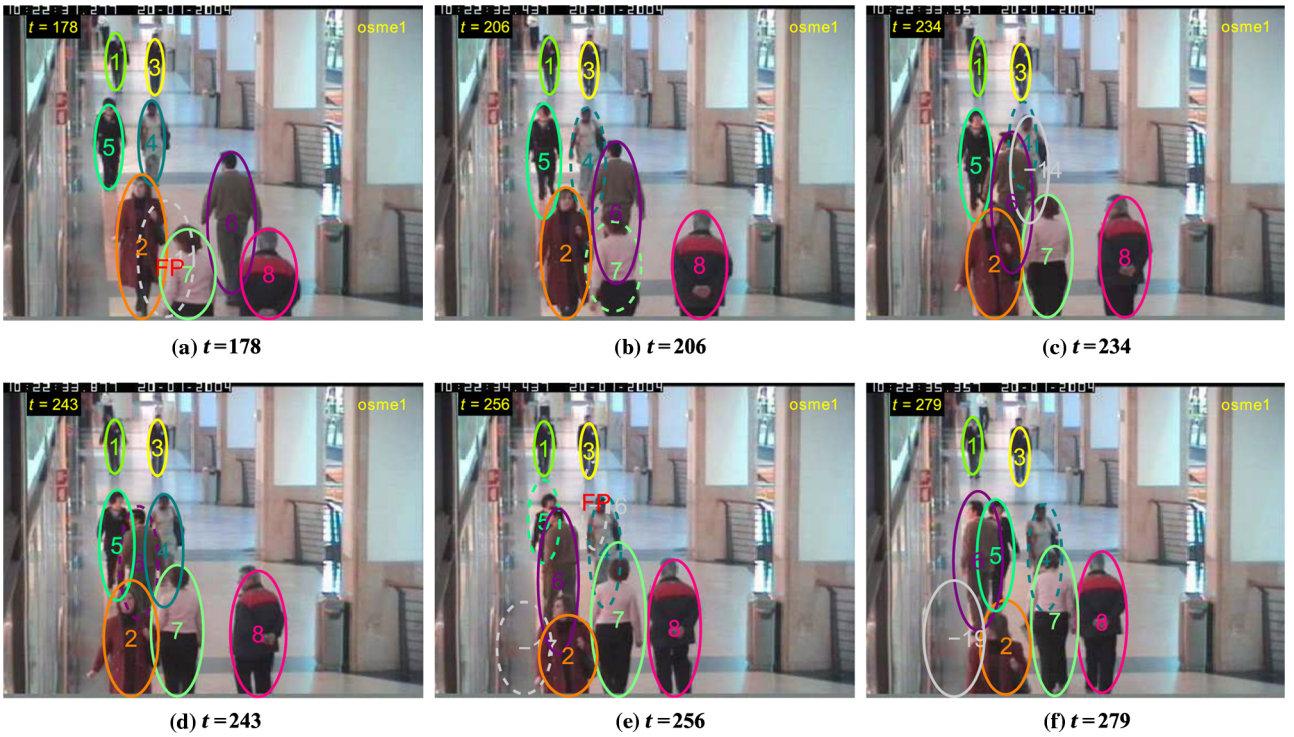
Our proposed detection scheme was able to locate multiple possibly interoccluded humans from static images. We evaluate our people detection scheme on the CAVIAR[28] dataset. This dataset has ground truth detection information available. We compare our results with the Dalal and Triggs[7] method which, as our detection scheme, is able to detect multiple humans in images with a moderate density of people. The Dalal and Triggs method uses HOG feature vectors and a trained SVM classifier to detect persons in images.

In Fig. 2, we show a comparison of our detection scheme with the HOG-SVM method[7] for the CAVIAR[28] dataset. For all of the test sequences, our detector achieved a better performance than the HOG-SVM method. This is mainly because our detection method is more robust to partial occlusions than the HOG-SVM method. Our method was designed to detect people with a visibility higher than



(a) Our detection scheme (b) HOG-SVM method

Fig. 2 Comparison of our detection scheme with the histogram of an oriented gradient-support vector machine method[7] for the CAVIAR[28] dataset.



(a) $t=178$ (b) $t=206$ (c) $t=234$

(d) $t=243$ (e) $t=256$ (f) $t=279$

Fig. 3 Results of our temporal association algorithm in the sequence *OneStopMoveEnter1* of the CAVIAR[28] dataset. In this crowd sequence, groups of people walk collectively along the corridor. Occlusion between persons with ID 4 and ID 6, and persons with ID 5 and ID 6, is successfully resolved. Timely detection of FP prevents the misleading correction of trajectories of people who walk close [frames (a) and (e)].

30%, whereas HOG-SVM can detect people with a visibility higher than 65%.

In the images of the CAVIAR sequences, our detection scheme achieved a precision of 87% and a recall of 62%, whereas the HOG-SVM method achieved a precision of 51% and a recall of 43%. Our detection scheme outperforms the HOG-SVM method in both precision and recall metrics because the FP and FN indices achieved by our scheme are lower. This happened because the detection decision of our scheme was taken by examining information from a shape detection filtering as well as from a motion detection filtering. Furthermore, the true positive metric of our detection scheme acquires a much higher value than the HOG-SVM method. Thus, our scheme achieved a high precision performance. This took place because our detection scheme was designed to deal with partial occlusions and to be tolerant to changes in the perceived 2D silhouette of the persons. On the other hand, the HOG-SVM method is unable to handle a large amount of within-class shape variations of the human body.

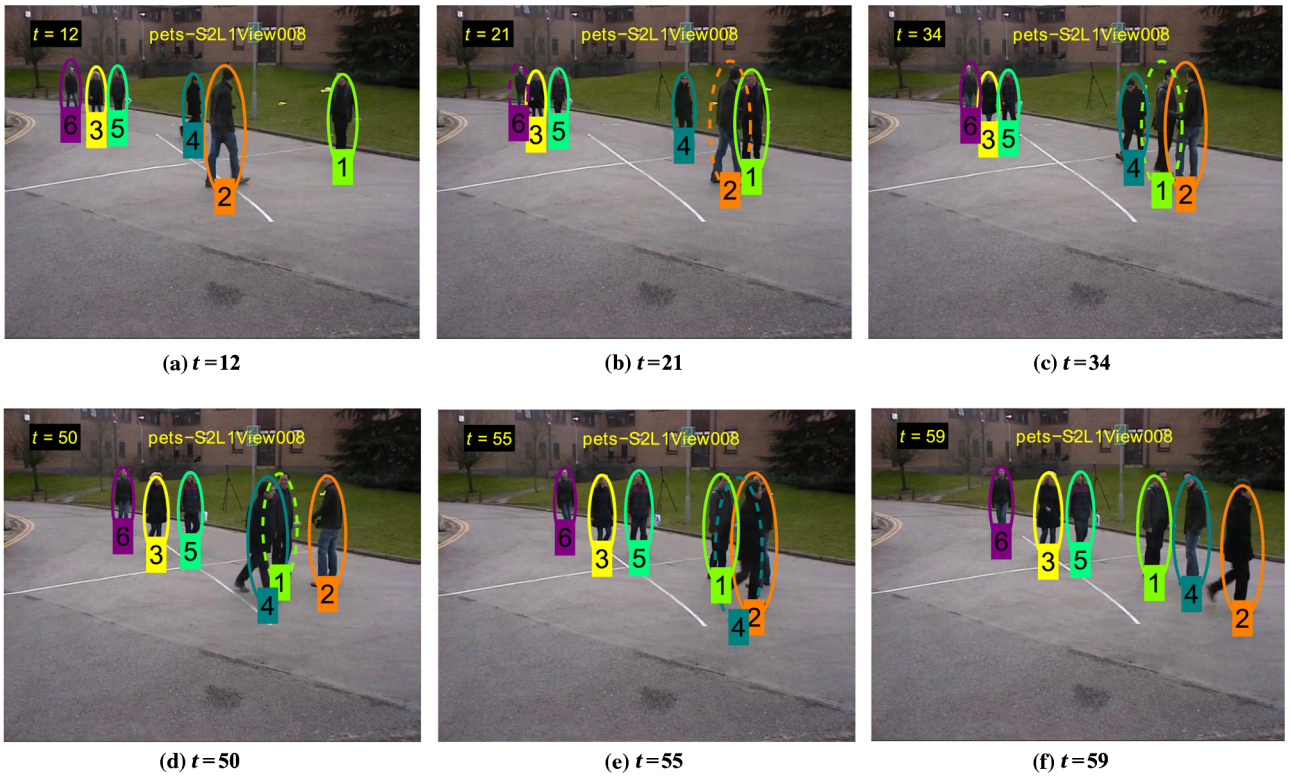
4.3 *Evaluation of Our Tracking Algorithm*

All of the test sequences show real sceneries where interactions and occlusions between individuals frequently occur. Complex situations of a variable number of interacting persons in the scenery are represented in the test sequences. There are also persons getting in and getting out from the field of vision of the camera. Additionally, there are severe occlusions that occur in long intervals of time. In some sequences, there are people wearing clothes with similar appearance features. There are also cases of individuals that change their trajectory when they interact with other people in the scenery.

Figure 3 shows visual examples of the results of our tracking algorithm for the sequence *OneStopMoveEnter1* of the CAVIAR[28] dataset. Figure 4 shows visual examples of the results of our tracking algorithm for the sequence *S2-L1-View_008* of the PETS[34] dataset. Figure 5 shows visual examples of the results of our tracking algorithm for the sequence *p3v1view1* of the UCO[35] dataset.

In Figs. 3 through 5, ellipses in the solid line style represent nodes of tracked targets that match with a detection measurement. Ellipses in the dashed line style represent the predicted successor of nodes for missing or occluded targets. Ellipses in a light gray color represent nodes that have to be confirmed by a detection measurement at subsequent frames. This happens in order to assign an identity to the new target or to show that the trajectory was generated by an FP in the detection measurements.

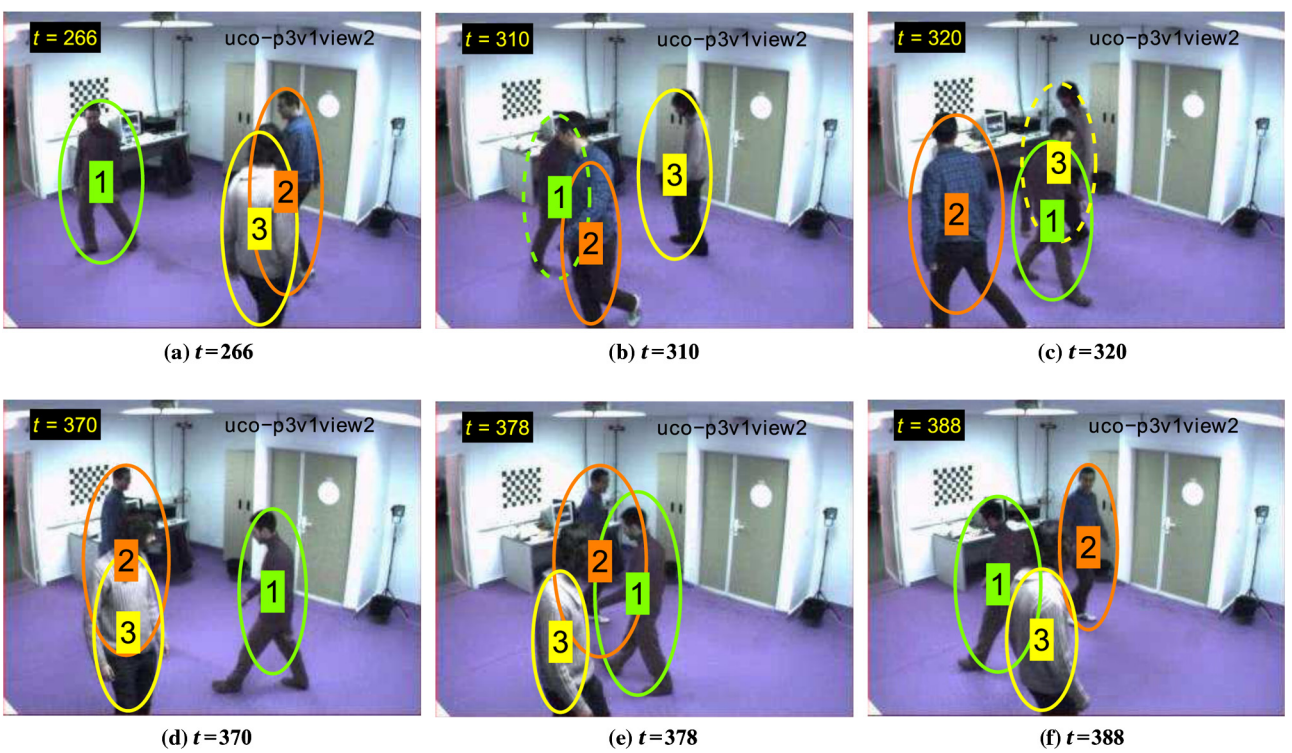
In Fig. 3, an FP is detected in a deferred way and its tracking immediately ends [Figs. 3(a) and 3(e)]. FN is acceptably handled by generating hypothesis' trajectories that keep the direction and speed of the tracked persons [Fig. 3(b), ID 4 and ID 7]. The occlusion between people with ID 5 and ID 6 is successfully resolved as shown in Figs. 3(d) through 3(f). The occlusion between people with ID 4 and ID 6 is also resolved as shown in Figs. 3(b) and 3(d). In Fig. 3(c), a trajectory was initialized for the measurement with ID −14. This measurement was not associated with the estimated trajectory of a person with ID 4 as there were significant variations in her attributes. However, one of the hypotheses built



(a) $t=12$ (b) $t=21$ (c) $t=34$

(d) $t=50$ (e) $t=55$ (f) $t=59$

Fig. 4 Results of our temporal association algorithm in the sequence *S2-L1-View_008* of the PETS[34] dataset. People are tracked correctly even when they are walking close [frames (a)–(d)]. Occlusion between persons with similar attributes is successfully resolved [frames (c)–(e), ID 4 and ID 1]. However, occlusion also produced identity switches between interacting persons with similar attributes [frame (f), ID 4 and ID 2]. This kind of error can be avoided by a better tuning of the parameters involved in the similarity association process.



(a) $t=266$ (b) $t=310$ (c) $t=320$

(d) $t=370$ (e) $t=378$ (f) $t=388$

Fig. 5 Results of our temporal association algorithm in the sequence *p3v1view1* of the UCO[35] dataset. Interactions and occlusion among persons are successfully resolved even when there are changes in the illumination of the scenery. Our algorithm can also track targets with partial visibility [frames (d)–(f)]. Keeping the tracking of objects in these situations is possible due to the robustness of the update in the people's attributes.

for the person with ID 4 at the beginning of the occlusion allows her to be correctly tracked in Fig. 3(d).

In Fig. 4, groups of persons with similar appearance attributes are walking at the same speed and direction (ID 3, ID 5, and ID 6). They are tracked correctly even though they are walking too close and there are missing people's measurements [Figs. 4(a) through 4(c)]. Complex interactions and occlusions between the persons with ID 1 and ID 2 are successfully resolved [Figs. 4(b) and 4(c)]. Occlusion of the person with ID 1, who remains motionless during the event, is correctly handled even when his occlusor with ID 4 has similar features [Figs. 4(d) and 4(e)]. Occlusion between the persons with ID 4 and ID 2 switches their identities during the event. This happened because they were placed too spatially close with very similar features of appearance and size. In order to correctly handle the association of identities in these conditions, for our algorithm, a better tuning of the parameters of the similarity association process is required.

In Fig. 5, partial and total occlusions commonly occur among individuals. The person with ID 1 keeps his identity although he is occluded over and over again by the other persons in the scene. The person with ID 2 keeps his identity whereas there are changes in his attribute of size during the tracking. The person with ID 3 is successfully tracked in spite of the fact that his appearance attributes vary due to the partial visibility of his body and changes in the illumination of the scene. Keeping the tracking of objects in these situations is possible due to the robustness of the update in the people's attributes.

Figure 6 shows sample results of tracking methods for the CAVIAR[28] dataset. As we can see in this figure, our method is able to track the persons that are partially occluded. Visual results of our method show the location of the persons more accurately than the other methods.

Table 1 shows the metrics we used to quantitatively evaluate our algorithm. These metrics were obtained from Song et al.[22] work. Table 2 shows a comparison of the results of

Fig. 6 Sample results of tracking methods for the CAVIAR[28] dataset. Frames of row 1 are from the *OneStopOneWait1* sequence. Frames of row 2, row 3, and row 4 are from the *OneStopMoveEnter1* sequence. Row 1 shows results of the tracking method using Bayesian combination of edgelet part detectors.[6] Row 2 shows results of a stochastic graph evolution framework used for tracking.[22] Row 3 shows results of the tracking method of global data association using networks flows.[18] Row 4 shows the results of our tracking algorithm.

Table 1 Evaluation metrics for object tracking.

Metric name	Definition
GT (ground truth)	Number of ground truth trajectories.
MT (mostly tracked)	Percentage of GT trajectories which are covered correctly by the tracking algorithm more than 80% of the time.
PT (partially tracked)	Percentage of GT trajectories which are covered correctly by the tracking algorithm between 20% and 80% of the time.
ML (mostly lost)	Percentage of GT trajectories which are covered correctly by the tracking algorithm less than 20% of the time.
Frag (Fragments)	The number of times that the ID of a tracked target changed along a GT trajectory.
IDS (ID switches)	The number of times that a tracked target changes its ID with another target.

Table 2 Comparison of tracking algorithms for the CAVIAR[28] dataset.

Reference	GT	MT (%)	PT (%)	ML (%)	Frag	IDS
Global data association using network flows[18]	140	85.7	10.7	3.6	20[a]	15[a]
Bayesian combination of edgelet part detectors[6]	140	75.7	17.9	6.4	35[a]	17[a]
Stochastic graph evolution framework[22]	75	84	12	4	6	8
Basic particle filter[22]	75	53.3	36	10.7	15	19
Our proposed tracking algorithm	72	88.9	11.1	0	21	6

[a]The number of fragments and identity switches were obtained using traditional metrics. The metrics we adopted are more stringent.

our algorithm for the CAVIAR[28] dataset with some of the state-of-the-art algorithms. Table 3 shows the evaluation of our tracking algorithm in the reference sequences of CAVIAR,[28] PETS,[34] and UCO.[35]

The suitable handling of trajectory's hypotheses allows our algorithm to produce a continuous and stable tracking of people when there are FNs in the detection measurements as well as events of occlusion. This quality is illustrated in the results of Tables 2 and 3 through the MT and PT evaluation indicators which measure the track completeness. The Frag and IDS indices are errors that indicate the lack of continuity of the tracks. In the evaluation of our algorithm, these indices are small in all of the reference sequences. This happened because we described the attributes of the people in detail, and also because we designed our temporal association algorithm to be robust to complex interactions and occlusion situations.

Table 3 Evaluation of our proposed tracking algorithm in surveillance sequences.

Reference	GT	MT (%)	PT (%)	ML (%)	Frag	IDS
CAVIAR	72	88.9	11.1	0	21	6
UCO	101	97	3	0	28	6
PETS	111	93.7	5.4	0.9	17	5

Compared with state-of-the-art algorithms, our algorithm performs well, especially if we consider that its nature is sequential, i.e., the temporal association is made without delay once the measurement information between two consecutive frames is available. This is worthy because we achieved better tracking results than trajectory-based algorithms such as those proposed by Zhang et al.[18] and Song et al.[22]

Experiments show that our algorithm is effective for tracking multiple people in sceneries, where there are varying illumination conditions as well as lack of restrictions about the clothing and motion directions of the interacting individuals. In general terms, our algorithm performs successfully during interaction and occlusion situations, such as: (1) people who walk in opposite directions during occlusions, (2) people who walk in duo toward the same destination, and (3) people who remain motionless during an occlusion for a long period of time.

5 Conclusions

This work addressed the problem of tracking multiple people under uncontrolled sceneries. It proposed a solution to the problem of partial and total occlusion using information acquired by a single camera.

Our temporal association algorithm builds a tracking graph to model the appearance, shape, and motion attributes of the tracked persons, as well as the interactions among them. The algorithm associates people detected in the current frame with available trajectories of people using a similarity matching process based on the appearance and spatial

motion similarity, and the location proximity. Our algorithm analyzes the spatio-temporal relationships between trajectories represented in the tracking graph to handle wrong or missing information in the detection stage. The algorithm also allowed the prediction of partial or total occlusions, and the estimation of the location of the people that were occluded in a period of time.

Our algorithm is robust to variations in the people's clothing appearance, size, and motion of the persons, and partial and total occlusions. This happens because the update of the person's attributes is carried out depending on how people interact in the scene and how they take part in the occlusion events.

Our algorithm was tested using detection measurements provided by our proposed people detection scheme, unlike our previous approach tested which used noisy detection measurements obtained by the ground truth. In the reference sequences of CAVIAR,[28] PETS,[34] and UCO,[35] our tracking algorithm achieved an overall precision of 93%. In these reference sequences, we demonstrated that our algorithm yields robust tracking of partially and totally occluded people, even when they are occluded over long periods of time. In the CAVIAR[28] sequences, we demonstrated that our detection-based algorithm outperforms even trajectory-based state-of-the-art algorithms.

As future work, we propose two research lines. As a first line, we will apply the proposed method to images sequences containing persons using a camera in movement. Initially, we can use a pan-tilt camera, which allows movement in the vertical and horizontal planes. This task requires the background model to adapt to the scene movement. It also requires a reasoning engine that allows reevaluating the tracked objects' positions and their interaction from the newly available information. As a second line, we plan to extend our method to track objects of different class (i.e., persons, vehicles, luggage, etc.) in a stationary scenery. This requires adapting the detection scheme to recognize an object among those of different classes. Then we need to define a specific data representation for each object class. Additionally, we need to adapt the interaction cases and the similarity metrics to properly deal with the interaction of objects of different classes.

Acknowledgments

The first author acknowledges support from National Council of Science and Technology (CONACyT) of Mexico to pursue graduate studies.

References

1. M. Wu et al., "Patches-based Markov random field model for multiple object tracking under occlusion," *Signal Process.* **90**(5), 1518–1529 (2010).
2. A. Senior et al., "Appearance models for occlusion handling," *Image Vision Comput.* **24**(11), 1233–1243 (2006).
3. B. Zhang, W. Tian, and Z. Jin, "Efficient hybrid appearance model for object tracking with occlusion handling," *Opt. Eng.* **46**(8), 087202 (2007).
4. Z. Wang et al., "Camshift guided particle filter for visual tracking," *Pattern Recognit. Lett.* **30**(4), 407–413 (2009).
5. R. Vezzani, C. Grana, and R. Cucchiara, "Probabilistic people tracking with appearance models and occlusion classification: the AD-HOC system," *Pattern Recognit. Lett.* **32**(6), 867–877 (2011).
6. B. Wu and R. Nevatia, "Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors," *Int. J. Comput. Vision* **75**(2), 247–266 (2007).
7. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *Comput. Vision Pattern Recognit.*, pp. 886–893, IEEE Computer Society (2005).
8. Z. Lin et al., "Hierarchical part-template matching for human detection and segmentation," *Comput. Vision Pattern Recognit.*, pp. 1–8, IEEE Computer Society (2007).
9. M. Andriluka, S. Roth, and B. Schiele, "People-tracking-by-detection and people-detection-by-tracking," *Comput. Vision Pattern Recognit.*, pp. 1–8, IEEE Computer Society (2008).
10. S. M. Khan and M. Shah, "Tracking multiple occluding people by localizing on multiple scene planes," *Pattern Anal. Mach. Intell.* **31**(3), 505–519 (2009).
11. R. Munoz-Salinas et al., "Multi-camera people tracking using evidential filters," *Int. J. Approx. Reason.* **50**(5), 732–749 (2009).
12. R. Kaucic et al., "A unified framework for tracking through occlusions and across sensor gaps," *Comput. Vision Pattern Recognit.*, Vol. 1, pp. 990–997, IEEE Computer Society (2005).
13. J. Ryu et al., "Camera placement for minimizing occlusion in object tracking systems," *J. Ubiquitous Convergence Technol.* **3**(1), 13–19 (2009).
14. A. Yilmaz, O. Javed, and M. Shah, "Object tracking: a survey," *ACM Comput. Surv.* **38**(4), 1–45 (2006).
15. W. Hu et al., "A survey on visual surveillance of object motion and behaviors," *Syst., Man Cybernet.* **34**, 334–352 (2004).
16. S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," *Aerosp. Electron. Syst. Mag.* **19**(1), 5–18 (2004).
17. Y. Bar-Shalom, *Tracking and Data Association*, Academic Press Professional, Inc., San Diego, California (1987).
18. L. Zhang, Y. Li, and R. Nevatia, "Global data association for multi-object tracking using network flows," *Comput. Vision Pattern Recognit.*, pp. 1–8, IEEE Computer Society (2008).
19. X. Wang et al., "Tracking interacting objects optimally using integer programming," in *Eur. Conf. on Computer Vision*, pp. 17–32, Springer International Publishing (2014).
20. B. Yang and R. Nevatia, "An online learned CRF model for multi-target tracking," *Comput. Vision Pattern Recognit.*, pp. 2034–2041, IEEE Computer Society (2012).
21. R. T. Collins, "Multitarget data association with higher-order motion models," *Comput. Vision Pattern Recognit*, pp. 1744–1751, IEEE Computer Society (2012).
22. B. Song et al., "A stochastic graph evolution framework for robust multi-target tracking," in *Eur. Conf. on Computer Vision*, pp. 605–619, Springer Berlin Heidelberg (2010).
23. W. Brendel, M. Amer, and S. Todorovic, "Multiobject tracking as maximum weight independent set," *Comput. Vision Pattern Recognit.*, pp. 1273–1280, IEEE Computer Society (2011).
24. M. Han et al., "Multi-object trajectory tracking," *Mach. Vision Appl.* **18**(3), 221–232 (2007).
25. J. Liu et al., "Automatic player detection, labeling and tracking in broadcast soccer video," *Pattern Recognit. Lett.* **30**(2), 103–113 (2009).
26. Y. Ma, Q. Yu, and I. Cohen, "Target tracking with incomplete detection," *Comput. Vision Image Understand.* **113**(4), 580–587 (2009).
27. C. Reta et al., "Occlusion model from human interaction analysis for tracking multiple people," in *IASTED Int. Conf. on Signal Processing, Pattern Recognition and Applications*, ACTA Press (2013).
28. CAVIAR, "Context aware vision using image-based active recognition benchmark data," EC Funded CAVIAR project/IST 2001 37540, http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/ (22 December 2014) (2005).
29. J. Yao and J. Odobez, "Multi-layer background subtraction based on color and texture," *Comput. Vision Pattern Recognit.*, pp. 1–8, IEEE Computer Society (2007).
30. UCLA Department of Statistics, "Dataset of people images," http://www.stat.ucla.edu/~ywu/AB/people200clusterRRCode.zip (22 December 2014) (2011).
31. Y. N. Wu et al., "Learning active basis model for object detection and recognition," *Int. J. Comput. Vision* **90**(2), 198–235 (2010).
32. G. Welch and G. Bishop, "An introduction to the Kalman filter," Technical Report TR 95-041, University of North Carolina at Chapel Hill, Department of Computer Science, (1995).
33. H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Res. Logist.* **52**(1), 7–21 (2005).
34. PETS, "Performance evaluation of tracking and surveillance benchmark data," *Computational Vision Group, Reading University*, UK, Winter-PETS 2009 in Conjunction with IEEE Computer Society, http://cs.binghamton.edu/~mrldata/pets2009 (22 December 2014) (2009).
35. UCO, "Multi-camera pedestrian videos data set," *Aplicaciones de la Visión Artificial (AVA) group*, University of Cordoba, Spain (2011).

Carolina Reta obtained her MS degree in computer science in 2009 and her PhD degree in computer science in 2014 from the National Institute for Astrophysics, Optics, and Electronics, Mexico. Her research interests include computer vision, image and video processing, data hiding, and data mining for image analysis.

Leopoldo Altamirano received his MS degree in electric engineering from CINVESTAV, Mexico, in 1991 and his PhD degree from Technische Universität München, Germany, in 1996. Since 1999, he has been a member of the National System of Researchers, level 2. He is a professor at the Computer Science Department and head of the Computer Vision Laboratory at the National Institute for Astrophysics, Optics, and Electronics, Mexico. His research lies in computer vision, sensor fusion, and industrial applications.

Jesus A. Gonzalez obtained his MS and PhD degrees in computer science and engineering in 1999 and 2001, respectively, from the University of Texas at Arlington. Since 2001, he has held a tenure track position from the National Institute for Astrophysics, Optics, and Electronics, Mexico. His areas of interest focus in data science, including machine learning, data mining, and structural knowledge representations. He applies his research to geographic information systems, remote sensing, medical applications, and bioinformatics.

Rafael Medina-Carnicer received his BS degree in mathematics from the University of Sevilla, Spain, and his PhD degree in computer science from the Polytechnic University of Madrid, Spain, in 1992. Since 1993, he has been a lecturer of computer vision at Cordoba University, Spain. His research is focused on edge detection, evaluation of computer vision algorithms, 3-D vision, and pattern recognition.