An Object Assignment Algorithm for Tracking Performance Evaluation

N. Saunier, T. Sayed, K. Ismail
Department of Civil Engineering
University of British Columbia
Vancouver, British Columbia, Canada

Abstract

Performance evaluation of detection and tracking methods is a crucial issue. However, despite the efforts of the research community, there is still a lack of widely adopted methods, measures and benchmark data for this purpose. Most contributions have embraced pixel-based method, and do not report their results in terms of objects. The latter requires the assignment of ground truth objects to the detected objects, which can be very complex. To the authors' knowledge, this paper is the first to describe an algorithm for the explicit unique assignment of objects, including one-to-one (correct) assignments, one-to-many and many-to-one assignments (over-segmentations and over-groupings), missed and false detections. Quantitative performance measures are also presented, and the approach is illustrated on a set of traffic videos recorded at two different locations.

1. Introduction

The problem of evaluating automatically the performance of methods is critical in all fields. Assessment must be systematic and objective to enable better understanding of their strengths and limitations, and to allow cross-comparison of different methods. Depending on the performance measures, the task can be more or less automated. In the field of computer vision, video surveillance is one of the most important applications with widespread use. This work is related to transportation applications, in particular to the development of systems for traffic monitoring and automated road safety analysis [10]. Such applications require not only the detection of moving objects in each frame, but also the tracking of moving objects in the field of view or a defined region of interest. For performance evaluation, it is usually easier to collect aggregated or instantaneous measures and compare them automatically to the results of the method under study, than to evaluate the object tracking performance since it requires to assign to each other the *detected objects* (detected by the method under study) and the objects manually identified, or *ground truth objects*. Yet, similar to [14], we advocate that measuring the performance "in terms of tracks rather than frames is a natural choice that is consistent to the expectations of the end-users".

There are subjective and quantitative methods to evaluate the performance of a tracking algorithm. Since a goal of this work is the automated comparison of tracking results, this paper deals with the latter, which requires the annotation of video data to create the ground truth. Or rather to create *a* ground truth since there are many types, and there can be significant variations between annotators [6]. The ground truth may have various data representations, e.g. bounding boxes or pixel-accurate contours, and definitions, regarding e.g. the way to deal with occluded objects, groups of objects with consistent motion like groups of pedestrians, stationary objects, the definition of entrance and exit of the field of view or a restricted region of interest. Since producing pixel-accurate object contours is very time-consuming, the ground truth used in this work consists in sequences of bounding boxes, although the approach can be adapted to other types of ground truth.

Many approaches have been proposed for tracking performance evaluation. The two important elements are: 1) the definition of similarity between ground truth and detected objects, along temporal and spatial dimensions, and 2) the way the correspondence or assignment between them is made. The latter seems to have been overlooked in the literature, since the assignments may be quite complex and difficult to entangle, without a "true" or unique solution. Many approaches assign simply each ground truth object to the most similar detected objects. Yet, if all objects are compared to each other, the possible assignments are

- correct assignment (one to one): one detected object is tracked by one ground truth object,
- over-segmentation (one ground truth object to many detected objects),
- over-grouping (many ground truth objects to one detected object),
- multiple assignments (many ground truth objects to many detected objects),
- missed detection (one ground truth object to zero detected object),

- false detection (zero ground truth object to one detected object).

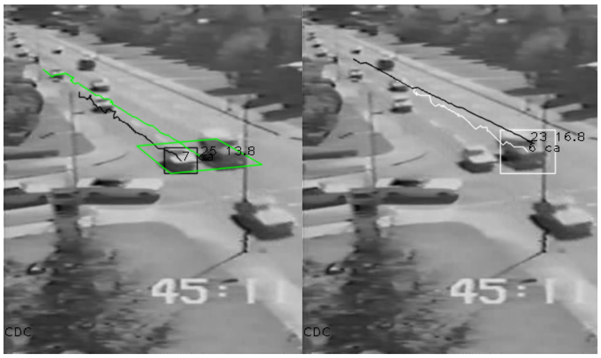

Figure 1: Example of two correct object assignments, ground truth object 7 to detected object 25 on the left, ground truth object 6 to detected object 23 on the right. In this typical case of close objects with consistent motion, a simple unique assignment may first assign the ground truth object 6 to the detected object 25, in which case it is very likely that the ground truth object 7 will not be assigned to the detected object 23, which will count respectively as one missed detection and one false detection.

Multiple assignments are bound to happen when two or more objects move consistently close to each other. This may be "aggravated" by additional missed or false detections. Simple unique assignments may result in the identification of more false and missed detections and was found to be inadequate for our needs (See Figure 1). This work aims at addressing these issues, i.e. not identifying more missed and false detections than are actually generated, and to provide the users of the applications detailed results in terms of the various possible assignments listed above, except multiple assignments which are typically the merging of the correct assignments, over-segmentations and over-groupings. The main reason for providing such detailed results is that these various types of errors do not have the same importance for different applications (more generally, there is no unique performance measure that suits all applications). To our knowledge, this work is the first to present a method to identify automatically explicitly and uniquely the one-to-many assignments of ground truth and detected objects.

The rest of the paper is organized as follows. The next section presents the related work. Section 3 describes a method for the assignment of ground truth and detected objects, as well as performance measures. Section 4 illustrates the method on video sequences, and section 5 concludes the paper.

2. Previous Work

Performance evaluation has generated a lot of interest over the last ten years. The best known is the Performance Evaluation of Tracking and Surveillance (PETS)[1] program, which organizes regular workshops and provides benchmark datasets. Other datasets are made available by the Context Aware Vision using Image-based Active Recognition (CAVIAR)[2] and the French Video Understanding Evaluation (ETISEO)[3] projects. Yet no standard has yet been widely adopted. The challenges are compounded by the lack of publicly available object tracking implementations for comparison.

Some very recent work [6, 7] proposes more exhaustive methodologies. As reported in [7], most contributions to the performance evaluation of detection and tracking algorithms have embraced pixel-based methods within a receiver-operating characteristic (ROC) framework. Yet this work deals only with frame-based performance measures. It relies on the F-measure and ROC-like analysis using precision-recall to present a comprehensive approach to the comparison and search for the optimal operating point of motion detection methods.

A large scale dataset and performance evaluation is presented in [6]. The objects are assigned "optimally" through the optimized search of all possible assignments. However, only one-to-one assignments are allowed. Two sets of performance measures are defined. The work described in [3] relies also on one-to-one assignments, although it mentions counting track fragmentation, i.e. the number of detected objects assigned to a ground truth object (it is nevertheless not clear how this is done). The original contribution of this work is the creation of synthetic video sequences for performance evaluation, although one can wonder whether the resulting videos are realistic and challenging enough.

The approaches described in [1, 8] are the closest to the method presented in this paper. The authors of [1] identify what they call "merge error scenarios" that correspond to over-segmentations and over-groupings. Yet, the performance measures are counted in frames. It is the same issue with [8]. Interestingly, this work makes use of a graph representing the similarity relationships between objects, but deal only with simple assignments of one-to-many objects, removing "invalid configurations" based on the spatial continuity of object assignments.

[4] presents a two-pass assignment scheme that processes first the ground truth objects, and may assign them to many detected objects, then processes the detected objects, and can assign them as well to many ground truth objects, without taking into account the previous assignments of the

[1] http://www.cvg.cs.rdg.ac.uk/slides/pets.html
[2] http://homepages.inf.ed.ac.uk/rbf/CAVIAR/
[3] http://www-sop.inria.fr/orion/ETISEO/

ground truth objects. Performance is reported in terms of the number of missed and false detections. [12] is a well-written article that presents a comprehensive set of metrics, yet resolves to the simple independent assignment of each ground truth object to the most similar detected object, then of each detected object to the most similar ground truth object, similarly to [4]. There is no illustration on experimental data. The method described in [14] can assign ground truth objects to many detected objects, which counts as a correct detection. However, it is not clear how the situation in which a ground truth object is assigned to more than one detected object is accounted for, since the detected objects do not meet the conditions to be detected as false positives. A ground truth objects assigned to two or more detected objects that exist simultaneously will not be detected by the measures of track fragmentation and ID change. Over-groupings cannot be identified in the proposed method.

As stated in the introduction, none of the existing work describes a method to identify explicitly and uniquely the assignment of ground truth and detected objects.

3. The Proposed Approach

3.1. An Algorithm for Object Assignment

The goal of the method presented in this section is to assign ground truth and detected objects. The "allowed" assignments are: correct assignment, over-segmentation, over-grouping, missed and false detections (not multiple assignments). This work is applied to a feature-based detection and tracking algorithm. Therefore, the detected objects are the result of the grouping of feature trajectories. At each instant, the position of a detected object is determined by a set of feature positions that can be matched to the ground truth bounding boxes. Nonetheless, the proposed approach is flexible and can accommodate various formulations of spatial matching between objects as have been proposed in the literature. The following list introduces the notations used in the remainder of this paper:

- GT_i denotes the i^{th} ground truth object trajectory, $GT_i(t)$ the position and size of the bounding box at time t.
- D_j denotes the j^{th} detected object trajectory, $D_j(t)$ the positions of the group of features constituting the object at time t.
- N_{frames} is the number of frames in the sequence.
- $\{GT_{i1}, GT_{i2}, ..., GT_{iN}, D_{j1}, D_{j2}, ..., D_{jM}\}$ denotes a set of N ground truth objects assigned to M detected objects. The algorithm results will be either one ground truth object assigned to any number of detected objects (0 to M), or a detected object assigned to any number of ground truth object (0 to N).
- N_{GT}, N_D, N_{CA}, N_{OS}, N_{OG}, N_{MD} and N_{FD} are respectively the number of ground truth objects, detected objects, correct assignments, over-segmentations, over-groupings, missed detections and false detections in a sequence.
- Let $Time(O)$ by the temporal interval during which an object O exists, let $Length(I)$ be the length of the temporal interval I (if $I = [a, b]$, $Length(I) = b - a$).

A good data representation of the assignment between ground truth and detected objects is an undirected graph, where vertices represent objects and edges represent object assignments (See Figure 2). Such a graph is very similar to a bipartite graph: the set of vertices is partitioned in two subsets, corresponding to ground truth and detected objects, and edges can connect only vertices of the two different subsets. It may not be a bipartite graph since some vertices may have no connection at all, corresponding to either a missed or false detection. Multiple assignments will be represented by connected components containing many ground truth and detected objects. The goal is therefore to remove some edges so that only the allowed assignments are left and the algorithm performance may be evaluated. This paper proposes an algorithm to do this assignment.

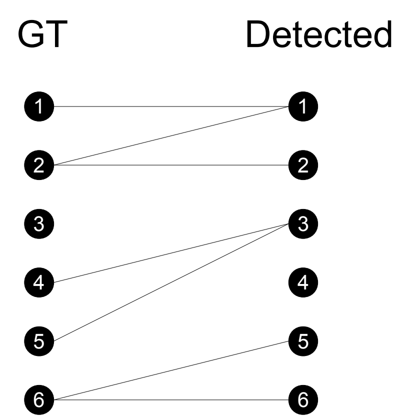

Figure 2: Example of assignment graph between 6 ground truth and 6 detected objects. $\{GT_6, D_5, D_6\}$ is an over-segmentation, $\{GT_4, GT_5, D_3\}$ is an over-grouping, GT_3 is a missed detection, D_4 is a false detection, and, as such, $\{GT_1, GT_2, D_1, D_2\}$ should probably be counted as two correct assignments, $\{GT_1, D_1\}$ and $\{GT_2, D_2\}$, after the edge linking GT_2 to D_1 is removed.

First the graph is built by adding an edge between a ground truth object and a detected object if they are considered similar enough (for convenience, objects and their corresponding vertices in the graph are used interchangeably). There are many ways to evaluate the similarity of ground truth and detected objects. In this work, a ground truth object GT matches a detected object D if the following spatial and temporal conditions are satisfied:

- if the types of the objects are available, they should be identical,
- temporal matching: to ensure that similar objects exist simultaneously for a temporal interval of a minimum length, the following condition is enforced $Length(Time(GT) \cap Time(D)) \geq \alpha\, max(Length(Time(GT)), Length(Time(D)))$
- spatial matching: many similarity measures or distances can be used, with a threshold, for the purpose of comparing object trajectories. The method and results presented in this work rely on matching objects in each frame, and counting a minimum number of matchings over the temporal interval of coexistence $Time(GT) \cap Time(D)$. The number of feature positions $D(t)$ located within the bounding box $GT(t)$ must be superior to a threshold $N_{features}$. The number of frames in which the previous condition is satisfied must be superior to a proportion β of $Length(Time(GT) \cap Time(D))$.

The above matching definition is called *complete matching*, as a reference to the temporal matching condition that enforces a large overlap of the two objects. It was also found useful to modify the temporal matching condition in order to identify partial matchings between objects. , so that objects detected for a short time, too short for temporal matching with a ground truth object, may be identified as partial detection rather than counted as false alarms. Ground truth objects may be assigned partially to detected objects rather than counted as missed detections. For that purpose, the temporal matching for GT and D is $Length(Time(GT) \cap Time(D)) \geq \alpha\, min(Length(Time(GT)), Length(Time(D)))$. This second matching definition is called *partial matching*. The parameters for object matchings are α, β and $N_{features}$ and could be different for the two matching definitions.

Once the graph is built (See an example of such a graph in Figure 2), it may need to be simplified by removing edges so that the graph is composed only of allowed assignments, i.e. there are no multiple assignments. Algorithm 1 performs this task. Let $N(O)$ be the set of adjacent objects to O, also called open neighbourhood in graph theory (if O is a ground truth object, $N(O)$ may contain only detected objects, and vice versa). Let $Isolated(O) \subset N(O)$ be the set of objects adjacent only to O, i.e. with degree 1.

Although one could pass first over the objects of one type, then the other, there is in fact no need to make any distinction. All the objects are assigned when the algorithm is finished, i.e. after one pass of the algorithm over each object (See the proof in the Appendix on page 6). The main goal of this algorithm is to maximize the number of correct assignments and minimize the number of false and missed detections by assigning them to over-segmentations or over-groupings if possible. An over-segmentation will be identified only between a ground truth object and detected objects adjacent only to this ground truth object (and similarly for over-groupings).

Algorithm 1 Algorithm for the assignment of ground truth to detected objects.

When objects are assigned, the edges linking them to all the objects other than the ones they are assigned to are removed.
for all objects O **do**
  if $N(O) = \emptyset$ **then**
    O is considered assigned. If O is a ground truth (resp. detected) object, a missed (resp. false) detection is counted.
  else
    if $Isolated(O) = \emptyset$ **then**
      if there exists an object $O' \in N(O)$ such that $Isolated(O') = \emptyset$ **then**
        O is assigned to O', and a correct assignment is counted.
    else if $Isolated(O) = \{O'\}$ **then**
      O is assigned to O', and a correct assignment is counted.
    else
      O is assigned to $Isolated(O)$. If O is a ground truth (resp. detected) object, an over-segmentation (resp. over-grouping) is counted.

When processing the graph described in Figure 2, starting for example with ground truth objects, nothing will be done for the first ground truth object GT_1 since $Isolated(GT_1) = \emptyset$, $D(GT_1) = \{D_1\}$ and $Isolated(D_1) = \{GT_1\}$. Next, when considering GT_2, the assignment $\{GT_2, D_2\}$ will be done since $Isolated(GT_2) = \{D_2\}$. As a result of this assignment, the edge linking GT_2 to D_1 will be removed. The rest of the assignments will be done in one pass over the remaining ground truth and detected objects (GT_1 will be assigned to D_1 when the latter is considered).

3.2. Performance Measures

The results of the assignment can be first used to visualize the method performance, visually inspecting the correct trackings and the errors. Another objective is the quantitative evaluation of the performance. A first set of measures is the number of assignments of each type. Proportions may be computed by normalizing by the maximum number that could be reached, i.e. by dividing the numbers of over-segmentations and missed detections by the number of ground truth objects, and by dividing the numbers of over-groupings and false detections by the number of detected objects.

For some use, for example the automated search for the parameters of a tracking algorithm that optimize its performance, a unique performance measure is convenient. For that purpose, costs may be associated with each type of error, based on the application. In some cases, over-segmentation may be less of an issue than missed objects, e.g. to detect interactions between objects, while in other false detections may not be tolerated, e.g. when a system acts upon the detection of objects. A cost function can be computed as

$$Cost = N_{OS}C_{OS} + N_{MD}C_{MD} + N_{OG}C_{OG} + N_{FD}C_{FD} \quad (1)$$

where C_{OS}, C_{OG}, C_{MD} and C_{FD} are respectively the costs associated with over-segmentations, over-groupings, missed detections and false detections. One can think of other forms of cost functions, e.g. non-linear with respect to some of the numbers, that could also take into account the number of assigned objects in over-segmentations and over-groupings, and with normalization to compare results on different datasets. The cost function defined in Equation 1 can be normalized in the following way.

$$\overline{Cost} = \frac{N_{OS}C_{OS} + N_{MD}C_{MD}}{N_{GT}} + \frac{N_{OG}C_{OG} + N_{FD}C_{FD}}{N_D} \quad (2)$$

Alternatively, one can also choose to count the over-segmentations and over-groupings as correct assignments with respectively some false and missed detections. Precisely, an over-segmentation (resp. over-grouping) with N detected (resp. ground truth) objects can be considered as one correct assignment and $N-1$ false (resp. missed) detections. This reduces the number of performance measures, at the expense of details. A corresponding simplified cost function can be used in this case: $Cost' = N_{TMD}C_{TMD} + N_{TFD}C_{TFD}$, with revised total numbers of missed and false detections N_{TMD} and N_{TFD} and corresponding costs C_{TMD} and C_{TFD}.

Again, these choices are heavily dependent on the application and the end user expectations, and it is difficult to provide further generic recommendations. It is possible to compute multiple performance evaluations with various matching parameters, or look for optimized evaluation parameters as in [7]. The approach advocated in practise by the authors of this work is to do at least two evaluations using the complete matching and partial matching definitions and to combine visual inspection of the results with systematic search through parameter settings.

Other potentially useful performance measures from the literature are the latency of detection [14] and the average matching distance for correct assignments.

4. Experimental Results

A set of traffic videos with mixed traffic has been exhaustively annotated. It contains 10 video sequences, for a total length of 5793 frames. 213 moving objects have been manually tracked. The feature-based algorithm described in [9], adapted from [2], is used for moving object detection and tracking. The matching parameters are selected trough trial and error, and visual inspection of the resulting assignments until the results are satisfactory for the end user. The following settings are chosen: $\alpha = 0.5$, $\beta = 0.5$ and $N_{features} = 1$. All the experimental results presented here rely on the partial object matching definition. A raw sample of the text output of the object assignment for one video sequence with 37 ground truth objects follows (assignments are separated by a space and denoted by ':'; on the left-hand side are the identifying number(s) (id) of ground truth object(s), and on the right-hand side are the id of detected object(s)):

```
correct: 11:13 14:25 17:24 23:35
over-segmentations: 1:0,1 9:8,9,10 15:18,
19,20,21 20:26,27,28,29 28:37,38,39
over-groupings: 33,34:46
missed: 0 2 3 4 5 6 7 8 10 12 13 16 18
 19 21 22 24 25 26 27 29 30 31 32 35 36
false: 2 3 4 5 6 7 11 12 14 15 16 17 22
 23 30 31 32 33 34 36 40 41 42 43 44 45
 47 48 49 50 51
```

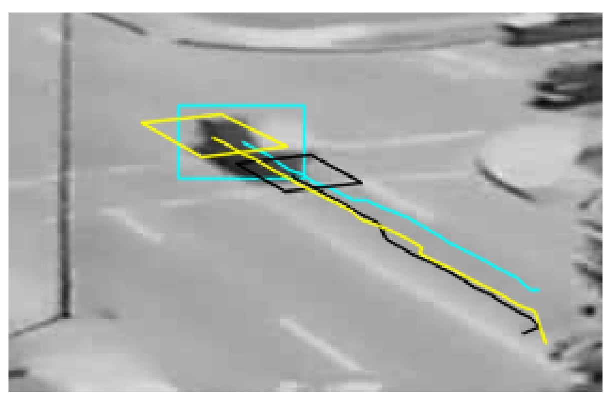

Figure 3: Example of display of an over-segmentation: the blue bounding box around the vehicle is the ground truth in this frame, and it is assigned to 2 detected objects in black and yellow (tracks represent past object positions).

The user can also inspect visually each type of assignment overlaid over the video data (See an example of over-segmentation in Figure 3). The main parameters of the feature-based object detection and tracking method are the connection distance $D_{connection}$ which controls the distance within which features are grouped, the segmentation distance $D_{segmentation}$ which is a threshold on the relative displacement of features to disconnect them, and the maximum distance $MaxDist$ which is a threshold on the maximum distance between connected features. A homography matrix has been estimated for the video sequences, and the parameters are therefore measured in world coordinates, in meters. The performance of the method can be analyzed for various settings of these three parameters. The same weights are used for the cost function, i.e. $C_{OS} = C_{OG} = C_{MD} = C_{FD} = 1$. Figure 4 illustrates the trade-off between over-segmentations and over-groupings, missed and false detections, by plotting the cost function, the numbers of correct assignments, over-segmentations, over-groupings, missed and false detections as a function of $D_{connection}$ for various values of $D_{segmentation}$, with $MaxDist = 10.5\ m$. The trade-off is clear for the numbers of missed and false detections: there is a general trend of a decrease in the number of missed detections and an increase in the number of false detections with increasing $D_{connection}$. The influence of $D_{segmentation}$ is less clear from these plots, although low values, e.g. $D_{segmentation} = 0.4\ m$, favour more missed detections and fewer false detections than higher ones, e.g. $D_{segmentation} = 2.8\ m$.

A search for optimal parameters can then be done. The approach followed on this dataset is to find the parameters minimizing the cost (with a small margin), and to select among them the parameters with the largest number of correct assignments (that should be as close as possible to the maximum over all parameters). For this dataset, this approach yields the following parameters: $D_{connection} = 7\ m$, $D_{segmentation} = 1.6\ m$ and $MaxDist = 10.5\ m$ (with a minimal cost of 113 and 109 correct assignments). Other parameter selections provide slightly higher costs with slightly smaller numbers of missed and false detections. This selection of parameters is different from the former reference settings of $D_{connection} = 4\ m$, $D_{segmentation} = 1.6\ m$ and $MaxDist = 6\ m$, obtained manually through trial and error, which yield a higher cost of 142 and a smaller number of correct assignments of 95, with many more over-segmentations (75 instead of 37) and similar numbers of missed and false detections.

A second dataset was labelled for about 1495 frames, with 61 ground truth moving objects. It is part of a study of the influence of the introduction of a scramble phase in a signalized intersection on pedestrian safety [5]. There is significant pedestrian traffic in this intersection and tracking proves very challenging for the feature-based tracking algorithm. Therefore, a search for the best parameters was done. In this case, the goal is to detect conflicts between pedestrians and vehicles, in which case the performance evaluation depends only on the number of missed detections. The numbers of missed and false detections are plotted as a function of $D_{connection}$ for various values of $D_{segmentation}$, with $MaxDist = 2\ m$ in Figure 5. Trends are fairly similar to the observations made for the first dataset. Minimizing the number of missed detections leads to the selection of $D_{connection} = 0.64\ m$, $D_{segmentation} = 0.6\ m$ and $MaxDist = 1.75\ m$ with 5 missed detections, 56 false detections, 37 correct assignments and a cost of 70. These parameter values are close to the choice made in [5] of $D_{connection} = 0.85\ m$ and $D_{segmentation} = 0.4\ m$ ($MaxDist$ was not optimized and set to $1\ m$). Following the same search approach as previously leads to the selection of $D_{connection} = 0.21\ m$, $D_{segmentation} = 0.4\ m$ and $MaxDist = 1\ m$ with 19 missed detections, 20 false detections, 31 correct assignments and a cost of 48.

5. Summary and Conclusions

This paper has presented a novel algorithm for the assignment of ground truth and detected objects. It enables a more detailed evaluation of tracking methods, in a way that is well suited for visual inspections by end users. It lends itself also to quantitative evaluation and computation of various performance measures. The approach has been illustrated on a set of traffic videos recorded at two different locations.

Future work will focus on the development and study of extra quantitative measures. In particular, the measures should quantify the spatial and temporal mismatches between ground truth and detected objects. They should account for each instant a ground truth (resp. detected) object has no match or is matched by more than one detected (resp. ground truth) object. Initial work was done on average measures of over-segmentation and over-grouping per object but was not included for lack of time. Per frame average over-segmentation and over-grouping should allow to discriminate simultaneous from consecutive over-segmentations and over-groupings. Other indicators of the literature will be adapted to the assignments produced by this approach and compared. This paper lays the groundwork for object-based tracking performance evaluation.

Appendix

Theorem 5.1. *When Algorithm 1 exits, all the objects are assigned.*

Proof. The case to be examined is that of objects that are

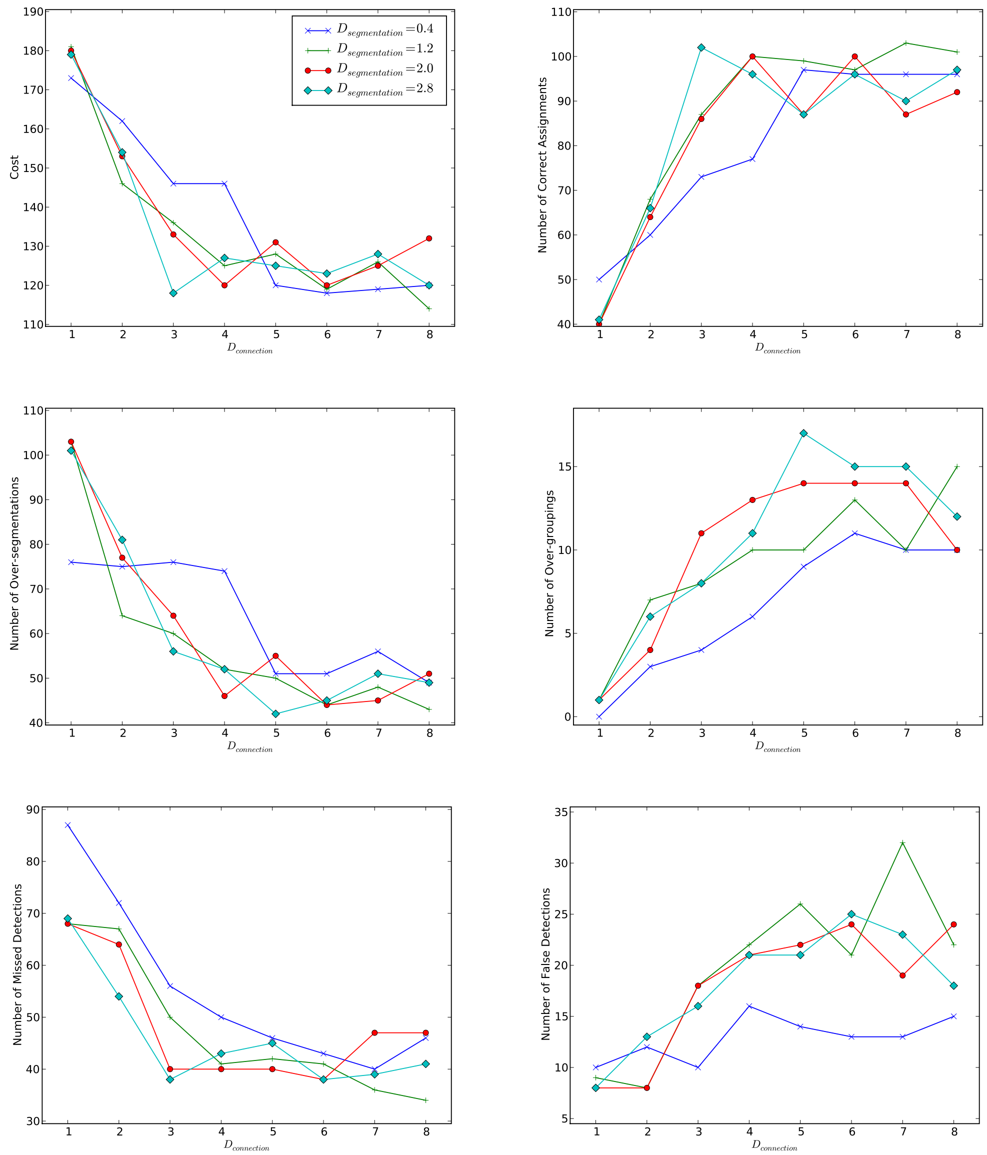

Figure 4: The cost function and the numbers of correct assignments, over-segmentations, over-groupings, missed and false detections are plotted as a function of $D_{connection}$ for various values of $D_{segmentation}$, with $MaxDist = 10.5\ m$ (the parameters are measures in meters).

not assigned after having been considered by the algorithm (each object is considered once in the main and only loop of the algorithm over all objects). Such an object O has the following properties:

$$N(O) \neq \emptyset \tag{3}$$

$$Isolated(O) = \emptyset \tag{4}$$

$$\forall O' \in N(O),\ Isolated(O') \neq \emptyset \tag{5}$$

The last condition (5) implies that all the objects $O' \in N(O)$ have not been considered yet by the algorithm (if they had, they would have been assigned to the non-empty set $Isolated(O')$).

There are two cases: either there exists an $O' \in N(O)$ such that $O \in Isolated(O')$, or $\forall O' \in N(O)$, $O \notin Isolated(O')$. In the first case, O' will be assigned to $Isolated(O')$ which includes O. In the second case, the algorithm will assign each $O' \in N(O)$ to $Isolated(O')$ except for the last object called O'_{last}. After these assignment(s), the edges connecting O to each object of $N(O) \setminus$

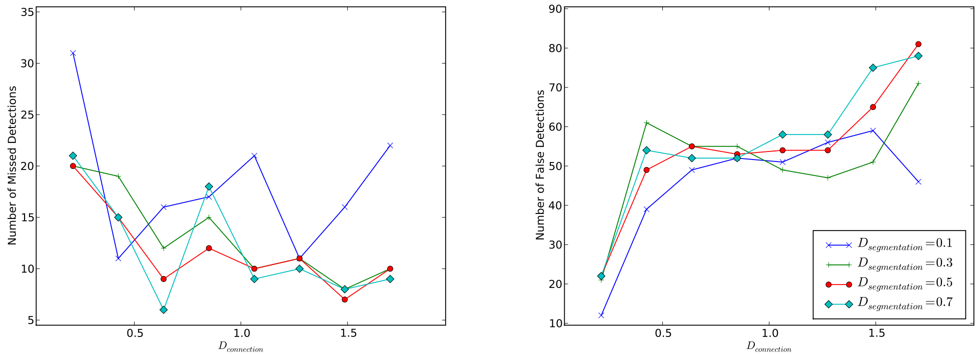

Figure 5: The cost function and the numbers of missed and false detections are plotted as a function of $D_{connection}$ for various values of $D_{segmentation}$, with $MaxDist = 2\ m$ (the parameters are measures in meters).

O'_{last} will be removed. At this point, $O \in Isolated(O'_{last})$ and $Isolated(O'_{last})$ will be assigned to O'_{last}. $\square$

References

[1] F. Bashir and F. Porikli. Performance evaluation of object detection and tracking systems. In *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*, June 2006.

[2] D. Beymer, P. McLauchlan, B. Coifman, and J. Malik. A real-time computer vision system for measuring traffic parameters. In *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, pages 495–501, Washington, DC, USA, 1997. IEEE Computer Society.

[3] J. Black, T. Ellis, and P. Rosin. A novel method for video tracking performance evaluation. In *Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS)*, pages 125–132, Nice, France, October 2003.

[4] L. M. Brown, A. W. Senior, Y. l. Tian, J. Connell, A. Hampapur, C. f. Shu, H. Merkl, and M. Lu. Performance evaluation of surveillance systems under varying conditions. In *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, January 2005.

[5] K. Ismail, T. Sayed, and N. Saunier. Automated pedestrian safety analysis using video data in the context of scramble phase intersections. In *Annual Conference of the Transportation Association of Canada*, Vancouver, B.C., October 2009.

[6] R. Kasturi, D. Goldgof, P. Soundararajan, V. Manohar, J. Garofolo, R. Bowers, M. Boonstra, V. Korzhova, and J. Zhang. Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 31(2):319–336, February 2009.

[7] N. Lazarevic-McManus, J.R. Renno, D. Makris, and G.A. Jones. An object-based comparative methodology for motion detection based on the f-measure. *Computer Vision and Image Understanding*, 111(1):74–85, 2008. Special Issue on Intelligent Visual Surveillance (IEEE).

[8] J.C. Nascimento and J.S. Marques. Performance evaluation of object detection algorithms for video surveillance. *Multimedia, IEEE Transactions on*, 8(4):761–774, August 2006.

[9] N. Saunier and T. Sayed. A feature-based tracking algorithm for vehicles in intersections. In *Third Canadian Conference on Computer and Robot Vision*, Québec, June 2006. IEEE.

[10] N. Saunier and T. Sayed. Probabilistic framework for automated analysis of exposure to road collisions. *Transportation Research Record*, 2083:96–104, 2008.

[11] N. Saunier, T. Sayed, and C. Lim. Probabilistic Collision Prediction for Vision-Based Automated Road Safety Analysis. In *The 10th International IEEE Conference on Intelligent Transportation Systems*, pages 872–878, Seattle, October 2007. IEEE.

[12] K. Smith, D. Gatica-Perez, J.-M. Odobez, and S. Ba. Evaluating multi-object tracking. In *IEEE Conference on Computer Vision and Pattern Recognition, Workshop on Empirical Evaluation Methods in Computer Vision (CVPR-EEMCV)*, San Diego, June 2005.

[13] M. Vlachos, G. Kollios, and D. Gunopulos. Elastic translation invariant matching of trajectories. *Machine Learning*, 58(2-3):301–334, February 2005.

[14] F. Yin, D. Makris, and S.A. Velastin. Performance evaluation of object tracking algorithms. In *10th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS2007)*, Rio de Janeiro, Brazil, October 2007.