# A Novel Method for Video Tracking Performance Evaluation

James Black
Digital Imaging Research Centre
Kingston University
Surrey KT1 2EE
J.Black@kingston.ac.uk

Tim Ellis
Digital Imaging Research Centre
Kingston University
Surrey KT1 2EE
T.Ellis@kingston.ac.uk

Paul Rosin
Department of Computer Science
Cardiff University
Cardiff CF24 3XF
Paul.Rosin@cs.cf.ac.uk

## Abstract

*This paper presents a methodology for evaluating the performance of video surveillance tracking systems. We introduce a novel framework for performance evaluation using pseudo-synthetic video, which employs data captured online and stored in a surveillance database. Tracks are automatically selected from the surveillance database and then used to generate ground truthed video sequences with a controlled level of perceptual complexity that can be used to quantitatively characterise the quality of the tracking algorithms.*

## 1. Introduction

Performance evaluation of image surveillance systems is an essential requirement, particularly when the system is deployed in a live environment. Our motivation for the work presented in this paper is to resolve some of the issues that arise when evaluating the performance of a video tracking algorithm. The evaluation issues include: how can we define ground truth for large datasets of video? What measures can be used to determine the complexity of a dataset along with the quality of its associated ground truth? What measures are appropriate to characterise tracking performance? The performance evaluation framework presented in this paper addresses each of these issues.

Our online surveillance system [1] comprises of a set of intelligent camera units with fixed camera views that utilise vision algorithms for detecting and tracking moving objects in 2D image coordinates. Each intelligent camera unit employs background subtraction [2] for motion detection and a partial observation-tracking algorithm [3] for object tracking and trajectory prediction. Tracked object data generated by each intelligent camera unit is stored in an on-line surveillance database. We will demonstrate how pseudo-synthetic video sequences can be generated from this data and then used within our performance evaluation framework. We choose to use pseudo-synthetic video to evaluate system performance, since it is possible to generate a large variety of datasets that represent a number of different tracking scenarios, which can vary in perceptual complexity. In addition, the ground truth is automatically acquired from the tracking data stored in the surveillance database. By adopting this approach it becomes practical to perform experiments over several hundred thousand frames of video data in order to quantitatively evaluate tracking performance.

The conventional approach for performance evaluation is to generate ground truth from pre-recorded video sequences. A number of semi-automatic tools are currently available for generating ground truth. The open development environment for evaluation of video systems (ODViS) [4] allows a user to generate ground truth for pre-recorded video. New tracking engines can be incorporated into the environment for evaluation within the ODViS framework. The Video Performance Evaluation Resource (ViPER)[5] provides a set of tools for ground truth generation, metrics for evaluation, and visualization of video analysis results. A number of metrics have been defined for tracker performance evaluation [5,6,7,8,9]. In [9] a number of metrics are used to evaluate tracking performance where ground truth is not available. They used a set of colour and motion metrics to assess the consistency of the tracked object between image frames. A number of metrics are defined for positional tracker evaluation in [8]. The main focus is on trajectory comparison to account for detection lag, or constant spatial shift. In [12] ground truth is automatically generated by using pre-determined cues such as shape and size on controlled test sequences.

The remainder of this paper is organized as follows: Section 2 describes the framework used to evaluate system performance using manual ground truth. Section 3 describes the method used to automatically select ground truth tracks from the surveillance database, and generate pseudo synthetic video sequences. Section 4 defines a set of surveillance metrics. Section 5 shows results obtained for performance evaluation using conventional pre-recorded and pseudo-synthetic video sequences. Section 6 is a discussion of what has been achieved by the current version of the evaluation framework and what new work is planned for the future.

## 2. Performance Evaluation

A typical approach to evaluating the performance of the detection and tracking system uses ground truth to provide independent and objective data (e.g. classification, location, size) that can be related to the observations extracted from the video sequence. Manual ground truth is conventionally gathered by a human operator who uses a 'point and click' user interface to step through a video sequence and select well-defined points for each moving object. The manual ground truth consists of a set of points that define the trajectory of each object in the video sequence (e.g. the object centroid). The human operator decides if objects should be tracked as individuals or classified as a group. The motion detection and tracking algorithm is then run on the pre-recorded video sequence and ground truth and tracking results are compared to assess tracking performance.

The reliability of the video tracking algorithm can be associated with a number of criteria: the frequency and complexity of dynamic occlusions, the duration of targets behind static occlusions, the distinctiveness of the targets (e.g. if they are all different colours), and changes in illumination or weather conditions. In this paper we express a measure for estimating the perceptual complexity of the sequence based on the occurrence and duration of dynamic occlusions, since this is the event most likely to cause the tracking algorithm to fail. Such information can be estimated from the ground truth data by computing the ratio of the number of target occlusion frames divided by the total length of each target track (i.e. the number of frames over which it is observed), averaged over the sequence (see section 4).

## 3. Pseudo Synthetic Video

As an alternative to manual ground truthing we propose using pseudo synthetic video to evaluate tracking performance. A problem for performance evaluation of tracking algorithms is that it is not trivial to accumulate datasets of varying perceptual complexity. Ideally, we want to be able to run a number of experiments and vary the perceptual complexity of the scene to test the tracking algorithm under a variety of different conditions. This is possible using manual ground truth but requires the capture of a large number of video sequences, which may not be practical at some surveillance sites.

The novelty of our framework is that we automatically compile a set of isolated ground truth tracks from the surveillance database. We then use the ground truth tracks to construct a comprehensive set of pseudo synthetic video sequences that are used to evaluate the performance of a tracking algorithm.

### 3.1 Ground Truth Track Selection

A list of ground truth tracks is initially compiled from the surveillance database. We select ground truth tracks during periods of low object activity (e.g. over weekends), since there is a smaller likelihood of object interactions that can result in tracking errors. The ground truth tracks are checked for consistency with respect to path coherence, colour coherence, and shape coherence in order to identify and remove tracks of poor quality.

Path Coherence: The path coherence metric [3] makes the assumption that the derived tracked object trajectory should be smooth subject to direction and motion constraints. Measurements are penalised for lower consistency with respect to direction and speed, while measurements are rewarded for the converse situation.

$$\varepsilon_{pc}=\frac{1}{N-2}\sum_{k=2}^{N-1}\left\{w_1\left(1-\frac{\overline{X_{k-1}X_k}\bullet\overline{X_kX_{k+1}}}{\left\|\overline{X_{k-1}X_k}\right\|\left\|\overline{X_kX_{k+1}}\right\|}\right)+w_2\left(1-\frac{2\sqrt{\left\|\overline{X_{k-1}X_k}\right\|\left\|\overline{X_kX_{k+1}}\right\|}}{\left\|\overline{X_{k-1}X_k}\right\|+\left\|\overline{X_kX_{k+1}}\right\|}\right)\right\}$$

Where $\overline{X_{k-1}X_k}$ is the vector representing the positional shift of the tracked object between frames k and k-1. The weighting factors can be appropriately assigned to define the contribution of the direction and speed components of the measure. The value of both weights was set to 0.5.

Colour Coherence: The colour coherence metric measures the average inter-frame histogram distance of a tracked object. It is assumed that the object histogram should remain constant between image frames. The normalised histogram is generated using the (r,g) colour space in order to account for small lighting variations. This metric has low values if the segmented object has similar colour attributes, and higher values when colour attributes are different. Each histogram contains 8x8 bins for the normalised colour components.

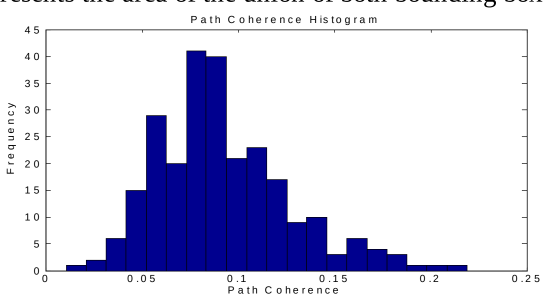
$$\varepsilon_{cc}=\frac{1}{N-1}\sum_{k=2}^{N}\sqrt{1-\sum_{u=1}^{M}p_{k-1}(u)\,p_k(u)}$$

Where $p_k(u)$ is the normalised colour histogram of the tracked object at frame k, which has M bins, and N is the number of frames the object was tracked over. This metric is a popular colour similarity measure employed by several robust tracking algorithms [10,11].

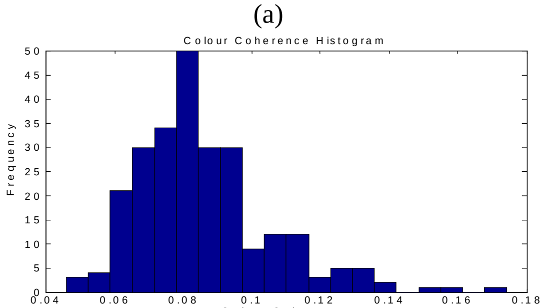
Shape Coherence: The shape coherence metric gives an indication of the level of agreement between the tracked object position and the object foreground region. This metric will have a high value when the localisation of the tracked object is incorrect due to poor initialisation or an error in tracking. The value of the metric is computed by evaluating the symmetric shape difference between the bounding box of the foreground object and tracked object state.

$$\varepsilon_{sc} = \frac{1}{N}\sum_{k=1}^{N}\frac{|R_f(k) - R_t(k)| + |R_t(k) - R_f(k)|}{|R_t(k) \cup R_f(k)|}$$

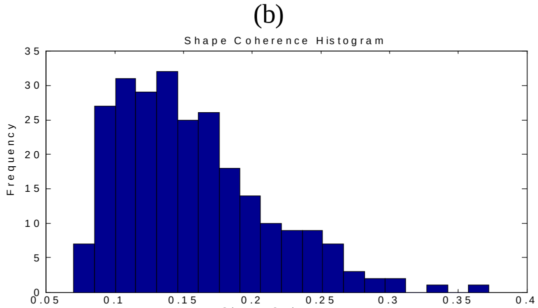
Where $|R_t(k) - R_f(k)|$ represents the area difference between the bounding box of the tracked object (state) and the overlapping region with the foreground object (measurement). The normalisation factor $|R_t(k) \cup R_f(k)|$ represents the area of the union of both bounding boxes.




(a)




(b)




(c)

Figure 1: Distribution of the average path coherence (a), average colour coherence (b), and average shape coherence of each track selected from the surveillance database.

Outlier ground truth tracks can be removed by applying a threshold to the values of $\varepsilon_{pc}$, $\varepsilon_{cc}$, and $\varepsilon_{sc}$. The distributions of the values are shown in the figure 1. It can be observed that a Gaussian distribution can adequately approximate each metric. The threshold is set so that the value should be within two standard deviations of the mean. The mean and standard deviations of $\varepsilon_{pc}$, $\varepsilon_{cc}$, and $\varepsilon_{sc}$ were (0.092, 0.086, 0.157) and (0.034, 0.020, 0.054) respectively. We also exclude any tracks that are short in duration and have not been tracked for at least N=50 frames, or have formed a dynamic occlusion with another track.

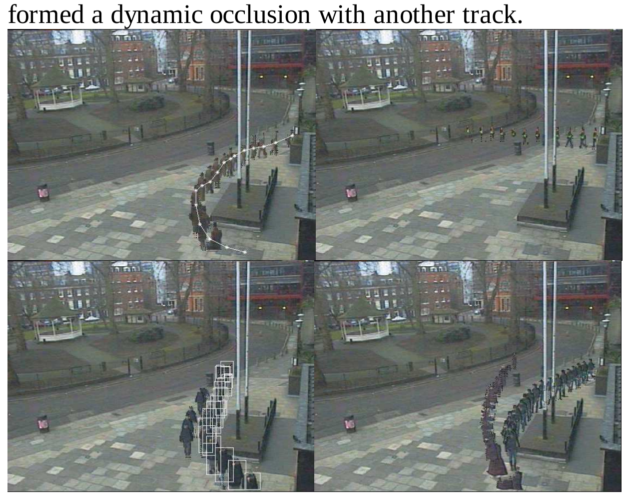


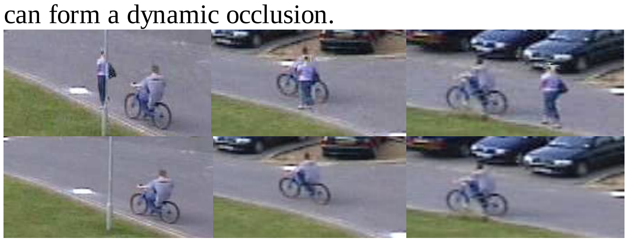
Figure 2: Example of outlier tracks identified during ground truth track selection.

In figure 2 some example outlier tracks are shown. The top left track was rejected due to poor path coherence, since the derived object trajectory is not smooth. The top right track was rejected due to poor colour coherence, which is a consequence of the poor object segmentation. The bottom left track was rejected due to poor shape coherence, where an extra pedestrian merges into the track. The tracked bounding boxes are not consistent with the detected foreground object. The bottom right track was rejected due to forming a dynamic occlusion with another track. It can be observed that in this instance the tracking failed and the objects switched identities near the bottom of the image. These examples illustrate that the metrics: path coherence, colour coherence, and shape coherence are effective for rejecting outlier ground truth tracks of poor quality.

## 3.2 Pseudo Synthetic Video Generation

Once the ground truth tracks have been selected they are employed to generate pseudo-synthetic videos. Each pseudo-synthetic video is constructed by replaying the ground truth tracks randomly in the generated video sequence. Two ground truth tracks are shown in left and middle images of figure 3, the tracked object is plotted every few frames in order to visualise the motion history of the object through the scene. When the two tracks are inserted in a pseudo-synthetic video sequence a dynamic occlusion can be created as shown in the right image of figure 3. Since the ground truth is known for each track we can determine the exact time and duration of the dynamic occlusion. By adding more ground truth tracks more complex object interactions are generated.

Figure 3: The left and middle images show two ground truth tracks. The right image shows how the two tracks can form a dynamic occlusion.



(a)

(b)

Figure 4: Examples of dynamic occlusions in a pseudo synthetic video sequence: The top and bottom rows in both figures represent the pseudo synthetic and original image frames, respectively (taken from PETS2001 dataset 2 (camera2)).

A number of steps are taken to construct each pseudo-synthetic video, since simple insertion of the ground truth tracks is not sufficient to create realistic video. Initially, a dynamic background video is captured for the camera view. This allows the pseudo-synthetic video to simulate small illumination changes that typically occur in outdoor environments. The framelets stored in the surveillance database consist of the foreground regions identified by the tracking algorithm (i.e. within the bounding box). When the framelet is replayed in the pseudo-synthetic video this improves the realism of dynamic occlusions. All the ground truth tracks are selected from a fixed camera view. This ensures the object motion in the constructed video sequence is consistent with the typical activity in the scene. 3D calibration information is used to ensure that framelets are plotted correctly during dynamic occlusions, according to their estimated depth from the camera. This gives the effect of an object occluding or being occluded by other objects based on their distance from the camera. This point is illustrated in figure 4, where a dynamic occlusion is simulated in a video sequence. The pseudo-synthetic and original image frames are shown to demonstrate how ground truth tracks can be used to construct realistic dynamic object occlusions. A pedestrian ground truth track is used to create a dynamic occlusion in figure 4a. In figure 4b a cyclist and pedestrian occlude a phantom vehicle, and the same vehicle then occludes a pedestrian later in the video sequence.

There are several benefits of using pseudo synthetic video: it is possible to simulate a wide variety of dynamic occlusions of varying complexity; pseudo-synthetic video can be generated for a variety of weather conditions; the perceptual complexity of each synthetic video can be automatically estimated; and ground truth can be automatically acquired. One disadvantage is that the pseudo synthetic video is biased towards the motion detection algorithm used to capture the original data, and few ground truth tracks will be generated in regions where tracking or detection performance is poor. In addition, the metrics described in section 3.1 do not completely address all the problems associated with motion segmentation. For example, the affects of shadows cast by moving objects, changes in weather conditions, the detection of low contrast objects, and the correct segmentation of an object's boundary. However, the pseudo-synthetic video is effective for evaluating the performance of tracking with respect to dynamic occlusion reasoning, which is the main focus of this paper.

## 3.3 Perceptual Complexity

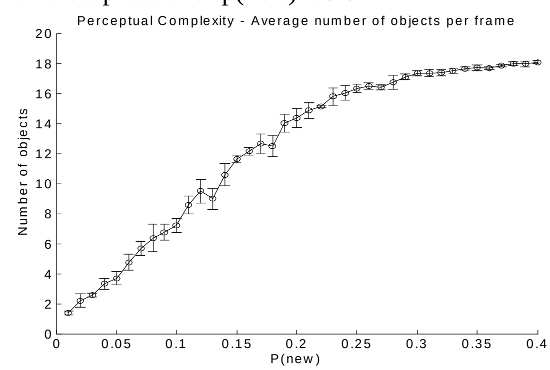
The perceptual complexity of each pseudo-synthetic video sequence is controlled by a set of tuneable parameters:

Max Objects (Max): The maximum number of the objects that can be present in any frame of the generated video sequence.

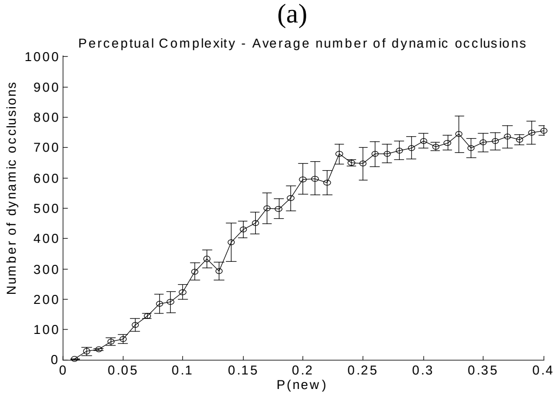
New Object Probability - p(new): The probability of creating a new object in the video sequence while the maximum number of objects has not been exceeded.

Increasing the value of p(new) results in a larger number of objects appearing in the constructed video sequence. This is illustrated in figure 5 where the three images demonstrate how the value of p(new) can be used to control the density of objects in each pseudo-synthetic video sequence. The images show examples for p(new) having the values 0.01, 0.10 and 0.20, respectively. These two parameters are used to vary the complexity of each generated video sequence. Increasing the values of the parameters results in an increase of object activity. We have found this model provides a realistic simulation of actual video sequences.

Figure 5: Perceptual Complexity: left – p(new)=0.01 image, middle – framelets plotted for p(new)=0.1, right – framelets plotted for p(new)=0.2.




(a)




(b)

Figure 6: (a) – plot of average no. of objects per frame by p(new), (b) – plot of average no. of dynamic occlusions by p(new); No. frames=1500, Max No. objects=20.

The number of dynamic occlusions in each pseudo synthetic video was determined by counting the number of occurrences where the bounding box of two or more ground truth objects overlap in the same image frame. We can count the number of dynamic occlusions (NDO), the average number of occluding objects (NOO), and the average duration of a dynamic occlusion (DDO) to provide a measure of the perceptual complexity [6]. Figure 6 demonstrates how p(new) can vary the perceptual complexity of each generated pseudo-synthetic video. Figure 6a and 6b are plots of p(new) by average number of objects per frame in the pseudo-synthetic video, and the average number of dynamic object occlusions respectively. The error bars on each plot indicate the standard deviation over the five simulations performed for each value of p(new). The values become asymptotic in both plots as the number of objects per frame approaches the maximum of 20, representing a complex and dense video sequence.

# 4. Surveillance Metrics

The surveillance metrics have been derived from a number of sources [4,5,6,7,8]. We first align the ground truth and results tracks by minimizing the trajectory distance metric that appears in [7]:

$$D_T(g,r) = \frac{1}{N_{rg}} \sum_{\exists i\ g(t_i)\wedge r(t_i)} \sqrt{(xg_i - xr_i)^2 + (yg_i - yr_i)^2}$$

Where $N_{rg}$ is the number of frames that the ground truth track and result track have in common, and $(xg_i, yg_i)$, $(xr_i, yr_i)$ is the location of the ground truth and result track at frame $i$ respectively.

Once the ground truth and results trajectories have been matched we use the following metrics to characterize the tracking performance:

Tracker Detection Rate (TRDR) = $\dfrac{\text{Total True Positives}}{\text{Total Number of Ground Truth Points}}$

False Alarm Rate (FAR) = $\dfrac{\text{Total False Positives}}{\text{Total True Positives} + \text{Total False Positives}}$

Track Detection Rate (TDR) = $\dfrac{\text{Number of true positives for tracked object}}{\text{Total number of ground truth points for object}}$

Object Tracking Error (OTE) = $\dfrac{1}{N_{rg}} \sum_{\exists i\ g(t_i)\wedge r(t_i)} \sqrt{(xg_i - xr_i)^2 + (yg_i - yr_i)^2}$

Track Fragmentation (TF) = Number of result tracks matched to ground truth track

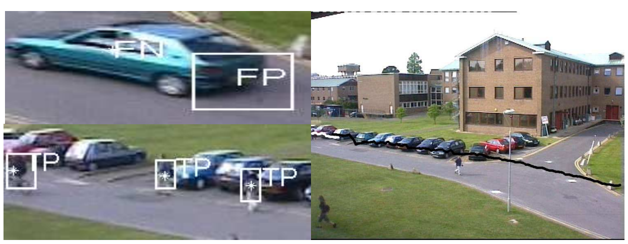
Occlusion Success Rate (OSR) =

$$\frac{\text{Number of sucessful dynamic occlusions}}{\text{Total number of number of dynamic occlusions}}$$

Tracking Success Rate (TSR) =

$$\frac{\text{Number of non-fragmented tracked objects}}{\text{Total number of number of ground truth objects}}$$

A true positive is defined as a ground truth point that is located within the bounding box of an object detected and tracked by the tracking algorithm. A false negative is a ground truth point that is not located with the bounding box of any object tracked by the tracking algorithm. A false positive is an object that is tracked by the system that does not have a matching ground truth point. These conditions are illustrated in figure 7. In figure 7(a) the vehicle in the top image has not been tracked correctly. The ground truth point for the vehicle is classified as a false negative. The bounding box of the incorrectly tracked vehicle is counted as a false positive. The three objects in the bottom image are counted as true positives, since the ground truth point is within the tracked bounding box.




(a) (b)

Figure 7: (a) Image to illustrate true positives, false negative and false positive, (b) Image to illustrate a fragmented tracked object trajectory.

The tracker detection rate (TRDR) and false alarm rate (FAR) characterise the tracking performance of the object-tracking algorithm. The track detection rate (TDR) indicates the tracking completeness of a specific ground truth track. The object tracking error (OTE) indicates the mean distance between the ground truth and the tracked object trajectory. The track fragmentation (TF) indicates how often a track label changes. Ideally, the TF value should be one, with larger values reflecting poor tracking and trajectory maintenance. The tracking success rate (TSR) summarises the performance of the tracking algorithm with respect to track fragmentation. The occlusion success rate (OSR) indicates how effective the tracking algorithm is with respect to occlusion reasoning. Figure 7(b) shows a tracked object trajectory for the pedestrian who is about to leave the camera view. The track is fragmented into two parts shown as black and white trajectories. The two track segments are used to determine the track detection rate, which indicates the completeness of the tracked object. As a consequence the ground truth object had a TDR, OTE, and TF of 0.99, 6.43 pixels, and 2 respectively.

# 5. Results

A number of experiments were run to test the performance of the tracking algorithm used by our online system. The tracking algorithm employs a partial-observation tracking model [3] for occlusion reasoning. We first generated purely manual ground truth for the second PETS2001 dataset (camera 2) using the point and click method described in section 2. We processed the data at a rate of 5fps. Table 1 provides a summary of the surveillance metrics report. The results demonstrate the robust tracking performance, since the track completeness is nearly perfect for all the objects. A couple of the tracks are fragmented due to poor initialisation or early termination. Figure 8 demonstrates what can happen when a tracked object is not initialised correctly. The left, and right images show the pedestrian exiting and leaving the parked vehicle. The pedestrian is partially occluded by other objects, so is not detected by the tracking algorithm until it has moved from the vehicle. The pedestrian relates to ground truth object 9.

An example of dynamic occlusion reasoning is shown in figure 9. The cyclist overtakes the two pedestrians, forming two dynamic occlusions and it can be noted that the correct trajectory is maintained for all three objects. The object labels in figure 9 have been assigned by the tracking algorithm and are different from the ground truth object labels.

We have also used the second PETS2001 dataset (camera 2) to construct a pseudo synthetic video by adding four additional ground truth tracks to the original video sequence. Table 2 summarises the differences in perceptual complexity between the original and pseudo synthetic video sequence. The number of dynamic object occlusions increases from 4 to 12, having the desired affect of increasing the complexity of the original video sequence. Table 2 also summarises the tracking performance for the original and pseudo synthetic sequences. These results validate our assumption that our object tracker can be used to generate ground truth for video with low activity.

In order to test the effectiveness of the tracking algorithm for tracking success and dynamic occlusion reasoning we generated several pseudo synthetic videos sequences. We automatically selected ground truth tracks from the surveillance database using the method described in section 3.1. We then generated five synthetic video sequences for each level of perceptual complexity. The value of p(new) was varied between 0.01 to 0.4 with increments of 0.01. Each pseudo synthetic video sequence was 1500 frames in length, which is equivalent to approximately 4 minutes of live captured video by our online system running at 7Hz. Hence, in total the system was evaluated with 200 different video sequences, totalling three hundred thousand image frames, or approximately 800 minutes of video.

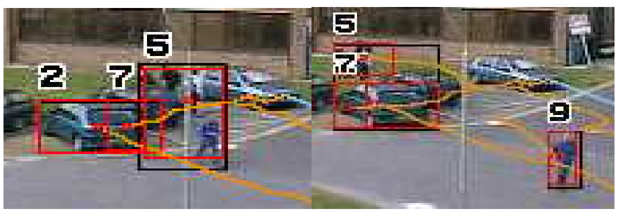


Figure 8: An example of how poor track initialisation results in low object track detection rate of the pedestrian leaving the vehicle.

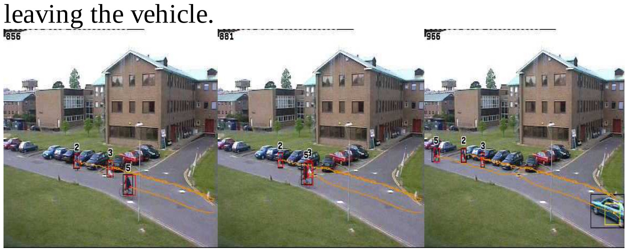
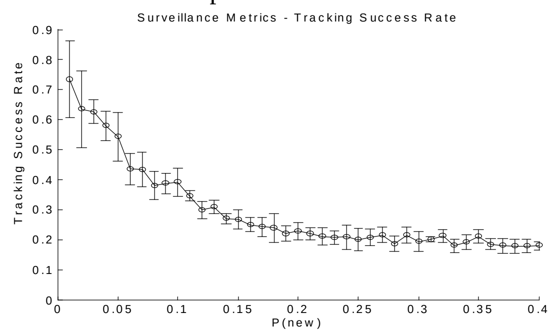
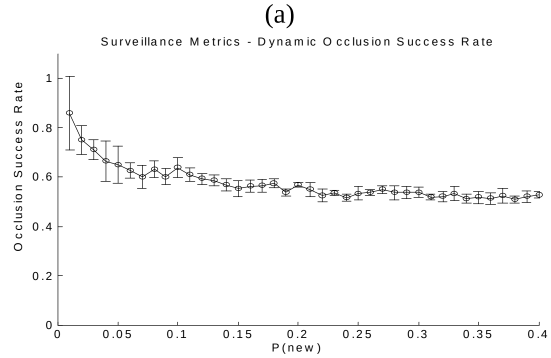


Figure 9: Example of dynamic occlusion reasoning for PETS2001 dataset 2 camera 2.

The synthetic video sequences were used as input to the tracking algorithm. The tracking results and ground truth were then compared and used to generate a surveillance metrics report as described in section 4. Table 3 gives a summary of the complexity of a selection of the generated video sequences. These results confirm that p(new) controls the perceptual complexity, since the number of objects, average number of dynamic occlusions and occluding objects increases from (12.8, 2.4, 2.0) to (357.2, 755.2, 3.24) respectively for the smallest and largest values of p(new). Table 4 summarises the tracking performance for various values of p(new). The object tracking error increases with the value of p(new), which represents a degradation of tracking performance with respect to occlusion reasoning. The occlusion success rate (OSR) and tracking success rate (TSR) decreases in value from (86%, 73%) to (53%, 18%) with the increasing value of p(new). When the number of objects per frame approaches the maximum this limits the number of dynamic occlusions created, hence increasing values of p(new) have a diminished affect of increasing the perceptual complexity. As a consequence the TSR and OSR become asymptotic once the number of objects per frame approaches the maximum of 20 as illustrated in the plots of figure (10). Larger values of p(new) and the maximum number of objects would result in more complex video sequences. Hence even with the bias present in the generated video sequences we can still evaluate the object tracking performance with respect to tracking success and occlusion reasoning, without exhaustive manual truthing, fulfilling the main objective of our framework for performance evaluation.



(a)



(b)

Figure 10: Plot of: (a) Tracking success rate, (b) occlusion success rate

# 6. Conclusion

We have presented a novel framework for evaluating the performance of a video tracking algorithm. The performance evaluation framework automatically selects ground truth tracks from a surveillance database in order to construct pseudo synthetic video sequences. We have compiled a comprehensive set of metrics, which can be used to measure the quality of the ground truth tracks, as well as characterise tracking performance. We recognise that the pseudo synthetic video will have a degree of bias to the motion detection algorithm used to capture the original data. However, the generated video sequences are effective for evaluating performance of occlusion reasoning, and can be used to evaluate other tracking algorithms. The main strength of our evaluation framework is that we can automatically generate a variety of different testing datasets. In this paper we have evaluated a tracking algorithm over three hundred thousand frames of video, without any human intervention or semi-automatic ground truth generation. In future work we plan evaluate other tracking algorithms within our framework using the results presented in this paper as a benchmark.

| Track | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| TP | 25 | 116 | 26 | 104 | 36 | 369 | 78 | 133 | 43 | 88 |
| FN | 0 | 2 | 0 | 5 | 0 | 5 | 1 | 1 | 1 | 2 |
| TDR | 1.00 | 0.98 | 1.00 | 0.95 | 1.00 | 0.99 | 0.99 | 0.99 | 0.98 | 0.98 |
| TF | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 |
| OTE | 11.09 | 7.23 | 8.37 | 4.70 | 10.82 | 11.63 | 9.05 | 6.43 | 8.11 | 11.87 |

TP: Number of true positives FN: Number of false positives TF: Track Fragmentation
TDR: Track Detection Rate OTE: Object Tracking Error

Table 1: Summary of surveillance metrics for PETS2001 dataset2 (camera 2)

| | TNO | NDO | DDO | NOO | TRDR | TSR | FAR | AOTE | | ATDR | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | mean | stdev | mean | Stdev |
| Dataset 2(Cam 2) | 10 | 4 | 8.5 | 2 | 0.99 | 8/10 | 0.01 | 8.93 | 2.4 | 0.99 | 0.010 |
| Pseudo Synthetic PETS Dataset | 14 | 12 | 8.58 | 2.08 | 1.00 | 9/13 | 0.01 | 1.36 | 2.09 | 1.00 | 0.002 |

NDO: No. of Dynamic Occlusions DDO: Duration of Dynamic Occlusion (frames)
NOO: Number of Occluding Objects TNO: Total Number of Objects

Table 2: Summary of perceptual complexity of the PETS2001 dataset2 (camera2) and object tracking metrics.

| | TNO | | NDO | | DDO | | NOO | |
|---|---|---|---|---|---|---|---|---|
| P(new) | mean | stdev | mean | stdev | mean | stdev | mean | stdev |
| 0.01 | 12.80 | 4.147 | 2.40 | 1.140 | 6.42 | 5.353 | 2.00 | 0.000 |
| 0.20 | 284.00 | 16.538 | 595.20 | 49.957 | 10.70 | 0.310 | 2.95 | 0.020 |
| 0.40 | 357.20 | 4.087 | 755.20 | 15.466 | 12.26 | 0.461 | 3.24 | 0.112 |

Table 3: Summary of the perceptual complexity of the 200 synthetic video sequences (300000 frames).

| | TRDR | FAR | OSR | | AOTE | | ATDR | | ATSR | |
|---|---|---|---|---|---|---|---|---|---|---|
| P(new) | | | mean | stdev | Mean | stdev | mean | stdev | mean | stdev |
| 0.01 | 0.91 | 0.08 | 0.86 | 0.149 | 3.21 | 1.466 | 0.90 | 0.049 | 0.73 | 0.129 |
| 0.20 | 0.91 | 0.09 | 0.57 | 0.010 | 12.64 | 0.599 | 0.76 | 0.008 | 0.23 | 0.029 |
| 0.40 | 0.90 | 0.09 | 0.53 | 0.014 | 14.12 | 0.581 | 0.72 | 0.006 | 0.18 | 0.015 |

Table 4: Summary of metrics generated using each synthetic video sequence.

# References


[1] T.J. Ellis, J. Black. A Multi-view surveillance system. IEE Intelligent Distributed Surveillance Systems, London, February 2003.

[2] M. Xu, and T.J. Ellis, "Illumination-Invariant Motion Detection Using Colour Mixture Models", British Machine Vision Conference (BMVC 2001), Manchester, September 2001, pp 163-172.

[3] M. Xu, and T.J. Ellis, "Partial Observation vs Blind Tracking through Occlusion", British Machine Vision Conference (BMVC 2002), Cardiff, September 2002, pp 777-786.

[4] C. Jaynes, S. Webb, R. Matt Steele, and Q. Xiong, "An Open Development Environment for Evaluation of Video Surveillance Systems", Proceedings of the Third International Workshop on Performance Evaluation of Tracking and Surveillance (PETS'2002), Copenhagen, June 2002.

[5] D. Doermann, and D. Mihalcik, "Tools and Techniques for Video Performance Evaluation", Proceedings of the International Conference on Pattern Recognition (ICPR'00), Barcelona, September 2000, pp 4167-4170.

[6] T.J. Ellis, "Performance Metrics and Methods for Tracking in Surveillance", Proceedings of the Third International Workshop on Performance Evaluation of Tracking and Surveillance (PETS'2002), Copenhagen, June 2002.

[7] A. Senior, A. Hampapur, Y. Tian, L. Brown, S. Pankanti, R. Bolle, "Appearance Models for Occlusion Handling" Proceedings of the Second International Workshop on Performance Evaluation of Tracking and Surveillance (PETS'2001), Hawaii, Kauai, December 2001.

[8] C.J. Needham, R.D. Boyle. Performance Evaluation Metrics and Statistics for Positional Tracker Evaluation. International Conference on Computer Vision Systems (ICVS'03), Graz, Austria, April 2003, pp 278-289.

[9] Ç. Erdem, B. Sankur, A.M. Tekalp. Metrics for Performance Evaluation of Video Object Segmentation and Tracking Without Ground-Truth. IEEE International Conference on Image Processing (ICIP'01), Thessaloniki, Greece, October 2001.

[10] D. Comaniciu., V. Ramesh, P. Meer. Real-Time Tracking of Non-Rigid Objects Using Mean Shift. IEEE Conference on Computer Vision and Pattern Recognition (CVPR'00), Hilton Head, South Carolina, June 2000, pp 2142-2151.

[11] K. Nummiaro, E. Koller-Meier, L. Van Gool. Color Features for Tracking Non-Rigid Objects. Special Issue on Visual Surveillance, Chinese Journal of Automation, May 2003, Vol. 29, No. 3, pp 345-355.

[12] P.L. Rosin, E. Ioannidis. Evaluation of Global Image Thresholding for Change Detection, Pattern Recognition Letters, 2003, Vol. 24, No. 14, pp 2345-2356.