

# PROBABILISTIC LEARNING OF SALIENT PATTERNS ACROSS SPATIALLY SEPARATED, UNCALIBRATED VIEWS

Pakorn KaewTraKulPong\*, Richard Bowden†

\*King Mongkut's University of Technology Thonburi,  
Faculty of Engineering, Toongkaru, Bangmod,  
Bangkok, 10400, Thailand  
ipakpong@kmutt.ac.th

† CVSSP, University of Surrey Guildford, Surrey, GU2 7XH, UK  
r.bowden@surrey.ac.uk

## Abstract

This paper presents a solution to the problem of tracking intermittent targets that can overcome long-term occlusions as well as movement between camera views. Unlike other approaches, our system does not require topological knowledge of the site or labelled training patterns during the learning period. The approach uses the statistical consistency of data obtained automatically over an extended period of time rather than explicit geometric calibration to automatically learn the salient reappearance periods for objects. This allows us to predict where objects may reappear and within how long. We demonstrate how these salient reappearance periods can be used with a model of physical appearance to track objects between spatially separate regions in single and separated views.

## 1 Introduction

Intelligent visual surveillance is an important application area for computer vision. Key to the users requirements is the ability to track objects across (spatially separated) camera scenes. However, extensive geometric knowledge about the site and camera position is typically required. Algorithms for tracking multiple objects through occlusion normally perform well for relatively simple scenes where occlusions by static objects and perspective effects are not severe. For example, Figure 1 shows a typical surveillance camera view with two distinct regions **A** and **B** formed by the presence of a static foreground object (tree) that obscures the ground from view.

In this work, regions or subregions are defined as separated portions grouped spatially within an image. A region can contain one or more paths which may cover an arbitrary number of areas. Tracking an object across regions in the scene, where the geometric relationship of the regions can not be assumed, possesses similar challenges to tracking an object across spatially-separated camera scenes (again where the geometric relationship among the cameras are unknown.) In a single view, the simplest solution to this problem is to increase the allowable number of consecutive frames that a target persists with no observation before tracking is

terminated. This process is shown in Figure 1 using a Kalman filter as a linear estimator.

By delaying the tracking termination, both the deterministic and the random components within the dynamics of the Kalman filter propagate over time [4], increasing the uncertainty of the predicted area in which the target may reappear (as shown in Figure 1c.) This increases the chance of matching targets undergoing long occlusions but also increases the chance of false matches. In situations where linear prediction can not be assumed (for example, the pathway changes direction behind large static objects), this will result in a model mismatch and the kinematic model assumed in most trackers will provide incorrect predictions. Furthermore this type of approach cannot be extended to multiple cameras without an explicit calibration of those cameras.

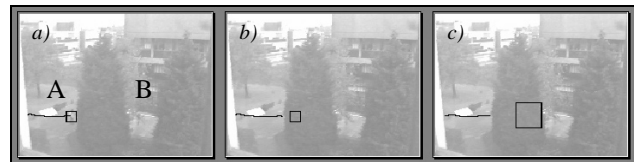


Figure 1 Tracking reappearance targets. a) a target is being tracked, b) the target is occluded but the search continues, c) uncertainty propagation increases over time.

## 2 Previous work

An approach often used to tackle the tracking of an object across multiple cameras is to ensure some overlap within the field of view of cameras is available. An example is the pilot military system reported in [1,2] where a birds-eye camera view is used to provide a global map. The registration of a number of ground-based cameras to the global map allows tracking to be performed across spatially separated camera scenes. The self-calibrated multi-camera system demonstrated in [3] assumes partially overlapping cameras. Epipolar geometry, landmarks as well as a target's visual appearance are used to facilitate the tracking of multiple targets across cameras and to resolve occlusions.

Tracking across non-overlapping views has been of recent interest to many researchers. Huang and Russel [5] develop a

system to track vehicles on a highway across spatially separated cameras. In their system, knowledge about the entrance/exit of the vehicles into the camera views must be provided along with transition probabilities. Kettner and Zabih [6] present a Bayesian framework to track objects across camera views. The user supplies topological knowledge of usual paths and transition probabilities. Javed et al. [7] present a more general solution to the problem by providing an update of inter-camera parameters and appearance probabilities. However, their method assumes initial correspondences of those parameters. Our method makes use of data obtained automatically for extended period to discover such relationship between camera views without user intervention.

In accumulating evidence of patterns over time we expect to discover common activities. These patterns can be modelled in a number of ways. They can be used to classify sequences as well as individual instances of a sequence or to discover common activities [8]. Howarth and Buxton [9] introduce a spatial model in the form of a hierarchical structure of small areas for event detection in traffic surveillance. The model is constructed manually from tracking data. Fernyhough et al. [10] use tracked data to build a spatial model to represent areas, paths and regions in space. The model is constructed using a frequency distribution collected from a convex-hull binary image of the objects. Thresholding of the frequency distribution filters out low distribution areas, i.e. noise. Johnson et al. [11] use flow vectors, i.e. the 2D position and velocity, collected from an image sequence over an extended period to build a probability distribution of the targets moving in the image. A neural network with competitive layers is used to quantise the data and represent the distribution. Its use is to detect atypical events which occurred in the scene. Makris and Ellis [12] use spline representations to model common routes and the activity of these routes. Entry-exit points and junctions are identified as well as their frequencies. Nair and Clark [13] use extended data to train two HMMs to recognise people entering and exiting a room in a corridor. Uncommon activity is identified as break-in by calculating the likelihood of the trajectories of the HMMs and comparing with a pre-defined threshold. Stauffer [14] use on-line Vector Quantisation as described in [11] to quantise all target information including position, velocity, size, binary object silhouette into 400 prototypes. Then they perform inference on their data to obtain a probability distribution of the prototypes encoded in a co-occurrence matrix. A Normalised cut [15] is then performed on the matrix which results in grouping similar targets in terms of the above features in a hierarchical form.

### 3 Relating Possible Reappearing Targets

In order to identify the same target reappearing after a long occlusion, features/characteristics of object similarity must be identified. The properties of reappearing targets are assumed as follows: a target should disappear from one area and reappear in another within a certain length of time; the reappearing target must occur only after that target has

disappeared. (There is no co-existence of the same target at any instance. This is according to the spatially separated assumption. For overlapped camera views, this assumption may not be applied.); and finally frequently occurring disappearances or reappearances will form consistent trends within the data.

If targets are moving at similar speeds along a path occluded by large static objects (such as the tree in Figure 1). Over a period of time, there should be a number of targets that disappear from a specific area of region **A** and reappear within another area of region **B** within a window of time. This can be called the *salient* reappearance period between the two areas. Both areas can be considered to be elements of the same path even though they are in different regions. The reappearance periods of these targets should be similar compared with random appearance of targets between other elements.

We start by automatically collecting data from our tracking algorithm [16] over an extended period in a single camera. The data consists of tracking details of targets passing into the field of view of the camera. Figure 4 shows trajectories of all objects collated. Notice due to the occlusion of the tree two separate regions are formed where no correspondence is known about the relationship between them. The linear dynamics of the tracker are insufficient to cope with the occlusion.



Figure 2 - Trajectory data and model of the training set.

The goal is to learn some reappearance relationship in an unsupervised fashion to allow objects which disappear to be successfully located and tracked when they reappear. This we term the salient reappearance period, and its construction is divided into two steps:

1. Extracting dominant paths: Trajectories in each subregion are classified into a number of paths. Only 'main' paths that consist of a large number of supporting trajectories are extracted.
2. Extracting salient reappearance periods among dominant paths: In this stage, a set of features common to reappearing targets is introduced. The features allow possible reappearance among pairs of paths to be calculated. A set of matches that show outstanding reappearance periods are chosen to train the matching models in the training phase.

### 3.1 Path Extraction

Paths can be effectively represented as a group of trajectories between two areas. However, some trajectories may begin and end in the same area. Paths of this type must be divided into different groups. This is done using a Normalised cut [15] of the 4D motion vectors formed through the concatenation of position and velocity. All paths which contain more than two percent of the total trajectories are kept for further processing and are shown in Figure 3. The figure shows how the normalised cut has broken trajectories down into groups of like trajectories.

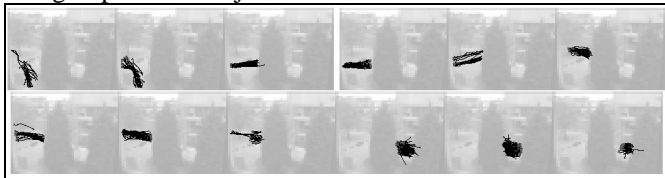


Figure 3 - Main paths in the first region.

### 3.2 Linking Paths

Paths are linked by building a fuzzy histogram of all possible links among trajectories of a pair of main paths from different regions within a period named 'allowable reappearance period'. The bin  $t$  of the histogram is calculated from

$$\tau_i = \sum_{\forall i,j} H_{ij}; (t-1) \leq (t_i^{end} - t_j^{start}) < t \quad (1)$$

where  $t_i^{start}$  and  $t_i^{end}$  are the time instances that target  $i$  starts and ends respectively.  $H_{ij}$  is the result from histogram intersection between the colour histogram of target  $i$  and that of target  $j$ .

$$H_{ij} = \sum_{k=1}^{11} \min(B_{ik}, B_{jk}) \quad (2)$$

The colour appearance histograms,  $B_j = \{B_{j1}, B_{j2}, \dots, B_{j11}\}$ , are constructed by taking the constituent pixels of the object and categorising each pixel as one of 11 basic colours using a consensus-colour conversion on Munsell colour space developed by Sturges et al [17]. This provides consistent labelling of colour across cameras without the burden of colour calibration.  $H_{ij}$  is within the range of 0 to 1 with 1 corresponding to a perfect match.

Using the linking scheme described previously on every pair of main paths between two regions (with only main paths from different regions permitted for the matches), a number of possible matches produces salient reappearance periods. Figure 4 shows the results. Two histograms for every pair of paths are produced. The one that has the maximum peak is selected. To check the validity of the outstanding peak it must exceed four times the noise floor level. The noise floor level was found by taking the median from non-empty bins of the histogram. Unimodality is assumed, therefore, a single peak is detected based on the maximum area under the bins that pass the noise level. The histograms are presented with the corresponding linked trajectories in Figure 4. The detected bins are shown with solid bars on each histogram, the noise level is also plotted in each histogram.

With the data automatically collected from the last process, a reappearing-target model can be formed for each pair of detected main paths. A simple model is calculated from the training data as follows. In each pair of detected main paths, the reappearance periods  $\{r_1, r_2, \dots, r_N\}$  between a pair of paths is represented compactly by their mean  $\mu_r$  and standard deviation  $\sigma_r$ .

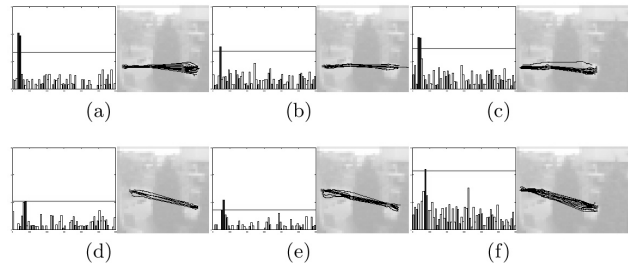


Figure 4 - Extracted salient reappearance periods among main paths between different regions

## 4 Path-based Target Recognition

Based on the model obtained in the training phase, target recognition can be performed. Target recognition is based on the same linking scheme. The process can be performed online using the same principle as the multiple hypothesis tracking which select the best hypothesis within the reappearance period or off-line after a batch collection of trajectories within the allowable reappearance period.

Similarity based on the salient reappearance periods is therefore calculated in the following manner. First, standard (zero mean and unity variance) normal random variable of each reappearance period is calculated using  $z = \frac{r - \mu_r}{\sigma_r}$ . Then the standard normal cumulative probability in the right-hand tail  $\Pr(Z > z)$  is determined using a look-up table. The similarity score is two times this value which gives the score in the range of 0 to 1. Another way of calculating this value is

$$2 \Pr(Z > z) = 1 - \int_0^z f(x|1) dx \quad (3)$$

where  $f(x|\nu)$  is the chi-square distribution with  $\nu$  degree of freedom.

$$f(x|\nu) = \frac{x^{\frac{(\nu-2)}{2}} e^{-\frac{x}{2}}}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})} \quad (4)$$

where  $\Gamma(\nu)$  is the Gamma function. However, a *no match hypothesis* or *null hypothesis* is also introduced, as it is possible to have no link. Hypothesis testing for this *null hypothesis* is required before any classification is performed. A score of 0.001 was set for the null hypothesis which corresponds to the standard value  $z$  of 3.09. Any candidates which are not 'null hypotheses' are selected based on their maximum score. Online recognition selects the best candidate at each time instance within the allowable reappearance period. This allows the tracker to change its link each time a better hypothesis is introduced. Batch recognition on the other hand collects all trajectories within the allowable

reappearance period and performs the classification based on the whole set.

In the first experiment, the training data was collected automatically by our target tracking algorithm [16] over a extended period of time constituting 1009 individual trajectories. An unseen test set of 94 trajectories was collected and hand labelled as ground truth. The recognition process is performed off-line by collecting the whole set of admissible targets according to the rules described in section 3. An example of the recognition with similarity score and its corresponding time line chart in Figure 5. The red line in the time line is the target of interest, while the green lines are the non *null hypothesis* candidates with plausible matches. The arrows depict the ground truth. It can be seen that target (b) was classified correctly to the ground truth as it obtained the best score during the recognition process.

Summary of the recognition on the whole test set is provided in Table 1. Since the model is based on the trend in a pair of main paths, if the target uses uncommon paths which has no trend in the data set, the target can not be recognised. This accounts for 2 out of the 6 misses. One of these misses was caused by a person who changed his mind during his disappearance and walked back to the same path in the same area, while the other was due to camouflage at the beginning of the trajectory.

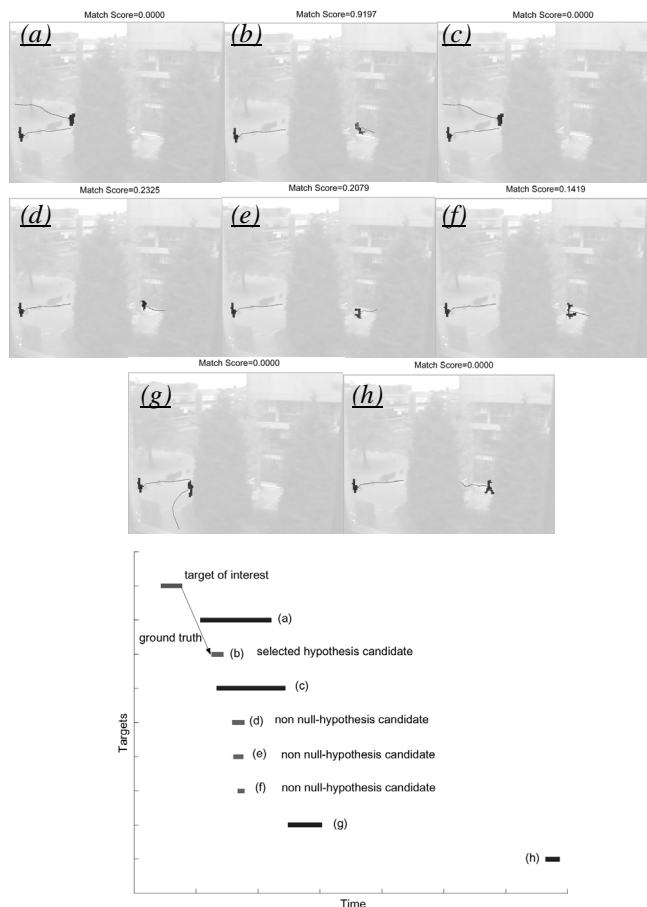


Figure 5 - A example of recognising reappearing target between different regions and the associated time line.

Items	Trajectories	%
Total training set	1009	
Total test set	94	100
Correct matches	84	89.36
Total incorrect matches	10	10.64
False detections	4	4.26
Misses	6	6.38
Misses (uncommon paths)	2	2.13

Table 1 - Matching results of an unseen data set of 10 minutes in a single camera view.

## 5 Learning and Recognising Trajectory Patterns across Camera Views

The target recognition presented thus far can also be extended to multiple cameras. The same process of trajectory data collection from two scenes with time synchronisation was performed. The set-up of cameras is shown in Figure 6. 2020 individual trajectories consisting of 56391 data points were collected for two cameras. An unseen test set of 133 trajectories was then hand labelled to provide ground truth.

Figure 7 shows some examples of the salient appearing periods and trajectories extracted between camera pairs for extracted main paths. Note that both of the examples are valid reappearing trends. However, the trend in Figure 7a is not as high due to a small sample set as well as an increased distance between the two main paths. It should be noted if any two regions are physically too distant, the prominence of the salient reappearance period is reduced.

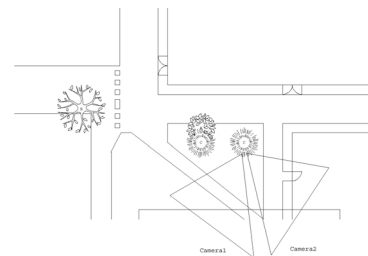


Figure 6 - Site map and camera layouts for recognising reappearing targets across camera views.

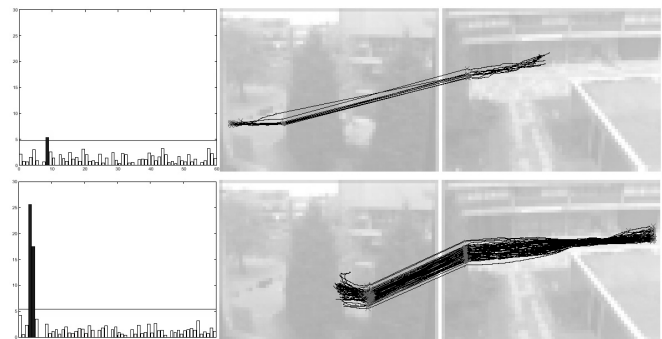


Figure 7 - Examples of extracting salient reappearance periods between main paths in different regions across camera views.

The results from pre-processing are then subjected to the same process as before and classification performed in the

same way as in the single view. The test data was collected and hand labelled. It consists of 133 targets. One example of the matching process is shown in Figure 8 along with its corresponding time-line chart. In this example, candidate (b) is the best match and is selected. The second best which is candidate (g) is also valid; however, it has a lower score due to the flatter peak in its training set. This higher variance is caused by the greater distance between the two main paths which increases the effect of variance of the targets speed. Only matches b & g are shown due to space limitations. Matching results from the two scenes are shown in table 2.

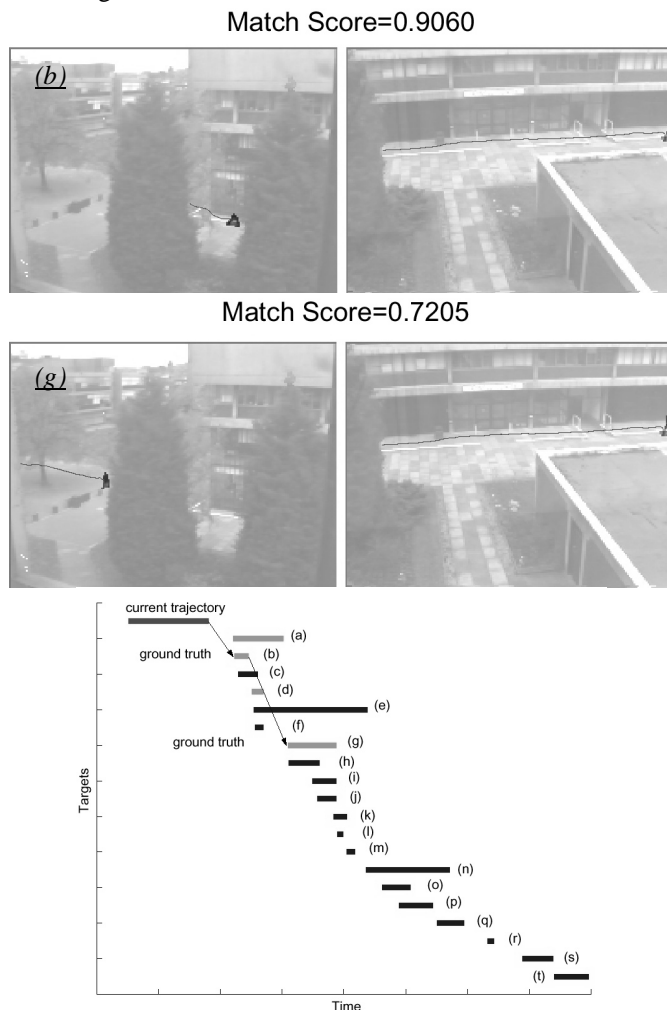


Figure 8 - Time line of an example of recognising reappearing target between different regions in spatially separated cameras.

Items	Trajectories	%
Total training set	2020	
Total test set	133	100
Correct matches	116	87.22
Total incorrect matches	17	12.78
False detections	8	6.02
Misses	9	6.77
Misses (uncommon paths)	2	1.50

Table 2 – Matching results of an unseen set of 10 minutes across camera views.

It can be seen from Tables 1 and 2 that the recognition rate of the proposed technique is high. However, the approach is still dependent on the assumption that the separated regions in the same or different views are close to each other. Future work will investigate the correlation between accuracy and distance for this technique.

## 6 Summary and Conclusions

In this paper, an approach for recognising targets after occlusion is proposed. It is based on salient reappearance periods discovered from long term data. By detecting and relating main paths from different regions and using a robust estimate of noise, salient reappearance periods can be detected with high signal-to-noise ratios. Off-line recognition is performed to demonstrate the use of this extracted salient reappearance period and the appearance model to associate and track targets between spatially separated regions. The demonstration is extended to regions between spatially separated views with minimal modifications. As the underlying process of reappearance is not the salient reappearance time but the average distance between paths, the performance of this recognising process is degraded if the average distance between paths is increased. These issues need further investigation.

## References

- [1] T. Kanade, R. Collins, A. Lipton, P. Anandan, P. Burt, L. Wixson. "Cooperative multi-sensor video surveillance." In DARPA Image Understanding Workshop, pages 3-10, 1997.
- [2] T. Kanade, R. Collins, A. Lipton, P. Burt, and L. Wixson. Advances in cooperative multi-sensor video surveillance. In Darpa Image Understanding Workshop, pages 3-24, 1998.
- [3] T.H. Chang, S. Gong, and E.J. Ong. Tracking multiple people under occlusion using multiple cameras. In BMVC00, 2000.
- [4] S. Blackman and R. Popoli. Design and analysis of modern tracking systems. Artech House, 1999.
- [5] T. Huang and S. Russell. Object identification in a bayesian context. In IJCAI97, 1997.
- [6] V. Kettaker and R. Zabih. Bayesian multi-camera surveillance. In CVPR99, 1999.
- [7] O. Javed, Z. Rasheed, K. Shafique and M. Shah. Tracking across multiple cameras with disjoint views. In ICCV03, 2003.
- [8] C. Stauffer and W.E.L. Grimson. Learning patterns of activity using real-time tracking. PAMI, 22(8):747-757, 2000.
- [9] R.J. Howarth and H. Buxton. Analogical representation of space and time. IVC, 10:467-478, 1992.
- [10] J.H. Fernyhough, A.G. Cohn, and D.C. Hogg. Generation of semantic regions from image sequences. In ECCV96, pages II:475-484, 1996.
- [11] N. Johnson and D. Hogg. Learning the distribution of object trajectories for event recognition. IVC, 14(8):609-615, 1996.
- [12] D. Makris and T. Ellis. Finding paths in video sequences. In BMVC01, 2001.
- [13] V. Nair and J.J. Clark. Automated visual surveillance using hidden markov models. In VI02, page 88, 2002.
- [14] C. Stauffer. Automatic hierarchical classification using time-based co-occurrences. In CVPR99, pages II:333-339, 1999.
- [15] J. Shi and J. Malik. Normalized cuts and image segmentation. PAMI, 22(8):888-905, August 2000.
- [16] P. KaewTraKulPong, R. Bowden, A Real-Time Adaptive Visual Surveillance System for Tracking Low Resolution Colour Targets In Dynamically Changing Scenes, IVC. 21(10):913-929, 2003.
- [17] J. Sturges and T.W.A. Whitfield. Locating basic colours in the munsell space. Color Research and Application, 20(6):364-376, 1995.