Department of Electrical
and
Computer Systems Engineering

Technical Report
MECSE-13-2007

Covert Human Behaviour Detection

P. Chakravarty, D. Rawlinson and R. Jarvis

MONASH
UNIVERSITY

Covert Human Behaviour Detection

Punarjay Chakravarty, Dave Rawlinson and Ray Jarvis
Monash University, Australia
punarjay.chakravarty,david.rawlinson,ray.jarvis,@eng.monash.edu.au

Abstract

This paper describes a method by which suspicious, covert human movements can be distinguished from ordinary activity. A network of fixed cameras observes an arbitrary indoor environment. The positions of moving, visible entities are separately tracked in a global planar map using multiple particle filters. Some objects can be visually identified as friendlies (guards). The motion of other objects is classified as covert (suspicious) or overt (normal) using two metrics that measure overall visibility of the chosen route and visibility with respect to guards' position. In other words, the system measures to what extent a person is trying to hide from the guards. This paper is the latest in a series of related works that together describe how a network of fixed cameras can learn to identify intruders and guide mobile robots to intercept.

1 Introduction

We introduce covert behaviour detection as a sub-system in the context of a larger system being developed by the authors: a networked surveillance system. A mobile robot, equipped with a wide-angle camera and a laser range finder comprises the mobile component of the surveillance system and is assisted by fixed cameras connected wirelessly to it. The robot is able to localize itself in a map of its environment using on-board sensing, and at the same time, the overhead cameras are able to track the robot as it moves about. This simultaneity of location information, both on the ground plane of the robot, and in the image plane of the camera, lends itself to the autonomous construction of a transformation mapping between the two planes [Rawlinson *et al.*, 2004]. Once this map is constructed, any point in the image plane is transformable to its corresponding point on the ground plane. When an intruder enters the environment, he/she is tracked by the overhead camera, which then passes on the ground plane coordinates of the intruder to the robot, which proceeds to pursue this target. Once the robot's on-board sensors are within range of the intruder, they take over and the robot continues to give chase using its onboard sensing [Chakravarty *et al.*, 2006].

The ability of the stationary cameras to determine the ground-plane coordinates of entities detected in their image planes makes possible the tracking of these entities on a map of the ground plane. The tracking is done using particle filtering, one particle filter per target. The trajectories of persons moving through the area is recorded. The visibility of the trajectory of an intruder is determined with respect to a sentry who is patrolling the environment. This visibility, calculated at each point on the intruder's trajectory is determined by tracing rays outward from the position of the sentry till they hit obstacles.

The method of determining the covert/overtness of *human* paths is the Dark Path algorithm [Marzouqi and Jarvis, 2006], hitherto used in the simulation of robot wargames. It measures, given a map of the environment, start and goal locations, the most covert route a robot can take through it.

Our proposal is to concentrate on analysis of the way the position of human targets changes over time, rather than scrutiny of characteristics of individuals' appearance, gait, size or shape. We also propose analysis of targets positions with respect to each other - how do the targets interact? In the "surveillance" literature, this topic is less comprehensively researched than recognition of individuals' characteristics but some pertinent related work on analysis of training data is described below.

[Hu *et al.*, 2004] train a neural network using normal trajectories and are subsequently able to use the network to detect trajectories that deviate from the learnt normal ones. This works well in situations where a large body of human traffic behaves consistently, as is often the case in public transport stations. However, it cannot be applied in low-traffic secure environments, where

no ordinary behaviour can be learnt.

[Brand *et al.*, 1997; Natarajan and Nevatia, 2007] use computer-synthesized agents to simulate the trajectories of people meeting and interacting with each other. These agents are used to train Coupled Hidden Markov Models and Coupled Hidden Semi Markov Models, mathematical modelling frameworks for classifying behaviours like "Follow", "Approach, Talk and Continue Separately", "Approach, Talk and Continue Together", etc.

[Stauffer and Grimson, 2000] classifies blobs tracked over time using image position, speed, direction and size. In contrast to Hidden Markov Models, which need a sequence of observations, this method can categorize a single observation into one or more classes: cars, trucks, pedestrians on the path, lawnmowers on the lawn, etc.

[Dee and Hogg, 2004; 2005] have come up with a system that does not use statistical models to determine the most commonly taken routes, are able to explain trajectories not previously seen and can deal with moving obstacles like cars in a car park. Their first system [Dee and Hogg, 2004] determines the goal-directedness of a person by using the direction of the person's motion, along with a model of possible goal locations for the person at each point on his/her trajectory, given the location of obstacles in the scene. Obstacles are marked off manually in the image plane, and like our work, ray-tracing is used to find out a polygonal area of visibility in an arc of one radian on either side of the direction of motion of the tracked entity. Obstacle vertices falling within this polygonal area are marked as goals. In subsequent work [Dee and Hogg, 2005], they outline an algorithm that uses a modified Hausdorff measure to compare points on the trajectory followed by a person with points on trajectories to known goal sites generated by the system.

All previous work in literature is based on building up a database of likely flows. The motion of a suspected intruder with respect to a guard as a defining characteristic for "suspiciousness" is what is novel in our work. The use of the Dark Path algorithm for classification of human behaviour is also a novel contribution.

The rest of the paper is divided as follows: Background modelling and shadow removal is used to extract the positions of entities in the camera images and this is discussed in section 2. Section 3 details the transformation of image plane coordinates to a map of the ground plane. The tracking of multiple targets on the ground plane is discussed in section 4. Metrics for the detection of covert behaviour are introduced in 5. Section 6 details the experiments performed.

2 Background Modelling and Foreground Extraction

N frames of the training phase are used to create, for each pixel, a bi-modal distribution [Haritaoglu *et al.*, 2000] of the background that consists of 3 parameters: the minimum pixel value N(x), the maximum pixel value M(x) and the maximum inter-frame difference value D(x) where the pixel location is indicated by the index x.

During the background subtraction (BGS) process, a pixel x from image I^t is segmented as foreground if:

$$|I^t(x) - M(x)| \geq kD(x) \vee |I^t(x) - N(x)| \geq kD(x) \quad (1)$$

where the modelled variance was empirically determined to be 4.

2.1 Shadow Removal

The appearance of shadows and highlights in video leads to the inaccurate extraction of sillhouettes, and consequently the inaccurate tracking of blobs. It is essential that the segmented blobs be tightly restricted by their bounding boxes for a good image plane to ground plane transformation (described in section 3). Highlights and shadows cause an increase, and decrease, respectively of the intensity of the observed image region compared to the background model. A simple method to check this, is to find out if the intensity of a foreground pixel I^t at time t, is within a threshold of the background pixel I_b at the same location (the background pixel value is the mean intensity over the training frames) [Schindler and Wang, 2006]. Thus, the condition for foreground pixel I^t to be a shadow is:

$$\beta \leq I^t / I_b \leq \gamma \quad (2)$$

where β and γ have been empirically determined to be 0.8 and 1 respectively.

3 Homography

The image to ground plane mapping is done in this paper by using homography [Criminisi *et al.*, 1999] based on manually marked corresponding measurements between the image and ground planes. Let the points in the ground plane be represented using the homogeneous vector $X = (X, Y, Z)$, and their counterparts in the image plane by $x = (x, y, 1)$. Under perspective projection, the points are related by:

$$X = Hx \quad (3)$$

Written in vector notation, we have:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (4)$$

From which, we get the following simultaneous equations:

$$X = h_{11}x + h_{12}y + h_{13} \quad (5)$$

$$Y = h_{21}x + h_{22}y + h_{23} \quad (6)$$

$$Z = h_{31}x + h_{32}y + h_{33} \quad (7)$$

Letting $Z = 1$ in equation 7:

$$h_{31}x + h_{32}y + h_{33} = 1 \quad (8)$$

Multiplying equation 8 by X and Y, we get:

$$h_{31}xX + h_{32}yX + h_{33}X = X \quad (9)$$

and

$$h_{31}xY + h_{32}yY + h_{33}Y = Y \quad (10)$$

Equation 5 can be re-written as:

$$h_{11}x + h_{12}y + h_{13} + 0.h_{21} + 0.h_{22} + 0.h_{23} - X = 0 \quad (11)$$

Equation 6 can be re-written as:

$$0.h_{11} + 0.h_{12} + 0.h_{13} + x.h_{21} + y.h_{22} + h_{23} - Y = 0 \quad (12)$$

Substituting the value of X from equation 9 into equation 11:

$$h_{11}x + h_{12}y + h_{13} + 0.h_{21} + 0.h_{22} + 0.h_{23} - xXh_{31} - yXh_{32} - Xh_{33} = 0 \quad (13)$$

Substituting the value of Y from equation 10 into equation 12:

$$0.h_{11} + 0.h_{12} + 0.h_{13} + x.h_{21} + y.h_{22} + h_{23} - xYh_{31} - yYh_{32} - Yh_{33} = 0 \quad (14)$$

Equations 13 and 14 can be re-arranged to give the homogeneous equation system for n points $\mathbf{A}\widetilde{h} = \widetilde{0}$

where $\mathbf{A} =$

$$\begin{bmatrix} x_i & y_i & 1 & 0 & 0 & 0 & -x_iX_i & -y_iX_i & -X_i \\ 0 & 0 & 0 & x_i & y_i & 1 & -x_iY_i & -y_iY_i & -Y_i \\ & & & & & \vdots & & & \end{bmatrix}_{2nx9}$$

and $\widetilde{h} = \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \\ h_{33} \end{bmatrix}$

The elements of the transformation vector $\widetilde{h}$ are obtained from the eigen vector corresponding to the least eigen value of A^TA. For each pair of points on the image plane and the ground plane, we fill up two rows of the matrix A, which becomes a $2n$x9 matrix, where n is the number of points. The Singular Value Decomposition (SVD) of A:

$$A = UDV^T \quad (15)$$

results in the eigenvalues being arranged in decreasing order along the diagonal of the matrix D and the corresponding eigenvectors along the columns of the matrix V. The values of the eigenvector associated with the smallest eigenvalue make up the values of the vector $\widetilde{h}$. With four points ($n = 4$), $\widetilde{h}$ becomes the null-vector of A, i.e., the column of V associated with a zero-valued eigenvalue in D. The elements of $\widetilde{h}$ are re-arranged to form the elements of the matrix H in equation 3. The ground plane location corresponding to any arbitrary point in the image plane can now be found using this equation. Note that one of the cameras is a wide-angle camera and the OpenCV library was used to get the distortion parameters of the camera and undistort the images before the aforementioned transformation procedure.

4 Multiple Target Tracking

The Particle Filter tracks a target by maintaining a set of probability-weighted particles which approximate the target probability density function. Arbitrary densities (not restricted to unimodal or Gaussian as in the Kalman Filter) can be tracked, and the larger the number of particles, the better this approximation of target density. One particle filter is initialized for each target. The complexity of this approach is linear in the number of targets: $O(Nn)$ where N is the number of targets and n is the number of particles per target.

For each target, the problem, which is one of estimating the current pose $\mathbf{x_t} = [x_t, y_t,]^{'}$ of the target in a map of the environment, given a history of measurements and past pose esimates, can be solved by using Bayesian Reasoning.

This approach assumes that the environment is Markov, i.e., a future pose is independent of all prior poses, conditional on the fact that the current pose is known. The background subtraction followed by image plane to ground plane transformation supplies the ground plane coordinates of measurements to the tracking algorithm, starting at time 0 upto time t: $(y_0...y_t)$. The pose of the target at time t is estimated using the posterior probability density function, called the *belief*, conditioned on the data:

$$\begin{aligned} Bel\,(x_t) &= p\,(x_t|y_{0...t}) \\ &= p\,(x_t|y_t, y_{t-1}, y_{t-2}, ...y_0) \end{aligned} \quad (16)$$

Equation 16 can be transformed by Bayes' rule to give the following recursive update equation:

$$Bel\left(x_t\right) = \frac{p\left(y_t|x_t, y_{t-1}, ...y_0\right) p\left(x_t|y_{t-1}, ..., y_0\right)}{p\left(y_t|y_{t-1}, ..., y_0\right)} \tag{17}$$

Using the Markov assumption that measurements y_t are conditionally independent of past measurements, given the knowledge of state x_t, we get:

$$\begin{aligned} Bel\left(x_t\right) &= \frac{p\left(y_t|x_t\right) p\left(x_t|y_{t-1}, ..., y_0\right)}{p\left(y_t|y_{t-1}...y_0\right)} \\ &= \eta_t p\left(y_t|x_t\right) \sum_{x_{t-1}} p\left(x_t|x_{t-1}, y_{t-1}, ...y_0\right) p\left(x_{t-1}|y_{t-1}, ...y_0\right) \\ &= \eta_t p\left(y_t|x_t\right) \sum_{x_{t-1}} p\left(x_t|x_{t-1}, y_{t-1}, ...y_0\right) p\left(x_{t-1}|y_{t-1}, ...y_0\right) \\ &= \eta_t p\left(y_t|x_t\right) \sum_{x_{t-1}} p\left(x_t|x_{t-1}\right) Bel\left(x_{t-1}\right) \end{aligned} \tag{18}$$

where $\eta_t = \frac{1}{p(y_t|y_{t-1},...y_0)}$ is a normalizing constant that ensures the probabilities sum to one.

Equation 18 is the basis for the particle filter algorithm, which cycles through the following steps:

1. Predict
   Particles are predicted based on a motion model (first term within summation sign in equation 18).

$$p\left(x_t|x_{t-1}\right) = Ax_{t-1} + \nu_t \tag{19}$$

   where A is the state transition matrix and ν is the system noise.

2. Update
   Particles are updated using a perception model (second term on right hand side of equation 18).

$$p\left(y_t|x_t\right) = \frac{1}{\sigma_m\sqrt{2\pi}} e^{\frac{-(\Delta m)^2}{2\sigma_m^2}} \tag{20}$$

   where Δm is the distance between the poses of the particle and the associated measurement. σ_m is the standard deviation of the measurement noise.

3. Resample
   Particles are resampled according to their probabilistic weights. Particles with higher weights have a higher probability of getting propagated to the next iteration. Systematic resampling [Arulampalam *et al.*, 2002] with $O(n)$ complexity is used, where n is the number of particles per filter.

4.1 Tracking Mechanism

N particle filters are used to track N targets. Each filter consists of a set of n particles. Each particle at time t has a state x_t , which evolves according to state transition equation 19. Each particle also has a probabilistic weight associated with it. For each filter, apart from the particle states and their associated weights, the following data is maintained:

1. The mean position of all the particles in the filter.
2. The standard deviation of the particles from the mean position.
3. The number of times in the past frames that the filters cumulative un-normalized probability has fallen below a threshold.

4.2 Data Association

The data association problem, one of associating the set of latest measurements with filters, is achieved in a modified nearest neighbour fashion. A measurement is assigned to the filter with the closest mean position, only if it is within 1.5 standard deviations of that filter. Measurements already associated with a filter are not associated again with any other filter, in what is known as a "hard" assignment. If a "soft" assignment procedure is followed, it has been observed from simulations, that one target tends to capture more than one particle filter when the filters are close together.

Given that the measurements, i.e., the ground plane locations of the targets are inferred from their camera image plane locations, it seems logical to utilize colour information in disambiguating targets that are close together or those that have disappeared in one camera, to appear in another. However, the colour responses of the two cameras used in the experimental setup were quite different, and including colour information to associate tracked targets to new measurements led to frequent loss of target track.

Once a measurement is assigned to a filter, the weights of all the particles in that filter are updated based on their distances from the measurement, as in equation 20.

4.3 Track Initialization

The track initialization procedure initializes a new particle filter around un-associated measurements that have been clustered over a few frames. Refer to [Chakravarty and Jarvis, 2005] for more details on the clustering procedure.

4.4 Track Deletion

The track deletion procedure uses the cumulative unnormalized probability, which is maintained for each filter as a measure of its confidence in tracking a target. If this

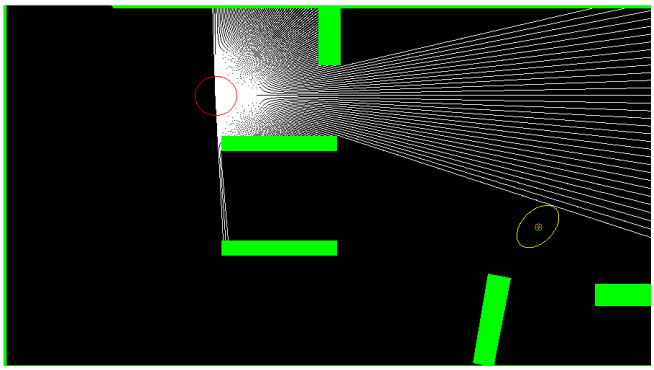

Figure 1: Visibility map with respect to sentry

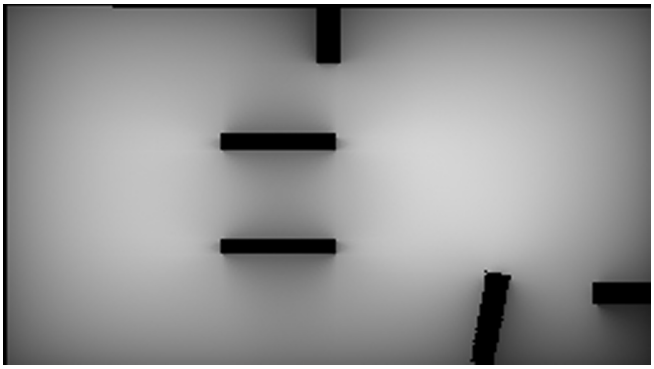

Figure 2: Generalized visibility map

cumulative weight falls below a pre-defined threshold for a majority of the last τ frames, then the particular filter is deactivated.

4.5 Track Continuation

The track continuation procedure implements the following steps (discussed earlier in this section):

1. Predict
2. Update
3. Re-sampling

5 Covert Behaviour Detection Metrics

Two metrics are used to determine the visibility of a tracked suspect at each point on his trajectory. These are described in the following sub-sections.

5.1 Visibility Map with respect to Sentry

A visibility map (Figure 1) is used to detect the covert behaviour of a suspect with respect to a sentry. It is the extent of a sentry's field of vision, given his/her position on a map of the environment. Ray-tracing is used to find out the visibility map of the sentry (rays are extended in each direction around the sentry's current position till an obstacle is hit; viewing angles are limited to 90 degrees on either side of the sentry's motion direction) at each point on his/her beat. Once an intruder is detected (the 2nd entity to enter the arena is assumed to be the intruder), if he/she is within the visibility map, the proximity to sentry is recorded at that point on the intruder's ground plane trajectory.

5.2 Generalized Visibility Map

A generalized visibility map (first used by [Marzouqi and Jarvis, 2006] for covert path planning for a mobile robot) is constructed to measure the visibility of the tracked suspect with regard to his environment, regardless of the position of the sentry. Rays are traced until they hit obstacles in all directions around each unoccupied cell in a map of the environment. The number of cells that are visible from the current cell are recorded for each cell in the environment. This is a time-consuming method, but can be done offline. For each position in the environment, a readout from this map (shown in Figure 2) gives the generalized visibility of that spot.

6 Experiments

An arena was set up with obstacles to hide behind. All parts of the arena were visible with either one of two cameras mounted overhead. The map of the arena and an image to ground plane mapping for each camera was provided to the system. Two actors were employed to play the roles of the sentry and intruder. Sixteen experiments (one experiment being a traversal of the intruder from one point in the arena to another) were recorded. The sentry was asked to patrol the environment at a relatively constant speed in all the experiments. The intruder was asked to pursue a covert path in 8 of the runs, and a normal path in the other 8.

Instructions for the covert path were as follows: "Move from the START to the GOAL without being seen by the guard. You can crouch and hide behind obstacles. Walk/crawl slowly: You must not run. Instead, plan your route to avoid the guard and exploit his blindspots. You can take as long as you like. The guard will repeat his patrol route continuously. Try to minimise any view of you the guard has, but don't worry if you're seen temporarily. Just keep going."

Instructions for the overt path: "Move from the START to the GOAL. Walk slowly: you must not run. You can take as long as you like. You may encounter another subject during your movement. You can choose any route you like."

A snapshot of the target tracking is displayed in Figure 3. Video streams from the 2 cameras show intruder and sentry positions, and particle filters track their respective positions on a map of the ground plane. Uncertainty ellipses for each filter are displayed.

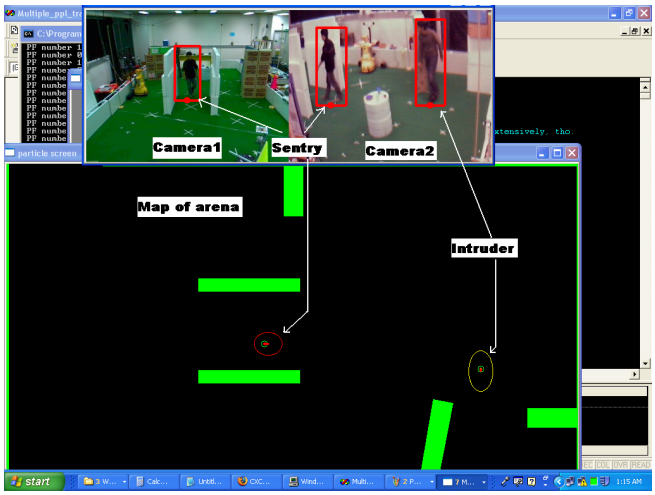

Figure 3: Snapshot of target tracking: Intruder and sentry positions in camera images and arena map. Particle Filter ellipses show tracking uncertainty in each dimension

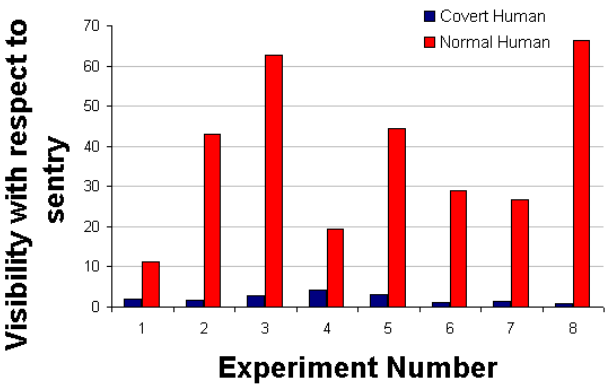

Figure 4: Visibility of intruder with respect to sentry (averaged over trajectory lengths) for 8 covert(blue) and 8 overt (red) experiments

Figure 4 displays the visibility value of the intruder with respect to the sentry averaged over the length of each experiment. Covert runs are shown in blue, and overt ones in red.

Figure 5 shows the average general visibility for the 16 trajectories. Again, blue and red bars are used to show the covert and overt trajectories respectively.

It can be seen that the covert (blue) trajectories show a smaller mean general visibility value in all the experiments except one for according to the general visibility metric, but significantly lower values for the visibility with respect to sentry metric.

Thus, using a network of fixed cameras, we can track the motion of individuals with enough precision to distinguish the behaviour of individuals having differing intentions. This result implies another conclusion, namely that persons' intent on evasion choose to move in a different way to "normal" individuals. We can also measure how their intent affects their interactions with others, for example, they may actively try to avoid being seen by others.

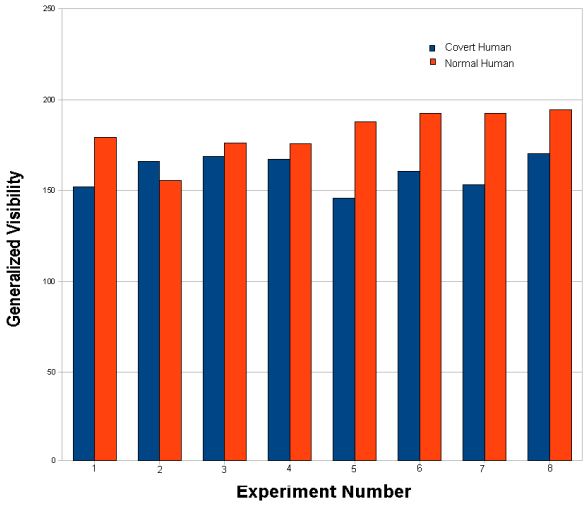

Figure 5: General visibility of intruder (averaged over trajectory lengths) for 8 covert(blue) and 8 overt (red) experiments

7 Conclusion

We have shown that a network of fixed cameras can simultaneously and collaboratively track the movements of a number of human targets in a low-traffic indoor environment. This is nothing new; in fact, we build on our earlier work that shows how the physical configuration of the camera network can be deduced by observing the motion of a mobile agent [Rawlinson *et al.*, 2004]. The arrangement of the cameras can also be manually estimated, as done here.

This work goes beyond previous research by attempting to infer the intentions of tracked human targets, from their movements. The work overcomes two principal difficulties, namely the acquisition of suitable data and autonomous analysis of that data to yield meaningful and useful results. In this case, the objective was to detect persons attempting to avoid or hide from trusted agents (guards) in secure environments.

Although there are a number of technologies that help to identify potentially threatening or unauthorised persons, few are able to exploit the low-resolution fixed camera networks that are prolific today. With such data it is hard to model facial appearance, gait, expression or other factors that might arouse suspicion. But it is relatively easy to track position.

The experiments conducted for this article are preliminary and need to be repeated under stricter double-blind experimental conditions, in order to ensure that realistic human behaviour is observed. Nevertheless, it would appear that a person who *tries* to avoid being seen by guards *does* move differently to a person walking normally. That is, the intent to hide changes behaviour in a way that can be autonomously and reliably detected by typical security equipment.

As implemented, the system analyses interactions between trusted agents (the guards) and unknown persons. In effect, it estimates the intent of one person (the intruder) by measuring their reaction to a third party (the guard). It is assumed that if the guard encounters an unauthorised person, he or she will establish their right to access that location.

Of course, a particularly sneaky intruder may avoid contact with guards altogether. And other times an authorised person might avoid all security guards by mere chance. How can these circumstances be differentiated? How can the system separate innocent events from deliberate evasion?

Fortunately, our results also show that the movements of a person attempting to avoid security guards can be distinguished from normal motion, even without an encounter or a near-miss. The motions of people trying to hide conform more closely to the Covert Path transform [Marzouqi and Jarvis, 2006].

Future work will include more carefully designed experiments to verify that the behaviour observed here is realistic and representative. Further, it needs to be shown that in larger and more complex environments different human intentions continue to have a measurable effect on human behaviour. The authors hope to analyse other aspects of human movement in an effort to determine other criteria that help to highlight suspicious persons. For example, the speed of motion may be significant. Not all such metrics will be intuitive: Although the authors believed that prolonged proximity to obstacles would be a good indicator of "hiding", our results suggest that this is not true. People tend to move close to obstacles for efficiency and convenience.

This work may have a number of direct applications in low-traffic secure environments, such as public places out of hours, and private business premises. Many false-alarms could be eliminated and security improved by causing the system to autonomously instruct guards to intercept individuals who have, by chance or design, thus far avoided inspection. Although our goal in creating this system was explicitly to detect "people trying to hide from security guards", similar analysis of peoples' relative movements could be applied to many other scenarios. The general element is the inference of human intent, from their movements with respect to other people. For example, CCTV cameras could be configured to detect people following others late at night, possibly prior to an assault.

References

[Arulampalam *et al.*, 2002] S. Arulampalam, S. Maskell, N. J. Gordon, and T. Clapp. A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking. *IEEE Transactions of Signal Processing*, 50:174–188, 2002.

[Brand *et al.*, 1997] M. Brand, N. Oliver, and A.P. Pentland. Coupled hidden markov models for complex action recognition. In *CVPR97*, pages 994–999, 1997.

[Chakravarty and Jarvis, 2005] P Chakravarty and R. Jarvis. Multiple target tracking for surveillance: A particle filter approach. In *ISSNIP*, 2005.

[Chakravarty *et al.*, 2006] P. Chakravarty, D. Rawlinson, and R. Jarvis. Person tracking, pursuit and interception by mobile robot. In *Australasian Conference on Robotics and Automation (ACRA)*, 2006.

[Criminisi *et al.*, 1999] A. Criminisi, I.D. Reid, and A. Zisserman. A plane measuring device. *IVC*, 17(8):625–634, June 1999.

[Dee and Hogg, 2004] H.M. Dee and D.C. Hogg. Detecting inexplicable behaviour. In *BMVC04*, 2004.

[Dee and Hogg, 2005] H. M. Dee and D. C. Hogg. Navigational strategies and surveillance. In *Proceedings of the IEEE International Workshop on Visual Surveillance*, pages 73–81, 2005.

[Haritaoglu *et al.*, 2000] I. Haritaoglu, D. Harwood, and L.S. Davis. W4: Real-time surveillance of people and their activities. *PAMI*, 22(8):809–830, August 2000.

[Hu *et al.*, 2004] W. Hu, D. Xie, T. Tan, and S.J. Maybank. Learning activity patterns using fuzzy self-organizing neural network. *SMC-B*, 34(3):1618–1626, June 2004.

[Marzouqi and Jarvis, 2006] M. Marzouqi and R. Jarvis. Covert path planning for autonomous robot navigation in known environments. In *Australasian Conference on Robotics and Automation (ACRA)*, 2006.

[Natarajan and Nevatia, 2007] P. Natarajan and R. Nevatia. Coupled hidden semi markov models for activity recognition. In *Motion07*, pages 10–10, 2007.

[Rawlinson *et al.*, 2004] D. Rawlinson, P. Chakravarty, and R. Jarvis. Distributed visual servoing of a mobile robot for surveillance applications. In *Australasian Conference on Robotics and Automation (ACRA)*, 2004.

[Schindler and Wang, 2006] K. Schindler and H. Wang. Smooth foreground-background segmentation for video processing. In *ACCV06*, pages II:581–590, 2006.

[Stauffer and Grimson, 2000] C. Stauffer and W. Grimson. Learning patterns of activity using real-time tracking. *PAMI*, 22(8):747–757, August 2000.