Annals of the BMVA Vol. 2010, No. 6, pp 1–12 (2010)

Performance Evaluation of Tracking for Public Transport Surveillance

V. Leung, J.Orwell and S.A. Velastin

Digital Imaging Research Centre,
Faculty of Computing, Information Systems and Mathematics,
Kingston University,
Kingston upon Thames, KT1 2EE
⟨leung.valerie@gmail.com; j.orwell,s.velastin@kingston.ac.uk⟩

Abstract

Starting with the end-user requirements, it is argued that the most appropriate performance metric with which to evaluate video-based tracking methods, in the public transport domain, is to measure the information gained through their use. This is equivalent to the reduction in uncertainty about a passenger's whereabouts, after tracking and appearance-based measurements have been taken into account. In this paper, we present a framework for the performance evaluation of such a system. Error propagation analysis is performed to investigate the impact of each system component on the final uncertainty, and allows an end-user requirement to be propagated through the system to determine the minimum performance requirement for each component. This proposed analysis framework is demonstrated on a simulated system.

1 Introduction

In public transport networks, the quality and quantity of available video surveillance data has increased significantly in recent years. Digital cameras, wireless transmission systems and effective digital compression standards have simultaneously reduced cost and increased the potential for higher quality signals. There is clearly the potential for automatic analysis of the data for various purposes.

This paper considers one particular objective for automatic analysis: the tracking of an individual (or 'target') as they use a small urban 'metro' network of stations and trains, from their point of entrance into the network, to the point when they leave. In this scenario, an operator manually 'tags' them in one camera, and the system subsequently displays the best available camera view of them, hence the application is called 'Tag and Track'.

In this analysis, a model is introduced for the transport network, the passengers' movement, and the processing of surveillance data. We propose that the appropriate performance

Annals of the BMVA Vol. 2010, No. 6, pp 1–12 (2010)

evaluation metric for such a system is the expected number of manual interventions necessary to maintain the correct view of the person. This has a direct relationship to the uncertainty associated with their position, and this uncertainty is reduced by the application of computer vision methods such as tracking and appearance-based recognition.

This model is applied to the Torino Metro network, and an analysis based on simulated data is carried out. The first result is the predicted overall efficacy for given levels of performance from specific computer vision components. The second result is the derivation of the error contribution from each component to allow the calculation of the performance levels needed for a given overall end-user requirement. A study of the marginal impact of changes in performance is also presented.

There are numerous applications of this analysis. Thus, in addition to the public transport (metro) scenario, it may be applied to any multi-sensor tracking problem, e.g. vehicles in a road system, or people in a shopping centre. The analysis can be applied to real-time systems, or to systems that track forward through archives ('where did this person go?'), or backwards through archives ('where did this person come from?') In each of these scenarios, the analysis may be used in several ways. First, it can be used to estimate the expected amount of manual intervention necessary to maintain correct tracks. Second, it allows comparisons of different scenarios e.g. levels of crowding or camera resolution. Finally, it provides a mechanism for understanding how improvements to specific components of the tracking and recognition solution will affect the overall performance.

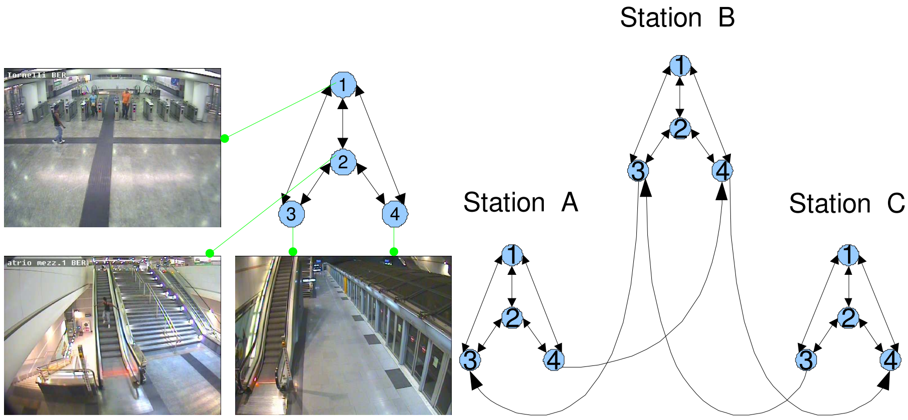
1.1 Previous work

Tracking pedestrians between multiple cameras for both overlapping and non-overlapping cases have been studied extensively recently [Kim and Davis, 2006, Javed et al., 2005, Fleuret et al., 2008, Makris et al., 2004, Rahimi et al., 2004, Khan and Shah, 2006]. For overlapping cameras, the geometric relations between the cameras are used to integrate the observations [Fleuret et al., 2008, Khan and Shah, 2006]. For non-overlapping cameras, the relationship between cameras are provided by observing the location and velocity of the objects traversing the scene [Makris et al., 2004, Rahimi et al., 2004], and trajectories of the pedestrians can be estimated by techniques such as Kalman filter and particle filter [Kim and Davis, 2006, Fleuret et al., 2008]. In both settings, the appearance of the target is modelled, typically using colour distributions, providing an independent term to the probabilistic expression. The 'Tag and Track' application can be viewed as a subset of the multi-object multi-camera problem.

The assessment of the performance of many tracking approaches is usually carried out on an algorithmic level, e.g. the spatial accuracy and continuity of the tracks, the ability to resolve occlusions etc., and many methods have been proposed for this purpose. [Black et al., 2003] proposed a method where ground truth is automatically generated, eliminating the laborious task of manual ground truth generation. Their tracking metrics include positional error and tracking success rate. [Brown et al., 2005] investigated a two-pass method to address track merging and fragmentation errors. The approach in [Needham and Boyle, 2003] focussed on trajectory comparison, taking into account temporal lag and spatial shift. In order to allow the different algorithms to be compared, efforts such as the PETS workshops[PETS] provide standard datasets for testing and evaluation; the tracking metrics include the number of track false positives and average positional error.

From a systems evaluation point of view, the iLIDS dataset [HOSDB] provided by the

Annals of the BMVA Vol. 2010, No. 6, pp 1–12 (2010)



(a) Single station; the numbers 1 to 4 represent the hall, the mezzanine, and the two platforms respectively

(b) Interconnections among stations; note that the platforms are connected with unidirectional arrows.

Figure 1: Layout of a camera network.

HOSDB (Home Office Scientific Development Branch) has recently been extended to multi-camera 'Tag and Track' scenarios. Their proposed evaluation metric is the F-measure [Van Rijsbergen, 1979], examining the number of instances that a track is successful. [Ning and Tan, 2006] proposed a framework for tracking a moving target in an environment with a heterogeneous camera setup. The evaluation metrics used are coverage and accuracy, and multiple candidate targets are returned for operator inspection to determine the correct match.

While these current evaluation methodologies can be applied to the 'Tag and Track' application here, many tend to focus on the algorithmic performance, which is insufficient in conveying the system performance to the operators. [Ning and Tan, 2006] provide the end-users with multiple candidate targets which is useful, but they have not provided a direct link between algorithmic performance and an operator-relevant metric.

In this paper, we argue that an information-theoretic evaluation approach is appropriate for two reasons. First, it can accommodate many algorithmic components in its formulation, and combine their contribution in a consistent manner. Second, it can summarise the constituent algorithmic performance into one metric that is relevant to the operators. The metric is appropriate because it provides an estimate of the potential for the time that can be saved by deploying the tracking system in a given environment. This is achieved by examining the expected rate of interactions required to correct the track produced by the system, compared with the expected rate of interactions required if no system was present. This expectation is based on prior distributions of activity, which are assumed to be valid for the prediction. If estimates are required for different sub-groups, then multiple metrics would need to be defined and calculated. Finally, the end-user may have different requirements, e.g. unbiased estimates of pedestrian density or route selection, which do not make use of the proposed metric, and would require an additional metric.

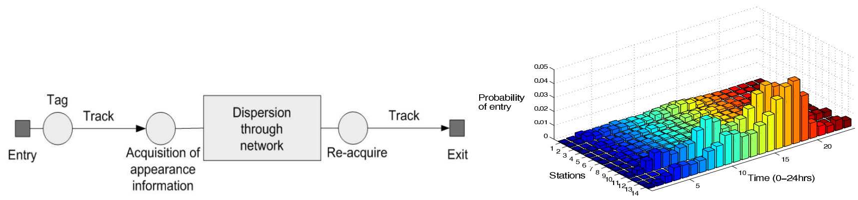
Annals of the BMVA Vol. 2010, No. 6, pp 1–12 (2010)

2 System representation

A metro train network can be represented as a directed graph of nodes and edges. Usually these represent stations and routes, however the graph can include further detail such as the areas within each station that passengers traverse in order to complete their journey. This results in a network such as that shown in Fig. 1, which includes the ticket hall, mezzanine and platforms for each station. Each edge in the graph can be assigned several attributes, such as the characteristic journey time between nodes and the expected density of passenger journeys between the respective nodes.

The model from which these attributes are generated can be simple or complex. For example one approach is to use Gaussian models for the distributions of journey times, passenger densities that vary according to the time of day but not the day of the week, and a fixed, empirically derived distribution between the possible journeys available to a passenger. More complex system behaviour can also be modelled, such as time-dependent distributions of choice of route and passenger speed (people walk quicker in the 'rush-hour'), and also mixtures of passenger behaviour (some passengers are 'loitering' and have no intention of travelling to another station).

Two categories of computer vision components are included in this system model: tracking, and appearance-based recognition. The former is intended to be used when consecutive cameras along a passenger's journey have overlapping or adjacent fields of view. The latter is used when there is a significant gap between the fields of view covered by these cameras. An illustration of the 'Tag and Track' system model is shown in Fig. 2(a), and an example of the entrance distribution constructed from statistics from the end-user is shown in Fig. 2(b).



(a) System diagram (b) Entry distribution

Figure 2: (a) System diagram. (b) An illustration of the entry distribution at the Torino metro network, showing the probability of entry at each station and time (discretised).

3 Evaluation of Automatic Tracking

The proposed method of performance evaluation is to measure the expected information gain that results from the application of any method to track or recognise the passengers. This is equivalent to the reduction in uncertainty (entropy) of the random variable used to describe the whereabouts of the person to be tracked. This section presents the notation for this framework; further details can be found in [Leung et al., 2008].

3.1 Uncertainty of localisation in a public transport network

In this framework, the following assumptions have been made. It is assumed that a closed world of N people use the network exactly one time each; that each observation refers to exactly one 'passenger'; and that the waypoints used to calculate the prior uncertainty $H(Y)$ are the entrance and exit of each passenger.

The set of entrance observations are represented as the set of events $\{x_1, \ldots, x_i, \ldots, x_N\}$ and the set of exit observations are represented as $\{y_1, \ldots, y_i, \ldots, y_N\}$. Each event has an associated time t and waypoint index u, defining when and where it took place: $x_i = x_i(t_i, u_i), y_j = y_j(t_j, u_j)$. The times t_i and t_j are between 0 and τ (e.g. over 24 hours) and the waypoints u_i and u_j take values between 1 and U. These entrance and exit events are samples from the underlying probability density functions (p.d.f.s), $p_x(t, u)$ and $p_y(t, u)$, respectively. Both these p.d.f.s represent a single entrance (or exit) event, happening at one of the waypoints. The entrance density function is specified externally (an illustration is shown in Fig. 2(b)), while the exit density function can be calculated from the convolution of the entrance p.d.f. with a dispersion p.d.f. θ, which specifies the relative likelihood of each of the routes and journey times that are available to the passenger.

The operator $I(\cdot)$ is used to denote the identity of the passenger associated with the event. The situation in which a passenger's entrance and exit were correctly recorded as x_i and y_j respectively is written as $I(x_i) = I(y_j)$. More generally, a random variable Y_i can be defined over the sample space of exit events $\{y_j\}$ to represent the probability $P(I_{ij})$ that y_j is the correct association with the entrance event x_i:

$$P(I_{ij}) = P(I(x_i(t_i, u_i)) = I(y_j(t_j, u_j))). \tag{1}$$

3.2 Prior uncertainty

The 'tagging' performed by the operator corresponds to the localisation x_i at time t_i and waypoint u_i. The uncertainty associated with the exit events $y_j(t_j, u_j)$ depends on two factors. The first is the relation between the times (t_i, t_j) and the waypoints referenced by (u_i, u_j). Some combinations of these values are impossible because of causality and physical constraints (e.g. south-bound trains can only reach stations south of the station from which a person boards), while some journeys are more likely than others. This relationship is expressed in the dispersion p.d.f. θ:

$$\theta(\alpha_{ij}, \mu_{ij}, \sigma_{ij}, t_j - t_i) = \begin{cases} C_\theta \frac{\alpha_{ij}}{\sigma_{ij}\sqrt{2\pi}} \exp\left\{\frac{-(t_j - t_i - \mu_{ij})^2}{2\sigma_{ij}^2}\right\} & \text{if}(t_j - t_i) > 0 \\ 0 & \text{otherwise} \end{cases}$$

where μ_{ij} is the mean travel time between the waypoints corresponding to the entry and exit events, with standard deviation σ_{ij}, and C_θ is a normalising constant. α_{ij} is the relative frequency with which passengers make the journey from waypoint u_i to waypoint u_j.

The dispersion model chosen here is one possibility; models of higher complexity can be used, and different dispersion models can be applied to represent different passenger states. For example, the 'commuter' would follow a different prior distribution from the 'loiterer'. The state can be set by the operator, or determined automatically by assessing the fit between the models and the observed data.

The second factor on which the uncertainty associated with the exit event $y_j(t_j, u_j)$ depend is the number of other passengers present at this point and time in the network. For example, if the passenger in question was the only one present on the network, then $P(I_{ij}) = 1$. On the other hand, if there were many other passengers exiting from waypoint u_j at around the time t_j, then the probability mass would be divided between (at least) these corresponding exit events. The probability $P(I_{ij})$, using only the prior information, is therefore:

$$P(I_{ij}) = \frac{P(y_j(t_j, u_j)|x_i(t_i, u_i))}{P(y_j(t_j, u_j)|x_i(t_i, u_i)) + (N-1)P(y_j(t_j, u_j))} \tag{2}$$

$$= \frac{\theta(\alpha_{ij}, \mu_{ij}, \sigma_{ij}, t_j - t_i)}{\theta(\alpha_{ij}, \mu_{ij}, \sigma_{ij}, t_j - t_i) + (N-1)P(y_j(t_j, u_j))}. \tag{3}$$

The first and second terms in the denominator correspond to the 'correct answer' and the 'clutter' (or the presence of other people) respectively. The entropy of Y_i is therefore:

$$H(Y_i) = -\sum_{j=1}^{N} P(I_{ij}) \log P(I_{ij}) \tag{4}$$

Finally, to obtain an expression for the overall uncertainty, we must generalize over the expected distribution of entrance events x_i (occurring at the various stations u_i and times t_i):

$$H(Y) = E_X[H(Y_i)] = \sum_{u'=1}^{U} \int_0^{\tau} \delta(u' = u_i) P(x_i(t_i, u_i)) H(Y_i) dt. \tag{5}$$

3.3 Reducing the uncertainty through measurements

The surveillance information is gained through measurements Z, resulting in a reduced uncertainty $H(Y|Z)$. These measurements can encompass appearance-based cues Z_A, tracking based on spatial continuity Z_T, and any other approach that is thought to deliver additional information about the whereabouts of a passenger, and provide a conditional probability of the correct re-acquisition, $P(I_{ij}|Z)$:

$$P(I_{ij}|Z) = \frac{P(y_j|x_i)P(y_j|Z_A)P(y_j|Z_T)^2}{P(y_j|x_i)P(y_j|Z_A)P(y_j|Z_T)^2 + \neg P(y_j|x_i) \neg P(y_j|Z_A) \neg P(y_j|Z_T)^2} \tag{6}$$

where $\neg P(y_j|Z_T) = 1 - P(y_j|Z_T)$, i.e. the symbol $\neg$ represents the cases where the exit localisation is from the other $N-1$ people in the network. The tracking terms are squared because of the two tracking components in the system (assumed to have the same parameters). Further measurements can be directly included in the formulation if they are independent of the other measurements; similarly, if only one type of measurement is considered, the other terms can be removed. The uncertainty of the localisation given the measurements is:

$$H(Y_i|Z) = -\sum_{j=1}^{N} P(I_{ij}|Z) \log P(I_{ij}|Z). \tag{7}$$

$H(Y|Z)$ uses the expected value of Eq.(7), as in Eq.(5).

Annals of the BMVA Vol. 2010, No. 6, pp 1–12 (2010)

Tracking measurements
The term $P(y_j|Z_T)$ is defined as the probability that a track is maintained successfully over one camera view (or a series of overlapping camera views) for the entire duration that the person is in view. This is equivalent to the probability that a measurement is correctly associated with the track over that time period. The following is summarised from [Mori et al., 1992]. Mathematically, the probability $P(y_j|Z_T)$ is defined as:

$$P(y_j|Z_T) = P_c^f, \quad P_c = \exp\{-C_m\beta\bar{\sigma}^m\}, \quad C_m = 2^{m-1}\pi^{(m-1)/2}\frac{\Gamma(\frac{m+1}{2})}{\Gamma(m/2+1)} \tag{8}$$

where P_c is the probability of correct data association. f is the number of frames and m is the dimension of the measurement space. $\Gamma(\cdot)$ is the Gamma function. The average innovations standard deviation, $\bar{\sigma}$, summarises the performance of the tracker:

$$\bar{\sigma} = \det(Q+R)^{\frac{1}{2m}} \tag{9}$$

where Q and R are the covariance matrices of the process and measurement noises respectively. The term β is the object density in the measurement space, defined as:

$$\beta = \frac{\nu}{B_m r^m}, \quad B_m = \frac{\pi^{m/2}}{\Gamma(\frac{m}{2}+1)}. \tag{10}$$

r is the radius of the m-ball in the measurement space, and ν is the mean number of people, modeled by a Poisson distribution [1].

Appearance-based measurements
The probability $P(y_j|Z_A)$ is the reduction in uncertainty that an appearance-based measurement *alone* can provide [2]. Z_A can be a scalar (from a single feature), or a vector of measurements of multiple features. Measurements of a single feature from a sample population will produce histograms similar to those in Fig. 3(b). The solid histograms represent the distance between the measurements given the correct and incorrect identity respectively, while the fitted models are shown respectively as a dotted line and crosses. Thus the more separable the two distributions are, the higher the probability of a correct match. The histograms shown have been generated using a population of 47 [Annesley et al., 2006]; more samples will result in smoother histograms. Extension to multi-dimensional features is achieved by fitting Gaussian Mixture Models to the multi-dimensional histograms.

4 Error propagation analysis

This section presents the error propagation analysis of the system, examining the relationship between the final performance and the constituent components. This result can be used to examine both the expected change in the output given a change in the input, and how good a particular system component has to be given a required level of output performance.

[1] The number of people in a given space is discrete, and the presence of each person is assumed independent.
[2] This reduction in uncertainty is a property of the appearance-based measurement, and is different from the overall reduction in uncertainty, given by the difference between $H(Y)$ and $H(Y|Z)$.

Annals of the BMVA Vol. 2010, No. 6, pp 1–12 (2010)

The final performance figure is the expected number of times an operator's attention is required, or the expected number of 'wrong guesses' that is made before the correct answer is given. We denote this figure of merit as $\overline{W}$, defined as:

$$\overline{W} = \frac{2^H - 1}{2}. \tag{11}$$

The derivation of this expression is straightforward by considering the case with a uniform prior, and the cumulative probabilities along the branches of a decision tree for a given population size. This result can be extended to the cases with non-uniform priors.

To examine the relationship between the input parameters and $\overline{W}$, error propagation can be applied to relate the change in $\overline{W}$, $\Delta\overline{W}$, with the change in each input parameter. Here, the input parameters can be divided into three parts : those related to the dispersion p.d.f. θ, the tracking measurements Z_T, and the appearance measurements Z_A. The respective parameters are enumerated by k_1, k_2 and k_3. The change in $\Delta\overline{W}$ introduced by the change in each parameter can then be computed by taking the partial derivative of $\overline{W}$ with respect to the particular parameter, multiplied by the change. The overall change $\Delta\overline{W}$ is therefore:

$$\begin{aligned}\Delta\overline{W} \;=\; & \sum_{k_1}\left|\frac{\partial\overline{W}}{\partial H}\cdot\frac{\partial H}{\partial P(I_{ij}|Z)}\cdot\frac{\partial P(I_{ij}|Z)}{\partial P(y_j|x_i)}\cdot\frac{\partial P(y_j|x_i)}{\partial\theta_{k_1}}\right|\cdot\Delta\theta_{k_1}+\\ & +\sum_{k_2}\left|\frac{\partial\overline{W}}{\partial H}\cdot\frac{\partial H}{\partial P(I_{ij}|Z)}\cdot\frac{\partial P(I_{ij}|Z)}{\partial P(y_j|Z_T)}\cdot\frac{\partial P(y_j|Z_T)}{\partial T_{k_2}}\right|\cdot\Delta T_{k_2}+\\ & +\sum_{k_3}\left|\frac{\partial\overline{W}}{\partial H}\cdot\frac{\partial H}{\partial P(I_{ij}|Z)}\cdot\frac{\partial P(I_{ij}|Z)}{\partial P(y_j|Z_A)}\cdot\frac{\partial P(y_j|Z_A)}{\partial A_{k_3}}\right|\cdot\Delta A_{k_3}.\end{aligned} \tag{12}$$

For this subsequent analysis, the partial derivatives w.r.t. T_{k_2} and A_{k_3} are examined to investigate how changes to the performance of these components affect the overall performance. (We do not examine partial derivatives of the parameters of θ, because these parameters describe the dynamics of the system under observation, and are therefore considered constant for this exercise.)

The calculation of the partial derivatives of Eq.(12) are now explained. The first two partial derivatives common to each additive term are:

$$\frac{\partial\overline{W}}{\partial H} = 2^{H-1}\ln 2 \tag{13}$$

$$\frac{\partial H}{\partial P(I_{ij}|Z)} = -\sum_{j=1}^{N} 1 + \log P(I_{ij}|Z). \tag{14}$$

In order to calculate the partial derivative $\frac{\partial P(I_{ij}|Z)}{\partial P(y_j|Z_T)}$, we refer to Eq.(6) and denote the denominator and the numerator as F and G respectively. The partial derivative is therefore:

$$\begin{aligned}\frac{\partial P(I_{ij}|Z)}{\partial P(y_j|Z_T)} \;=\; & \frac{2P(y_j|x_i)P(y_j|Z_A)P(y_j|Z_T)}{F} - \\ & -\frac{2G(P(y_j|x_i)P(y_j|Z_A)P(y_j|Z_T) - \neg P(y_j|x_i)\neg P(y_j|Z_A)\neg P(y_j|Z_T))}{F^2}.\end{aligned} \tag{15}$$

Annals of the BMVA Vol. 2010, No. 6, pp 1–12 (2010)

For the partial derivative $\frac{\partial P(y_j|Z_T)}{\partial T_{k_2}}$, we will use the measurement noise as an example. Assuming that the process and the measurement noise covariances are identity matrices scaled by s_Q and s_R respectively, the partial derivative w.r.t. s_Q is:

$$\frac{\partial P(y_j|Z_T)}{\partial s_R} = -fP_c^{f-1}C_m\beta m\overline{\sigma}^{m-1}\exp\{-C_m\beta\overline{\sigma}^m\} \times (3s_Q^2 + 6s_Qs_R + 3s_R^2). \tag{16}$$

The term $\frac{\partial P(I_{ij}|Z)}{\partial P(y_j|Z_A)}$ can easily be derived and is not included here.

In order to define $\frac{\partial P(y_j|Z_A)}{\partial A_{k_3}}$, recall that the term $P(y_j|Z_A)$ is the reduction in uncertainty that an appearance-based measurement can provide, and this information is encapsulated by the histograms in Fig.3(b). The term $\frac{\partial P(y_j|Z_A)}{\partial A_{k_3}}$ can then be interpreted as the separability of the two histograms. Separability in this context depends on the weights of the two distributions, the number of standard deviations between their means, as well as the underlying distribution type (Gaussian, Poisson etc). We propose the use of the expected reduction in uncertainty given the appearance-based measurement as the measure of separability. Hence, $P(y_j|Z_A) = E[\Delta H_A]$. Here we have not defined an analytic expression for $E[\Delta H_A]$, but a look-up table (LUT) with sufficient resolution has been generated. This is illustrated in Fig.3(a), showing the expected reduction in uncertainty for a weighting parameter $\rho = |P(1) - 0.5| + |P(0) - 0.5|$ varying from 0 to 0.9, and the number of standard deviations between the means (n_{SD}) varying from 0 to 8, for Gaussian distributions. The error propagation terms are therefore:

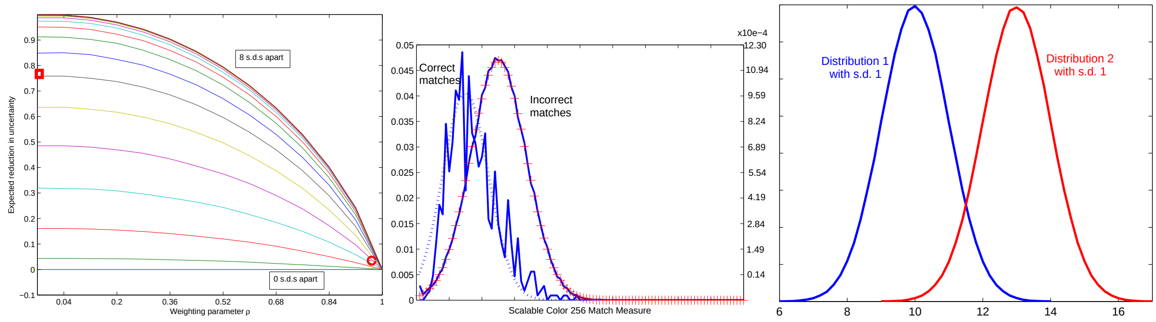
$$\frac{\partial P(y_j|Z_A)}{\partial \rho} = \frac{\partial E[\Delta H_A]}{\partial \rho}, \quad \frac{\partial P(y_j|Z_A)}{\partial n_{SD}} = \frac{\partial E[\Delta H_A]}{\partial n_{SD}}. \tag{17}$$

The use of the error propagation analysis in the forwards and backwards directions are illustrated next.

4.1 Impact of input changes on performance

This forwards problem answers questions such as: what is the impact on the output given a change in a particular input? Assuming that there are 1000 people using the metro per hour and the system parameters listed in Table 1, the system is simulated. The parameters in Table 1 correspond to a 3-dimensional measurement in a volume radius of 5 metres. Assuming a measurement rate of 5Hz and that a track is maintained over 10 frames, a process noise covariance of 0.5 and a measurement noise covariance of 0.5 are found to be appropriate. Using these parameters, the terms $P(y_j|Z_T)$ and $P(y_j|Z_A)$ are calculated using Eq.(8) and the LUT in Fig.3(a) respectively. $P(y_j|x_i)$ is calculated through the Monte Carlo simulation of the system model described. Using Eq.(6), $P(I_{ij}|Z) = 0.0110$ and through simulation, the uncertainty H (Eq.(7)) is 0.3868. Therefore, given this system configuration, the expected number of wrong guesses, $\overline{W}$, is 0.1537 in an hour.

The partial derivatives are calculated using these values together with the system parameters. To investigate the system sensitivity to the measurement noise (which is a parameter that can be modified using different sensors/sensor configuration), the derivative w.r.t. s_R is calculated. Let us examine two cases: when $\Delta s_R = -0.25$, i.e. when the measurement noise is halved, and when $\Delta s_R = 0.5$, i.e. when the noise is doubled. Using the terms above in



(a) Look-up table (b) Histograms (c) Distributions 3 s.d.s apart

Figure 3: (a) The expected reduction in uncertainty for a weighting parameter varying from 0 to 0.9, and the number of standard deviations between the means varying from 0 to 8, for Gaussian distributions. (b) Normalised histogram for correct and incorrect matches, as a function of the distance measure for the MPEG-7 color descriptor Scalable Color. Solid lines: from experiments. Dotted line/crosses: fitted curves. The prior between correct and incorrect matches is reflected in the respective vertical scales are on the left and right of the graph. They have been displayed together for clarity. These histograms correspond to the dot in (a). (c) Two equal-prior distributions with 3 standard deviations between the means. These histograms correspond to the square in (a).

Eq.(12) gives:

$$\begin{aligned}\Delta\overline{W} &= \left|\frac{\partial\overline{W}}{\partial H}\cdot\frac{\partial H}{\partial P(I_{ij}|Z)}\cdot\frac{\partial P(I_{ij}|Z)}{\partial P(y_j|Z_T)}\cdot\frac{\partial P(y_j|Z_T)}{\partial s_R}\right|\cdot\Delta s_R \\ &= \left|0.6537\cdot 5.5064\cdot 0.1578\cdot -0.4452\right|\cdot\Delta s_R \qquad (18)\end{aligned}$$

Therefore, halving the measurement noise variance reduces $\overline{W}$ by 0.0632, while doubling the measurement noise variance increases $\overline{W}$ by 0.1265.

Table 1: Metro system parameters.

Parameter	Symbol	Value
Dimension of measurement space	m	3
Volume radius for data association	r	5
Variance of process noise	s_Q	0.5
Variance of measurement noise	s_R	0.5
Numframes to track over	f	10

4.2 Required component performance level for overall end-user requirement

This backwards problem starts with the end-user requirement, and examines how good each or a particular constituent component has to be to meet this requirement. This analysis is useful for two reasons. First, it relates the system performance to component performance,

Annals of the BMVA Vol. 2010, No. 6, pp 1–12 (2010)

and provides a 'target' level of performance for algorithmic development. If a component is 'good enough', then effort should be focussed on other parts of the system. Second, in system design, if a system is not performing to a desired level, it can be difficult to attribute this under-performance to a component without an analysis tool.

Here we consider the case where only appearance-based measurements are used, illustrating that a subset of the model can be used. Starting from the value of the expected number of wrong guesses $\overline{W} = 10$, which from requirements elicitation from our project end-users is within the range of acceptable performance, the uncertainty H is 4.3923 and $P(I_{ij}|Z)$ is required to be 0.4640 for a throughput of 1000 people per hour. From the simulations, $P(y_j|x_i) = 0.2015$, therefore $P(y_j|Z_A)$ is 0.7743. This means that the expected reduction in uncertainty, $E[\Delta H_A]$, is also 0.7743. Referring to Fig. 3(a), at equal prior, 3 standard deviations between the means of the distributions are required to achieve the desired performance. Fig. 3(c) shows an example of these distributions. This can be used as a criterion in the selection and assessment of feature(s) and sets a minimum performance level for algorithm evaluation.

5 Conclusions

We have presented a framework for evaluating video-based tracking methods in the public transport domain, using the information gained through their use as an evaluation metric. This metric has two distinct advantages: it summarises and abstracts from the implementation details of the video-based tracking methods, and it links the technical performance of the system directly with the end-users' requirements, which is the frequency with which operator attention is required, i.e. the expected number of wrong guesses. A comprehensive mathematical framework of the system has been presented, and error propagation analysis has been carried out to allow the investigation of system sensitivity and end-user driven analysis; examples have been given illustrating their use. This framework is directly applicable to existing 'Tag and Track' methods in transport networks, by modelling the topology of the network of interest, as well as each technical component analytically. While the final performance of each method will undoubtedly depend on the system parameters, their results can be compared directly by using the proposed approach.

Acknowledgements

The work is funded under the CARETAKER project (European Union IST 4-027231). Our thanks to Gruppo Torinese Trasporti for giving permission for images of their stations to be used in this publication.

References

J. Annesley, V. Leung, A. Colombo, J. Orwell, and S.A. Velastin. Fusion of multiple features for identity estimation. pages 534–539. The Institution of Engineering and Technology Conference on Crime and Security, June 2006.

J. Black, T. Ellis, and P. Rosin. A novel method for video tracking performance evaluation.

pages 125–132. Joint IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS), October 2003.

L.M. Brown, A.W. Senior, Y-L. Tian, J. Connell, and A. Hampapur. Performance evaluation of surveillance systems under varying conditions. IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance (PETS), October 2005.

F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua. Multi-camera people tracking with a probabilistic occupancy map. *IEEE. Trans. Pattern Analysis and Machine Intelligence*, 30(2):267–282, 2008.

HOSDB. http://scienceandresearch.homeoffice.gov.uk/hosdb/cctv-imaging-technology/video-based-detection-systems/i-lids.

O. Javed, K. Shafique, and M. Shah. Appearance modeling for tracking in multiple non-overlapping cameras. volume 2, pages 26–33. CVPR, 2005.

S.M. Khan and M. Shah. A multiview approach to tracking people in crowded scenes using a planar homography constraint. pages 133–146. ECCV, May 2006.

K. Kim and L.S. Davis. Multi-camera tracking and segmentation of occluded people on ground plane using search-guided particle filtering. pages 98–109. ECCV, May 2006.

V. Leung, J. Orwell, and S.A. Velastin. Performance evaluation of re-acquisition methods for public transport surveillance. pages 705–712. IEEE Tenth Int. Conference on Control, Automation, Robotics and Vision, December 2008.

D. Makris, T.J. Ellis, and J.K. Black. Bridging the gaps between cameras. pages 205–210. CVPR, 2004.

S. Mori, K.-C. Chang, and C.-Y. Chong. Performance Analysis of Optimal Data Association with Applications to Multiple Target Tracking. In *Multitarget-Multisensor Tracking: Applications and Advances*, volume II, chapter 7, pages 183–234. Artech House, 1992.

C.J. Needham and R.D. Boyle. Performance evaluation metrics and statistics for positional tracker evaluation. pages 278–289. Int. Conference on Computer Vision Systems (ICVS), April 2003.

N. Ning and T. Tan. A framework for tracking moving target in a heterogeneous camera suite. pages 1–5. IEEE Ninth Int. Conference on Control, Automation, Robotics and Vision, December 2006.

PETS. http://www.cvg.cs.rdg.ac.uk/slides/pets.html.

A. Rahimi, B. Dunagan, and T. Darrell. Simultaneous calibration and tracking with a network of non-overlapping sensors. pages 187–194. CVPR, 2004.

C.J. Van Rijsbergen. *Information Retrieval*. Butterworth-Heinemann Ltd., 2nd edition, 1979.