Performance Evaluation of Tracking for Public Transport Surveillance

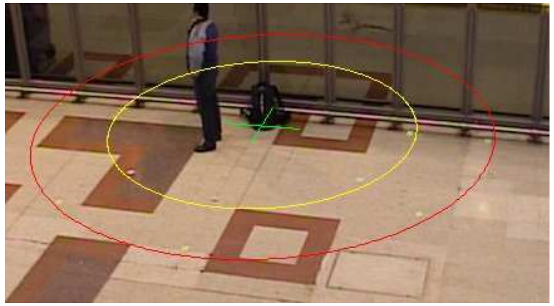
J.Orwell, V.Leung, A.Colombo, S.A.Velastin
Digital Imaging Research Centre
Kingston University, London

Objective

- For operators ('end-users') of video surveillance systems:
  - It is useful to keep ***track*** of individuals (or groups of people) in one station, or over the transport network
    - Someone behaving suspiciously
    - Lone woman at night
  - It is sometimes necessary to look at recordings of video surveillance, to find the time and place that an individual entered (or exited) the transport network
- Automatic tracking and recognition methods may be useful
- How can these methods be evaluated, to provide meaningful and useful results to the operators?

Evaluation of Automatic Event Detection

- Detection of events to generate an alert/alarm
  - Left Luggage
  - People exiting through the wrong door
  - Fighting, running



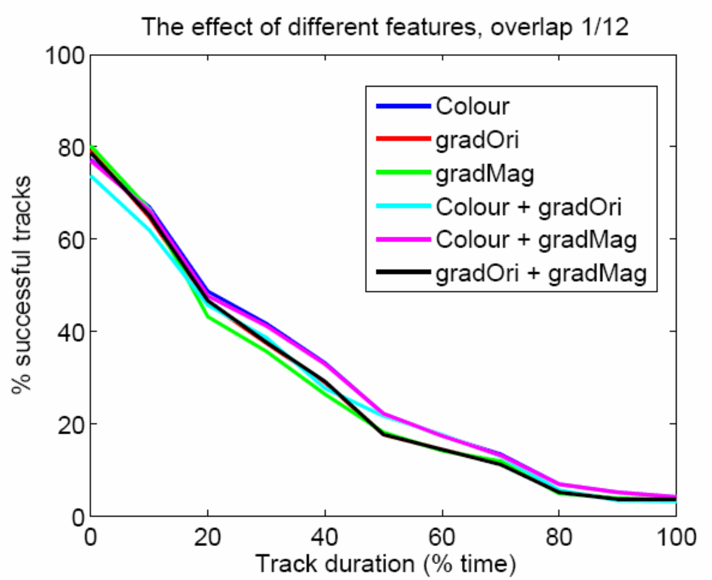
- ROC analysis, precision/recall, F-measure
  - 'trade-off' between false positive and false negative

- This provides to the end-user, an estimate of:
  - How many times a day this event type will be missed
  - Frequency of 'false alarms' that must be dismissed

Evaluation of tracking methods

Possible metrics for tracking:

- **Accuracy** of track
- **Continuity** of track
- Proportion of total time tracking is **successful**
- Content-based retrieval **metrics** e.g. ANMRR



BUT these do not relate directly to the operator's priorities:

- How much operator time the proposed method will save?
- How the proposed method will interact with the normal controls to move and switch between cameras?
- How can the operator interact with the system to select from multiple hypotheses?

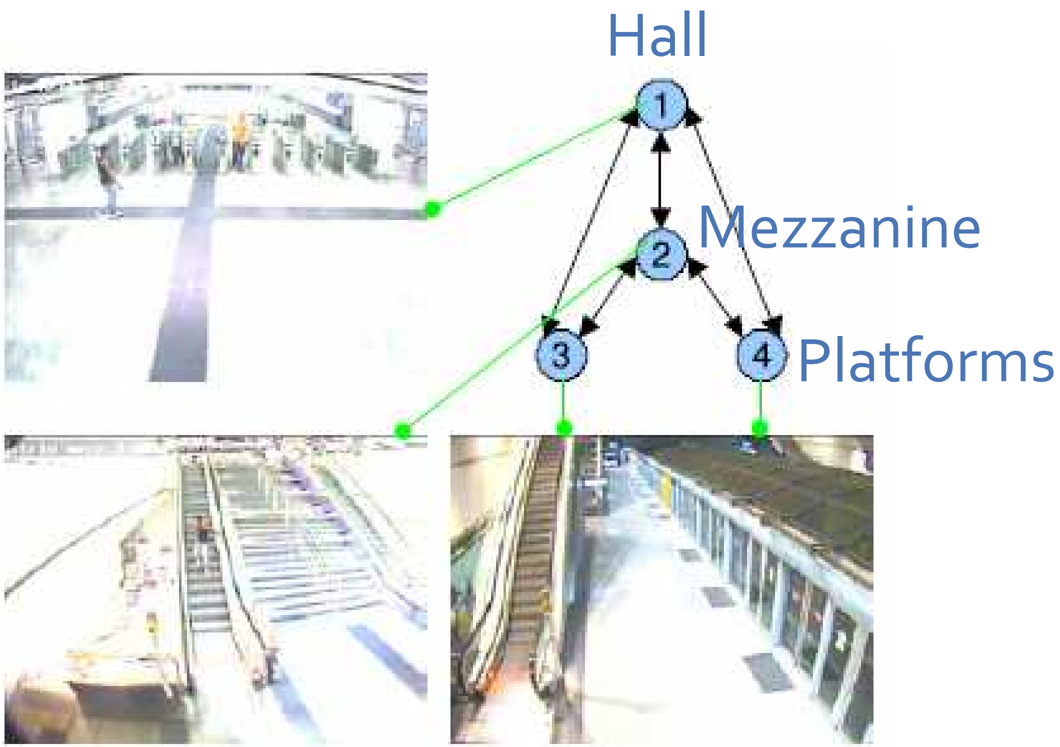
Operator Evaluation Requirements

From the evaluation, the end-user demands:

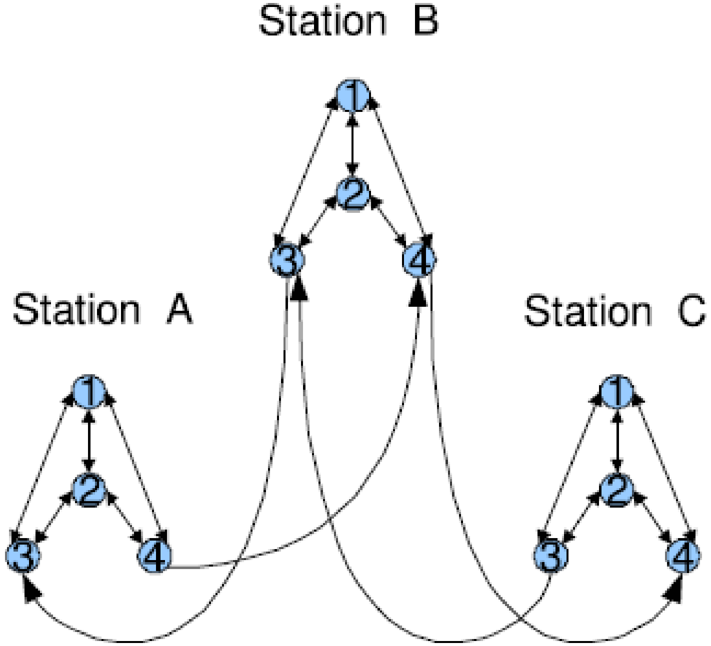
- A close correlation with the benefits to the end-user
- An indication of the difficulty of the tracking scenario
- The accommodation of tracking systems that provide multiple hypotheses
- The performance to be evaluated at key way-points (e.g. entry, exits, turnstiles)

The proposal: to measure the performance of a tracking/recognition system, by estimating the reduction in uncertainty (equivalently, gain in information) about a passenger’s whereabouts that it provides.

Representation of a metro network



Single station

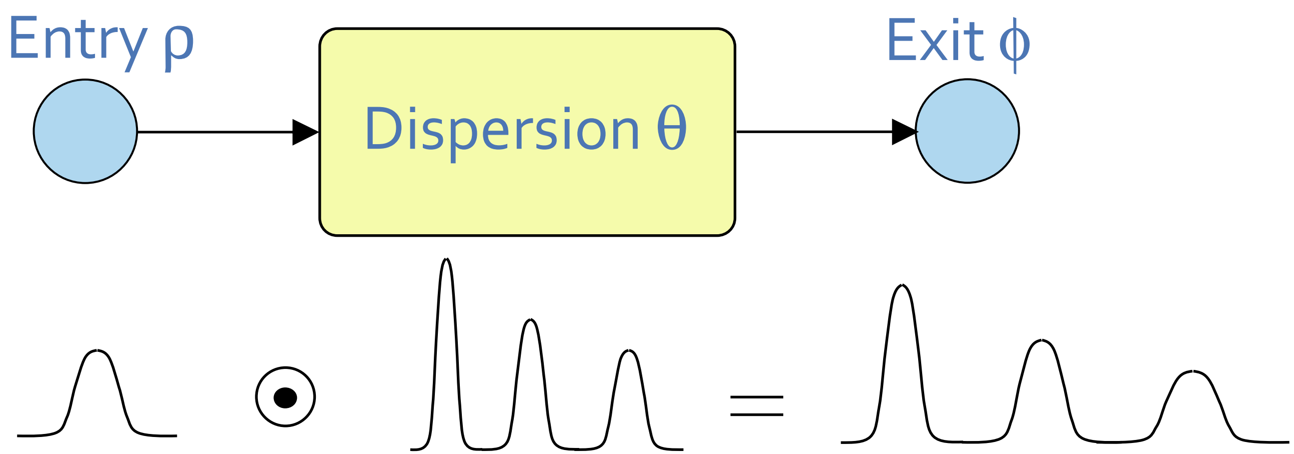


The uncertainty of passenger whereabouts:

Different models:

1. No model: they could be anywhere in the system
   - Location is a continuous random variable across space and time
2. Assume the positions of the *n* passengers have been **detected**, and that the target is one of these passengers
   - Location is a discrete random variable in space and time: the passenger is one of the *n*
3. Some predictions can be made about where the passenger goes, using **prior statistics**
4. Observations: more information provided by tracking and appearance recognition components

Prior statistics only (1)



Given that a person entered the system at time t_1 station s_i, what is the probability of correctly identifying them, at their exit point?

Prior statistics only (2)

- Probability of correct identification using prior statistics

The 'winning' distribution

Localisation at exit

$$p(L_2|L_1) = \frac{\theta}{\theta + (n-1)\phi}$$

Localisation at entrance

Density of other people

Background clutter

- Entropy

$$H(L_2|L_1) = \underbrace{-p(L_2{=}a|L_1{=}a)\log p(L_2{=}a|L_1{=}a)}_{\text{Right answer}} \underbrace{- p(L_2{\neq}a|L_1{=}a)\log p(L_2{\neq}a|L_1{=}a)}_{\text{Wrong answer}}$$

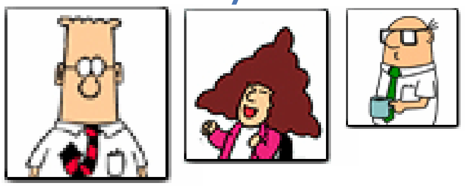
Prior statistics + observations

- Provided by recognition algorithms
- Using Bayes' rule:

$$p(L_2|L_1, Z) = \frac{p(L_2|L_1)\, p(L_2|Z)}{p(L_2|L_1)\, p(L_2|Z) + p(\neg L_2|L_1)\, p(\neg L_2|Z)}$$

Can be used to present multiple hypotheses to operators

Most likely matches:



- Entropy:

$$H(L_2|L_1,Z) = -p(L_2{=}a|L_1{=}a,Z) \log p(L_2{=}a|L_1{=}a,Z) - p(L_2{\neq}a|L_1{=}a,Z) \log p(L_2{\neq}a|L_1{=}a,Z)$$

Experiments

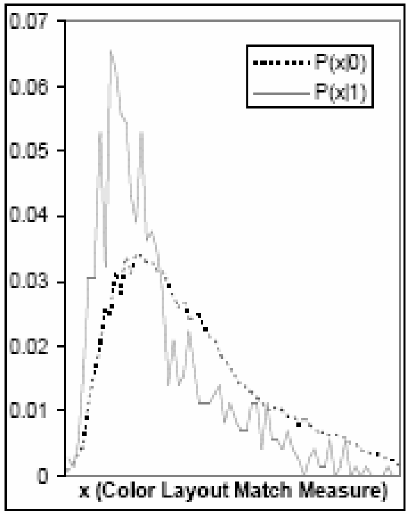
- Examined two entry distributions:
  - Uniform: over three stations
  - Mixture of Gaussians: 2 Gaussians at each of four stations
- Dispersion p.d.f. θ , to simulate choice of destinations and the expected duration of journey :

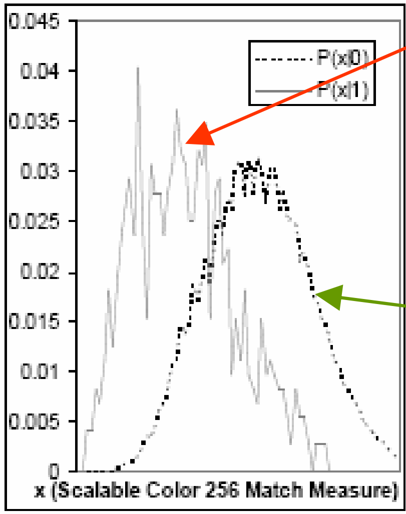
$$\theta(\alpha_{ij},\mu_{ij},\sigma_{ij},t_2-t_1)=\begin{cases} C\dfrac{\alpha_{ij}}{\sigma_{ij}\sqrt{2\pi}}\exp\left\{-\dfrac{((t_2-t_1)-\mu_{ij})^2}{2\sigma_{ij}^2}\right\} & \text{if } (t_2-t_1)>0 \\ 0 & \text{Otherwise} \end{cases}$$

- Ran simulations of up to 100 passengers in the network
- Observations based on earlier work using MPEG-4 Color Descriptors

Sample observations

- Our previous work[1]:
  - Examined MPEG-4 descriptors for re-identifying people
  - Used distance between the attributes of the query subject and the other subjects to form p.d.f.s:





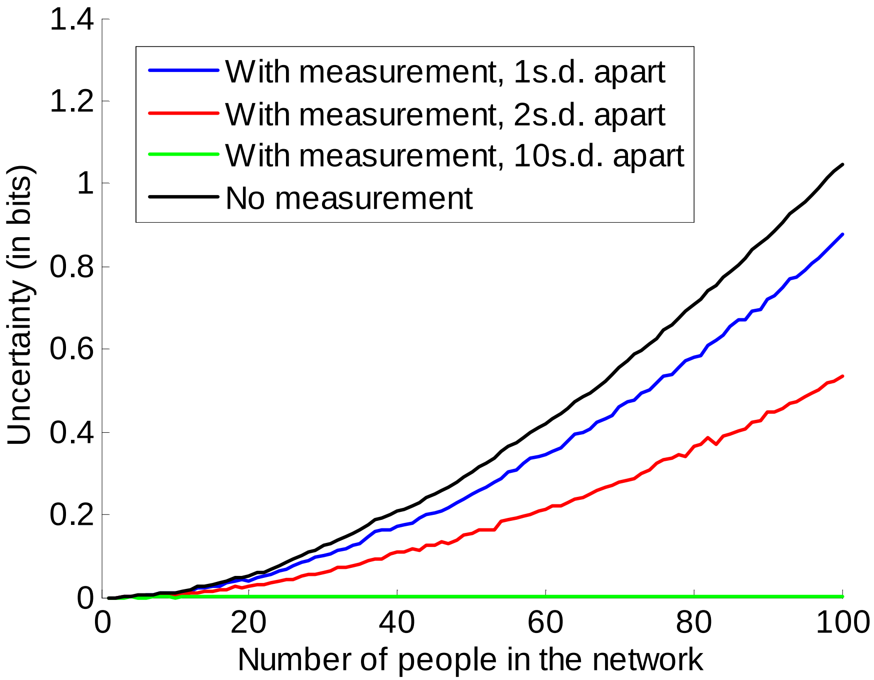
Prob. of measurement given it's the right person

Prob. of measurement given it's the wrong person

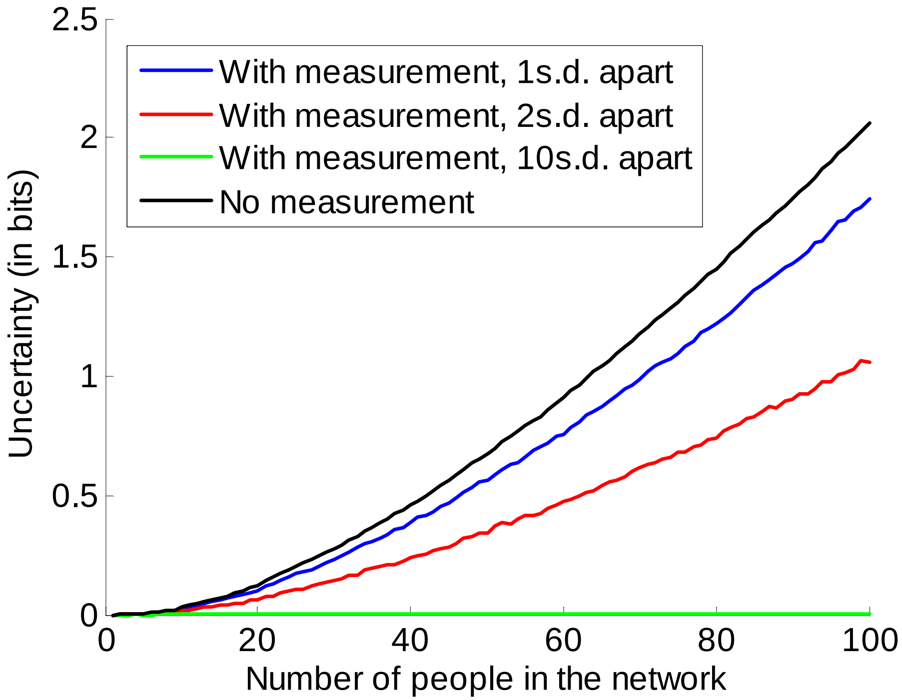
1. Annesley J., Leung V., Colombo A., Orwell J. and Velastin S.A., "Fusion of Multiple Features for Identity Estimation", ICDP '06

Results

Uniform entry distribution



MoG entry distribution



Without measurement: highest uncertainty

The more separated the p.d.f.s the less the uncertainty

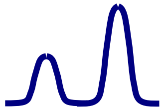
Completely separated p.d.f.s: no uncertainty

What this means for the 'end user'

With 100 people using the network and uncertainty H, the expected number of incorrect identifications is $(2^H-1)/2$

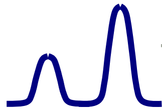
No prior model, no observations

H=6.6
On average, 49 incorrect identifications per passenger



Prior model, no observations

H=2
On average, 1.5 incorrect identifications per passenger



Prior model, observations

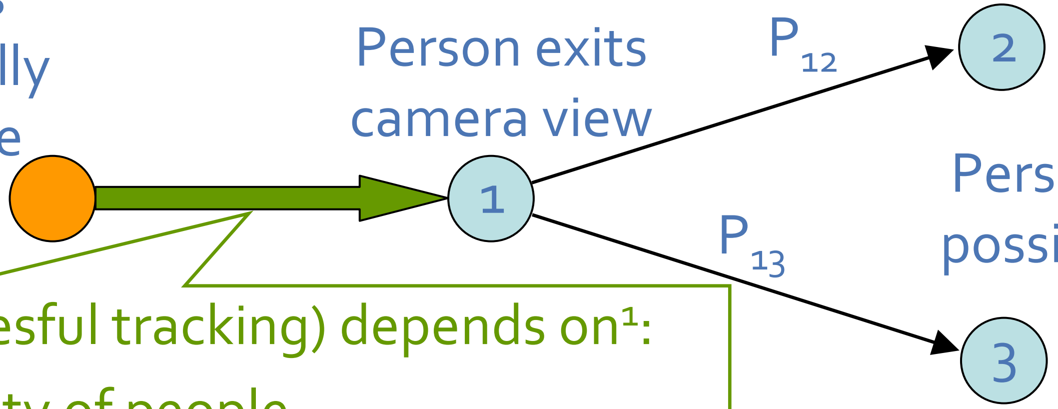
1 s.d. H=1.7
→ 1.1 incorrect ids per passenger

2 s.d. H=1
→ 0.5 incorrect ids per passenger

With tracking information

Person enters system, manually initialised to be tracked

Person exits camera view



Person re-emerges in possible camera views

P(succesful tracking) depends on[1]:

- Density of people
- Tracker accuracy
- Dimension of measurement space

- Hand-over: Initial confidence in particular individual will dissipate with every handover having non-zero uncertainty
- Combine with appearance-based measurements

1. Mori S., Chang K.C. and Chong C.Y., "Performance analysis of optimal data association with applications to multiple target tracking", In *Multitarget-Multisensor Tracking: Applications and Advances*, Vol. II, 1992.

Conclusions

- Proposed a metric of a visual surveillance system that can indicate the tracking/recognition performance to an operator
- Information-theoretic approach:
  - Uncertainty of system with prior information only
  - Reduction in uncertainty with side information
    - In the form of appearance-based measurements

Future Work

- Include information provided by tracking in framework
- Perform actual tracking and recognition experiments to compare with theoretical calculations

Acknowledgements

- Funded under CARETAKER project (EU IST 4-027231)
- Thanks to Gruppo Trasporti Torino (GTT) and Agenzia per I Trasporti Autoferrotramviari del Commune di Roma SPA (ATAC) for the permission to use the images from their networks