Tracking With Intent

Rolf Baxter¹, Michael Leach¹,², Neil Robertson¹

¹ Vision Lab, Institute of Sensors, Signals and Systems, Heriot-watt University, Edinburgh
² Roke Manor Research (Chemring Technology Solutions), Romsey, Hampshire
Motivation: The end goal

- Identifying anomalous behaviour in the proverbial haystack to enhance situation awareness

[WPAFB 2009 Dataset]
**Approach:** Build better models of normality by using more of the signal

- Specifically, we propose the use of *intentional priors*
  - Priors that are indicative of future intent
  - Could be derived from different signals (e.g. car indicator, AIS, pattern of life)
  - Could be context sensitive

- This talk focuses on person tracking in video with head-pose priors
  - Relevant to automated visual surveillance (e.g. base protection)
  - People perform a broad range of behaviours so represent challenging targets
  - Concept is extensible to other real-world targets
Motivating Example

**Intuition:** Head-pose is an informative intentional prior

- Head pose can provide both spatial and social context
- We can use head-pose to build better person trackers
Related Work

- Recent work has shown that performing head-pose estimation within the outdoor built environment is reasonable
  - Odobez at Idiap [5]
  - Benfold at Oxford [6]
  - Our latest work at Heriot-Watt (IEEE Sig.Proc.Letters, to appear)
- A number of trackers consider social context when making target predictions (e.g. Pelligrini et al. [7])

- Sankaranarayanan et al. fused person tracking with a PTZ facial tracking system, but did not use head-pose to predict target location [8].

- **No prior work has used head-pose to aid target tracking**
  - ➔ The video signal is being under utilised
Is the signal present?

- Is head-pose is well correlated with direction of travel?

- Analysis: 3 datasets - Caviar, PETS and Oxford [10,11,6]

- Using automatic detections [12] and ground truth head-pose annotations
The Kalman Filter

Time update ("predict")

1) Project the state ahead
\[ \hat{x}_t^- = F_{t-1} \hat{x}_{t-1} + Bu_{t-1} \]

2) Project the error covariance ahead
\[ P_t^- = F_{t-1} \hat{x}_{t-1} F_{t-1}^T + Q_t \]

Measurement update ("correct")

1) Compute the innovation
\[ \epsilon_t = z_t - H \hat{x}_t^- \]

2) Compute the Kalman Gain \( K_t \)
\[ K_t = P_t^- H^T (HP_t^- H^T + R)^{-1} \]

3) Update estimate with \( Z_t \)
\[ \hat{x}_t = \hat{x}_t^- + K_t \epsilon_t \]

4) Update the error covariance
\[ P_t = (I - K_t H_t) P_t^- \]


State vector: \[ x_t = [x, y, \dot{x}, \dot{y}]^T \]

Observation matrix: \[ H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \]

Mixing component: \[ \alpha_t = (1 + \exp(-\rho(s_t - \tau)))^{-1} \]
\[ \gamma_t = 1 - \alpha_t \]

Motion model: \[ F_t = \begin{bmatrix} 1 & 0 & \gamma_t & 0 \\ 0 & 1 & 0 & \gamma_t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \]

Intentional prior: \[ B_t = [\alpha_t dx, \alpha_t dy, \alpha_t dx, \alpha_t dy]^T \]

\[ d_t \::\: \text{Distance}(\hat{x}_{t-1}, \hat{x}_t) \]
\[ \theta^h_t \::\: \text{head-pose direction at } t \]

\[
\begin{align*}
& dx = d_t \cos \theta_t \\
& dy = d_t \sin \theta_t
\end{align*}
\]
Mixing component: \[ \alpha_t = (1 + \exp(-\rho(s_t - \tau)))^{-1} \]

**Strength of prior:**

\[ S_t :: | \sum_{k=\max(0,t-10)}^{t} Bin(\theta^h_k) - Bin(\theta^v_k) | \]

\[ \theta^v_k :: \text{Travel direction} \quad \theta^h_k :: \text{Head−pose direction} \quad \rho :: \text{Sensitivity} \quad \tau :: \text{Base weight} \]

**Head-pose binning:**

![Head-pose binning images]
Evaluation

- 5 core trajectories:
  - Gaussian noise added to positions and head-poses

- Comparative baseline:
  - Standard Kalman Filter (i.e. without head-pose)
Sigmoid optimisation
Sigmoid optimisation

Best performers: High base-weight for head-pose

Worst performers: Low base-weight for head-pose
Performance with Deep BN. head-pose

- Median improvements:
  - Ben: 7.21%
  - Cav: 19.5%

- Optimising the head-pose classifier for the scene could improve this.
Conclusions & Future Work

- Shown how to integrate an intentional prior into the Kalman Filter (KF)

- Validation has shown that using head-pose intentional priors we can make better target predictions

- Key next steps:
  - How do we integrate anomaly detection?
  - How do we learn and use different kinds of intentional prior & context?
Questions
References


References


Deep Belief Net. Classifier

- Unlike competing approaches, we do not use motion or body information to ‘classify’ head-pose.

- Poorest performance collates with fewest training examples.