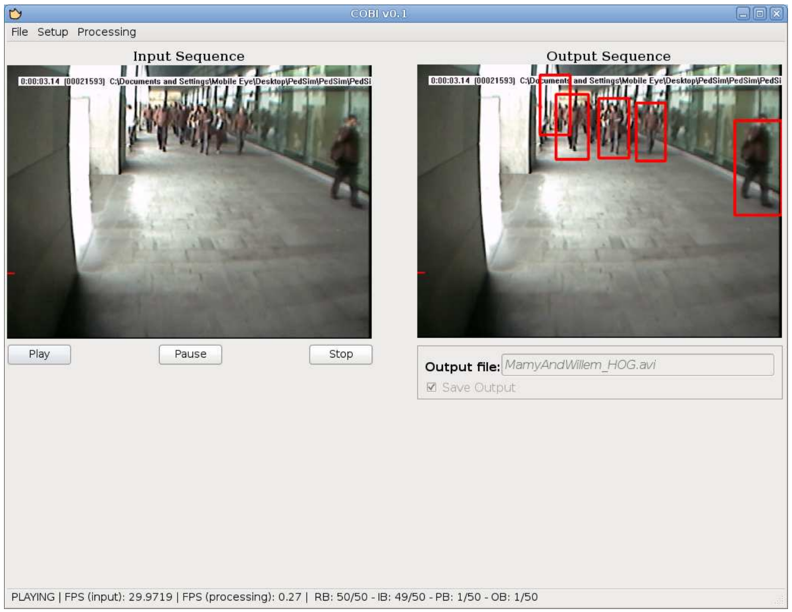
Behavioural Pedestrian Tracking

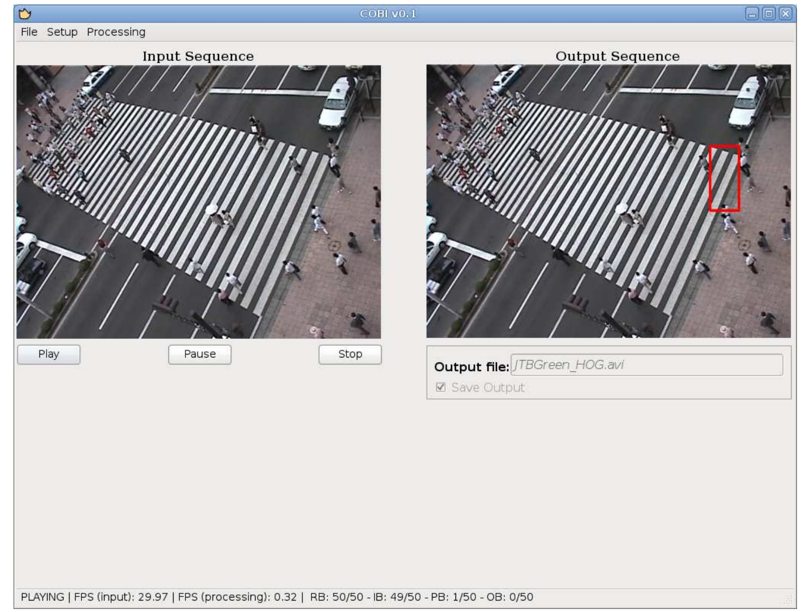
Javier Cruz, Michel Bierlaire and Jean-Philippe Thiran

{javier.cruz, michel.bierlaire, jean-philippe.thiran}@epfl.ch

École Polytechnique Fédérale de Lausanne

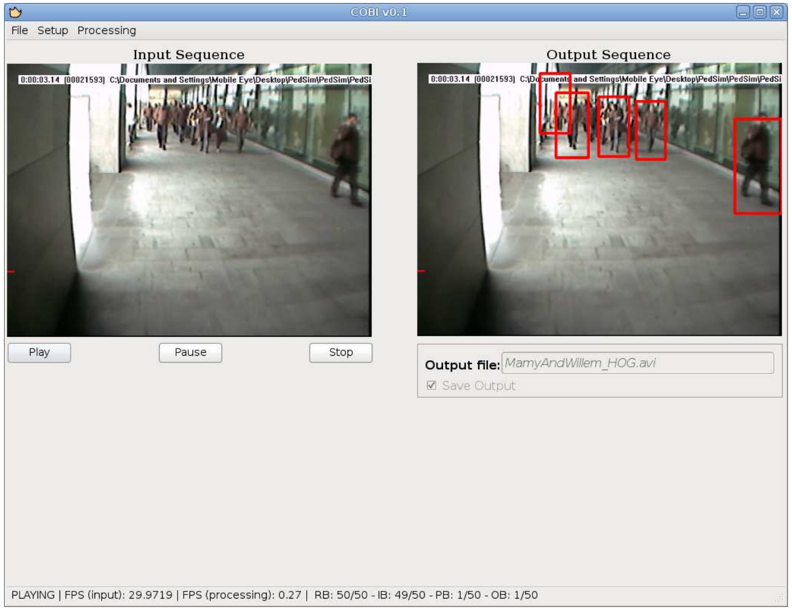
Motivation

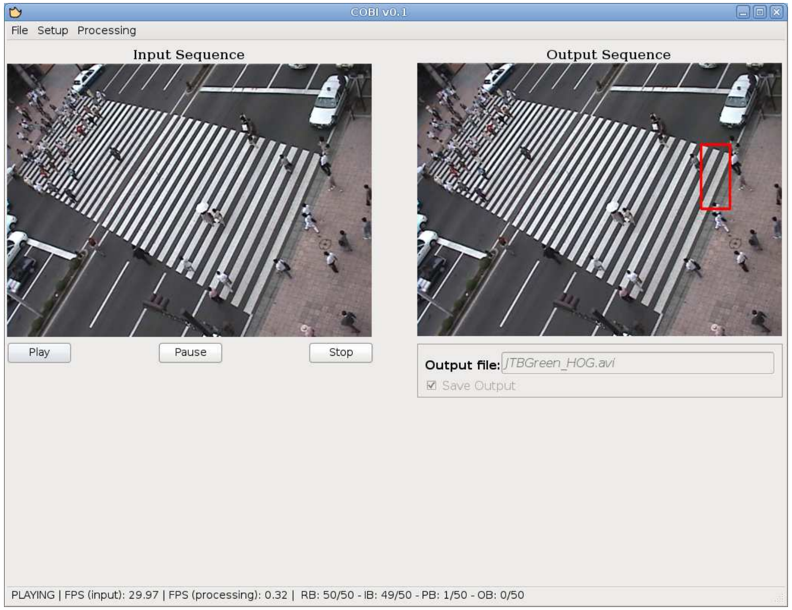

COBI v0.1
File Setup Processing
Input Sequence
Output Sequence
Play
Pause
Stop
Output file: MamyAndWillem_HOG.avi
Save Output
PLAYING | FPS (input): 29.9719 | FPS (processing): 0.27 | RB: 50/50 - IB: 49/50 - PB: 1/50 - OB: 1/50


COBI v0.1
File Setup Processing
Input Sequence
Output Sequence
Play
Pause
Stop
Output file: JTBGreen_HOG.avi
Save Output
PLAYING | FPS (input): 29.97 | FPS (processing): 0.32 | RB: 50/50 - IB: 49/50 - PB: 1/50 - OB: 0/50

Motivation

Motivation





TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Outline

- Introduction
- Visual tracking
- Pedestrian Visual Tracking and Detection
- Questions and future work

Introduction

Common pedestrian tracking systems: Detection and Inter-Frame Tracking

Detection:

- Haar/HOG feature + Boosting
- Background substraction
- Model based detection (skeleton models, silhouettes, etc.)

Tracking:

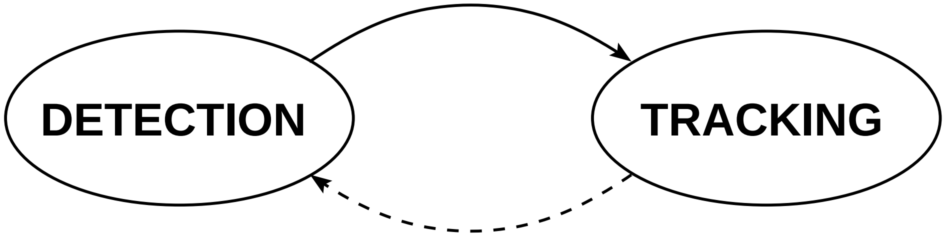
- Kalman filter
- Condensation algorithm
- Mean-shift
- Covariance tracking

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

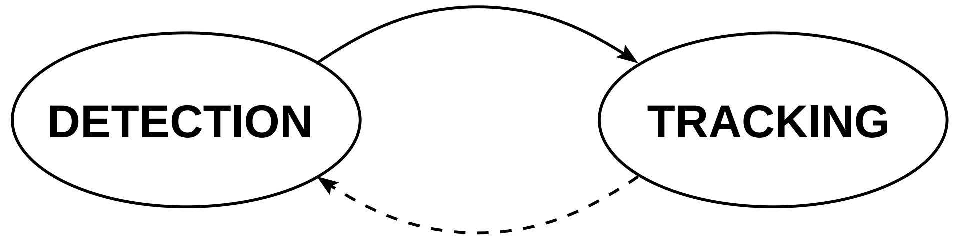
Introduction

Usual approach:

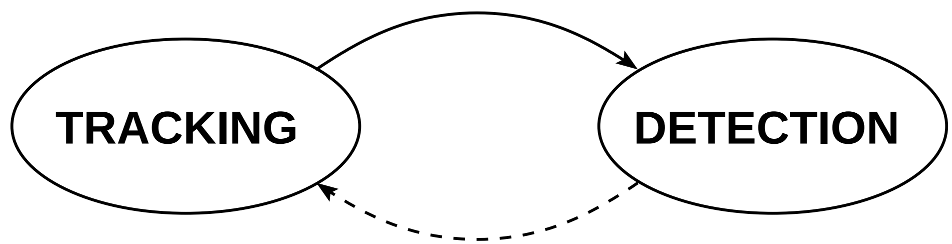


Introduction

Usual approach:



What about doing a tracking that may end in a detection?



Visual Tracking

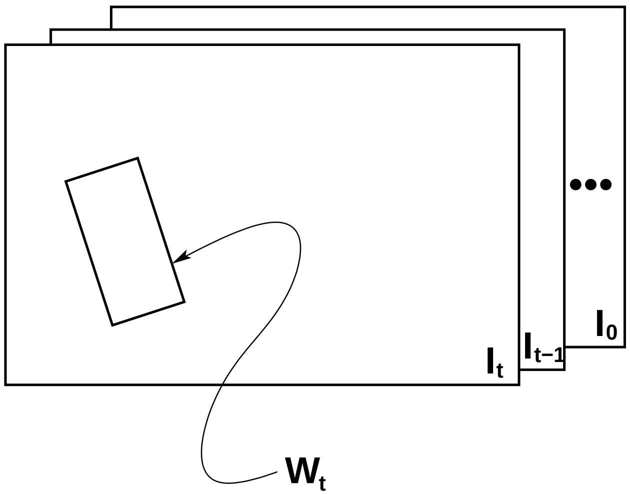
Ross, D., Lim, J. and Lin, R.-S. (2008), *Incremental Learning for Robust Visual Tracking*. International Journal of Computer Vision 77: 125-141.

- On-Line low-dimensional subspace representation (incremental PCA)
- Gaussian variables
- Particle filtering

Visual Tracking

Definitions:

- I_t: frame t
- $X_t = (x_t, y_t)$: position
- φ_t: rotation
- s_t: scale
- $r_t = \frac{w_t}{h_t}$: aspect ratio
- $W_t = f(I_t, X_t, \varphi_t, s_t, r_t)$: a patch in I_t
- $\Psi_t = \{X_t, \theta_t\} = \{X_t, \varphi_t, s_t, r_t\} \in \Theta_t$



Visual Tracking

Everything is supposed to be Gaussian:

- $X_t \sim \mathcal{N}(X_{t-1}, \sigma_X)$
- $\varphi_t \sim \mathcal{N}(\varphi_{t-1}, \sigma_\varphi)$
- $s_t \sim \mathcal{N}(s_{t-1}, \sigma_s)$
- $r_t \sim \mathcal{N}(r_{t-1}, \sigma_r)$

$$\Psi_t = \{X_t, \varphi_t, s_t, r_t\} \in \Theta_t$$

Visual Tracking

$$\Psi_t^* = \arg\max_{\Psi_t \in \Theta_t} p(I_t|\Psi_t)p(\Psi_t|\Psi_{t-1})$$

Observation:

$$p(I_t|\Psi_t) = \mathcal{N}(W_t; \mu, UU^\top + \varepsilon I)\mathcal{N}(W_t; \mu, U\Sigma_o^{-2}U^\top)$$

Dynamics:

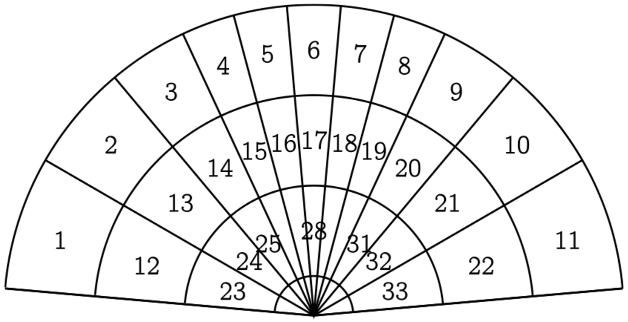
$$p(\Psi_t|\Psi_{t-1}) = \mathcal{N}(\Psi_t; \Psi_{t-1}, \Sigma_\Psi)$$

Visual Tracking

Robin, T., Antonini, G., Bierlaire, M., and Cruz, J. (2009), *Specification, estimation and validation of a pedestrian walking behavior model*. Transportation Research Part B: Methodological 43(1):36-56.

Takes into account:

- Direction
- Destination
- Speed
- “Leader-follower”
- “Collision avoidance”



Choice Set

Visual Tracking

This forces some assumptions:

- Camera is calibrated
- Camera is fixed
- Pedestrians walking in normal conditions
- Destination known!!

Visual Tracking

Pedestrian model + Gaussian:

- $X_t \sim$ pedestrian walking behaviour model (PWBM)
- $\varphi_t \sim \mathcal{N}(\varphi_{t-1}, \sigma_\varphi)$
- $s_t \sim \mathcal{N}(s_{t-1}, \sigma_s)$
- $r_t \sim \mathcal{N}(r_{t-1}, \sigma_r)$

$$\Psi_t = \{X_t, \theta_t\} = \{X_t, \varphi_t, s_t, r_t\} \in \Theta_t$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Visual Tracking

$$\Psi_t^* = \arg\max_{\Psi_t \in \Theta_t} p(I_t|\Psi_t)p(\Psi_t|\Psi_{t-1})$$

Observation:

$$p(I_t|\Psi_t) = \mathcal{N}(W_t; \mu, UU^\top + \varepsilon I)\mathcal{N}(W_t; \mu, U\Sigma_o^{-2}U^\top)$$

Dynamics:

$$p(\Psi_t|\Psi_{t-1}) = \text{PWBM}(X_t; X_{t-1})\mathcal{N}(\theta_t; \theta_{t-1}, \Sigma_\theta)$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Visual Tracking

Considered models:

- M_G a Gaussian model
- M_P the Pedestrian model

Bayesian Model Averaging over $\mathcal{M} = \{M_G, M_P\}$

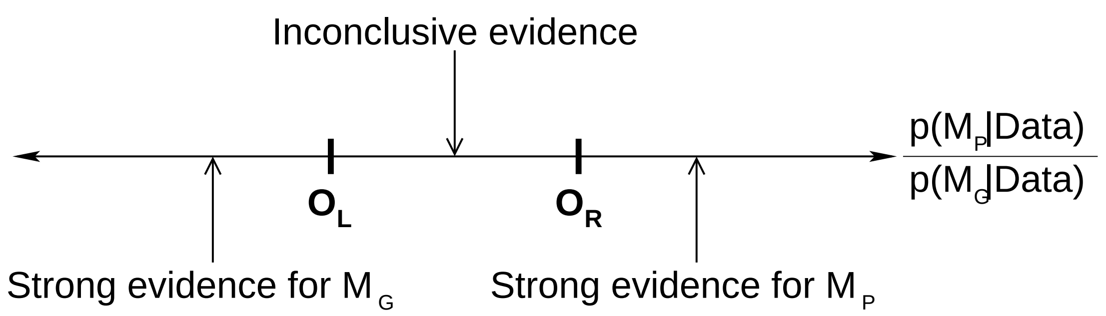
Visual Tracking

Modified probabilities of the Visual Tracking Algorithm:

$$
\begin{aligned}
p(\Psi_t|\Psi_{t-1}, I_t) &= \sum_{M_j \in \mathcal{M}} p(\Psi_t|M_j, \Psi_{t-1}, I_t) p(M_j|\Psi_{t-1}, I_t) \\
p(M_i|\Psi_{t-1}, I_t) &= \frac{p(\Psi_{t-1}, I_t|M_i) p(M_i)}{\sum_{M_j \in \mathcal{M}} p(\Psi_{t-1}, I_t|M_j) p(M_j)} \\
p(\Psi_{t-1}, I_t|M_i) &= \int_{\Theta_t} p(\Psi_{t-1}, I_t|\Psi_t, M_i) p(\Psi_t|M_i) d\Psi_t
\end{aligned}
$$

Visual Tracking

Interpretation of the ratio of posterior model probabilities (The Occam's Window):



Common values are $O_R = 20$ and $O_L = O_R^{-1}$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Pedestrian Visual Tracking and Detection!

Proposed algorithm:

Initialization: set priors, $\mathcal{M} = \{M_P, M_G\}$
Foreground detection in I_0
for $t = 1$ to N **do**
    for each obtained or tracked blob **do**
        Apply “Visual Tracking Algorithm” with modified probabilities
        Classify blob according to model posteriors
        if M_P is chosen **then**
            $\mathcal{M} = \{M_P\}$
        end if
    end for
    Foreground detection in I_t
end for

Questions and Future work

- What kind of priors?
- “Occam’s window” per frame or with memory?
- Particle filters, Gibbs sampling, Metropolis-Hastings algorithms?
- Implement and test
- Generalization to N models
  - Action detection
  - Other “objects”

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE