Multi-Human Tracking, and Path Learning by Following a Person

Aijun Bai
http://home.ustc.edu.cn/~baj/

University of Science and Technology of China

October 7, 2014

Outline

Outline

Figure 1 : Yancheng, Hefei and Pittsburgh.

Figure 2 : Yancheng and Hefei.

Figure 3 : University of Science and Technology of China.

Figure 4 : My lab in USTC.

Outline

1. Maps and Pictures
2. Multi-Human Tracking via Particle Filtering over Sets
3. Robot Learning a Path by Following a Person
4. Demo

Multi-human tracking for human-robot interaction

- For successful HRI, a robot needs to know at least
  - The number of humans around
  - Each human's state information
- The problem is tackled in a tracking-by-detection framework
  - A human detector runs on each frame separately
  - A human tracker estimates the true situation given sequential detections

Multi-human tracking challenges

- Missing knowledge about the number of humans
- Missing detection-to-target association
  - Detections do not carry identification information
- Inevitable false and missing detections
  - Computer vision algorithm limitations
  - People occlusions
- Real-time constraint

Human detection results on CoBot

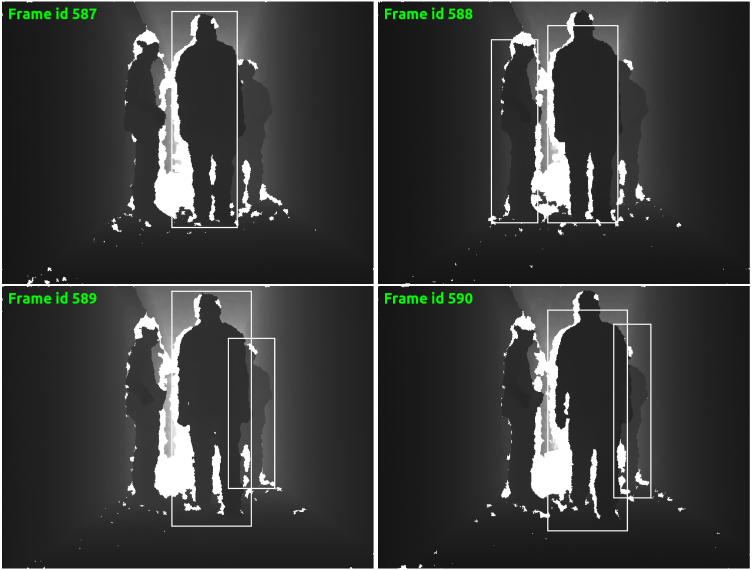

Figure 5 : Successive raw detections on CoBot.

Human detection results on CoBot (cont'd)

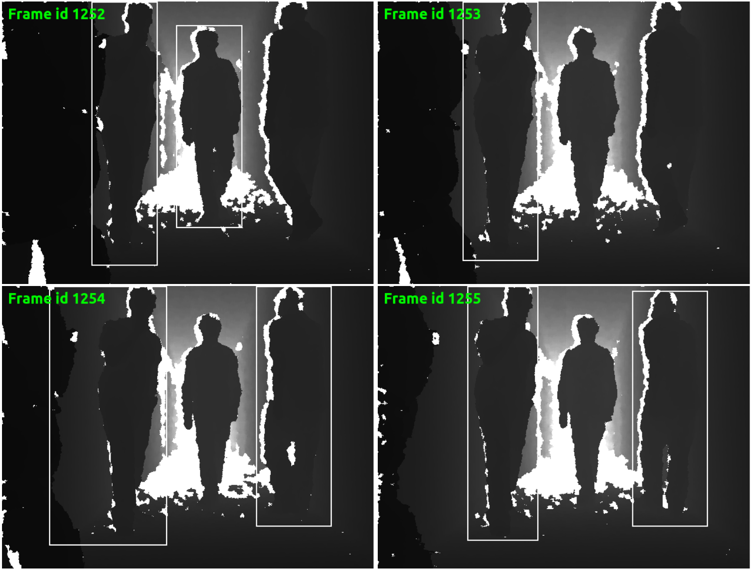

Figure 6 : Another example of successive raw detections on CoBot.

Motivation

- Most approaches in multi-object tracking (MOT) domain
  - Assume a hypothesis on data-association
    - Each detection is associated with a target
  - Perform Bayesian filtering separately
    - Kalman filter
    - Particle filter
- Drawback: it is difficult to recover from wrong assumptions
- Idea: track multi-objects in joint space without performing detection-to-target association before particle filtering
  - $\{s_0, s_1\}$
  - $[s_0, s_1]$, $[s_1, s_0]$

Hidden Markov model formalization

- HMM formalization
  - State: a set of humans $S = \{s_0, s_1, \ldots, s_n\}$
  - Observation: a set of human detections $O = \{o_1, o_2, \ldots, o_m\}$
  - Transition function: $\Pr(S' \mid S)$
  - Observation function: $\Pr(O \mid S)$
- Multi-human tracking
  - Observation history $\rightarrow$ joint state
  - I.e., $\Pr(S_t \mid O_0, O_1, \ldots, O_t)$

A set as a random variable - example

- Two coins are tossed
- Possible results as vectors
  - $[\mathsf{H}, \mathsf{H}]$, $[\mathsf{H}, \mathsf{T}]$, $[\mathsf{T}, \mathsf{H}]$, $[\mathsf{T}, \mathsf{T}]$
    * $\Pr([\mathsf{H}, \mathsf{H}]) = \Pr([\mathsf{H}, \mathsf{T}]) = \Pr([\mathsf{T}, \mathsf{H}]) = \Pr([\mathsf{T}, \mathsf{T}]) = \frac{1}{4}$
- Possible results as sets
  - $\{\mathsf{H}\}, \{\mathsf{H}, \mathsf{T}\}, \{\mathsf{T}\}$
    * $\Pr(\{\mathsf{H}\}) = \Pr(\{\mathsf{T}\}) = \frac{1}{4}$
    * $\Pr(\{\mathsf{H}, \mathsf{T}\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

A set as a random variable - theorem

- $S = \{X_0, X_1, \ldots, X_n\}$ is a set of n random variables
- The probability of observing $S = \{x_0, x_1, \ldots, x_n\}$ is

$$\Pr(S) = \sum_{\sigma \in A_n} \Pr(X_1 = x_{\sigma(1)}, \ldots, X_n = x_{\sigma(n)}) \qquad (1)$$

  - A_n is the set of all permutations of $\{1, 2, \ldots, n\}$
  - Value-to-variable association is unknown when observing $S = \{x_0, x_1, \ldots, x_n\}$
  - The joint probability counts all possibilities

A set as a random variable - corollaries

- $\mathcal{O} = \{o_0, o_1, \ldots, o_n\}$ is a set of n distinct objects
  - Sampling without replacement for k times, the result is $S = \{o_{(0)}, o_{(1)}, \ldots, o_{(k)}\}$
  - The probability of observing S is

$$\Pr(S) = k!\frac{1}{n(n-1)\cdots(n-k+1)} = \frac{1}{\binom{n}{k}} \tag{2}$$

- Random variable X follows a distribution function $f_X(x)$
  - $S = \{X_0, X_1, \ldots, X_n\}$ is a set of n X-variables
  - The probability of observing $S = \{x_0, x_1, \ldots, x_n\}$ is

$$\Pr(S) = n! \prod_{1 \leqslant i \leqslant n} f_X(x_i) \tag{3}$$

Motion model

- A state is a set of humans $S = \{s_0, s_1, \ldots, s_n\}$
  - Independence between humans
- A single human state is a vector $s = (x, y, \dot{x}, \dot{y})$
- A random acceleration motion model is defined as
  - Position: $(x, y) \leftarrow (x, y) + (\dot{x}, \dot{y})\tau + \frac{1}{2}(\ddot{x}, \ddot{y})\tau^2$
  - Velocity: $(\dot{x}, \dot{y}) \leftarrow (\dot{x}, \dot{y}) + (\ddot{x}, \ddot{y})\tau$
  - Time interval: τ
  - Random acceleration: $(\ddot{x}, \ddot{y}) = (p \cos\theta, p \sin\theta)$
    * Dash power: $p \sim \mathcal{N}(0, \sigma_p^2)$
    * Dash direction: $\theta \sim \mathcal{U}(0, 2\pi)$
- Different intentions/activities as different motion models

Observation model - detection

- An observation is a set of detections $O = \{o_1, o_2, \ldots, o_m\}$
- A detection is a 3-tuple $o = (x, y, c)$
  - Position in world frame: (x, y)
  - Confidence value: $c \in [0, 1]$
    - Reflects the classification confidence of the underlying detector
      - E.g., comes from margin distances of SVMs
    - Can also be a default value (e.g., 0.5)
      - A general formulation

Observation model - single human case

- State: $s = (x, y, \dot{x}, \dot{y})$
- Detection: $o = (x', y', c)$
- Observation function:

$$\Pr(o \mid s) = \Pr(c \mid \mathbf{1}) \Pr(x', y' \mid x, y) \tag{4}$$

  - Probability of having confidence c:
    $\Pr(c \mid \mathbf{1}) = Beta(c \mid 2, 1)$
  - Probability of observing (x', y'):
    $\Pr(x', y' \mid x, y) = \mathcal{N}(x', y' \mid x, y, \mathbf{\Sigma})$

Observation model - single false detection case

- State: $\varnothing$
- Detection: $o = (x', y', c)$
- Observation function:

$$\Pr(o \mid \varnothing) = \Pr(c \mid \mathbf{0}) f_b(x', y') \tag{5}$$

  - Probability of having confidence c:
    $\Pr(c \mid \mathbf{0}) = Beta(c \mid 1, 2)$
  - Probability of observing (x', y'): $f_b(x', y') = \frac{1}{|V|}$
    - Field of view: V

Observation model - matched case

- State: $S = \{s_0, s_1, \ldots, s_n\}$
- Observation: $O = \{o_1, o_2, \ldots, o_n\}$
- Joint observation function:

$$\Pr(O \mid S) = \sum_{\psi \in \Psi_S^O} \prod_{s \in S} \Pr(\psi(s) \mid s) \tag{6}$$

  - Ψ_S^O is all possible assignments from S to O

Observation model - general case

- State: $S = \{s_0, s_1, \ldots, s_n\}$
- Observation: $O = \{o_1, o_2, \ldots, o_m\}$
- Missing and false detections: $F \subseteq O$ and $M \subseteq S$
- All F-M pairs: $O \circ S = \{\langle F_0, M_0 \rangle, \langle F_1, M_1 \rangle, \ldots, \langle F_z, M_z \rangle\}$
- Human birth and death rates: ν and $|S|\xi$
- The overall joint observation function:

$$\Pr(O \mid S) = \sum_{\langle F, M \rangle \in O \circ S} \Pr(O - F \mid S - M)$$

$$\cdot \, (\nu\tau)^{|F|} e^{-\nu\tau} \prod_{o \in F} \Pr(o \mid \emptyset) \frac{(|S|\xi\tau)^{|M|} e^{-|S|\xi\tau}}{|M|!} \frac{1}{\binom{|S|}{|M|}} \qquad (7)$$

Observation model - approximation

- Intractable number of all terms:
  $\sum_{0 \leqslant i \leqslant \min\{|O|,|S|\}} \binom{|O|}{i}\binom{|S|}{i} i! = \Omega\left(\left(\frac{\max\{|O|,|S|\}}{e}\right)^{\min\{|O|,|S|\}}\right)$
- Approximation with prunings:
  - False-missing pruning: find F-M pairs in probability decreasing order until a threshold
    * A priority queue
  - Assignment pruning: find matched state-to-observation assignments in probability decreasing order until a ratio-threshold
    * Murty's algorithm (Murty, 1968)
      · Finds top-k assignments of a general assignment problem

Particle filtering over sets (PFS)

- A particle as a joint state $X = \{s_0, s_1, \ldots, s_n\}$
- A set of N particles to approximate $\Pr(S_t \mid O_t)$:

$$\mathcal{P}_t = \{\langle X_t^{(0)}, w_t^{(0)} \rangle, \langle X_t^{(1)}, w_t^{(1)} \rangle, \ldots, \langle X_t^{(N)}, w_t^{(N)} \rangle\} \tag{8}$$

  - $\sum_{i=1}^{N} w = 1$
- Particle filtering steps:
  - Propose: $\hat{X}_t \sim \pi(\cdot \mid X_{t-1}, O_t)$
  - Update: $w_t \leftarrow w_{t-1} \frac{m_t o_t}{p_t}$
    - Motion weight: $m_t = \Pr(\hat{X}_t \mid X_{t-1})$
    - Observation weight: $o_t = \Pr(O_t \mid \hat{X}_t)$
    - Proposal weight: $p_t = \pi(\hat{X}_t \mid X_{t-1}, O_t)$
  - Normalize and resample

Particle refinement - motivation

- Motion based proposal distribution
  - Propose: $\pi(\hat{X}_t \mid X_{t-1}, O_t) = \Pr(\hat{X}_t \mid X_{t-1})$
  - Update: $w_t \leftarrow w_{t-1} o_t$
    - Observation function: $o_t = \Pr(O_t \mid \hat{X}_t)$
    - Probability of $\hat{X}_t$ matching O_t if a new human detected in O_t: extremely small
- Idea: particle refinement according to observation

Particle refinement - single new detection

- Detection: $o = (x', y', c)$
- Probability of not being a false detection:
  $\Pr(\mathbf{1} \mid c) = \frac{\Pr(c|\mathbf{1})\Pr(\mathbf{1})}{\Pr(c|\mathbf{1})\Pr(\mathbf{1}) + \Pr(c|\mathbf{0})\Pr(\mathbf{0})} = c$
  - Beta assumptions
  - $\Pr(\mathbf{1}) = \Pr(\mathbf{0}) = 0.5$
- Probability of originating from state $s = (x, y, \cdot, \cdot)$:
  $\Pr(s \mid o) = \eta \Pr(o \mid s) \Pr(s) = \mathcal{N}(x', y' \mid x, y, \mathbf{\Sigma})$
  - Gaussian assumption
  - $\Pr(s) = \frac{1}{|V|}$
- Single state mixture proposal distribution $\pi_s(\cdot \mid o)$:
  - $\pi_s(\varnothing \mid o) = 1 - c$
  - $\pi_s(s \mid o) = c\mathcal{N}(x', y' \mid x, y, \mathbf{\Sigma})$

Particle refinement - to identify new detections

- Particle: $X = \{s_0, s_1, \ldots, s_n\}$
- Observation: $O = \{o_0, o_1, \ldots, o_m\}$
- Data association: $\varphi = \langle F, M, \psi \rangle$
  - False detections: $F \subseteq X$
  - Missing detections: $M \subseteq O$
  - Assignment from $X - M$ to $O - F$: $\psi \in \Psi_{X-M}^{O-F}$
- Observation function: $\Pr(O \mid X) = \sum_{\varphi} \Pr(O, \varphi \mid X)$
- Best association:
  $\varphi^* = \langle F^*, M^*, \psi^* \rangle = \operatorname{argmax}_{\varphi} \Pr(O, \varphi \mid X)$
- Most likely set of new detections: F^*

Particle refinement - general case

- Refinement based proposal $\hat{X}_t \sim \pi_r(\cdot \mid X_{t-1}, O_t)$:
  (a) Sample $X'_t \sim \Pr(\cdot \mid X_{t-1})$ using only motion model
  (b) Find best data-association $\varphi^* = \langle F^*, M^*, \psi^* \rangle$ given X'_t
  (c) Propose new humans $X' = \{s \mid s \sim \pi_s(\cdot \mid o), o \in F^*\}$
  (d) Propose a refined particle $X''_t \leftarrow X'_t \cup X'$
  (e) Return $\hat{X}_t \leftarrow \operatorname{argmax}_{X \in \{X'_t, X''_t\}} \Pr(O_t \mid X)$

Bayesian density estimation - motivation

- Particle X_{t-1} and refined proposal $\hat{X}_t$
  - Motion weight: $m_t = \Pr(\hat{X}_t \mid X_{t-1})$
  - Proposal weight: $p_t = \pi(\hat{X}_t \mid X_{t-1}, O_t)$
- Density estimation (Biswas et al., 2011):
  - Particles: $\mathcal{P} = \{X_0, X_1, \ldots, X_N\}$
  - Proposals from motion model:
    $\mathcal{P}' = \{X' \mid X' \sim \Pr(\cdot \mid X), X \in \mathcal{P}\}$
  - Proposals from particle refinement:
    $\mathcal{P}'' = \{X'' \mid X'' \sim \pi_r(\cdot \mid X'), X' \in \mathcal{P}'\}$
  - Motion weight: $\Pr(X'' \mid X) \approx \Pr(X'' \mid \mathcal{P}')$
  - Proposal weight: $\pi_r(X'' \mid X) \approx \Pr(X'' \mid \mathcal{P}'')$

Bayesian density estimation - $\Pr(X \mid \mathcal{P})$

- Samples: $\mathcal{P} = \{X_0, X_1, \ldots, X_N\}$
- Query: $X = \{s_0, s_1, \ldots, s_n\}$
- Single human distribution: $f_s(\cdot \mid \mathcal{P})$
  - All humans: $\mathcal{H}(\mathcal{P}) = \{s \mid s \in X, X \in \mathcal{P}\}$
  - Kernel density estimation (KDE): $f_s(s \mid \mathcal{P}) \approx \frac{1}{|\mathcal{H}(\mathcal{P})|} \sum_{s' \in \mathcal{H}(\mathcal{P})} \phi(x - x')\phi(y - y')$ (ignoring velocities)
- Posterior probability of having n humans: $\Pr(|X| = n \mid \mathcal{P}) = \mathcal{NB}(n; r, p) = \binom{n+r-1}{n} p^n (1-p)^r$
  - Number of failures: $r = \sum_{X \in \mathcal{P}} |X| + \alpha_0$
  - Success rate: $p = \frac{1}{1+N+\beta_0}$
- Overall density estimation: $\Pr(X \mid \mathcal{P}) = n! \Pr(|X| = n \mid \mathcal{P}) \prod_{s \in X} f_s(s \mid \mathcal{P})$

Human identification

- Joint posterior distribution in terms of sets: $\mathcal{P}_t$
  - $\mathcal{P}_t = \{\{s_0, s_1\}, \{s_2, s_3, s_4\}, \{s_5\}, \ldots\}$
- Human identification process: recognize and report each potential human individually
  - Input: $\{X_0, X_1, \ldots, X_N\}$, where $X_i = \{s_0, s_1, \ldots, s_n\}$
  - Output: $\{h_1, h_2, \ldots, h_z\}$
- An identified human/identity: $h = (s, c, \rho)$
  - State pool: $\mathcal{H}(h) \subseteq \mathcal{H}(\mathcal{P}_t)$
  - Estimated state: $s = \frac{1}{|\mathcal{H}(h)|} \sum_{s' \in \mathcal{H}(h)} s'$
  - Confidence: $c = \frac{|\mathcal{H}(h)|}{N} \in [0, 1]$

Human identification - formalization

- List of identities at cycle $t-1$: $L_{t-1} = \{h_0, h_1, \ldots, h_k\}$
- Particle $\mathcal{P}_t$ and observation O_t
- Proposals: h_o with $\mathcal{H}(h_o)$ initially empty for $o \in O_t$
- New identities according to O_t: $L_{O_t} = \{h_o \mid o \in O_t\}$
- List of candidates at cycle t: $C_t = L_{t-1} \cup L_{O_t}$
  - Labeling: $\mathcal{H}(h) = \{s \mid l(s) = h, s \in \mathcal{H}(\mathcal{P}_t)\}$ for $h \in C_t$
  - State distributions: $\mathbf{P} = \{f_h \mid h \in C_t\}$
- Human identification:

$$\mathbf{P}^* = \underset{\mathbf{P}}{\operatorname{argmax}} \max_{l} \Pr(\mathcal{P}_t, l \mid \mathbf{P}) \tag{9}$$

Human identification - expectation-maximization

- E-step: $l^{(k)} = \operatorname{argmax}_l \Pr(\mathcal{P}_t, l \mid \mathbf{P}^{(k-1)})$
  - Labeling method that maximizes $\prod_{s \in X} f_{l(s)}(s)$
  - N number of assignment subproblems
- M-step: $\mathbf{P}^{(k)} = \operatorname{argmax}_{\mathbf{P}} \Pr(\mathcal{P}_t, l^{(k-1)} \mid \mathbf{P})$
  - Maximal likelihood estimation (MLE)
  - Consideration of current observation O_t
- $l \to \mathbf{P} \to l' \to \mathbf{P}' \to \cdots \to l^* \to \mathbf{P}^*$

Human identification - M-step approximation

- M-step: $l \to \mathbf{P}$
- Observation state pool: $\mathcal{H}(o) \subseteq \mathcal{H}(\mathcal{P}_t)$ for $o \in O_t$
  - Most likely data association: $\langle F^*, M^*, \psi^* \rangle = \varphi^* = \operatorname{argmax}_\varphi \Pr(O_t, \varphi \mid X)$
  - Update: $\mathcal{H}(\psi^*(s)) \leftarrow \mathcal{H}(\psi^*(s)) \cup s$ for all $s \in X$
- MLE: $f_h(s) = \sum_{o \in O_t} f_h(s, o) + f_h(s, \emptyset) = \sum_{o \in O_t} \Pr(s \mid o) f_h(o) + f_h(s, \emptyset) = \sum_{o \in O_t} \mathbf{1}[s \in \mathcal{H}(o)] f_h(o) + \mathbf{1}[\forall o : s \notin \mathcal{H}(o)] f_h(\emptyset)$
  - $f_h(o) = \Pr(o \mid h) \Pr(h) \approx \frac{1}{N} |\mathcal{H}(o) \cap \mathcal{H}(h)|$
  - $f_h(\emptyset) = \Pr(\emptyset \mid h) \Pr(h) \approx \frac{1}{N} \left| \mathcal{H}(h) - \bigcup_{o \in O_t} \mathcal{H}(o) \cap \mathcal{H}(h) \right|$
- Set of state distributions: $\mathbf{P} = \{f_h \mid h \in C_t\}$

Human identification - EM process

- Initialize $l^{(0)}$ according to the converged/final labeling from the last cycle
  - Consideration of state deletion, addition and repetition in the particle filtering step
- Iteratively propose new set of distributions $\mathbf{P}^{(k)}$ and new labeling $l^{(k+1)}$ until convergence or a maximal number of steps is reached
- Construct L_t as: $L_t = \{l(s) \mid s \in \mathcal{H}(\mathcal{P}_t)\}$
  - $|L_t| \leqslant |C_t|$

Benchmark experiment

- PETS2009 dataset (Ferryman & Shahrokni, 2009)
  - Frame rate at ≈ 7 fps
  - 795 frames
  - Raw detection
    - Bounding box (x, y, h, w) in image frame
    - Confidence value c
  - Camera calibration
    - Image frame $\to$ world frame

Benchmark experiment - parameters

	Parameter	PETS2009
λ	Human birth rate (1/s)	0.0
μ	Human death rate (1/s)	0.02
σ_p	Dash power deviation (m^2/s)	1.0
ν	False detection rate (1/s)	6.0
ξ	Missing detection rate (1/s)	2.0
τ	Update time interval (s)	0.14
T'	Assignment pruning threshold	0.1
T''	False-missing pruning threshold	0.001
$\boldsymbol{\Sigma}$	Observation covariance	$0.5\mathbf{I}$
α_0	Initial Gamma α parameter	2.0
β_0	Initial Gamma β parameter	1.0
A'	Min. area of bounding box (m^2)	0.5
A''	Max. area of bounding box (m^2)	2.5
R	Min. conf. of reported identities	0.4
N	Number of total particles	128
H	Max. number of EM steps	10

Table 1 : Parameters used in PETS2009 experiment.

CLEAR MOT metrics

- Multiple Object Tracking Accuracy (MOTA)

$$\text{MOTA} = \left(1 - \frac{\sum_t (g_t + a_t - 2n_t + m_t)}{\sum_t g_t}\right) \times 100\% \quad (10)$$

  - g_t: the number of ground truths
  - a_t: number of reported targets
  - n_t: number of matched reported targets
  - m_t: number of ID-switch errors

CLEAR MOT metrics (cont'd)

- Multiple Object Tracking Precision (MOTP)

$$\mathrm{MOTP} = \left(1 - \frac{\sum_t \sum_{1 \leqslant i \leqslant n_t} d_t^{(i)}}{\sum_t n_t}\right) \times 100\% \tag{11}$$

  - n_t: number of matched reported targets
  - $d_t^{(i)}$: distance between reported target i and its matched ground truth

Quantitative results in PETS2009 dataset

Algorithm	MOTA	MOTP	IDS	MT	FM
PFS[1] (**proposed**)	93.1%	76.1%	3.6	18.0	16.0
PFS[12] (**proposed**)	90.6%	74.5%	4.8	17.6	20.4
Milan (2014)	90.6%	80.2%	11	21	6
Milan et al. (2013)	90.3%	74.3%	22	18	15
Segal & Reid (2013)	92%	75%	4	18	18
Segal & Reid (2013)[2]	90%	75%	6	17	21
Zamir et al. (2012)[2]	90.3%	69.0%	8	-	-
Andriyenko & Schindler (2011)	81.4%	76.1%	15	19	21
Breitenstein et al. (2011)[2]	56.3%	79.7%	-	-	-

[1]averaged over 16 runs.

[2]evaluated within tracking region not cropped.

Table 2 : Quantitative results in PETS2009 dataset.

Tracking results in PETS2009 dataset

Figure 7 : Qualitative tracking results in PETS2009 dataset.

Real robot demonstration

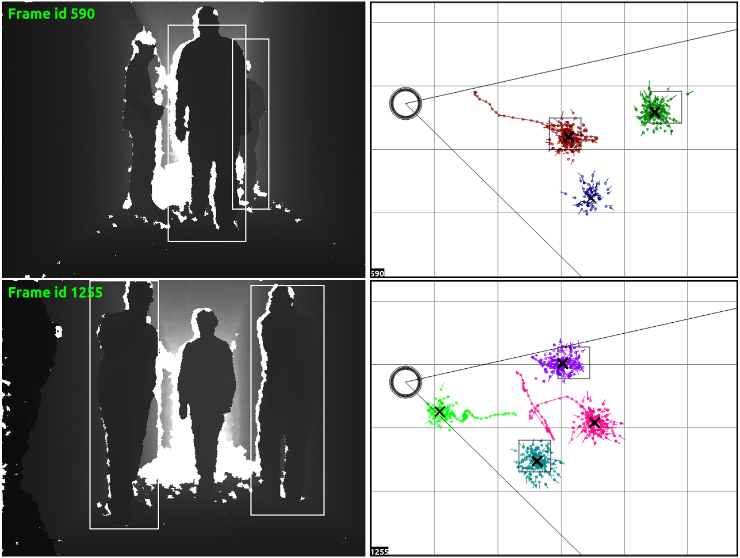

Figure 8 : Qualitative tracking results on CoBot.

Summary

- A particle filtering over sets (PFS) approach to MOT
  - A set formalization in joint space
  - No direct observation-to-target association
- Future work
  - Scale up to crowded environments
  - Integrate with intention/activity recognition

Outline

The problem

- Follow a human in an environment without a map
- Learn the path/trajectory in the environment
- Replay the learned path/trajectory reliably

Two primary ideas

- Idea I: POMDP formalization
- Idea II: Hierarchical framework

POMDP formalization

- State: robot state $s = (x, y, \theta)$ and landmarks $L = \{l_0, l_1, \ldots, l_n\}$
- Action: motion commands (dx, dy, dr)
- Transition function: motion transition with noises
- Observation: landmarks $O = \{l_0, l_1, \ldots, l_m\}$
- Observation function: $\Pr(O \mid s, L)$ with Gaussian assumptions

Following, learning and replaying in POMDP

- Follow a person: move towards tracked person
- Learn the path
  - Action-observation history $\rightarrow$ trajectory and landmarks
  - I.e., $\Pr(s_0, s_1, \ldots, s_t, L \mid a_0, o_0, a_1, o_1, \ldots, a_t, o_t)$
  - SLAM-like techniques
    - Kalman filter in joint space
    - Rao-Blackwellized Particle Filter in joint space
  - Minimal landmarks: $L^* = \operatorname{argmin}_{U(f(s_0,s_1,\ldots,s_t,L \mid a_0,o_0,a_1,o_1,\ldots,a_t,o_t)) \geqslant \epsilon} |L|$
- Replay the path
  - Follow the learned trajectory $(s_0, s_1, \ldots, s_t)$
  - Localize according to the landmarks $\{l_0, l_1, \ldots, l_n\}$

Hierarchical framework

- Assumptions: low-level skills/macro-actions
  - Move along a corridor (with local perception)
  - Turn left/right at a corner (with local perception)
- A trajectory as a high-level plan
  - A sequence of macro-actions (no branches/loops)
  - E.g., Move-forward $\rightarrow$ Turn-left $\rightarrow \ldots \rightarrow$ Turn-right
  - Each macro-action
    * Initial predicate
      · In-corridor / at-corner / reached-a-landmark
    * Termination predicate
      · Moved-for-3-meters / moved-for-10-seconds / passed-2-blocks / out-of-the-corner / reached-a-landmark

Following, learning and replaying with hierarchy

- Follow a person and learn the path
  - Track human's activity history
    - Move-forward / Turn-left / Turn-right
      - Observation history $\rightarrow$ activity history
      - Activity recognition: integrate in the tracker by associating different motion models with different activities
  - Invoke the respective sequence of macro-actions
  - Count execution status for current macro-action
  - Figure out the termination predicate when current macro-action needs to be terminated
- Replay the path: Execute the learned high-level plan given low-level macro-actions

Summary

- Two primary ideas of following, learning and replaying
  - POMDP formulation
    - Reliable landmark detections
      - Plants on ground
      - Pictures on walls
      - AR tags for test purposes
  - Hierarchical framework
    - Reliable low-level skills
      - Move forward along a corridor
      - Turn left/right at a corner

Outline

Demo

- CoBot following, learning and replaying

References I

Andriyenko, A., & Schindler, K. (2011). Multi-target tracking by continuous energy minimization. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, (pp. 1265–1272). IEEE.

Biswas, J., Coltin, B., & Veloso, M. (2011). Corrective gradient refinement for mobile robot localization. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, (pp. 73–78). IEEE.

Breitenstein, M. D., Reichlin, F., Leibe, B., Koller-Meier, E., & Van Gool, L. (2011). Online multiperson tracking-by-detection from a single, uncalibrated camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(9), 1820–1833.

Ferryman, J., & Shahrokni, A. (2009). PETS2009: Dataset and challenge. In *2009 Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS-Winter)*, (pp. 1–6). IEEE.

Milan, A. (2014). *Energy Minimization for Multiple Object Tracking*. PhD, TU Darmstadt, Darmstadt. URL `http://tuprints.ulb.tu-darmstadt.de/3463/`

Milan, A., Schindler, K., & Roth, S. (2013). Detection-and trajectory-level exclusion in multiple object tracking. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, (pp. 3682–3689). IEEE.

Murty, K. G. (1968). An algorithm for ranking all the assignments in order of increasing cost. *Operations Research*, *16*(3), 682–687.

Segal, A. V., & Reid, I. (2013). Latent data association: Bayesian model selection for multi-target tracking. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, (pp. 2904–2911). IEEE.

Zamir, A. R., Dehghan, A., & Shah, M. (2012). Gmcp-tracker: Global multi-object tracking using generalized minimum clique graphs. In *Computer Vision–ECCV 2012*, (pp. 343–356). Springer.