

Tracking a Varying Number of People with a Visually-Controlled Robotic Head

Yutong Ban, Xavier Alameda-Pineda, Fabien Badeig, Sileye Ba and Radu Horaud

Perception team, INRIA Grenoble Rhône-Alpes, France

26/09/2017, IROS 2017, Vancouver, Canada

Outline

- 1 Multi-Person Tracking Problem
- 2 Variational Tracking Model
- 3 Experimental Results
- 4 Conclusions

Multi-Person Tracking Problem

Multi-Person Tracking Problem

Problem:

- Unknown and time-varying number of moving persons
- Persons may appear/disappear
- Person detectors are not 100% reliable

Solution:

- Combine person detection with person tracking
- Efficient solution based on variational Bayesian algorithm
- Person tracking and visual servoing are combined to optimize long trajectories

Existing Multi-Person Tracking Methods

Sampling methods:

- Markov chain Monte Carlo (MCMC) Sampling, e.g. [Choi et al. 2015]
- Probability Hypothesis Density (PHD) filters, e.g. [Vo et al. 2003]

Problems:

- No closed-form solutions
- Non deterministic methods
- Algorithms with high computational cost

Our Solution: Variational Bayesian framework

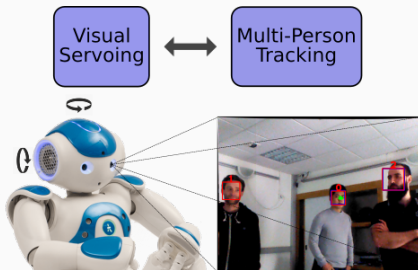
Variational Bayesian:

- Mathematical framework for approximating intractable posterior distributions, e.g. the filtering distribution associated with multiple person tracking.

Why choose variational Bayesian methods?

- Closed-form solution
- Efficient algorithms
- Deterministic method

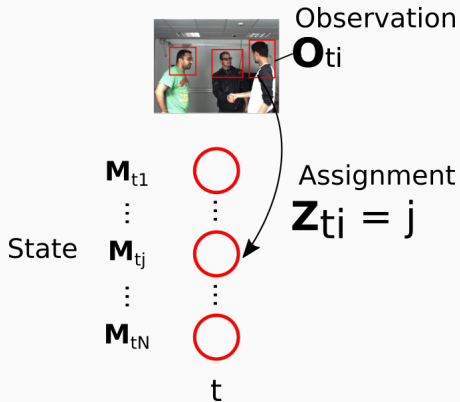
Multi-Person Tracking and Visual Servoing



- Problems arise whenever persons move out of field of view, e.g. identity changes when a person reappears
- Tracking can be more robust if visual servoing is used
- We propose a formal method to incorporate motor information into tracking

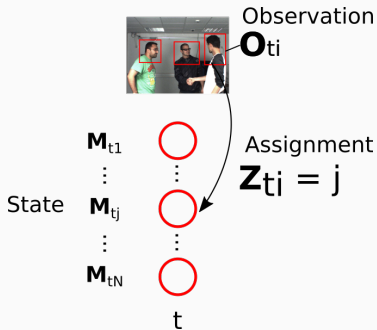
Variational Tracking Model

Notations



- Observation: detected face position
- Assignment:
 $Z_{ti} = j$ assigns face detection i to the identity j
- State: the tracked person's position and velocity

Notations



Goal:

$$\max p(\mathbf{Z}_t, \mathbf{M}_t | \mathbf{o}_{1:t})$$

Joint filtering distribution

$$p(Z_t, \mathbf{M}_t | \mathbf{o}_{1:t}) \propto$$

$$p(\mathbf{o}_t | Z_t, \mathbf{M}_t) p(Z_t) \int p(\mathbf{M}_t | \mathbf{M}_{t-1}) \sum_{Z_{t-1}} p(Z_{t-1}, \mathbf{M}_{t-1} | \mathbf{o}_{1:t-1}) d\mathbf{M}_{t-1}.$$

Joint filtering distribution

$$p(Z_t, \mathbf{M}_t | \mathbf{o}_{1:t}) \propto$$

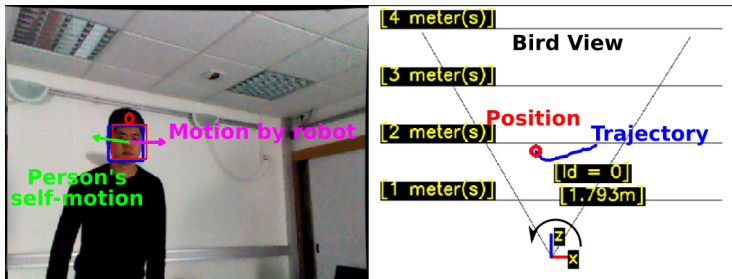
$$\underbrace{p(\mathbf{o}_t | Z_t, \mathbf{M}_t)}_{\text{Joint observation likelihood}} p(Z_t) \int p(\mathbf{M}_t | \mathbf{M}_{t-1}) \sum_{Z_{t-1}} p(Z_{t-1}, \mathbf{M}_{t-1} | \mathbf{o}_{1:t-1}) d\mathbf{M}_{t-1}.$$

Joint filtering distribution

$$p(Z_t, \mathbf{M}_t | \mathbf{o}_{1:t}) \propto$$

$$p(\mathbf{o}_t | Z_t, \mathbf{M}_t) p(Z_t) \int \underbrace{p(\mathbf{M}_t | \mathbf{M}_{t-1})}_{\text{state dynamics}} \sum_{Z_{t-1}} p(Z_{t-1}, \mathbf{M}_{t-1} | \mathbf{o}_{1:t-1}) d\mathbf{M}_{t-1}.$$

State dynamics



Predicted person location is the combination of:

- person self-motion
- motion by robot
- previous location

Joint filtering distribution

$$p(Z_t, \mathbf{M}_t | \mathbf{o}_{1:t}) \propto$$

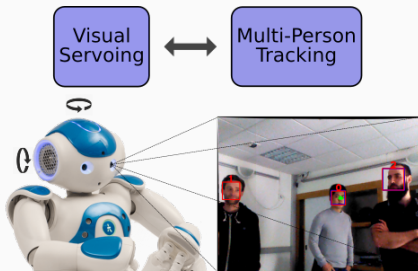
$$p(\mathbf{o}_t | Z_t, \mathbf{M}_t) p(Z_t) \int p(\mathbf{M}_t | \mathbf{M}_{t-1}) \underbrace{\sum_{Z_{t-1}} p(Z_{t-1}, \mathbf{M}_{t-1} | \mathbf{o}_{1:t-1})}_{\text{filtering distribution at } t-1} d\mathbf{M}_{t-1}.$$

- Using variational factorization, we have

$$p(\mathbf{Z}_t, \mathbf{M}_t | \mathbf{o}_{1:t}) \approx q(\mathbf{Z}_t)q(\mathbf{M}_t)$$

- Variational Expectation Maximization (VEM) algorithm derived a closed-form expression for each factorized distribution

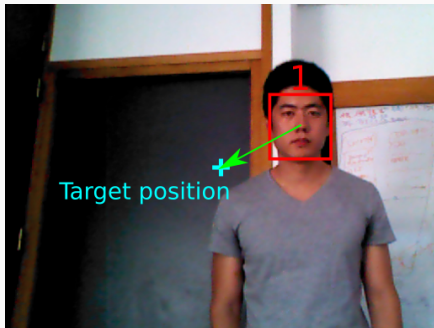
Combine Multi-Person Tracking with Visual Servoing



Visual Servoing

Using visual information, we can:

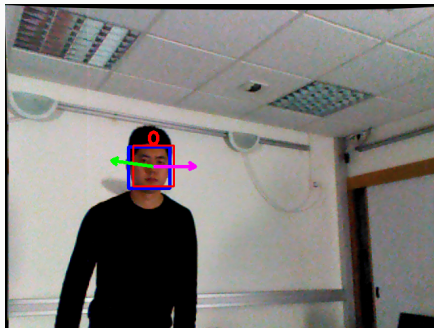
- Control the robot-head velocity to keep the selected person at the target position



keep person at target position

Using motor information, we can:

- Predict the velocity due to robot-motion. (distinguish person's self-motion from robot-motion)
- Give to the tracker additional person information even if the person is out of field-of-view.

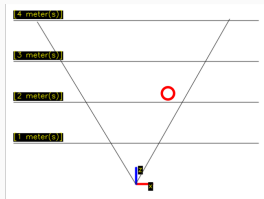


green arrow: person's self-motion
magenta arrow: motion due to robot-motion

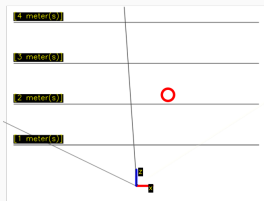
Visual Servoing

Using motor information, we can:

- Predict the velocity due to robot-motion. (distinguish person's self-motion from robot-motion)
- Give to the tracker additional person information even if the person is out of field-of-view



(a)



(b)

Robot rotates head, making the person out of the field-of-view

Experimental Results

<https://team.inria.fr/perception/research/mot-servoing/>

Quantitative Evaluations on NAO-MPVS Dataset

Sequence	Ctrl	MOTA(\uparrow)	MOTP(\uparrow)	FP(\downarrow)	FN(\downarrow)	IDs(\downarrow)
NAO-MPVS-1-1	w/o	92.1	67.2	9	9	0
	w/	91.3	68.7	10	10	0
NAO-MPVS-2J-1	w/o	52.8	67.1	93	207	2
	w/	81.6	68.0	30	88	0
NAO-MPVS-3J-1	w/o	35.8	62.3	159	433	19
	w/	63.1	62.1	83	268	0
Overall	w/o	48.8	65.0	261	649	21
	w/	73.1	65.3	123	366	0

Table 1: Comparison of tracking result without (**w/o**) and with (**w**) control using MOT metrics on three sequences of increasing complexity. \uparrow : the higher the better \downarrow : the lower the better

Conclusions

- A multi-person tracker incorporates motor information to encompass the effect of the robots ego-motion
- Variational approximation allows to effectively solve the problem with low computational cost
- NAO implementation : 10 FPS (due to hardware limitations)

For more information:

<https://team.inria.fr/perception/research/mot-servoing/>

Thanks for your attention

Questions?