Multiple Pedestrian Tracking in 3D With a Monocular Camera

Jan Krejčí[1]

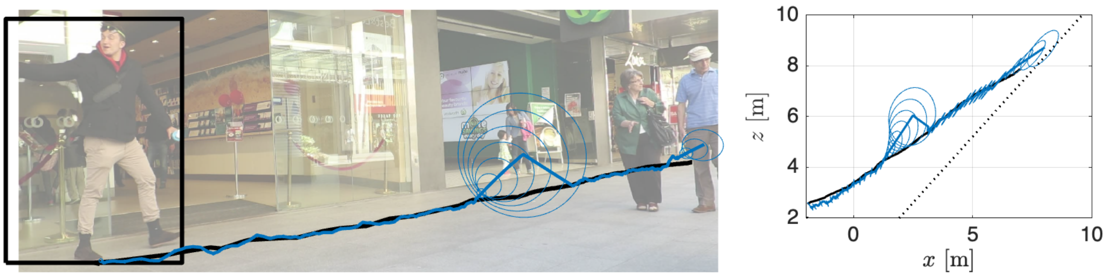
1 Introduction

The omnipresent low-cost monocular camera equipment calls for investigation if it can be used in safety-critical tasks such as traffic control or autonomous driving. Addressing such tasks must include high-quality pedestrian tracking under minimum prior information. Traditional solutions, however, either assume pedestrians move on a known ground plane or perform tracking only within the 2D image plane. To overcome this, Krejčí et al. (2024) proposed a single-pedestrian tracking algorithm working straight in 3D. This contribution shows that this algorithm can be used to track multiple pedestrians using the recent *random finite sets of trajectories* paradigm and illustrates the benefits of 3D tracking over traditional 2D tracking.

2 Tracking With a Monocular Camera

A simple way to describe a pedestrian in a video image is to surround it with a *2D bounding box*, described with its location, width, and height as depicted in Figure 1a. Traditional tracking algorithms stay with such a realm and develop a 2D tracking algorithm, whose example is the *2D-Model-Based Filter* described in Krejčí et al. (2024), taken as a benchmark.

Using the pinhole camera model, the 2D bounding box is a *projection* of a corresponding box in 3D. Krejčí et al. (2024) took this model into account and proposed a first-principle 3D tracking algorithm. Its outputs can be depicted both in 2D image and in 3D, see Figure 1.



(a) Tracking results in 2D. **(b)** Tracking results in 3D topview.

Figure 1: Results of the 3D tracking by Krejčí et al. (2024) with error covariance ellipses (——) and the true pedestrian path (——). Faster R-CNN detections were used, and they were unavailable in the middle of the image. The uncertainty is largest in the direction of the line of sight.

The input of single-pedestrian tracking algorithms are *detections*. These are noisy 2D boxes provided by a visual detection network that processes each image individually using machine learning techniques. There are, however, usually many detections per image that need to be *associated* to the tracked pedestrians. Moreover, detections are occasionally missing and there may be extra detections that do not belong to any real or already-tracked pedestrian.

[1] student of the doct. deg. prog. Applied Sciences, field of study Cybernetics, e-mail: jkrejci@students.zcu.cz

3 Bayesian Multiple Pedestrian Tracking

In the field of multi-object estimation, objects such as pedestrians in the scene are modeled as a finite random set. Both the cardinality and the individual elements of the set are random, which allows us to seek for the posterior probability law of the set using Bayes rule given the detections. A state-of-the-art algorithm stemming from this field is the *trajectory Poisson multi-Bernoulli mixture* (TPMBM) filter described by Xia et al. (2019). The TPMBM filter models entire trajectories of the objects (pedestrians) and naturally provides their *smoothed* estimates.

The TPMBM filter can be used with any single-object tracking algorithm. Here, both the aforementioned 2D and 3D tracking algorithms are used. The performance is demonstrated using the publicly available MOT17 dataset by Milan et al. (2017). In particular, the MOT17-09 video with included detections from Faster R-CNN visual detection network is used.

Results using performance metrics standard to both computer vision and estimation communities, see Milan et al. (2017) and Xia et al. (2019), respectively, are given in Table 1. The TPMBM filter using the 3D tracking algorithm clearly outperforms the case when the 2D tracking algorithm is used in all the metrics. Moreover, as opposed to the 2D model, the 3D model enables the TPMBM filter to address short-term occlusions, which is exemplified in Figure 2. It should be emphasized that the TPMBM filter uses no Re-ID module.

Tracking in	MOTA (↑)	HOTA (↑)	IDF1 (↑)	LP (↓)	GOSPA (↓)	ID Switches (↓)
2D	55.42	48.33	57.43	1841.39	71.03	32
3D	57.95	51.34	60.87	1783.46	69.47	26

Table 1: Results with the TPMBM filter using either the 2D or 3D tracking algorithm.

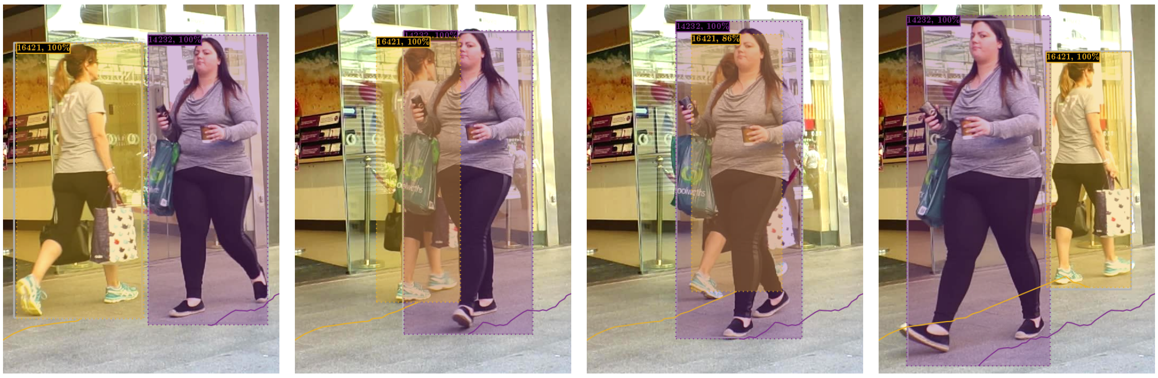


Figure 2: Example output using the 3D tracking algorithm when facing short-term occlusions.

References

Krejčí, J., Kost, O., Straka, O., and Duník, J. (2024) *Pedestrian Tracking with Monocular Camera using Unconstrained 3D Motion Model*. Available from: `https://arxiv.org/abs/2403.11978` [Accessed May 10 2024].

Xia, Y., Granström, K., Svensson, L., García-Fernández, Á. F., and Williams, J. L. (2019) Multiscan Implementation of the Trajectory Poisson Multi-Bernoulli Mixture Filter. *Journal of Advances in Information Fusion*, Volume 14, No. 2, pp.213-235.

Milan, A., Leal-Taixe, L., Reid, I., Roth, S., and Schindler, K. *MOT16: A Benchmark for Multi-Object Tracking*. Av. fr.: `https://arxiv.org/abs/1603.00831` [Ac. May 13 2024]