Real-Time Multi-Person Tracking with Time-Constrained Detection

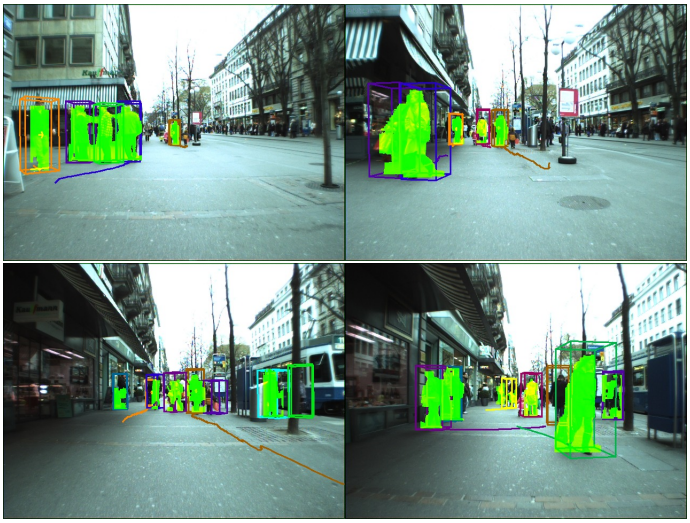
Dennis Mitzel
mitzel@umic.rwth-aachen.de
Patrick Sudowe
sudowe@umic.rwth-aachen.de
Bastian Leibe
leibe@umic.rwth-aachen.de

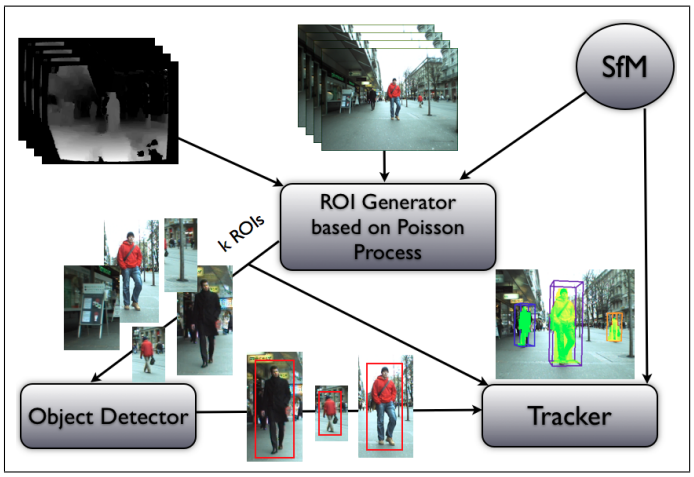
UMIC Research Centre
RWTH Aachen University
Aachen, GERMANY

In this paper we address the problem of vision-based multi-person tracking in busy urban environments using a camera setup mounted on a moving vehicle, *e.g.* an autonomous mobile robot. Recent years have seen considerable progress in this area, fueled by the development of advanced tracking-by-detection approaches [1, 3, 5]. However, those approaches require a robust object detector, which is triggered in each frame to detect all target objects in the scene.

In automotive scenarios the detector is usually restricted to evaluation of a small number of pre-selected ROIs based on stereo depth [3]. Recent approaches targeted at mobile robotics have adopted similar strategies [1, 2]. However, such approaches risk losing detections if the corresponding regions are missed by the ROI selection stage. What makes matters worse, the question which ROIs to select is usually addressed independently for every frame [4]. This results in a suboptimal selection strategy, either risking to lose important detections, or spreading the detector's time budget over many regions that have already been verified as containing or not containing an object before.

In contrast, we consider the case of an object detector with a fixed time budget *in the context of a tracking system*. We also assume that the detector can only process a small number of ROIs in each frame, but we balance the ROI selection over time, such that at each time instant, only those ROI candidates are considered for which attention is most urgently required in order to produce stable tracking results. The question we pose is: given a detector with a budget to attend to k ROIs in each frame and a cheap low-level tracking system to follow ROI candidates over time, which ones should be selected? To address this question, we propose the following approach. We first create ROI candidates from a depth map of the scene and from already existing object trajectories. These candidates are associated and tracked over time using local depth and appearance information. We then model the selection process of k ROIs to be verified by the detector using a statistical Poisson process model. Briefly stated, this model associates each tracked ROI candidate with a low probability of causing an important event. For regions in the background, this event means that the region now contains a person, despite of this having previously been verified as not being the case. For regions on tracked person trajectories, the event indicates a tracking failure that causes the low-level tracker to drift. In both cases, the occurrence of an event has the consequence that the region should be attended to and be verified by the object detector. Since we cannot predict where those events will happen, we model their probability of occurrence using a Poisson process. The result of this process indicates the *urgency* by which the detector should attend





to a region in order to limit the probability of the unfavorable event. In our approach, the urgency of a region is additionally moderated by its *utility* for maintaining tracking performance, which gives preference to regions close to the robotic platform.

Once the selected ROIs have been verified by the detector, its output is converted to 3D world coordinates using the camera position from Structure-from-Motion (SfM), together with an estimate of the ground plane. We then integrate the 3D measurements in a multi-hypothesis tracking approach similar to [5]. As our experimental results will demonstrate, our approach reaches state-of-the-art performance with high tracking quality, even with a significantly reduced time budget for the detector. We experimentally investigate the time budget required for a robust system-level performance and show that employing the stochastic Poisson process model optimizes ROI selection, such that only three detector evaluations per frame are sufficient for obtaining a highly robust tracking system.

In summary, our paper makes the following contributions: (1) We demonstrate how ROI selection can be optimized in general by employing a Poisson process model and how this model can be adapted for a tracking-by-detection approach. (2) In order to satisfy the conflicting goals of detecting new objects while stabilizing already existing tracks, we propose a two-tiered realization of the Poisson process model that takes into account a track's accumulated uncertainty. (3) We experimentally show that the proposed framework achieves robust multi-person tracking performance even with few ROI detector evaluations, making it possible to reduce detector evaluation to a minimum.

[1] M. Bajracharya, B. Moghaddam, A. Howard, S. Brennan, and L. Matthies. Results from a real-time stereo-based pedestrian detection system on a moving vehicle. In *ICRA*, 2009.

[2] M. Bansal, S. H. Jung, B. Matei, J. Eledath, and H. S. Sawhney. A real-time pedestrian detection system based on structure and appearance classification. In *ICRA*, 2010.

[3] D. Gavrila and S. Munder. Multi-Cue Pedestrian Detection and Tracking from a Moving Vehicle. *IJCV*, 2007.

[4] D. Geronimo, A.D. Sappa, D. Ponsa, and A.M. Lopez. 2D-3D-based On-Board Pedestrian Detection System. *CVIU*, 2010.

[5] B. Leibe, K. Schindler, and L. Van Gool. Coupled Object Detection and Tracking from Static Cameras and Moving Vehicles. *PAMI*, 2008.