Robotics and Autonomous Systems 103 (2018) 68–82

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Multiple target tracking in occlusion area with interacting object models in urban environments ☆

Jiun-Fu Chen [a], Chieh-Chih Wang [c], Cheng-Fu Chou [a,b,*]

[a] Department of Computer Science and Information Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei 10617, Taiwan
[b] Graduate Institute of Networking and Multimedia, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei 10617, Taiwan
[c] Department of Electrical and Computer Engineering, National Chiao Tung University, No. 1001, Ta Hsueh Rd., HsinChu 30010, Taiwan

ARTICLE INFO

Article history:
Received 17 March 2017
Received in revised form 18 December 2017
Accepted 9 February 2018
Available online 16 February 2018

Keywords:
Multitarget tracking
Interaction
LIDAR

ABSTRACT

Multiple target tracking in crowded urban environments is a daunting task. High crowdedness complicates motion modeling, and occlusion makes tracking difficult as well. Based on the variable-structure multiple-model (VSMM) estimation framework, this paper extends an interacting object tracking (IOT) scheme with occlusion detection and a virtual measurement model for occluded areas. IOT is composed of a scene interaction model and a neighboring object interaction model. The scene interaction model considers the long-term interactions of a moving object and surroundings, and the neighboring object interaction model considers three short-term interactions. With these interacting object models, the motion feature of a moving object can be represented with the weight of each model. A virtual measurement model is proposed to exploit the motion feature with the IOT scheme under occlusion. The proposed approach was validated using a stationary 2D LIDAR. To verify in occlusion, a 3D LIDAR based benchmark system was developed to extract occluded moving segments. The ample experimental results show that the proposed IOT scheme tracks over 57% of occluded moving objects in an urban intersection.

1. Introduction

Moving object tracking is an important issue in many aspects. Concerning the environment and the number of tracked objects, there are three different levels of tracking in general.

1. Level 1: Tracking in a free space. The moving object moves freely within the environment.
2. Level 2: Tracking in a constrained space. Objects must follow certain rules when moving. For example, road vehicles follow traffic rules.
3. Level 3: Tracking with multiple moving objects. For instance, vehicles and pedestrians moving in a metropolis. Not only are rules considered, but we also focus on the interactions with the surrounding moving objects and environment.

Table 1 shows the usage models for different tracking levels in Galceran et al. [1], interacting object tracking (IOT) [2–4], and the proposed virtual measurement model.

Level-three tracking – that is, multiple target tracking – is a critical capability for surveillance, scene understanding, intelligent transportation systems, and robotics. The ability to track multiple moving objects in urban environment is the basic requirement to provide driving safety information from intelligent traffic infrastructures. However, estimation difficulty rises with the level number. In this paper, we exploit an advanced scheme: the variable-structure multiple-model (VSMM) estimation framework [5].

Highly crowded urban environments complicate motion modeling with their many occlusion areas. Different motion patterns from a wide variety of moving objects make motion modeling difficult. Usually, occlusions occur because of a limited sensor field of view and because of surrounding objects. In urban environments, temporally- and spatially-varying surrounding objects introduce dynamic occluded areas which further complicate the tracking moving targets task. Failure to track under occlusion could lead to traffic accidents for any moving entity in the scene. Although multiple sensors, inter-vehicle communication, and vehicle to infrastructure communication can reduce occlusion, the problem remains for non-communicated nodes on the road such as pedestrians and single vehicles.

In this paper, we propose a framework to perform multiple interacting object tracking in urban areas using a stationary laser scanner based on the given segmentation results without any classification. The framework is composed of a virtual measurement model for tracking in occlusion and interacting object models that describe the interactions between the nearby moving objects and

☆ This paper is an extended work of our previous work described in Wang et al. (2007), Wan et al. (2008) and Lin et al. (2011).

* Corresponding author.
E-mail addresses: jeffchen@cmlab.csie.ntu.edu.tw (J.-F. Chen), bobwang@ieee.org (C.-C. Wang), ccf@csie.ntu.edu.tw (C.-F. Chou).

https://doi.org/10.1016/j.robot.2018.02.004
0921-8890/

the environment. In our initial study [2–4], the evaluation was conducted using a manually-labeled benchmark; we propose a benchmark-generating system that uses 3D LIDAR to provide the occluded information in 2D. To quantify the occlusion between 2D and 3D sensors and to build the benchmark system, an occluded area detection module is also proposed to extract occluded grids.

The interacting object models are composed of the scene interaction model and the neighboring object interaction model for long-term and short-term interactions, respectively. The scene interaction model is represented using a stationary occupancy map and moving object maps for the monitored scene. The occupancy-motion grids are utilized to store the collected moving object information such as speed and moving direction, and the k-means clustering algorithm [6] is applied to cluster the samples in order to provide predictions with means and covariances. The neighboring object interaction model is extended from [2] from a simple following interaction to three kinds of interactions: following, attracting, and repelling. These interacting object models not only solve the challenging modeling problem but also yield higher-level scene understanding. These interacting object models are consistent in occlusion and the weights of the motion models are utilized as the representation of a moving object's motion feature. We also find that the motion features of moving objects tend to vary little, with small changes in the weights of motion models. Thus, these interacting object models are applied to compute the virtual measurement model with the stored motion features for tracking in occluded space.

Only a few works address the observation and motion modeling issues of interactions among the tracked objects and the scene implicitly. Khan et al. [7] propose a Markov chain Monte Carlo (MCMC)-based particle filter to track interacting ants, in which interactions are modeled through a Markov random field motion prior. Their interaction potential is based only on static poses which yield no higher-level scene understanding. Smith et al. [8] use a simple interaction model to penalize object overlapping. Sullivan and Carlsson [9] propose constructing an interaction graph and then apply a two-stage clustering scheme to label the identity of the target. Instead of modeling interactions explicitly, these studies use the term "interaction" to describe situations in which the target and adjacent objects share the common measurements and cannot be correctly labeled.

For scalability, Kostavelis and Gasteratos [10] categorize semantic mapping-related works as either indoor or outdoor and single-scene or large-scale. Sengupta et al. [11] exploit two conditional random fields to provide a street-level semantic map from visual imagery. This work, however, focuses on static scenes or objects. Because it does not provide information on moving objects, existing work is unable to model moving objects in such environments. Wolf et al. [12] not only focus on the semantic terrain mapping problem, but also introduce a solution to the semantic activity mapping problem using supervised learning techniques, namely hidden Markov models (HMMs) and support vector machines (SVMs). In their activity-based semantic mapping, HMMs and SVMs are approaches for determining the spatial usage of dynamic entities in urban environments. Nevertheless, the final semantic map from [12] is not for directly tracking moving objects: it only summarizes how the moving objects utilize space.

Regarding the occlusion, Nashashibi and Bargeton [13] compensate for occlusion issue by determining about the occluded areas and use object classification to identify the occluded part, and then the confidence level estimation is used in Kalman filter-based tracking. Wyffels et al. [14] utilize occlusion as negative information in which they assume the tracked targets will exist in the next step and treat missing targets as negative information to estimate the occlusion or to remove nonexistent tracks. Most such works attempt to balance the effects of occlusion and utilize information to remove fake tracks.

Yu et al. [15] proposed to estimate objects in occluded areas. By splitting and merging shadow regions through time, they create a graph of the shadow information spaces through time to find occluded targets. Then, given the probability disturbance of the observation, the max flow or probability mass propagation algorithm is applied to find the number of hidden agents. [15] topologically estimates occluded objects, while [1] attempts to precisely track occluded objects inside shadow regions. Galceran et al. [1] introduce a road network model and a driving behavior model to perform estimation under occlusion. The road network model is composed of a discrete set of policies with a prior map of the environment, while the driving behavior model is composed of prescribe commands of steering-wheel-angle and forward-speed pairs. A hybrid Gaussian mixture model is applied to capture the estimates under occlusion within multiple hypotheses. However, this approach is suitable only for multi-lane roads or for single-lane intersections. The limited road network model is not easily applied to multi-lane intersections, and the fact that it does not take interactions between moving targets into account makes [1] it unsuitable for heavy traffic intersections.

As shown in Table 1, interacting object tracking and the proposed approach better handle level-3 tracking; the approaches are even better for level-2 tracking than [1], because its road network model models only the lane area for vehicles, which in contrast is part of the scene interaction model in [3] and in the proposed approach, which models the long-term intensions of all moving objects of the environment to provide information for multitarget tracking. Considering that moving objects do not change behavior suddenly and in fact interact with their surroundings, the proposed approach exploits both short-term and long-term interactions in managing the data association problem, thereby providing a precise estimate in occluded areas. The key contributions of this work are providing feasible approaches for modeling short-term and long-term interactions in urban environments, providing a laser based interacting object tracking framework with occlusion detection, an automatic benchmark generated system with 3D LIDAR data for verification, and successfully tracking in challenging scenarios.

The remainder of the paper is organized as follows. In Section 2 we review the VSMM estimation framework, and describe how we integrate the basic maneuver and interaction models. The scene interaction model and the neighboring object model are described in Sections 3 and 4, respectively. In Section 5 we describe occlusion detection and use the virtual measurement model to estimate for occluded areas. The experimental results and performance evaluation with a 3D LIDAR benchmark system are in Section 6. In Section 7 we conclude and discuss future work.

2. Variable-structure multiple-model estimation

In this section, we review the theoretical foundations of the variable-structure multiple-model (VSMM) estimation framework [5] and describe in detail how we integrate the moving models, the proposed scene interaction model and the proposed neighboring object interaction model.

2.1. Theory

The tracking problem can be solved with Bayesian approaches such as the Kalman filter and the particle filter. As the true motion mode is often unavailable in many applications, online motion modeling is needed. Moving object tracking can be formalized in probabilistic form as:

$$P(x_k, s_k \mid Z_k) \propto P(z_k \mid x_k, s_k) \sum_{s_{k-1}} \int P(x_k, s_k \mid x_{k-1}, s_{k-1}) P(x_{k-1}, s_{k-1} \mid Z_{k-1}) dx_{k-1} \tag{1}$$