REIHE 08
MESS-,
STEUERUNGS-
UND REGELUNGS-
TECHNIK

Fortschritt-Berichte VDI

Sascha Steyer, M.Sc.,
München

NR. 1272

Grid-Based Object Tracking

BAND
1 | 1

VOLUME
1 | 1

Lehrstuhl für Steuerungs- und Regelungstechnik
Technische Universität München
Univ.-Prof. Dr.-Ing./Univ. Tokio Martin Buss

Grid-Based Object Tracking

Sascha Jannik Steyer

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender: Prof. Dr.-Ing. Hans-Georg Herzog

Prüfer der Dissertation:

1. Priv.-Doz. Dr.-Ing. habil. Dirk Wollherr

2. Prof. Dr.-Ing. Marcus Baum

Die Dissertation wurde am 02.11.2020 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 12.04.2021 angenommen.

REIHE 08
MESS-, STEUERUNGS- UND REGELUNGS-TECHNIK

Fortschritt-Berichte VDI

Sascha Steyer, M.Sc.,
München

NR. 1272

Grid-Based Object Tracking

BAND
1|1

VOLUME
1|1

vdi verlag

VDI verlag

Steyer, Sascha

Grid-Based Object Tracking

Fortschritt-Berichte VDI, Reihe 08, Nr. 1272. Düsseldorf: VDI Verlag 2021.

196 Seiten, 66 Bilder, 2 Tabellen.

ISBN 978-3-18-527208-0, ISSN 0178-9546

71,00 EUR/VDI-Mitgliederpreis: 61,30 EUR

Für die Dokumentation: Autonomes Fahren – Objekterkennung – Objektverfolgung – Occupancy Grid Mapping – Sensordatenassoziation – Sensordatenfusion – Umfelderfassung – Umgebungswahrnehmung – Zustandsschätzung

Keywords: Autonomous Vehicles – Data Association – Environment Perception – Moving Object Detection – Object State Estimation – Object Tracking – Occupancy Grid Mapping – Sensor Data Fusion

Mobile robots require an accurate environment perception to plan intelligent maneuvers and avoid collisions. This thesis presents a novel multi-sensor environment estimation strategy that fully combines tracking moving objects and mapping the static environment. The basic idea is to fuse and accumulate measurement data by a dynamic occupancy grid model, whereas moving objects are extracted subsequently based on that generic low-level grid representation. Overall, this work results in a robust and consistent estimation of arbitrary objects and obstacles, which is demonstrated in the context of autonomous driving in complex unstructured environments.

Bibliographische Information der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet unter www.dnb.de abrufbar.

Bibliographic information published by the Deutsche Bibliothek (German National Library)
The Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliographie (German National Bibliography); detailed bibliographic data is available via Internet at www.dnb.de.

ISBN 978-3-18-527208-0, ISSN 0178-9546

Foreword

This thesis is the result of my research within the last years as a doctoral student at the Technical University of Munich in conjunction with the Department of Automated Driving of the BMW Group in Munich. I am very grateful to all the people who supported me during this time and made this work even possible.

First of all, I would like to thank PD Dr. habil. Dirk Wollherr, my doctoral advisor at the Institute of Automatic Control Engineering (LSR) of the Technical University of Munich, for his great support and guidance as well as the scientific freedom during my entire doctoral time.

I would also like to express my special thanks to Dr. Georg Tanzmeister for his outstanding supervision and mentoring on site at the BMW Group. He also supervised me before during my internship at the BMW Group Research and Technology, while he was a doctoral student himself, and my subsequent master thesis there, which all eventually originated this work.

I am also thankful to all the other great colleagues at the BMW Group that I have worked with over the past years. Thank you all for the supportive and friendly atmosphere, various interesting discussions, and the overall very exciting working time. Special thanks to the other PhD students in our group, in particular Jens Schulz, Constantin Hubmann, Christian Pek, and Kai Stiens, but also to Dr. Dominik Kellner and the other team members of the feature team I have been part of.

This work has also been supported by student projects, for which I would like to thank my former students Vinzenz Dallabetta, Christian Lenk, David Raba, and Max Pittner.

Lastly, I want to express my deepest gratitude to my family and Britta for their endless support, encouragement, and love.

As a final remark, I sincerely hope that the work of this thesis will make a small contribution to making driving safer and more automated in the future, with the ultimate goal of minimizing the number of road fatalities – in memory of M. B.

Munich, 2020 *Sascha Steyer*

Contents

Notations

The following overview of the notations of this thesis is limited to the most common abbreviations and symbols regarding the overall context of this work. Some of the symbols are also used as part of subscripts and superscripts.

Abbreviations

CTRA	Constant Turn Rate and Acceleration (motion model)
DST	Dempster–Shafer Theory of evidence [34, 35, 111]
FOV	Field Of View
GPU	Graphics Processing Unit
GTAM	Grid-based Tracking And Mapping [118]
ID	Identifier
OCS	Object Coordinate System
RMSE	Root Mean Square Error
UKF	Unscented Kalman Filter [55]
VRU	Vulnerable Road User (pedestrians, cyclists, etc.)
Cell	Individual grid cell c of the discretized grid representation
Map	Accumulated dynamic grid map $\mathcal{M}_t$ of the occupancy grid mapping
Particle	Individual particle filter hypothesis χ of the low-level particle tracking
Track	Individual object track instance τ of the high-level object tracking

Symbols

General

$t, \Delta t$	Time instance t and time difference Δt between two time instances
v	Velocity component
x, y	Position components, partly summarized as 2-D position vector $x = [x_x, x_y]$
z	Measurement
η	Normalization coefficient
Γ	Specified threshold
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2 (evaluation: $\mathcal{N}(\,\cdot\,; \mu, \sigma^2)$)
$\mathcal{U}(\,\cdot\,,\,\cdot\,)$	Uniform distribution with specified lower and upper bounds of the interval
$p(\cdot), p(\cdot\,\vert\,\cdot)$	Probability density function p, with conditional probability $p(\cdot\,\vert\,\cdot)$
$E_{(\cdot)}[\cdot]$	Expected value of a random variable
$[\cdot]^\mathsf{T}$	Transpose of a vector or matrix
R_φ	Rotation matrix with angle φ

Evidential Occupancy Grid Representation

c	Cell of the discretized grid structure
d_c	Grid cell size (square with $d_c \times d_c$)
x_c	Position of the center of cell c
$\mathcal{A}(c)$	Spatial area (2-D interval) of the cell c with center position x_c
$\mathcal{G}$	Grid structure, describing the set of all individual cells c, i.e., $c \in \mathcal{G}$
F, S, D	Hypotheses of freespace (F), static occupancy (S), dynamic occupancy (D)
O	Hypothesis of occupancy, abbreviation for $\{S, D\}$, i.e., $O = \{S, D\}$
$\{F, D\}$	Hypothesis of passable area (freespace or dynamic occupancy)
$\{S, D\}$	Hypothesis of occupancy (static or dynamic occupied)
Θ, Θ_z	Frame of discernment (full set), with $\Theta = \{F, S, D\}$, $\Theta_z = \{F, O\}$
2^Θ	Hypotheses power set of all combinations of the different hypotheses of Θ
θ	Arbitrary hypothesis subset of Θ, with $\theta \in 2^\Theta$
$m(\cdot)$	Basic belief assignment (evidence mass) of a specified hypothesis set
$m(\cdot \mid \cdot \oplus \cdot)$	Evidence mass based on the evidential combination ($\oplus$) of two sources
$\mathrm{bel}(\cdot)$	Belief of a specified hypothesis set
$\mathrm{pl}(\cdot)$	Plausibility of a specified hypothesis set
$\zeta(\cdot\,,\,\cdot)$	Conflict mass between two sources

Dynamic Grid Mapping and Low-Level Particle Tracking

$\mathcal{M}_t$	Accumulated dynamic grid map with the individual evidence masses $m(\cdot)$
$\overline{\mathcal{M}}_t$	Predicted dynamic grid map
$\widehat{\mathcal{M}}_t$	Particle-based map prediction of the dynamic part
$\mathcal{M}'_t$	Adapted map prediction of the non-dynamic part
$\mathcal{M}_{z,t}$	Fused measurement grid
$\mathcal{M}_{z^*,t}$	Fused measurement grid enhanced by additional occupancy classification
f_D	Function for the particle-based convergence toward dynamic occupancy (D)
β_S, β_D	Sensor-based occupancy classification coefficients for S and D, respectively
γ_D	Assignment uncertainty parameter of $\{D\}$ based on $\{F, D\}$ and $\{S, D\}$
$\lambda^{(\cdot)}_{(\cdot)}$	Individual terms of the adapted evidential occupancy filtering
χ	Individual particle hypothesis (position x_χ, velocity v_χ, occupancy value o_χ)
$\mathcal{X}, \hat{\mathcal{X}}$	Total population of all particles $\chi \in \mathcal{X}$ and corresponding prediction $\hat{\mathcal{X}}$
$\lvert\mathcal{X}^c\rvert$	Number of particles in a cell c
$n_{\max}$	Maximum number of particles per grid cell
Δn_χ	Difference between the number of existing particles and its target in a cell
ν^c_t	Mean particle-based 2-D velocity estimate of a grid cell c
v_R, v_T	Radial and tangential velocity component, respectively
w_χ	Particle (velocity) weight based on radar and camera measurement data

Object Extraction and Association

$\mathcal{C}_{\tau,t}$	Set of dynamic occupied cells associated to a track τ
$\mathcal{C}_{\tau,t}^{k}$	Sub-cluster of the set $\mathcal{C}_{\tau,t}$
$\mathcal{G}_{D,t}$	Set of currently dynamic occupied grid cells above a threshold $\Gamma_{\min}^{D}$
$\mathcal{G}_{D,t}^{\zeta_\emptyset}$	Set of unassociated dynamic occupied grid cells, with $\mathcal{G}_{D,t}^{\zeta_\emptyset} \subseteq \mathcal{G}_{D,t}$
$\mathcal{G}_{O,t}$	Set of currently occupied grid cells above a threshold $\Gamma_{\min}^{O}$
τ	Individual object track, $\tau \in \mathcal{T}_t$
$\mathcal{T}_t$	Set of currently tracked objects
$f_a(c)$	Association function for each grid cell c
l_χ	Particle label that connects a particle χ to a track τ
$\mathcal{X}_t^{\tau}$	Particle population of all particles linked to a track τ
$\zeta_\emptyset$	Symbol for denoting no association
N_ε^{c}	ε-neighborhood of a cell c for density-based clustering

Object State Estimation

a	Acceleration state component, $a = \dot{v}$
g_τ	Geometric state of a track τ (bounding box size)
l	Length of the bounding box
$s_{\tau,t}$	Dynamic state of a track τ at time t
w	Width of the bounding box
φ	Orientation state component
ω	Turn rate state component, $\omega = \dot{\varphi}$
$\mathcal{I}_{\varphi}^{\mathcal{X}}$	Particle-based confidence interval of the assumed object orientation
k	Specific object class instance, $k \in \mathcal{K}$
$\mathcal{K}$	Set of possible classes (car, truck, pedestrian, cyclist, motorcycle, other)
$r_F(\mathcal{A})$	Ratio of the freespace evidence within a defined grid area $\mathcal{A}$
ϑ_z^{e}	Edge visibility of the measurement box, $e \in \{\text{front, rear, left, right}\}$
$\overline{\nu}$	Weighted mean cell velocity vector of all associated cells

Abstract

Mobile robots require an accurate environment perception to plan intelligent maneuvers and avoid collisions with traffic participants or obstacles. To obtain a robust model of the current surroundings, measurement data of multiple sensors have to be processed in different ways. This includes several tasks, such as abstracting object instances, data association, temporal filtering, and sensor data fusion. Popular approaches are commonly based on a sensor-individual object tracking with an early-stage object abstraction and a late-stage sensor fusion. However, that procedure causes significant information loss of the raw measurement data for the subsequent processing steps, potentially resulting in object ambiguities and thus in an error-prone high-level object fusion. This is particularly critical for autonomous driving in complex urban scenarios with densely moving traffic, partial occlusions, and unstructured parts of the static environment.

This thesis proposes a novel multi-sensor environment estimation strategy – the *Grid-Based Object Tracking*. The fundamental idea of this work is to fuse and temporally filter measurement data by a low-level environment model based on the generic concept of dynamic occupancy grids, whereas the object estimation is performed subsequently based on that grid representation. That way, objects are extracted robustly and consistently by using the full information of the fused and temporally accumulated data of all sensors, without requiring any early-stage object abstraction. Moreover, this approach fully comprises and combines tracking moving objects and mapping the static environment, taking into account all measurement detections, which overall ensures that all occurring objects and obstacles are reliably contained in the resulting environment representation.

In this work, measurement data of lidar, radar, and camera sensors are processed. They are modeled in a uniform occupancy grid representation with uncertainties using an evidential Dempster–Shafer model, which is further extended by separate velocity and classification grid layers. The sensor fusion is performed cell-wise based on the grid cell discretization. For the temporal accumulation, a new dynamic grid mapping approach combined with a low-level particle filter tracking is proposed, resulting in accurate cell velocity estimates and a differentiation of static and dynamic occupancy over time. Finally, moving objects are extracted based on the dynamic grid estimation, which thereby serves as a track-before-detect strategy that enables a generic detection of arbitrary-shaped moving objects primarily by identifying their cell-wise occupancy motion. To utilize the full potential of the overall approach, also new concepts for the object state estimation and association are presented that benefit from the grid-based representation as well, e.g., by evaluating the current object visibility based on the additional freespace information of the grid.

The proposed approach is evaluated with real sensor data and test vehicles, demonstrating that it is well suited for real-time multi-sensor environment perception applications, especially in the context of autonomous driving in challenging urban environments.

Zusammenfassung

Mobile Roboter benötigen eine genaue Umgebungswahrnehmung zur intelligenten Fahrmanöverplanung und Kollisionsvermeidung mit Verkehrsteilnehmern oder Hindernissen. Um ein robustes Modell der aktuellen Umgebung zu erhalten, müssen Messdaten mehrerer Sensoren verarbeitet werden. Dies beinhaltet verschiedene Schritte, wie etwa die Abstraktion von Objekten, die Datenassoziation, die zeitliche Filterung und die Sensordatenfusion. Bekannte Ansätze basieren üblicherweise auf einem sensorindividuellen Objekt-Tracking mit einer frühen Abstraktion der Messdaten zu Objektinstanzen und einer späten Fusion der Sensordaten. Allerdings führt diese Vorgehensweise zu einem deutlichen Informationsverlust der Messdaten für die darauffolgenden Verarbeitungsschritte, was potenziell Objekt-Mehrdeutigkeiten und somit eine fehleranfällige Sensorfusion auf Objektebene zur Folge haben kann. Dies ist insbesondere für autonomes Fahren in komplexen urbanen Szenarien mit dichtem Straßenverkehr, teilweisen Verdeckungen und einer unstrukturierten statischen Umgebung kritisch.

Diese Thesis stellt eine neuartige Strategie zur Multi-Sensor Umgebungsschätzung vor – genannt *Grid-Based Object Tracking*. Die Grundidee dieser Arbeit besteht darin, Messdaten in einem Low-Level Umgebungsmodell zu fusionieren und zeitlich zu filtern, basierend auf dem generischen Konzept von Dynamic Occupancy Grids. Die Objektschätzung erfolgt hingegen erst aufbauend auf dieser Grid-Repräsentation. Dies ermöglicht eine robuste und konsistente Extraktion von Objekten, da hierbei die vollständigen Informationen der fusionierten und zeitlich akkumulierten Messdaten aller Sensoren betrachtet werden, ohne dass diese zuvor zu Objektinstanzen abstrahiert werden müssen. Zudem kombiniert dieser Ansatz vollumfänglich die Schätzung von sich bewegenden Objekten mit der Schätzung der statischen Umgebung und berücksichtigt dabei alle Messdetektionen. Insgesamt wird somit sichergestellt, dass in der resultierenden Umgebungsrepräsentation alle auftretenden Objekte und Hindernisse zuverlässig enthalten sind.

In dieser Arbeit werden Messdaten von Lidar-, Radar-, und Kamerasensoren verarbeitet. Diese werden in einer einheitlichen Occupancy Grid Darstellung mit Unsicherheiten mithilfe eines Dempster–Shafer Evidenzmodells sowie weiterer Grid-Ebenen für Geschwindigkeits- und Klassifikationsinformationen modelliert. Die Sensorfusion erfolgt dabei zellweise basierend auf der Grid-Zelldiskretisierung. Für die zeitliche Akkumulation wird ein neuer Dynamic Grid Mapping Ansatz vorgestellt, der mit einer Partikelfilter-basierten Dynamikschätzung kombiniert ist. Hieraus resultiert eine genaue Schätzung von Geschwindigkeiten der Grid-Zellen sowie eine Unterscheidung von statischer und dynamischer Belegung über die Zeit. Bewegte Objekte werden dann aufbauend auf dieser Dynamic-Grid-Repräsentation extrahiert, die dabei als Track-Before-Detect Strategie dient und eine generische Detektion beliebiger Objekte primär durch die Erkennung der Bewegung derer Occupancy-Zellen ermöglicht. Um das volle Potenzial dieses Ansatzes zu entfalten, werden zudem neue Konzepte für die Objektzustandsschätzung und -assoziation vorgestellt, die ebenfalls von der Grid-basierten Darstellung profitieren, z. B. wird die Sichtbarkeit der Objekte anhand der zusätzlichen Freiraum-Information des Occupancy Grids bestimmt.

Der Ansatz wurde mit realen Sensordaten und Testfahrzeugen evaluiert; die Ergebnisse demonstrieren eine erfolgreiche Anwendung als Multi-Sensor-Umgebungsschätzung, insbesondere im Kontext des autonomen Fahrens in komplexen urbanen Umgebungen.