Selected Topics in Computer Vision

Stefan Breuers

Multi-Object Tracking and Person Analysis from Mobile Robot Platforms

Band 4

Herausgeber: Prof. Dr. Bastian Leibe

Lehr- und Forschungsgebiet Informatik 8
Computer Vision Group

RWTH AACHEN UNIVERSITY

Multi-Object Tracking and Person Analysis from Mobile Robot Platforms

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von M. Sc.
Stefan Breuers
aus Meerbusch, Deutschland

Berichter: Prof. Dr. Bastian Leibe
Prof. Dr. Horst-Michael Groß

Tag der mündlichen Prüfung: 02.12.2019

Diese Dissertation ist auf den Internetseiten der Hochschulbibliothek online verfügbar.

Selected Topics in Computer Vision

herausgegeben von
Prof. Dr. Bastian Leibe
Lehr- und Forschungsgebiet Informatik 8
(Computer Vision)
RWTH Aachen University

Band 4

Stefan Breuers

Multi-Object Tracking and Person Analysis from Mobile Robot Platforms

Shaker Verlag
Düren 2020

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Zugl.: D 82 (Diss. RWTH Aachen University, 2019)

Printed in Germany.

ISBN 978-3-8440-7283-9
ISSN 2198-3372

Shaker Verlag GmbH • Am Langen Graben 15a • 52353 Düren
Phone: 0049/2421/99011-0 • Telefax: 0049/2421/99011-9
Internet: www.shaker.de • e-mail: info@shaker.de

Abstract

Multi-object tracking is a broad and very active field of research in the area of computer vision. Finding the trajectories of multiple persons in a scene is an important key component in video analysis, surveillance, autonomous driving, as well as mobile robotics. The latter application has led to several international research projects, e.g., developing social service platforms, on whose results this thesis is based on.

First, we study common approaches for image-based 2D multi-object tracking and analyze exemplary methods with regard to the errors they make. We propose a classifier that learns the situations where false positive tracks appear, based on bounding box context features. The individual characteristics allow for a combination of the trackers' output and we show that this leads to an improved general result. This not only indicates that there is still potential to improve individual methods, but also that multi-object trackers have different strengths and we always need to take a full look on all the evaluation measures.

When analyzing the results of those trackers it is therefore important to keep the application scenario in mind. As mentioned above, we have a look at robot platforms and examine how well recent multi-object tracking approaches perform in those 3D world situations. For this, we present a highly modular detection-tracking pipeline. We discuss important design choices, considering the chosen data association or the use of multi-modal detectors, where complex methods or more input, respectively, does not always lead to better tracking performance.

We then extend the above pipeline to also integrate person analysis modules as another modular level. By using the unique trajectories, we can apply temporal filtering on the analysis output of each tracked person. On the example of head and body pose estimation, we show that this way, we get a smoothed, improved result of those attributes. Additionally, it is possible to run those filters with a certain stride, resulting in a huge performance boost when dealing with those expensive deep learning methods.

Finally, we also explore a new multi-object tracking approach building on top of this successful deep learning framework. While existing methods often use deep appearance or motion models to help the data association step, we try to completely sidestep the dependency on a detector and therefore the need for data association. In order to do so, we make use of a strong re-identification model based on triplet loss inside an optimal Bayes filter, which forms the theoretical foundation of many tracking methods. By modeling track states as full probability maps, we can operate directly on the image input, taking a step towards an end-to-end image-to-track approach.

Zusammenfassung

Multi-Objekt-Tracking ist ein breites und sehr aktives Forschungsgebiet im Bereich der Computer-Vision. Das Auffinden der Trajektorien mehrerer Personen in einer Szene ist eine wichtige Schlüsselkomponente in der Videoanalyse, der Überwachung, dem autonomen Fahren sowie der mobilen Robotik. Letztere Anwendung hat zu mehreren internationalen Forschungsprojekten geführt, die z.B. soziale Dienstplattformen entwickeln, auf deren Ergebnissen diese Arbeit basiert.

Zunächst werden gängige Ansätze für bildbasiertes 2D Multi-Objekt-Tracking untersucht und exemplarische Methoden hinsichtlich ihrer Fehler analysiert. Wir schlagen einen Klassifikator vor, der Situationen lernt, in denen falschpositive Tracks auftreten, basierend auf Bounding Box Kontext Merkmalen. Die einzelnen Charakteristika ermöglichen eine Kombination der Ausgabe der Tracker und wir zeigen, dass dies zu einem verbesserten Gesamtergebnis führt. Dies zeigt nicht nur, dass es noch Verbesserungspotenzial für einzelne Methoden gibt, sondern auch, dass Multi-Objekt-Tracker unterschiedliche Stärken haben und wir immer alle Bewertungsmaßnahmen in Betracht ziehen müssen.

Bei der Analyse der Ergebnisse dieser Tracker ist es daher wichtig, das Anwendungsszenario im Auge zu behalten. Wie bereits erwähnt, werfen wir einen Blick auf Roboterplattformen und untersuchen, wie gut aktuelle Multi-Objekt-Tracking Ansätze in diesen 3D-Welt Situationen funktionieren. Hierfür stellen wir eine hochgradig modulare Detektions-Tracking-Pipeline vor. Wir diskutieren wichtige Designentscheidungen unter Berücksichtigung der gewählten Datenassoziation oder des Einsatzes multimodaler Detektoren, bei denen eine komplexe Methode bzw. mehr Eingangsdaten nicht immer zu einer besseren Tracking-Performance führen.

Wir erweitern dann die oben genannte Pipeline, um auch Personenanalysemodule als weitere modulare Komponente zu integrieren. Durch die Verwendung der eindeutigen Trajektorien können wir eine zeitliche Filterung auf die Analyseausgabe jeder getrackten Person anwenden. Am Beispiel der Schätzung von Kopf- und Körperhaltung zeigen wir,

dass wir auf diese Weise ein geglättetes, verbessertes Ergebnis dieser Attribute erhalten. Darüber hinaus ist es möglich, diese Filter mit einem gewissen Schrittwert auszuführen, was zu einem enormen Leistungsschub im Umgang mit diesen teuren Deep Learning Methoden führt.

Schließlich untersuchen wir auch einen neuen Multi-Objekt-Tracking Ansatz, der auf diesem erfolgreichen Deep-Learning-Framework aufbaut. Während bestehende Methoden oft tiefe Erscheinungs- oder Bewegungsmodelle verwenden, um die Datenassoziation zu unterstützen, versuchen wir, die Abhängigkeit von einem Detektor und damit die Notwendigkeit der Datenassoziation vollständig zu umgehen. Dazu nutzen wir ein starkes Re-Identifikationsmodell, das auf triplet loss basiert, innerhalb eines optimalen Bayes-Filter, welcher die theoretische Grundlage für viele Tracking Methoden bildet. Durch die Modellierung von Track Zuständen als vollstndige Wahrscheinlichkeitsverteilungen können wir direkt auf den Eingangsbildern arbeiten und einen Schritt in Richtung eines Ende-zu-Ende Bild-zu-Track-Ansatzes machen.

Acknowledgments

First of all, I would like to thank my supervisor Prof. Bastian Leibe for the opportunity and resources he has granted me in order to pursue my Ph.D., as well as Prof. Horst-Michael Groß for agreeing to be my second supervisor.

A big shout out to all my colleagues in the lab, especially Dennis Mitzel ("Boom!") for introducing me to the lab and his "house of cards" code framework; Lucas Beyer for awesome project meetings, twice the joint work and Mana ("Kotol giff!"); Alexander Hermans (not only) for food tips; Patrick Sudowe for all the pieces of advice during lunch time; Markus Mathias for an awesome road and skiing trip, besides the help and motivation on my first published paper, of course; Dan Jia and Sabarinath Mahadevan for taking over some of my work; and all of my soccer pals, including Aljoša Ošep, Wolfgang Mehner, Paul Voigtlaender and Jonathon Luiten.

Another thanks goes to the project partners in all these years, especially Timm Linder for the joint work ("The best thing about a deadline is the sunrise on the way home."). All those world-wide integration and review meetings were such a challenging, but always fun experience.

Thanks to all my students, who contributed to this work with their own effort: Shishan Yang, Kinan Halloum, Kersten Schuster, Antonia Breuer, Judith Hermanns, Neng Qian.

This is to my friends from university, Sebastian Freitag, Mario Claer, Tobias Baumgartner and Jonathan Meyer for fun-filled discussions, sleepless gaming nights and all the shared time... see you at the next "Regulärer".

I would also like to thank my parents Barbara and Josef and my brothers Jan and André for their support and relaxing family time in between.

To my wife Jessica: Thank you so much for your love and kind words, especially in the final weeks of this thesis. You endured my downs, cheered me up and shared my highs. I am so glad to have you by my side.

Thank you.

Contents