Multi-Object Tracking without dynamic models and hard association metrics

Christian Alexander Holz, Christian Bader, Matthias Drüppel

***Abstract*—In diesem Paper wird die Entwicklung innovativer Machine Learning (ML)-basierter Methoden zur Multi Object Tracking (MOT) im Kontext von Advanced Driver Assistance Systems (ADAS) untersucht. Angesichts der zunehmenden Komplexität und Anforderungen an präzise und effiziente Objektverfolgungssysteme in der Automobilindustrie, fokussiert sich diese Arbeit auf die Integration von ML-Techniken in etablierte Tracking-Verfahren. Zentrale Beiträge umfassen die Entwicklung und Evaluierung von drei spezialisierten neuronalen Netzwerken: Single Prediction Network (SPENT), Single Association Network (SANT), und Multi Association Network (MANTa). Diese Netzwerke zielen darauf ab, ML-Methoden mit einem traditionellen Kalman-Filter-Framework zu kombinieren und bieten einen innovativen Ansatz zur Bewältigung der MOT Herausforderungen. Die Vorteile eines Tracking-by-Detection (TbD) Frameworks, wie der modulare Aufbau, wurden mit den Vorteilen von Machine Learning (ML) Verfahren kombiniert. Die Ergebnisse zeigen ein modularen, robusten und wartbaren Tracker, welcher das Potenzial der ML-Integration in ADAS unterstreicht.**

***Index Terms*—Article submission, IEEE, IEEEtran, journal, LaTeX, paper, template, typesetting.**

I. Introduction

THIS [1] Die fortwährende Evolution von Advanced Driver Assistance Systems (ADAS) hat die Notwendigkeit einer präzisen und zuverlässigen Multi Object Tracking (MOT) ins Rampenlicht gerückt. In komplexen und dynamischen Umgebungen, wie sie im städtischen Verkehr vorkommen, ist es entscheidend, die Positionen und Bewegungen mehrerer Objekte gleichzeitig und genau zu erfassen. Die Herausforderung hierbei liegt nicht nur in der Erfassung und Verfolgung einzelner Objekte, sondern auch in der Berücksichtigung ihrer Interaktionen und gegenseitigen Beeinflussungen, insbesondere bei Verdeckungen und plötzlichen Bewegungsänderungen.

In dem häufig verwendeten Paradigma des Tracking-by-Detection (TbD) fusioniert ein Tracker erkannte Sensorobjekte (SO), um konsistente Objektspuren über die Zeit zu erstellen. Eine Schlüsselherausforderung dabei ist, eingehende Messungen (Sensor Objekten (SO)) den entsprechenden bestehenden Spuren zuzuordnen bzw. neue Objektspuren zu Initialisieren. Diese Datenassoziation wird in den meisten existierenden Methoden basiert auf Ähnlichkeitswerten durchgeführt, die zwischen den Messungen und den bestehenden Spuren berechnet werden. Diese Ähnlichkeitswerte können auf der letzten Erkennung beruhen oder aus historischen Erkennungen aggregiert werden. Für die Zustandsvorhersage haben sich bei TbD Ansätzen in vielen Anwendungen, Kalman-Filter und dessen Varianten als effektiv erwiesen. Diese stoßen jedoch bei komplexeren Szenarien, insbesondere bei nicht-linearen Bewegungsmustern und Interaktionen mehrerer Objekte, an ihre Grenzen. In dieser Arbeit stellen wir einen neuartigen MOT Ansatz vor, der Machine Learning (ML) nutzt, um diese Herausforderungen zu überwinden. Wir konzentrieren uns insbesondere auf die Entwicklung und Implementierung von Neural Networks (NN), die eine präzisere und flexiblere Objektverfolgung datenbasiert ermöglichen können, ohne auf umständliche Heuristiken und Hyperparameter angewiesen zu sein.

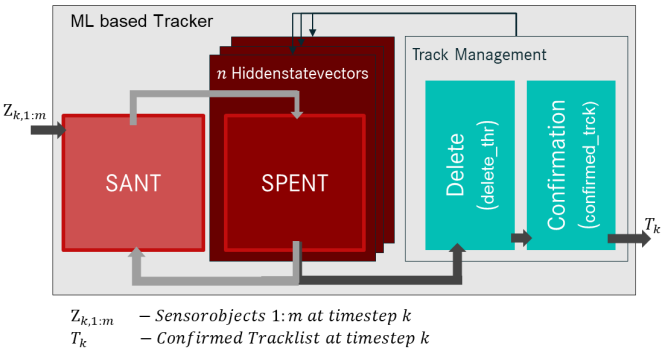

Fig. 1. Schematische Darstellung der beiden integrierten Netzwerke SPENT und SANT in einen Tracking-by-Detection (TbD) Framework.

Unser Hauptbeitrag liegt in der Entwicklung und Evaluierung des Single Prediction Network (SPENT), des Single Association Network (SANT) und des Multi Association Network (MANTa). Im Vergleich zum Kalman-Filter, ist SPENT in der Lage, Systemzustände einzelner Objekte vorherzusagen ohne den Bedarf eines vor der Laufzeit definierten Zustands- bzw. Beobachtungsmodells. SPENT bietet das Potenzial, insbesondere in Bezug auf die Anpassungsfähigkeit an verschiedene Szenarien und die Fähigkeit, Nichtlinearitäten effektiv zu handhaben.

Viele herkömmliche Tracking-Systeme nutzen statische Methoden zur Datenassoziation, die oft auf einfachen Heuristiken oder festen Schwellenwerten basieren. SANT hingegen nutzt maschinelles Lernen, um diese Prozesse zu automatisieren und sich besser an unterschiedliche Szenarien anzupassen. Somit ersetzt SANT innerhalb des TbD MOT Verfahrens die Berechnung einer Abstandsmetrik sowie den Hungarian-Algorithm zur entsprechenden Zuordnung.

C. Holz is with Master of Engineering, Duale Hochschule Baden-Württemberg (DHBW), Stuttgart, Germany

C. Bader is with Graduate School of Mathematics, Nagoya University, Nagoya, Japan

M. Drüppel is with the Center for Artificial Intelligence, DHBW Stuttgart, Stuttgart, Germany

Darüber hinaus diskutieren wir die Integration von SPENT und SANT in ein bestehendes Tracking-System und demonstrieren die Performance auf Basis mehrere Tests und Vergleiche mit etablierten Methoden. Diese Arbeit bietet somit wertvolle Einblicke und einen bedeutenden Fortschritt für die Entwicklung von Advanced Driver Assistance Systems (ADAS) und trägt zur Weiterentwicklung der Technologien für autonomes Fahren bei.

II. Related work

a) Multi-Object Tracking: Die Verfolgung mehrerer Objekte ist eine zentrale Herausforderung in der Computer Vision. Die meisten bestehenden Ansätze zur Verfolgung mehrerer Objekte, einschließlich des hier vorgestellten, basieren auf dem Ansatz der "Tracking-by-Detection (TbD)". Offline-Methoden [xx, ...] verarbeiten dabei das gesamte Videomaterial auf einmal in einem Stapelverarbeitungsprozess. Diese Methoden sind jedoch für die meisten Echtzeit-Anwendungen, wie beispielsweise ADAS, ungeeignet. In solchen Anwendungen ist es entscheidend, den Zustand von Objekten unmittelbar nach neuen Erkennungen vorherzusagen. Daher setzen die meisten neueren Ansätze zur Verfolgung mehrerer Objekte auf Online-Methoden, die nicht auf zukünftige Bildinformationen angewiesen sind [xx, ...]. Online-Methoden verwenden verschiedene Merkmale, um die Ähnlichkeit zwischen den erkannten Objekten und den existierenden Spuren zu schätzen. Dies kann auf Grundlage von vorhergesagten Positionen oder Ähnlichkeiten im Erscheinungsbild geschehen [xx, ...]. Während einige Ansätze [xx, ...] nur die jüngste Erkennung, die einer Spur entspricht, berücksichtigen, integrieren andere Methoden zeitliche Informationen in eine Spurhistorie. Verfahren nutzen beispielsweise rekurrente neuronale Netze, um zeitliche Informationen zu aggregieren [xx, ...]. Wie von Mertz et al. [xx] wurde auch in dieser Arbeit das Ziel verfolgt einen datenbasierten Ansatz zu entwickeln, welcher lernen kann, das kombinatorische Non Deterministic Polynomial Time (NP) hard Optimierungsproblem der Datenassoziation vollständig zu lösen. Mertz et al. [xx] nutzen als Inputdaten für das entwickelte DA Netzwerk eine Distanzmatrix auf Basis des euklidischen Abstandsmaßes, und ersetzt somit einen Assoziationsalgorithmus wie z.B. den Hungarian Algorithm (HA). Es ist anzunehmen, dass bei der Erstellung der Groundtruth (GT) Trainingsdaten (Distanzmatrizen), sowie bei der Evaluierung, das euklidische Abstandsmaß als Basisberechnung verwendet wurde. Dies ist jedoch nicht explizit aufgeführt. Es lässt sich somit die Behauptung aufstellen, dass durch diesen Berechnungsschritt dem Netzwerk die Möglichkeit genommen wird, einer anderen Assoziationslogik zu folgen bzw. diese datenbasiert zu lernen.

Im Rahmen dieser Arbeit wurde daher die These aufgestellt, dass durch ein nicht definiertes Abstandsmaß ein Gated Recurrent Unit (GRU) basiertes Assoziationsnetzwerk die Zuordnung von der zeitlichen Speicherkomponente und somit von der Historie verstärkt gebildet werden kann. In dieser Arbeit wurde daher das Ziel verfolgt ein Assoziationsnetzwerk zu entwickeln, welches die Zuordnung eines oder mehrerer SO zu einer bestehenden Anzahl an Tracks ohne definiertes Abstandsmaß lösen soll.

III. Tracking with Prediction and Association Networks

Wir wenden das Paradigma des Tracking-by-Detection (TbD) an, bei dem ein Tracker die Objekterkennungen fusioniert, um Objektspuren zu erzeugen, die über die Zeit konsistent sind. Ba-Tuong Vo et al. [xx] stellt beispielsweise einen Framework für die Untersuchung von Tracking Ansätzen zur Verfügung, welche dem TbD Paradigma folgen. Wir schlagen einen Tracker vor mit jeweils einem Long Short-Term Memory (LSTM) Netzwerk für die Prädiktion und Assoziation der Sensorobjekte (SO) zu den bestehenden Tracks.

Das vorgeschlagene Sinlge Prediction Network (SPENT) verarbeitet die Erkennungen in einem zeitlichen Fenster, das aus Objektmessungen, wie Position, Objektdimension, relativer Geschwindigkeit und Objekttyp besteht und sagt einen festdimensionalen Zustandsvektor für jede Erkennung voraus. Die ausgegebenen Merkmale dienen dem Single Association Network (SANT) als Input, um Zieltrajektorien zu bilden. Siehe Fig. 1 für einen Überblick über unseren Ansatz. Im weiteren Verlauf dieses Abschnitts werden wir die beiden vorgeschlagenen Module im Detail erläutern.

A. Single Prediction Network (SPENT)

Die meisten bestehenden Verfolgungsmethoden verknüpfen eingehende Erkennungen paarweise mit Objektzuständen, die durch ein einfaches Bewegungsmodell, z. B. ein Modell mit konstanter Geschwindigkeit, unter Verwendung eines Kalman-Filters vorhergesagt werden. Neuere Arbeiten haben jedoch gezeigt, dass die Aggregation zeitlicher Informationen sowie von Kontextinformationen die Verfolgung mehrerer Objekte verbessern kann, indem zusätzlich zu den paarweisen Ähnlichkeiten zwischen den Erkennungen Informationen höherer Ordnung genutzt werden [xx]. Unserer Ansatz sieht es vor, auf die Bewegungsmodelle zu verzichten und stattdessen die Hiddenstates der LSTM Schicht als objektspezifisches Parameterset zu nutzen. Die initialen Werte der Hiddenstates der LSTM Schicht werden im Tracking anhand der erhaltenen Messdaten aktualisiert. Durch diese Aktualisierung erfolgt somit eine interne Korrektur über den Sequenzverlauf.

...

IV. Experimental Evaluation

... KITTI-Car Benchmark.

V. Conclusion

The conclusion goes here.

Acknowledgments

This should be a simple paragraph before the References to thank those individuals and institutions who have supported your work on this article.

Appendix
Proof of the Zonklar Equations
Proof of the First Zonklar Equation

Appendix goes here.

PROOF OF THE SECOND ZONKLAR EQUATION

And here.

REFERENCES

[1] G. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, no. 8, pp. 114–117, 1965.