

Masterthesis  
Fachhochschul-Studiengang  
Master Informatik

# Implementierung eines Verfahrens zur automatisierten Verkehrsmittelerkennung

Transportation Mode Detection

ausgeführt von

**Michael Zangerle, BSc**  
**1310249004**

zur Erlangung des akademischen Grades  
Master of Science in Engineering, MSc

Dornbirn, im Juli 2015

**Betreuer: Prof. (FH) DI Thomas Feilhauer**



# Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich vorliegende Masterthesis selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Stellen sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder in gleicher noch in ähnlicher Form einer anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Dornbirn, am 29. Juli 2015

Michael Zangerle, BSc



---

# Zusammenfassung

---

Zusammen  
ergänzen



---

# Abstract

---

Abstract e  
zen





# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Ziele dieser Arbeit . . . . .	1
1.2	Motivation und Nutzen . . . . .	2
1.3	State of the art . . . . .	3
1.3.1	Daten . . . . .	4
1.3.2	Verkehrsmittel . . . . .	4
1.4	Weiterer Aufbau der Arbeit . . . . .	4
<b>2</b>	<b>Daten</b>	<b>7</b>
2.1	Trainingsdaten . . . . .	8
2.1.1	Akquirierung . . . . .	8
2.1.2	Struktur . . . . .	8
2.1.3	Entscheidungsbaum . . . . .	8
2.2	Neue Aufzeichnungen . . . . .	8
2.2.1	Akquirierung . . . . .	8
2.2.2	Struktur . . . . .	8
2.3	GIS-Daten . . . . .	8
2.3.1	Akquirierung . . . . .	8
2.3.2	Struktur . . . . .	8
<b>3</b>	<b>Der Prototyp</b>	<b>9</b>
3.1	Funktionalitäten . . . . .	10
3.2	Aufbau und Architektur . . . . .	10
<b>4</b>	<b>Filter</b>	<b>11</b>
4.1	Geschwindigkeitsfilter . . . . .	12
4.2	Höhenfilter . . . . .	12
4.3	Distanzfilter . . . . .	12
4.4	Aufbereitung ohne GIS-Daten . . . . .	13
4.4.1	Auswahl . . . . .	13
4.5	Aufbereitung mit GIS-Daten . . . . .	14
4.5.1	Auswahl . . . . .	14

4.5.2	Abstand zu Bushaltestellen . . . . .	14
4.5.3	Abstand zu Gleisen . . . . .	14
4.6	Segmentierung . . . . .	15
4.6.1	Terminologie . . . . .	15
<b>5</b>	<b>Analyse</b>	<b>17</b>
5.1	Entscheidungsbaum . . . . .	18
5.2	Nachbearbeitung . . . . .	18
<b>6</b>	<b>Ausblick</b>	<b>19</b>
	<b>Literaturverzeichnis</b>	<b>21</b>
	<b>Quellcodeverzeichnis</b>	<b>23</b>
	<b>Tabellenverzeichnis</b>	<b>25</b>
	<b>Abbilungsverzeichnis</b>	<b>27</b>
	<b>Anhang</b>	<b>29</b>

## Todo list

Zusammenfassung ergänzen . . . . .	i
Abstract ergänzen . . . . .	iii
Auswertung Überblick . . . . .	5
vervollständigen der einleitung für den hauptteil daten . . . . .	7



# Einleitung

Der Mensch produziert täglich eine extrem große Menge an Daten. Allein auf Youtube werden pro Minute 300 Stunden Video-Material veröffentlicht und täglich hunderte Millionen Stunden von Videos konsumiert. [Youtube, 2015] Hochgerechnet auf das gesamte Internet und die gesamte Bevölkerung ergibt dies eine unvorstellbar große Menge an Daten die bewusst oder auch unbewusst generiert werden.

Sehr viel an Daten wird auch durch diverse Fitnessgadgets, Smartwatches, Smartphones sowie Navigations-Geräten und Ähnlichem generiert. Abseits von Fitnesswerten sind all diese Geräte im Stande GPS-Spuren aufzuzeichnen. Dies bedeutet, dass man genau nachvollziehen kann, wann man wo unterwegs war. Mit ein wenig Rechenarbeit kann man auch die Geschwindigkeit und viele andere Werte berechnen sofern dies diese Geräte nicht schon selbst machen.

Genau auf diesen GPS-Daten basiert diese Arbeit. Dabei ist es nicht wichtig von welcher Person diese Daten stammen sondern, dass sich mit Hilfe dieser aufgezeichneten Daten feststellen lässt, wann ein Individuum sich auf welcher Strecke mit welchem Verkehrsmittel fortbewegt hat.

## 1.1 Ziele dieser Arbeit

Das Hauptziel dieser Arbeit ist es, einen Prototypen zu erstellen, welcher anhand von aufgezeichneten GPS-Spuren und in Kombination mit verschiedenen Methodiken versucht das benutzte Verkehrsmittel mit einer möglichst hohen Wahrscheinlichkeit zu bestimmen. Diese aufgezeichneten GPS-Spuren enthalten dabei keinerlei Informationen über die jeweilige Person. Deshalb erfolgt die Auswertung ausschließlich über die GPS-Spur sowie über öffentlich zugängliche Daten wie zum Beispiel Busstationen und Gleise. Jede dieser Aufzeichnung kann mehrere Verkehrsmittel beinhalten und von unterschiedlicher Länge und Dauer sein.

Es soll weiters, für jede Person die ein Gerät besitzt welches in der Lage ist eine GPS-Spur im GPX-Format aufzuzeichnen, möglich sein, diese Spur analysieren zu lassen. Dies bedeutet, dass keine speziellen Geräte oder andere Sensoren benötigt werden und dass sich dieser Prototyp mit möglichst geringen Anpassungen auch auf andere Regionen anwenden lässt. Zum Aufzeichnen der in dieser Arbeit verwendeten GPS-Spuren, wurde mehrere Smartphones mit der App “MyTrack“ sowie mehrere GPS-Geräte benutzt.

Im Zuge dieser Arbeit wird auch untersucht welches Level an Genauigkeit sowohl mit als auch ohne geografischen Zusatzinformationen im Raum Vorarlberg erreicht werden kann. Dabei soll es nach der Analyse die Möglichkeit geben, die automatisch bestimmten Verkehrsmittel manuell zu korrigieren sollten diese nicht mit der Realität übereinstimmen. Weiters sollen diese manuellen Änderungen in die Auswertung mit einfließen.

Schlussendlich soll mit Hilfe der Auswertungen auch eine Aussage über die erzielte Genauigkeit mit und ohne den verwendeten Zusatzinformationen gemacht werden. Außerdem soll eine Aussage darüber getroffen werden können, ob noch weitere Zusatzdaten für eine noch genauere Bestimmung benötigt werden würden und welche Daten dies sein könnten.

### **Nichtziele**

In dieser Arbeit werden die diverse Schlussfolgerungsmodelle nicht betrachtet oder verglichen sondern es wird auf ein von anderen Arbeiten vielversprechendes Modell gesetzt. Dieses Modell ist der Entscheidungsbaum. Außerdem werden die Daten zur Analyse bzw. Auswertung nicht in Echtzeit betrachtet sondern in Form einer GPS-Spur an den Prototypen übergeben.

Die Frage, an welcher Position man das Gerät zur Aufzeichnung am besten trägt um möglichst genaue GPS-Daten zu erhalten, sowie die Frage nach dem Energiverbrauch der App bzw. wie eine möglichst energieschonende App und die dazugehörige Kommunikation aufgebaut sein könnte bleibt unberücksichtigt. Auch eine umfassende Behandlung der Themen Sicherheit und Privatsphäre würden den Rahmen der vorliegenden Arbeit sprengen und bleibt daher unberührt.

## **1.2 Motivation und Nutzen**

Schon früher wurde versucht Aufzeichnungen über die Verkehrswege von verschiedenen Menschen zu sammeln. Aber die Protokollen in Papierform sowie die Telefonbefragungen waren zu aufwändig und die Menschen nicht zuverlässig genug. Darum ist es von entscheidendem Vorteil eine App oder ein Gerät zur Verfügung zu haben, welches die Vorgänge des Aufzeichnens möglichst genau für einen übernimmt. [Zheng et al., 2010]

Wird eine Auswertung mit einer für das Zielgebiet aussagekräftigen Anzahl an Personen durchgeführt, so kann das Resultat für verschiedenste Zwecke verwendet werden. Auch ohne Analyse kann rein durch die Betrachtung der gesammelten GPS-Spuren festgestellt werden, welche Routen besonders häufig benutzt werden.

Zieht man nun verschiedene Werte aus der Auswertung hinzu kann auch festgestellt werden wo sich zum Beispiel verkehrstechnische Engstellen befinden und welche Routen sehr populär sind oder Aussagen über die allgemeine Verkehrssituation machen. Durch die gesammelten Daten könnten sich auch Simulationen für anstehende Bauvorhaben machen lassen und auch versucht werden eine Vorhersage für bestimmte Situationen zu tätigen. Diese Aspekte können unter anderem für das Verkehrsministerium, den öffentlichen Personennahverkehr oder auch für die Stadtplanung sehr interessant sein.

Eine ganze Reihe von Apps lässt sich mit den Auswertungen erstellen. Diese Apps könnten die Auswertungen in soziale Medien zu integrieren, für Fitnessanalysen verwendet werden, einen einfachen Rückblick über die eigene Fortbewegung ermöglichen oder für Umweltbewusste errechnen wie viel CO<sub>2</sub> sie produziert oder gespart haben. Ein Reisetagebuch könnte daraus genau so Nutzen ziehen wie eine App die beim Autofahren Auskunft über die aktuell billigste Tankstelle in näherer Umgebung gibt oder eine App die einfach nur Vorschläge für alternative Routen zu einem bekannten Ziel anbietet.

Zusammenfassend kann man sagen, dass die Verwendungsmöglichkeiten für solche Daten umfangreich sind und sich am besten unter den Begriffen kontextorientierte, geographische Apps zusammenfassen lassen. Nicht zuletzt öffnen sich mit solchen Daten aber auch umfangreiche Möglichkeiten für die Werbebranche.

### 1.3 State of the art

Zu den Meilensteinen auf diesem Gebiet der Forschung zählt sicher die Arbeit von Yu Zheng in welcher er unter anderem auf die Erkennung von den Abschnitten mit nur einem Verkehrsmittel eingeht. Weiters verwendete er in seiner Arbeit auch einen mit dem von den GPS-Spuren gesammelten geographischen Wissen aufgebauten Graphen welcher zur weiteren Auswertung verwendet wurde. [Zheng et al., 2010]

Mit der Frage wie geographische Daten in eine solche Analyse miteinbezogen werden können hat sich auch Leon Stenneth beschäftigt. Dabei hat er nicht nur fixe Daten wie Gleise und Busstationen sondern auch aktuelle Buspositionen miteinbezogen. [Stenneth et al., 2011]

Sowohl Stenneth als auch Zheng haben in ihren Arbeiten detailliert erklärt wieso sie welche Attribute (Geschwindigkeit, Beschleunigung, ...) für die Bestimmung des Verkehrsmittels verwendet haben und sie haben diese auch durch Versuche nach ihrer Wichtigkeit gereiht.

Außerdem haben beide und auch Sasank Reddy mehrere Schlussfolgerungsmodelle (Entscheidungsbaum, Bayessches Netz, Markov Modelle, Random Forest, ...) betrachtet und miteinander verglichen. [Reddy et al., 2010, Stenneth et al., 2011, Zheng et al., 2010]

Wie man mit Verbindungsabbrüchen umgeht und zwischen ähnlichen Verkehrsmitteln unterscheiden kann hat unter anderem auch Filip Biljecki beschäftigt. Desweiteren baut er für die unterschiedlichen Kategorien von Verkehrsmittel ein hierarchisches Modell auf, welches ihm helfen soll bessere Entscheidungen zu treffen. [Biljecki et al., 2013]

### 1.3.1 Daten

Ein wesentlicher Unterschied zwischen all den betrachteten Publikationen sind die verwendeten Daten. Einzig die GPS-Spuren bilden eine gemeinsame Basis. Manche untersuchten die Verwendung von GSM- und WIFI-Informationen [Reddy et al., 2010], stützten sich auf Zusatzinformationen durch weitere Sensoren wie zum Beispiel ein Beschleunigungssensor [Reddy et al., 2010, Nadine Schüssler et al., 2011]. Andere wie Leon Stenneth verwendeten Live-Informationen von den öffentlichen Verkehrsmittel und kombinierten diese mit GIS-Informationen, seien es Busstationen, Bahnstrecken, das Straßennetz oder Parkplätze [Stenneth et al., 2011].

### 1.3.2 Verkehrsmittel

Ein weiterer Unterschied zwischen den Publikationen sind die betrachteten Verkehrsmittel. Hierbei reicht die Spanne der unterschiedenen Verkehrsmittel von "Gehen und Motorisiert"(siehe [Reddy et al., 2010]) bis hin zu "Gehen, Zug, U-Bahn, Rad, Auto, Straßenbahn, Bus, Fähre, Segelboot und Flugzeug" (siehe [Biljecki et al., 2013]).

## 1.4 Weiterer Aufbau der Arbeit

Der Hauptteil der vorliegenden Arbeit gliedert sich in fünf große Abschnitte:

Im ersten wird auf die Akquirierung der GPS-Daten eingegangen. Dies umfasst sowohl die gesammelt GPS-Spuren und deren Struktur, sowie die verwendeten GIS-Daten. Dabei geht es einerseits um deren Herkunft als auch darum wie diese extrahiert wurden und wie auch diese Daten aufgebaut sind. Weiters wird ein Überblick über den entstandenen Prototypen sowie dessen Handhabung und Architektur vermittelt.

Der zweite Abschnitt behandelt den entstandenen Prototypen der die übergebenen GPS-Spuren analysiert. Dabei werden einerseits dessen Funktionalitäten erklärt sowie der



grundlegende Ablauf für den Benutzer dargelegt. Weiters wird auch auf die Architektur des Prototyps sowie auf dessen Konfiguration eingegangen.

Der dritte Abschnitt befasst sich sowohl mit dem Filtern als auch dem Aufbereiten der Daten sowie das Aufteilen von GPS-Spuren in Teile in denen nur ein Verkehrsmittel verwendet wurde. Filtern der Daten bedeutet in diesem Zusammenhang, dass Ausreißer aus den GSP-Spuren entfernt werden. Diese Ausreißer können aufgrund von Geschwindigkeitssprüngen als auch unwahrscheinlich große Sprünge im dreidimensionalen Raum sein. Mit Aufbereiten der Daten ist gemeint, dass Werte zu GPS-Punkten für die spätere Analyse berechnet werden. Diese Werte können sowohl die Geschwindigkeit, Beschleunigung, Distanz, Zeit als auch der Abstand zu der nächsten Bushaltestelle sein. Schlussendlich sollen die gefilterten und erweiterten Daten verwendet werden um eine einfache und sichere Aufteilung der GPS-Spuren anhand der Verkehrsmittel zu ermöglichen.

Aufbauend auf die Resultate aus dem dritten Abschnitt befasst sich der vierte Abschnitt mit der tatsächlichen Erkennung von der Verkehrsmittel anhand der berechneten Werte und den einzelnen Abschnitten der GPS-Spur. Dabei wird auch genauer auf den Entscheidungsbaum sowie dessen Training eingegangen. Die aus dem Entscheidungsbaum gewonnenen Erkenntnisse werden schlussendlich ein letztes Mal überprüft um sinnfreie bzw. sehr unwahrscheinliche Wechsel zwischen Verkehrsmitteln zu verhindern.

Im fünften und letzten Abschnitt des Hauptteils befasst sich mit der Auswertung der gewonnenen Erkenntnisse aus den vorhergehenden Abschnitten sowie den Testläufen mit neuen GPS-Spuren.

Auswertung  
Überblick



# Daten

Dieser Abschnitt behandelt die Akquirierung und die Struktur der verwendeten Daten für das Training des Entscheidungsbaums, der aufgezeichneten GPS-Spuren sowie der verwendeten GIS-Daten. Im Zusammenhang mit dem Aufzeichnen der GPS-Spuren wird auch dargelegt welches Smartphone und welche App dazu verwendet wurden.

## 2.1 Trainingsdaten

Für die Erstellung der Schlussfolgerungsmodelle konnten GPS-Aufzeichnungen aus einem Projekt von Sebastian Nagel "Möglichkeitsstudie zum Projekt: Mobilitäts-Tracker" verwendet werden. Diese Daten wurden mit verschiedenen Geräten und mehreren Personen aufgezeichnet und beinhalten alle in dieser Arbeit betrachteten Verkehrsmittel. [Sebastian Nagel, 2011]

1 car 59 0.35542168674698793 2 walk 51 0.3072289156626506 3 bike 43 0.25903614457831325  
4 train 8 0.04819277108433735 5 bus 5 0.030120481927710843

### 2.1.1 Akquirierung

### 2.1.2 Struktur

### 2.1.3 Entscheidungsbaum

## 2.2 Neue Aufzeichnungen

### 2.2.1 Akquirierung

### 2.2.2 Struktur

## 2.3 GIS-Daten

### 2.3.1 Akquirierung

### 2.3.2 Struktur

# Der Prototyp

## 3.1 Funktionalitäten

## 3.2 Aufbau und Architektur

# Filter

## 4.1 Geschwindigkeitsfilter

## 4.2 Höhenfilter

## 4.3 Distanzfilter



## 4.4 Aufbereitung ohne GIS-Daten

### 4.4.1 Auswahl

Durchschnittliche Geschwindigkeit

Maximale Geschwindigkeit

Durchschnittliche Beschleunigung

Maximale Beschleunigung

Höhenmeter

Distanz

## 4.5 Aufbereitung mit GIS-Daten

### 4.5.1 Auswahl

Durchschnittliche Geschwindigkeit

Maximale Geschwindigkeit

Durchschnittliche Beschleunigung

Maximale Beschleunigung

### 4.5.2 Abstand zu Bushaltestellen

### 4.5.3 Abstand zu Gleisen

## 4.6 Segmentierung

### 4.6.1 Terminologie



# Analyse

## 5.1 Entscheidungsbaum

## 5.2 Nachbearbeitung

# Ausblick





# Literaturverzeichnis

- [Biljecki et al., 2013] Biljecki, F., Ledoux, H., and van Oosterom, P. (2013). Transportation mode-based segmentation and classification of movement trajectories. *International Journal of Geographical Information Science*, 27(2):385–407.
- [Nadine Schüssler et al., 2011] Nadine Schüssler, Lara Montini, and Christoph Dobler (2011). Improving post-processing routines for gps observations using prompted-recall data. In *9th International conference on survey methods in transport*.
- [Reddy et al., 2010] Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M., and Srivastava, M. (2010). Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks (TOSN)*, 6(2):13.
- [Sebastian Nagel, 2011] Sebastian Nagel (2011). Möglichkeitsstudie zum Projekt: Mobilitäts-Tracker.
- [Stenneth et al., 2011] Stenneth, L., Wolfson, O., Yu, P. S., and Xu, B. (2011). Transportation mode detection using mobile phones and GIS information. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 54–63. ACM.
- [Youtube, 2015] Youtube (2015). Youtube statistics.
- [Zheng et al., 2010] Zheng, Y., Chen, Y., Li, Q., Xie, X., and Ma, W.-Y. (2010). Understanding transportation modes based on GPS data for web applications. *ACM Transactions on the Web (TWEB)*, 4(1):1.



# Listings



# Tabellenverzeichnis



# Abbildungsverzeichnis





# Anhang