Institut für Telematik
Technische Universität Hamburg

Forschungsprojekt und Seminar

Computer Science

Handgestenerkennung mit Entscheidungsbäumen

von

Tom Dymel

Februar 2021

Betreut von

Dr. Venzke
Institute of Telematics, Hamburg University of Technology

Erstprüfer	Prof. Dr. Volker Turau Institute of Telematics Hamburg University of Technology
Zweitprüfer	Dr. Venzke Institute of Telematics Hamburg University of Technology

Inhaltsverzeichnis

Verwendete Symbole iii

1 Einleitung 1

2 Techniques and Methods 3
- 2.1 Decision Trees . . . 3
- 2.2 Ensemble . . . 3
  - 2.2.1 Voting . . . 3
  - 2.2.2 Bagging . . . 3
  - 2.2.3 Random Forest . . . 3
  - 2.2.4 Boosting? . . . 3
- 2.3 Generation . . . 3
  - 2.3.1 Greedy approximation (CARD?) / (Boosting?) . . . 3
  - 2.3.2 Cherry Picking? . . . 3
  - 2.3.3 Scikit-Learn? . . . 3

3 State of the Art 5
- 3.1 Gesture candidate . . . 5
- 3.2 Feed Forward Neural Network . . . 5
  - 3.2.1 Performance . . . 5
  - 3.2.2 Resource usage . . . 5
  - 3.2.3 Optimization . . . 5
- 3.3 Recurrent Neural Network . . . 5
- 3.4 Related Work . . . 5

4 Achievements 7
- 4.1 Infrastructure . . . 7
  - 4.1.1 Parsing gestures . . . 7
  - 4.1.2 Feature extraction . . . 7
  - 4.1.3 Training . . . 7
  - 4.1.4 Arduino . . . 7
  - 4.1.5 Utility Tools . . . 7
  - 4.1.6 Workflow? . . . 7
- 4.2 DymelData . . . 7
  - 4.2.1 Configurations . . . 8
  - 4.2.2 NullGestures . . . 8
  - 4.2.3 Synthetic Brightness Dataset . . . 8
- 4.3 Features . . . 8
  - 4.3.1 Requirements . . . 8
  - 4.3.2 Curse of dimensionality . . . 8
  - 4.3.3 Unpromising features . . . 8

4.3.4 Brightness distribution . . . 8
4.3.5 Motion History . . . 8
4.3.6 Center of Gravity Distribution . . . 8
4.4 Performance Evaluation . . . 9
4.4.1 Testsets . . . 9
4.4.2 Feasible solution . . . 9
4.4.3 Considering NullGestures . . . 9
4.4.4 Brightness Distribution . . . 9
4.4.5 Motion History . . . 9
4.4.6 Brightness Distribution and Motion History . . . 10
4.4.7 Center of Gravity Distribution Float Ansatz . . . 10
4.4.8 Center of Gravity Distribution Integer Ansatz . . . 10
4.4.9 Center of Gravity Distribution Float and Integer Ansatz . . . 10
4.4.10 Comparison to previous work . . . 10
4.5 Execution Time Evaluation . . . 10
4.5.1 WCEP and WCET . . . 10
4.5.2 AVR Compiler for AtMega328p . . . 10
4.5.3 Optimization Level . . . 10
4.5.4 Feature extraction . . . 11
4.5.5 Tree Evaluation . . . 11
4.5.6 Forest Evaluation . . . 11
4.5.7 Total Execution Time . . . 11
4.6 Size Evaluation . . . 11
4.6.1 CCP (TODO: Abrev.) . . . 11
4.6.2 Minimum Leaf Sample Size . . . 11

5 Conclusion **13**

A Content of the DVD **15**

Verwendete Symbole

$\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$	Regular sets of numbers
$\mathcal{NP}, \mathcal{P}$	Complexity classes
$\mathcal{V} = \{v_0, \ldots, v_{N-1}\}$	Set of N nodes v_i belonging to a network with sink v_0
ϱ	Node density, i.e., the average number of nodes within another node's communication range
$(v_i, v_j) \in \mathcal{E}$	Set of bidirectional communication links in the network
$G = (\mathcal{V}, \mathcal{E})$	Graph representation of a wireless sensor network
$\mathcal{N}_i = \{v_j \in \mathcal{V} \mid i \neq j \wedge (v_i, v_j) \in \mathcal{E}\}$	The set of bidirectional communication partners of node v_i
$\mathcal{T} \subseteq \mathcal{E}, \quad \lvert\mathcal{T}\rvert = N-1$	Routing tree rooted in the sink
$\mathcal{T}_i$	Subtree rooted in (and including) node v_i
$\mathcal{C}_i, \quad C_i = \lvert\mathcal{C}_i\rvert$	The set and number of children of node v_i in $\mathcal{T}$
$\mathcal{F} = \{v_i \in \mathcal{V} \mid \mathcal{C}_i = \emptyset\}$	Set of leafs in $\mathcal{T}$ and the number of leafs
$\mathcal{F}_i$	Set of leafs in the subtree $\mathcal{T}_i$ of $\mathcal{T}$

Kapitel 1

Einleitung

Machinelles Lernen (ML) gewann in den vergangenen Jahren an Popularität, u.a. durch die Fortschritte in parallelen Rechnen, sinkende Speicherpreise, schnelleren Speicher und den Zugang zu Bibliotheken, wie zum Beispiel Scikit-Learn, Keras und PyTorch, welche den Einstieg in maschinellen Lernen erleichtern (TODO: Quelle?). Ein namenhaftes Beispiel für das Potential von maschinellen Lernen ist die AlphaGo-KI, die einen Sieg gegen den besten menschlichen Spieler erringen konnte in dem Spiel Go, welches als besonders schwierig für Computer zu meistern galt durch den enormen Suchraum von möglichen Aktionen (TODO: Quelle).

Ein häufiges Anwendungsgebiet in eingebetteten Systemen ist die optische Gestenerkennung, die zur kontaktlosen Interaktion mit technischen Geräten u. a. genutzt wird (Todo: Quelle). Die eingesetzten Micro-Controller sind jedoch häufig nicht ausreichend leistungsstark, um ein trainiertes Model in passabler Zeit auszuführen (TODO: Quelle). Gründe dafür sind Kosten oder Anforderungen an die Batterielanglebigkeit (TODO: Quelle). Eine Lösung ist das Auslagern der Modelle in einen leistungsstarken Rechenkluster, indem die nötigen Eingabedaten auf dem Mikro-Controller gesammelt werden und anschließend an den Rechenkluster gesendet werden (TODO: Quelle). Dies erzeugt allerdings Latenz und eine Abhängigkeit zu einer solchen Infrastruktur. Ein alternativer Ansatz ist das lokale Ausführen der Modelle, was allerdings die Komplexität des Modells limitiert um eine passable Ausführungszeit zu gewährleisten.

In dieser Arbeit wird die Handgestenerkennung mit Entscheidungsbäumen auf dem Arduino-Board ATmega328P untersucht, das mit 9 Fotowiderständen ausgestattet ist, die in einer 3x3 Matrix angeordnet sind. Der ATmega328P verfügt über eine 8-Bit APU, 2 kB RAM, 32 kB Flash-Speicher und operiert unter 16 MHz. Im Vergleich zu vorherigen Arbeiten, die sich mit Künstlichen Neuronalen Netzen (KNN) auseinander gesetzt haben, versprechen Entschei-

dungsbäume einen geringeren Rechenaufwand.

Um dieses Ziel zu erreichen, werden Entscheidungsbaum basierte Klassifizierer auf einem leistungsstarken Computer trainiert und anschließend in eine bestehende Infrastruktur eingebettet, die Gestenkandidaten als Folge von Bildern identifiziert. Es werden insgesamt zwei Komponenten hinzugefügt. Die erste Komponente extrahiert Features, welche an die zweite Komponente, der Klassifizierer, weitergegeben werden um die Handgesten zu erkennen. Der Klassifizierer wird mit den vorherigen Arbeiten verglichen im Hinblick auf Ausführungszeit, Resourcenverbrauch und Erkennungsgenauigkeit unter verschiedenen Verhältnissen, wie Geschwindigkeit, Licht und Entfernung. Zusätzlich wird untersucht inwiefern `Nullgesten`, i. e. invalide Gesten, erkannt werden können und welche Konsequenzen es auf die Erkennungsgenauigkeit hat.

Zunächst werden in Kapitel 2 Entscheidungsbäume eingeführt, verschiedene Ensemble-Methoden erläutert und auf die Generierung von den Modellen eingegangen. In Kapitel 3 wird erläutert, wie Gestenkandidaten extrahiert werden und welche Ansätze vorherige Arbeiten bereits verfolgt hatten. Anschließend präsentiert Kapitel 4 den Kern der Arbeit. Zuallererst wird die Infrastruktur vorgestellt, welche im Rahmen dieser Arbeit angefertigt wurde, um Klassifizierer zu trainieren und zu evaluieren. Anschließend wird das Datenset `Dymel` vorgestellt, was zusätzlich zu dem Datenset von `Klisch` erstellt wurde, um Nullgesten zu und verschiedene Helligkeitsstufen zu untersuchen. Danach werden die Features vorgestellt, die im Laufe dieser Arbeit untersucht wurden. Zuletzt wird die Erkennungsgenauigkeit, die Ausführungsgeschwindigkeit und der Resourcenverbrauch evaluiert. In Kapitel 5 werden die Schlussfolgerungen dargestellt.

Kapitel 2

Techniques and Methods

2.1 Decision Trees

2.2 Ensemble

2.2.1 Voting

2.2.2 Bagging

2.2.3 Random Forest

2.2.4 Boosting?

2.3 Generation

Talk about NP-Complete etc.

2.3.1 Greedy approximation (CARD?) / (Boosting?)

2.3.2 Cherry Picking?

2.3.3 Scikit-Learn?

Kapitel 3

State of the Art

3.1 Gesture candidate

3.2 Feed Forward Neural Network

3.2.1 Performance

3.2.2 Resource usage

3.2.3 Optimization

Compression

Pruning

3.3 Recurrent Neural Network

3.4 Related Work

TODO!

Kapitel 4

Achievements

4.1 Infrastructure

4.1.1 Parsing gestures

4.1.2 Feature extraction

4.1.3 Training

Scikit-Learn

Ensemble

4.1.4 Arduino

Export Tree as C code

Feature extraction

4.1.5 Utility Tools

Gesture Recorder

Playground

Gesture Simulation

4.1.6 Workflow?

4.2 DymelData

TODO: Motivation @ consistent real world evaluation

4.2.1 Configurations

4.2.2 NullGestures

Explain what it is in general

Corner

Same side in and out

Rotations

4.2.3 Synthetic Brightness Dataset

4.3 Features

4.3.1 Requirements

Invariance to velocity

Development over time

Sense of direction

Sense of position

4.3.2 Curse of dimensionality

4.3.3 Unpromising features

Explain that they had been tried out and did work out. Explain why

4.3.4 Brightness distribution

Explain whaat it is. Note this as "discretisized"position. Explain what requirements it fullfills. Explain different variants.

4.3.5 Motion History

Explain origin and how it works. Explain what requirements it fullfills and why.

Implementation

4.3.6 Center of Gravity Distribution

Motivate.?

Center of Gravity

Integer variant

Adding development over time

Window size

Also talk about the squishing?

4.4 Performance Evaluation

Talk about what was searched for, i.e. not all sizes had been tried out. Talk about different testset sizes? Talk about the greedy nature of the training algorithm? (Was that done in Cherry Picking?)

4.4.1 Testsets

Motivate. Explain what they are used for. Explain whats in there?

Metrics

Talk about used metrics

Klisch

DymelNull

DymelGesture

4.4.2 Feasible solution

4.4.3 Considering NullGestures

4.4.4 Brightness Distribution

Show graphs about: Best solution, Best feasible solution, With and WithOUT considering null gestures. Talk when it starts to generalize more poorly(?)

4.4.5 Motion History

Show graphs about: Best solution, Best feasible solution, With and WithOUT considering null gestures. Talk when it starts to generalize more poorly(?)

4.4.6 Brightness Distribution and Motion History

Show graphs about: Best solution, Best feasible solution, With and WithOUT considering null gestures. Talk when it starts to generalize more poorly(?)

4.4.7 Center of Gravity Distribution Float Ansatz

Show graphs about: Best solution, Best feasible solution, With and WithOUT considering null gestures. Talk when it starts to generalize more poorly(?)

Talk about brightness distribution!!

4.4.8 Center of Gravity Distribution Integer Ansatz

Show graphs about: Best solution, Best feasible solution, With and WithOUT considering null gestures. Talk when it starts to generalize more poorly(?)

Talk about brightness distribution!!

4.4.9 Center of Gravity Distribution Float and Integer Ansatz

Show graphs about: Best solution, Best feasible solution, With and WithOUT considering null gestures. Talk when it starts to generalize more poorly(?)

Here state what was expected and state how it worked out! Talk about brightness distribution!! ==> Alternative approach => Stacking (?)

4.4.10 Comparison to previous work

4.5 Execution Time Evaluation

Talk about constrained micro controller. Say that more frames potentially give more insight about fast gestures or that a longer battery life can be achieved. Say that we only consider COCD here, because that has the best feasible solution.

4.5.1 WCEP and WCET

4.5.2 AVR Compiler for AtMega328p

4.5.3 Optimization Level

Say that we considered O2 since the feasible solution can be compiled in O2. And why its a good Kompromiss between O3 and Os

4.5.4 Feature extraction

4.5.5 Tree Evaluation

4.5.6 Forest Evaluation

Here mainly talk about voting.

4.5.7 Total Execution Time

4.6 Size Evaluation

Explain that scikit learn provides some hyperparameters that can be changed. Elaborate which exist. Say that I considered here two parameters because all ensemble methods supported that. Maybe other reasons.

4.6.1 CCP (TODO: Abrev.)

4.6.2 Minimum Leaf Sample Size

Kapitel 5

Conclusion

Do a summary. Should be brief but also name everything and its result more or less. Say the improvement compared to the old approach. (Or downsides).

Mention future work. => Stacking

Content of the DVD

In this chapter, you should explain the content of your DVD.