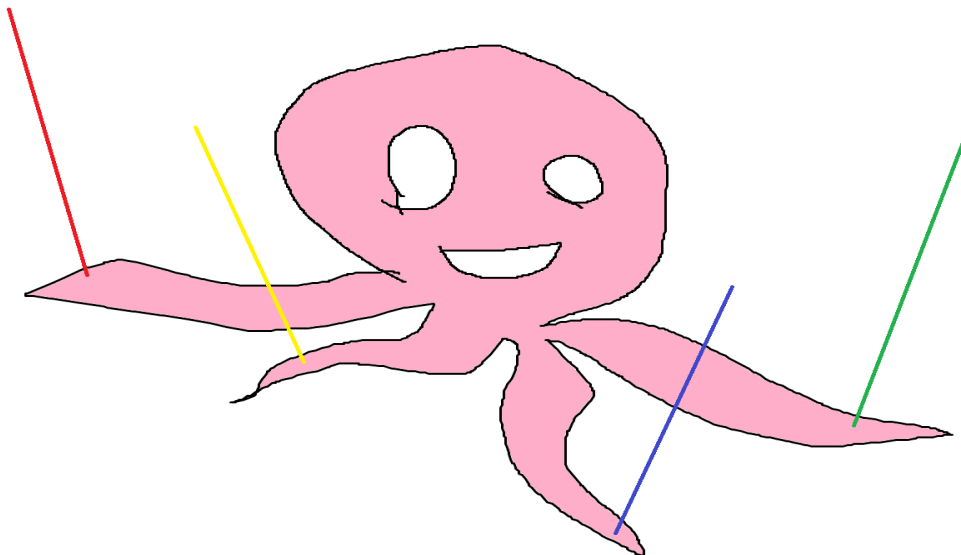


Projektmappe für den COSIMA-Wettbewerb 2018

Der Gestikulaser

Die neue Technologie der Gestenerkennung

Oktober 2018



Verfasst von

Christoph Behr
Cailing Fu
Nicole Grubert
Daniel Wolff

Inhaltsverzeichnis

1. Einleitung	3
2. Stand der Technik	4
3. Aufbau	5
3.1. Oktokommander	6
3.1.1. Transimpedanzverstärker	7
3.2. Detektormodul	8
3.3. Software	8
4. Marketing und Kosten	11
A. Ausblick	13

1. Einleitung

Was wurde gemacht? Was war die Motivation? Inwiefern besitzt das Projekt eine Alltagsrelevanz?

Gestenerkennung ist immer wieder ein Thema, welches viel Aufmerksamkeit erregt. Und obwohl der Mensch so einfach typische Gesten erkennen kann, bleibt es für den Computer eine große Herausforderung, zuverlässig die Gesten eines Menschen zuzuordnen.

Mit Gestikulaser wollen wir ein neues System entwickeln, um die Handgesten eines Menschen zu erkennen und diese individuell auf den Nutzer ab zu stimmen. Dabei soll es nicht, wie die meisten heute typischen Systeme, mit einer Kamera arbeiten, sondern durch Lichtreflexionen einer Hand die Geste zuordnen. Denn damit ist die Gestenerkennung nicht nur Tageslicht unabhängig, sondern kann auch in vollkommener Dunkelheit betrieben werden.

Mit Hilfe eines modularen Stecksystems mit mehreren Komponenten soll es möglich sein, ein individuelles Muster des Systems an zu fertigen. Somit soll in Zukunft nicht nur die Handgesten-Steuerung möglich sein, sondern auch eine ganzkörper Gestensteuerung wie sie in einem Cave verwendet wird ermöglicht werden.

Im Alltag soll der Gestikulaser dann eingesetzt werden, um verschiedenste Dinge, wie ein ferngesteuertes Auto, eine Smart Home Einrichtung oder eine Drohne zu steuern.

2. Stand der Technik

Wird vielleicht gelöscht. Außer irgendjemand hat Lust zu recherchieren.

3. Aufbau

Die Funktionsweise des Gestikulasers besteht aus der Detektion und Weiterverarbeitung von Lichtsignalen, welche mit einer Geste erzeugt werden. Dabei wird infrarotes Licht von einer LED Quelle durch die Hand reflektiert und mit Hilfe von mehreren Photodioden detektiert. Die durch die Photodioden erhaltenen Daten werden dann in einer Software weiterverarbeitet, welche mit Hilfe von Machine Learning die tatsächliche Handgeste erkennt. Von dort aus kann dann jedes beliebige Endgerät angesteuert werden.

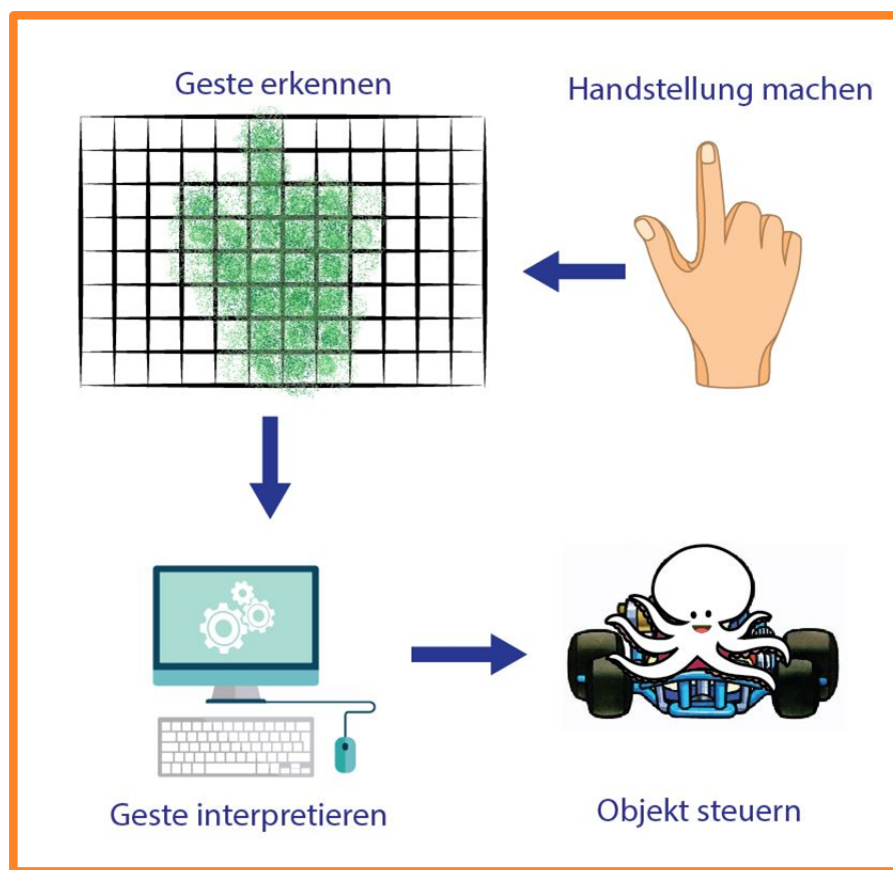


Abbildung 3.1.: Erklärung!

Der Gestikulaser selbst besteht aus einer Platte, der Photoplatte, welche aus verschiedenen Steckmodulen zusammen gesteckt werden kann. Die Steckmodule bestehen aus einzelnen kleinen Boxen, in welche die Elektronik integriert ist. In der Mitte befindet sich der Oktokommander, welcher durch weitere Detektormodule erweitert werden kann. Auf jedes der Module befinden sich vier Photodioden um das reflektierte Licht zu messen.

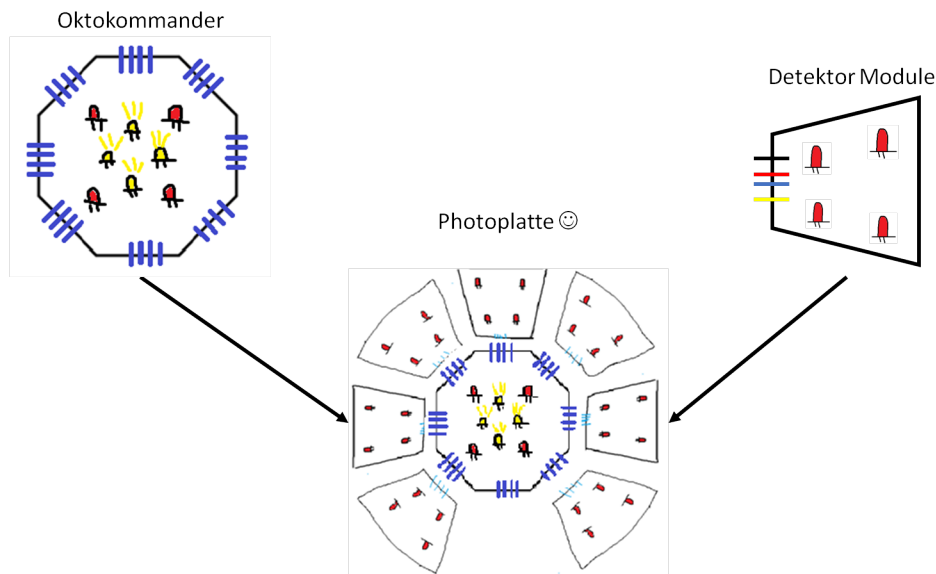


Abbildung 3.2.: Erklärung!

3.1. Oktokommander

Der Oktokommander ist das Steuersystem der gesamten Photoplatte. In ihr befinden sich ein Arduino Micro, welcher die Signale aller Photodioden bearbeitet sowie einen i2c Expander. Durch den i2c Expander ist es möglich, jedes der Photodioden eine eigene Adresse zu zu weisen. Dies ist nötig, um die voneinander unabhängigen Signale der Photodioden richtig zuzuordnen zu können. Zusätzlich dazu befinden sich auf dem Oktokommander vier Infrarot LEDs sowie vier infrarot Photodioden. Um die vier infrarot Photodioden an zu steuern wird zusätzlich zum i2c Expander noch ein i2c Multiplexer sowie die dazu gehörende Verstärkerschaltung (siehe auch Abschnitt ??) benötigt.

Erklärung der Komponenten an Hand des Bildes

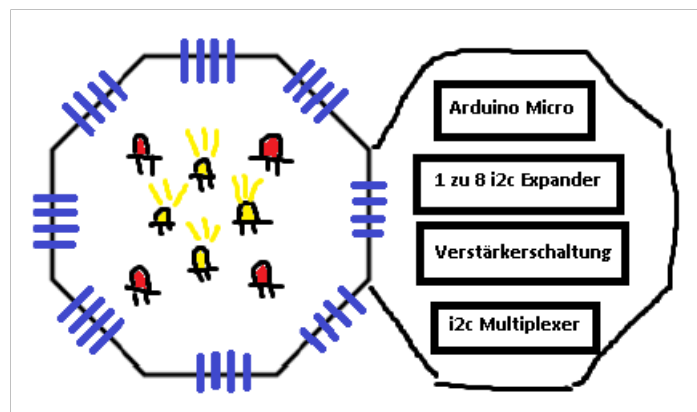


Abbildung 3.3.: Erklärung!

Um den Oktokommander mit den Detektormodulen erweitern zu können wurden USB-

2A Schnittstellen verwendet. Durch diese Schnittstellen können bis zu 7 Detektormodule, jeweils eines pro Seite eingekoppelt werden. Um eine größere Platte zu konstruieren können noch weitere Detektormodule an den bereits vorhandenen Detektormodule angeschlossen werden. **Erklärung der Komponenten an Hand des Bildes**

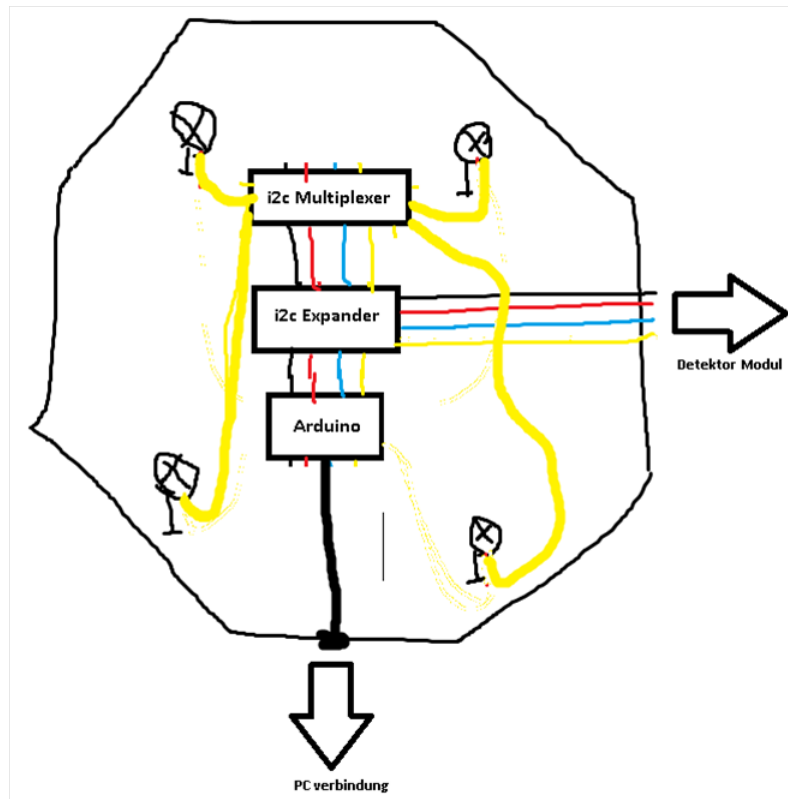


Abbildung 3.4.: **Erklärung!**

3.1.1. Transimpedanzverstärker

Die Verstärkerschaltung besteht aus zwei sogenannten Transimpedanz Verstärkern. Die Photodiode fungiert in ihm als Konstantstromquelle, die einen kleinen Strom an das System abgibt. Mit Hilfe des Widerstandes R_1 und des Kondensators C_1 kann der Operationsverstärker (LM358) eine vom Mikrocontroller verarbeitbare hohe Spannung ausgeben. Wichtig ist hier die möglichst genaue Wahl der Kapazität C_1 ohne die das System anfangen würde zu Schwingen. Es ergibt sich:

$$C_1 = \sqrt{\frac{C_I}{R_1 \cdot GBP}}$$

Hierbei steht GBP für *Gain Bandwidth Product*. Für den Operationsverstärker LM358 ist das GBP mit 1 MHz im Datenblatt angegeben. C_I ist die Summe der beiden Sperrschichtkapazitäten der Photodiode und des Operationsverstärkers. Typische Anwendungsgrößen für R_1 liegen bei etwa 1 MΩ. Daraus lässt sich die Bandbreite des Systems

berechnen:

$$f_{3\text{dB}} = \sqrt{\frac{GBP}{2\pi \cdot R_1 \cdot C_I}}$$

Für unser System liegt die Bandbreite bei etwa 73 kHz.

Abbildung 3.5.: Schematische Darstellung des Transimpedanzverstärkersystems

3.2. Detektormodul

Das Detektormodul besteht lediglich aus vier Photodioden und einem USB-2A Eingang. Dadurch kann das Detektormodul an den Oktokommander angeschlossen werden. Zur Ansteuerung der Photodioden wurde auch hier ein i2c Expander verwendet. **Erklärung der Komponenten der Bilder**

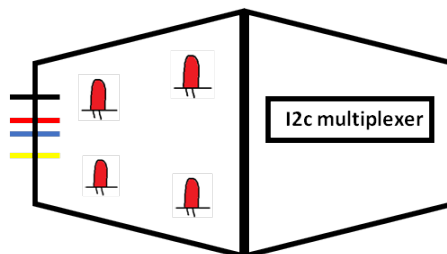


Abbildung 3.6.: **Erklärung!**

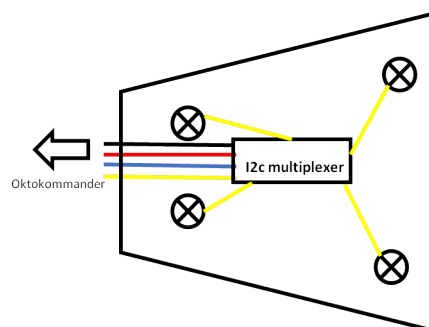


Abbildung 3.7.: **Erklärung!**

3.3. Software

Die Software-Seite des Gestikulasers ist für die Zuordnung der Gesten zuständig. Die von den Detektormodulen gemessenen Reflektionsmuster werden an den Mikrocontroller im

Oktokommander und von da aus an einen Computer weitergeleitet, wo sie verarbeitet werden. Bei der dazu eingesetzten Software wurde komplett auf open-source verfügbare Programme und Bibliotheken gesetzt. Der Code auf den Microcontrollern wurde mit Hilfe der **Arduino IDE** entwickelt. Die Verarbeitung am Computer erfolgt mit Hilfe von **Python** Skripten. Für den Machine Learning Teil wurde die **TensorFlowTM** Bibliothek verwendet. Die Software kann in zwei Teile unterteilt werden, die im Folgenden genauer erläutert werden.

Trainingsphase

Der erste Teil der Software kommt während der Trainingsphase zum Einsatz. Hier werden die von den Photodioden gemessenen Daten aus dem Microcontroller im Oktokommander zunächst ausgelesen und gemeinsam mit einem Label, das der gerade aufgenommenen Geste entspricht, in eine Datei geschrieben. Die auf diese Weise gesammelten Daten werden dann im nächsten Schritt verwendet, um ein mathematisches Modell zu erstellen, das im Live-Betrieb angewendet wird, um die vom Nutzer gemachte Geste, einer der bekannten zuordnen zu können. Angenommen es stehen Daten von N Photodioden zur Verfügung, dann können diese Daten in dem Vektor $x \in \mathbb{R}^N$ zusammengefasst werden. Für den Anfang soll ein Modell entwickelt werden, welches einem beliebigen gemessenen Datensatz x eine der m vordefinierten Gesten G_1, \dots, G_m zuordnet. Dazu stellt der Nutzer vor Aufzeichnung der Daten in dem Python Skript ein, was für eine Geste G_j er als nächstes aufnehmen will. Nachdem für jede Geste ausreichend Daten aufgenommen wurden, kann als nächstes das mathematische Modell erstellt werden, um die Sensordaten zu verarbeiten. Ziel ist es, ein gegebenes Datum x_i , das im Live-Betrieb gemessen wird, eindeutig einer Geste G_j zuzuordnen. Damit Datum x_i der Geste G_j zugeordnet wird, muss die Bedingung

$$p(G_j|x_i) \geq p(G_k|x_i) \quad \forall k = 1, \dots, m \quad k \neq j$$

erfüllt sein. Dabei ist $p(G_j|x_i)$ die bedingte Wahrscheinlichkeit für Geste G_j gegeben die Photomessdaten x_i . Diese Bedingung lässt sich leicht überprüfen, allerdings ist die bedingte Wahrscheinlichkeit $p(G|x)$ unbekannt. Aus diesem Grund wird ein neuronales Netz mit Hilfe der zuvor aufgezeichneten Daten $(x, G) \in \mathbb{R}^N \times \mathbb{R}$ trainiert, das die unbekannte bedingte Wahrscheinlichkeit annähert. Der Aufbau und das Training des neuronalen Netzes erfolgen mit Hilfe vordefinierter Methoden von **TensorFlowTM**. Nach erfolgreich abgeschlossenem Training wird das neuronale Netz in eine Datei exportiert und kann nun im Live-Betrieb verwendet werden.

Live-Betrieb

Im Live-Betrieb wird das zuvor trainierte Modell nun verwendet, um die nun unbekannten Reflektionsmuster, die von den Photodioden gemessen und über den Mikrocontroller an den PC weitergeleitet wurden, einer bekannten Geste zuzuordnen. Nachdem die gemessenen Daten einer Geste zugeordnet wurden, wird der mit dieser Geste verknüpfte Steuerungsbefehl bestimmt und an das angeschlossene Endgerät weitergeleitet. Dieser Ablauf ist in Abbildung 3.8 dargestellt.

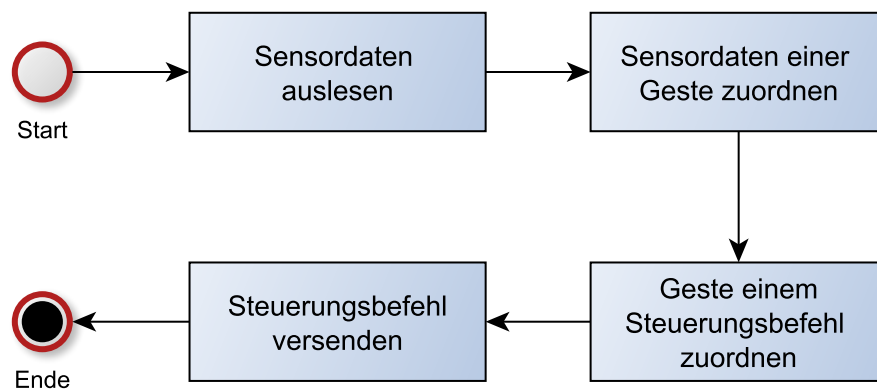


Abbildung 3.8.: Ablaufdiagramm der Software im Live-Betrieb: Die aus dem Microcontroller ausgelesenen Sensordaten werden zur Verarbeitung an das Python Skript übergeben, wo sie einer Geste und diese einem Steuerbefehl zugeordnet wird. Dieser Steuerungsbefehl wird schließlich an das anzustuernde Endgerät weitergegeben.

4. Marketing und Kosten

Neben der Entwicklung unseres Produktes wurde im Laufe der Entwicklungszeit zusätzliche Öffentlichkeitsarbeiten geleistet. So wurde die Website <https://www.gestikulas.de/> ins Leben gerufen sowie der Instagram Account **laserharptos** mit über 100 Abonnenten betreut. Im Zuge dessen wurde auch Sponsoren angeworben, welche Interesse an der Unterstützung unseres Projektes hatten.

Folgende Sponsoren haben uns bei diesem Projekt unterstützt:

Aconity3D GmbH



Fraunhofer ILT



TOS RWTH Aachen University



Würth Elektronik GmbH & Co. KG



Die Produktionskosten unseres Gestikulasers können wie folgt aufgliedert werden:

Oktokommander

Produktname	Kosten /Stk	Anzahl	Gesamtkosten	Kostenträger
Arduino Micro	19.99 €	1	19.99 €	TOS
I2C-Multiplexer TCA9548A	7.80 €	1	7.80 €	ILT
I2C ADS1015	11.20 €	1	11.20 €	ILT
Infrarot LED		4		Würth
LED Treiber		2		Würth
Infrarot Photodiode		4		ILT
OP-Verstärker	1.85 €	4	7.40 €	TOS
UBS 2 TypA -Mount		7		Würth
Passive Bauteile	———	—	1 €	TOS
			60 €	

Detektormodul

Produktname	Kosten /Stk	Anzahl	Gesamtkosten	Kostenträger
I2C ADS1015	11.20 €	1	11.20 €	ILT
Infrarot Photodiode		4		ILT
OP-Verstärker	1.85 €	4	7.40 €	TOS
UBS 2 TypA -Plug		1		Würth
UBS 2 TypA -Mount		1		Würth
Passive Bauteile	——	-	1 €	TOS
			20 €	

Dabei wurde ein Oktokommander und sieben Detektormodule aufgebaut. Die Gesamtkosten für die Produktion belaufen sich also auf 200€.

Darüber hinaus sind Kosten von BLA € für Unterkunft und Anreise entstanden sowie T-Shirts in Wert von BLA €. Die T-Shirts wurden von Aconity3D gesponsert.

A. Ausblick

Erweiterung um Sensorhandschuh zur verbesserten Erkennung der Daten. Überarbeitung des Designs um die Module besser bzw. anders zusammen zu stecken.

An dieser Stelle möchten wir einen Ausblick geben, wie der von uns entwickelte Gestikulaser weiterentwickelt werden könnte.

Hier wäre zunächst einmal eine Verfeinerung der aktuell durchführbaren Gestenerkennung denkbar, die nicht nur die Stellung der Hand, sondern auch die Krümmung der einzelnen Finger berücksichtigt. Zu diesem Zweck wurde bereits ein Prototyp für einen Sensorhandschuh entwickelt.

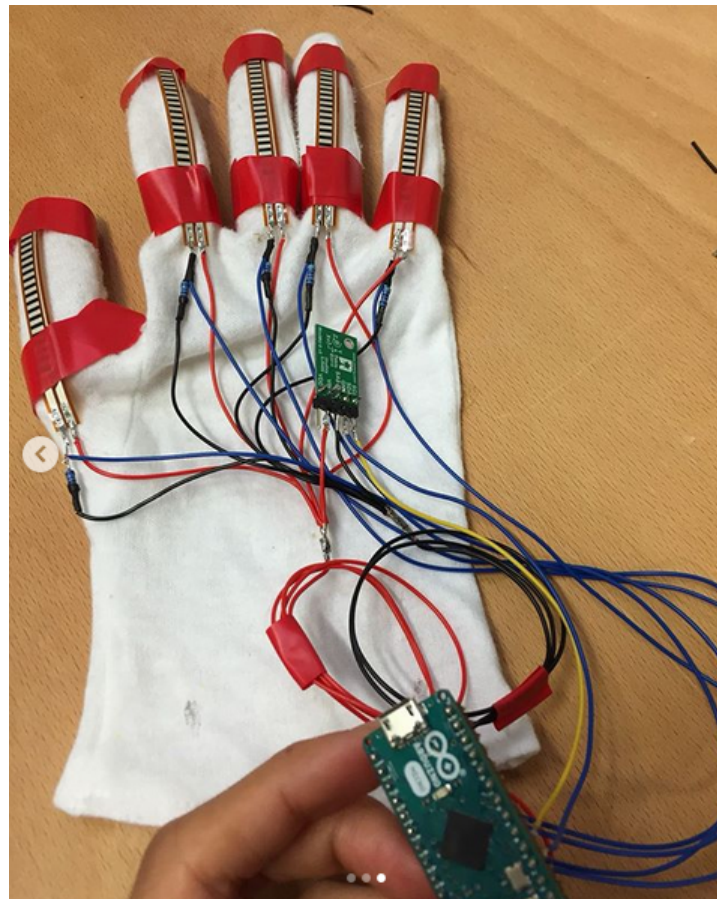


Abbildung A.1.: Aktueller Prototyp des Sensorhandschuhs. Es sind Biegesensoren für alle Finger, ein Gyroskop und ein Beschleunigungssensor im Einsatz. Bei der Verarbeitung der Sensordaten kommt ein Arduino Micro zum Einsatz.

Dieser könnte während der Trainingsphase vom Nutzer dazu verwendet werden, feinere Gesten aufzunehmen, wobei neben den von den Detektormodulen detektierten Photoströmen zusätzlich die Daten der auf dem Sensorhandschuh befindlichen Module gespeichert werden. Aktuell kommen hier Biegesensoren für jeden Finger, sowie ein Gyroskop und ein Beschleunigungssensor zum Einsatz. Diese verfeinerten Gestendaten stellen nun natürlich eine neue Herausforderung an das Modell, das zur Auswertung der Photoströme im Live-Betrieb eingesetzt wird: Statt wie vorher den eingehenden Photoströmen nur das Klassenlabel einer Geste zuzuordnen, muss es nun in der Lage sein, auf Basis der gemessenen Photoströme die wahrscheinliche Lage der Hand, sowie die Krümmung der Finger vorauszusagen. Dieses Problem ist mathematisch deutlich schwieriger zu lösen, was nicht zuletzt an der deutlich höheren Anzahl an Parametern liegt, die bestimmt werden müssen, um eine ausreichende Aussagekraft zu gewährleisten. Aufgrund der höheren Anzahl an Parametern werden zudem auch mehr Daten benötigt, um das Modell zu trainieren.

Ebenfalls denkbar wäre eine Erweiterung auf dynamische Gesten, bei denen der Nutzer die Hand bewegt. In diesem Fall müssen die Daten der Photodioden im Live-Betrieb in Form Zeitreihe aufgenommen werden und das neuronale Netzmodell muss diese Zeitreihe mit einer Geste verknüpfen. Auch hier ergibt sich ein mathematisch deutlich komplexeres Problem, das ein aufwändigeres Training und viele Trainingsdaten erfordert. Die Erweiterung um dynamische Gesten könnte ebenfalls mit der Verwendung des Sensorhandschuhs kombiniert werden.