

# 1 Einführung

## 1.1 Problemstellung

Im Rahmen zur Bestimmung der Aufmerksamkeit im Unterricht sind verschiedene Parameter relevant, einer wichtiger ist die Blickrichtung des einzelnen Schülers.

Um mit möglichst geringem Aufwand an Hardware eine Messung der ganzen Klasse vorzunehmen, wird nur eine einzelne fest montierte Kamera eingesetzt.

Ziel ist es nun, anhand des Filmmaterials die gesamte Klasse auf einmal auszuwerten, um von allen die Blickrichtungen während einer Schulstunde zu bestimmen.

## 1.2 Hardware

Als Messinstrument wird nur eine einzelne Farbkamera verwendet. Das Videomaterial der Schulklasse wurde mit einer unbekannten Videokamera aufgezeichnet, daher sind nur die Werte des Filmes ( $640 \times 480$  Hz) bekannt.

Für die Messungen im Versuch wurde die Explorer 4K Action Camera mit der 2.7K Einstellung verwendet. Die Kamera verwendet eine  $170^\circ$  Weitwinkel-Linse mit einem  $2688 \times 1520$  Video mit 29FPS.

## 1.3 Software

Software-Elemente auf fremder Quelle die eingesetzt wurden.

### 1.3.1 OpenCV

Open Source Computer Vision, eine Bibliothek von Algorithmen zur Bildverarbeitung in Echtzeit unter der Lizenz von BSD (Berkeley Software Distribution)

### 1.3.2 ElSe

Ellipse Selection for Robust Pupil Detection in Real-World Environments, ein Algorithmus zur Bestimmung der Umrisse der Pupille im Bild eines Auges. Der ursprüngliche ElSe-Algorithmus ist für Infrarotbilder ausgelegt, wurde für diese Anwendung aber zu Farbbilder modifiziert.

### 1.3.3 OpenFace

Ein Open-Source Echtzeitverfahren auf Basis von CLNF zur Bestimmung und Analyse von Gesichtsmerkmalen in Grau-Bilder und Videos. Dabei werden 68 Signifikante Punkte im Gesicht bestimmt und auf Basis jener Position und Orientierung ermittelt.

### 1.3.4 MTCNN Face Detection

Multi-task Cascaded Convolutional Networks, ein Algorithmus zur Detektion von Gesichtern und Bestimmung von 5 Gesichts-Landmarks. Dabei wird ein CNN auf verschiedene Skalierungen des Bildes angewendet um so zuverlässig ein Gesicht zu erkennen.

## 1.4 Gesetzte Ziele

Da die Anwendung auf Aufnahmen eines Unterrichtes arbeiten soll, ergeben sich folgende Bedingungen:

- Normale Brillen, Kontaktlinsen und ähnliches sind erlaubt.

- Bewegung/Sprechen usw. der Schüler ist erlaubt.
- Es soll gleichzeitig auf Distanzen zwischen  $2.5 - 8m$  zur Kamera funktionieren, ohne dass sich diese bewegt oder gezoomt werden kann.
- Möglichst alle Blickrichtungen der Schüler sollen möglichst genau erfasst werden.

Ein deutsches Klassenzimmer hat  $55 - 65m^2$ , da noch Abstand zur Tafel usw. beachtet werden muss ergibt sich, wenn sich die Kamera an der Tafel befindet, einen Abstand zu den Schülern von  $2.5 - 8m$  zur Kamera auf einer Breite von  $6m$ . Somit muss der Linsenwinkel mindestens  $100^\circ$  betragen.

## 1.5 Randbedingungen

Des Weiteren werden folgende Annahmen gemacht:

- Die Szene ist innerhalb eines Gebäudes stattfinden, mit ausreichend gleichmäßiger Beleuchtung.
- Die Überführung zwischen Welt- und Kamerakoordinatensystem bekannt.
- Die Kamera befindet sich vor der Klasse, so dass die Blickrichtung nach vorn von den Schülern möglichst zur Kamera ist.
- Die Gesichter der Schüler sind komplett sichtbar.

Natürlich sind auch alle inneren Parameter der Kamera bekannt.

## 2 Theorie & Grundlage

### 2.1 OpenFace

- Funktion
- 64 Landmarks & 2 mal 28 für Auge
- Veröffentlichte Genauigkeit
- Bestimmung der Gesichtsorientierung
- Bestimmung Augen-Orientierung

### 2.2 MTCNN Face Detection

- Bestimmung der Boxen um die Gesichter
- 5 Landmarks
- Image-Pyramide
- Funktionsablauf

### 2.3 ELSE

- Funktion
- Verwendung

## 2.4 Skalieren von Bildern

- Verschiedene Verfahren
- Ansprüche für Berechnung

## 2.5 Berechnung

- Kamera  $\rightarrow$  Welt
- Bestimmung der Ziele

# 3 Implementierung

## 3.1 Ablauf der Implementierung

Für die Bestimmung der Kopfposition und Orientierung wird ein mehrstufiges Verfahren verwendet. Am Anfang müssen alle Gesichter die im aktuellen Frame vorhanden sind detektiert werden. Dazu wird die MTCNN Face detection verwendet, mit diesem Verfahren auch kleinste Gesichter erkannt werden. 3.1.1 Für die weiteren Berechnungen muss bekannt sein welchen Bereich das Gesicht in Frame einnimmt und um welches es sich handelt. Der Bereich wird vom MTCNN als Box geliefert, als Personenzuordnung wird ein Matching zum vorigen Frame verwendet.

Damit auch eine Berechnung auf den kleineren Gesichtern stattfinden kann werden alle zu kleinen Bildbereiche hochskaliert. Dabei muss wegen Ungenauigkeiten die gefundene Box etwas vergrößert werden und sollte dann auf eine Mindestbreite gebracht werden. 3.1.2

Diese Bildbereiche werden nun mit OpenFace weiterverarbeitet um die Position der Landmarks im Bild zu bestimmen. Dadurch kann das CNN sich auf jede einzelne Person einstellen um so bessere Ergebnisse zu erreichen. Außerdem können alle gefundenen Personen gleichzeitig (parallel) bestimmt werden. 3.1.3

Bei großen Bildern wird nun ElSe auf den Augenbereich angewendet um die Position der Pupille noch exakter zu ermitteln, damit die Blickrichtung genauer wird. Dazu muss allerdings die Differenz zwischen ElSe-Ergebnis und OpenFace-Ergebnis betrachtet werden um Fehler zu erkennen. 3.1.4

Nun wird auf Basis der Landmarks und der Kameraparameter die Position und Orientierung des jeweiligen Gesichtes bestimmt und können für weitere Anwendungen verwendet werden. 3.1.5

### 3.1.1 Detektion der Gesichter

To Do!

- Nicht-Gesichter
- Überschneiden von Bereichen
- Füllen von Fehlenden Bereichen
- Qualität der 5 Landmarks

### 3.1.2 Skalierung auf Mindestgröße

To Do!

- Mindestens 130 Pixel Breit

- ca. 30% größere Boxen
- Grenzen bei der Mindestgröße

### **3.1.3 Bestimmung der Landmarks**

To Do!

- Falsche Gesichter
- Lernen
- Bild & Film Detection
- Verbesserung durch Farbkorrektur

### **3.1.4 Verbesserung der Augen**

To Do! ElSe

- Mindestens 10Pixel? großes Bild der Augen
- Ergebnis innerhalb der Augen
- ableiten von Iris und Pupille
- Mittlung der beiden Augen

### **3.1.5 Bestimmung der Position & Orientierung**

To Do!

- Kameraparameter und ihre Auswirkung
- Bestimmung der Position und Orientierung

## **4 Ergebnisse**

### **4.1 Erreichte Werte**

- Distanzen
- Winkel
- Zuverlässigkeit
- Auswerten der Messung

## 4.2 Fehleranalyse

Mit entsprechend hochauflösenden Kameras können auch bessere Resultate auf größerer Distanz erreicht werden. Gerade die Bestimmung der Blickrichtung ist meist nicht möglich, da die Augenpartie viel zu klein für eine Berechnung ist.

Da Bewegung erlaubt ist kassiert es immer wieder das Teile des Gesichtes verdeckt werden oder sich der Kopf zweit wegrehen. Aber auch die Frisuren spielen eine Rolle, da dadurch einige Landmarks im Gesicht verdeckt werden. Dadurch sind in diesen Fällen keine weiteren Brechungen möglich da kein Tracking mehr möglich ist. Eine Lösungsansatz wären Landmarks in Profilbildern zu detektieren und das verwenden von weiteren Kameras aus anderen Perspektiven.