OSTBAYERISCHE
TECHNISCHE HOCHSCHULE
REGENSBURG

INFORMATIK UND
MATHEMATIK

Emotion Recognition

P2 - Deep Learning Eye Catcher

Bericht zum Laborpraktikum
im Sommersemester 2019 von

Lukas Schulz
Matrikelnummer: 1234567

Noor Alrabea
Matrikelnummer: 12345678

Markus Hofer
Matrikelnummer: 123456789

Benjamin Bauer
Matrikelnummer: 12345678

Mona Ziegler
Matrikelnummer: 3091609

Anna Will
Matrikelnummer: 12345678

Julia Karnaukh
Matrikelnummer: 12345678

David Wolf
Matrikelnummer: 12345678

Fakultät Informatik und Mathematik
Ostbayerische Technische Hochschule Regensburg
(OTH Regensburg)

Prüfer: Prof. Dr. Christoph Palm
Abgabedatum: xx. Juni 2019

Inhaltsverzeichnis

1 Einleitung

1.1 Motivation und Einführung

Künstliche Intelligenz ist bereits in vielen Situationen des Lebens alltäglich geworden. So steckt hinter der Gesichtserkennung im Smartphone oder Sprachassistenten wie Amazons "Alexa" Künstliche Intelligenz. Der große Vorteil und Nutzen an KI liegt vor allem in der Mustererkennung und Zuordnung zu vorgegebenen Kategorien. In der Medizin spielt das vor allem bei der Verarbeitung der Ergebnisse Bildgebender Diagnoseverfahren eine große Rolle. Eine KI kann Muster in beispielsweise Röntgenbildern erkennen, und so den Arzt bei der Diagnosefindung unterstützen. Daran wird auch im Labor des ReMIC (Regensburg Medical Image Computing) geforscht und verschiedene Anwendungen zur Bildanalyse entwickelt. Hierbei liegt der Fokus immer mehr auf Deep-Learning Ansätzen. Um nun mehr interessierte Studierende der OTH über diese Tätigkeit zu informieren, soll eine einfache Deep-Learning Anwendung zur Emotionserkennung entwickelt werden. Um so viele Studierende wie möglich dafür zu begeistern, wurde entschieden, ein Spiel zu entwickeln.

1.2 Das Programm

Neben der Tür des ReMIC gibt es ein "Schaufenster", hinter diesem ist ein Bildschirm aufgebaut, der im Moment genutzt wird um allgemeine Informationen über das Labor anzuzeigen. In Zukunft soll dort zwischen den Vorlesungszeiten das entwickelte Programm spielbar sein. Hierfür wird eine Kamera über den Bildschirm aufgebaut.
Das Spiel wird gestartet, wenn das Gesicht eines Spielers in einem ausgewiesenen Bereich erkannt wurde. Dem Spieler wird nun eine zufällig ausgewählte Emotion vorgegeben, die er nun nachmachen soll. Das Programm ist in der Lage sieben Emotionen zu erkennen: Neutral, Angewidert, Wütend, Glücklich, Überrascht, Traurig, Ängstlich. Die erreichte Punktzahl pro Runde ergibt sich daraus, wie gut die Emotion durch das Programm erkannt werden konnte. Die höchste pro Runde erreichbare Punktzahl ist 100, für eine zu 100% erkannte Emotion.

Falls kein oder mehrere Gesichter erkannt wurden, soll eine Fehlermeldung auf dem Bildschirm ausgegeben werden.
Das Spiel endet, wenn der Spieler den von der Kamera erfassten Bereich verlässt. Danach werden die Bilddaten des Spielers gelöscht.

1.3 Künstliche Intelligenz und Deep Learning

Traditionell liegen die Stärken einer Maschine bei rechenaufwändigen, für den Menschen kaum oder nur sehr langsam lösbare Aufgaben. Auch große Datenmengen stellen dabei für einen Computer keine große Herausforderung dar. Im Gegensatz dazu sind Probleme, die der Mensch intuitiv oder situationsabhängig lösen würde, durch konventionelle Programmierung kaum erfassbar. Der Programmierer müsste jede mögliche Situation abfangen, in vielen Fällen ist dies aufgrund des Umfangs nicht effizient umsetzbar. Bei der Emotionserkennung beispielsweise hat der Mensch keine Schwierigkeiten, in verschiedenen Gesichtern Emotionen zu erkennen und zu unterscheiden, obwohl sich die Ausprägung verschiedener Merkmale dabei in jedem Gesicht stark unterscheiden kann. Um eine solche Aufgabe maschinell lösen zu können, wird die Verwendung einer künstlichen Intelligenz notwendig.

Ein Vielversprechender Ansatz zur Entwicklung einer künstlichen Intelligenz ist das Deep Learning. Dabei werden neue, komplexe Probleme gelöst indem sie aus bekannten, einfachen Problemen zusammengesetzt werden. Hierfür werden neuronale Netze verwendet. Diese sind vom Aufbau und der Struktur der Verschaltung des menschlichen Gehirn nachempfunden. Es gibt Knotenpunkte (= Neuronen), die mit anderen Knotenpunkten verbunden sind, und so Informationen weitergeben. Beim Deep Learning sind diese Knoten in Schichten angeordnet. Die erste Schicht ist sichtbar, und umfasst die rohen Eingangsdaten. Diese werden abstrahiert, auf eine nächste versteckte Schicht abgebildet, wieder verarbeitet und an die nächste Schicht weitergegeben. Dies geschieht einige Male, bis die Daten an die letzte Ebene weitergegeben werden. Dort erhält man ein prozentuales Ergebnis, zu welcher Kategorie die Eingangsdaten zugeordnet werden können. Wie genau die versteckten Schichten arbeiten, an welchen Punkten sie sich orientieren, oder worauf die letztendliche Entscheidung basiert, ist von außen nicht ersichtlich. Die internen Parameter müssen nicht von einem Programmierer festgelegt werden, sondern werden vom neuronalen Netz selbst über ein Trainingsverfahren gelernt [1].

Zum Training wird in der Praxis meist Supervised Learning angewendet. Dabei werden neben den Daten, die in eine Kategorie eingeordnet werden sollen, auch der gewünschte Output übergeben. Nun sollen die internen Parameter so angepasst werden, dass sich der Fehler minimiert. Um dieses Minimum zu finden, wird häufig das Gradientenabstiegsverfahren verwendet. Hierfür berechnet der Lernalgorithmus für jeden Parameter einen Gradienten, der angibt, in wie weit sich der fehlerhafte Output verbessert oder verschlechtert, falls der Parameter leicht erhöht wird. Der Parametervektor wird dann in die entgegengesetzte Richtung zum Gradienten angepasst, um so einen minimalen Wert für den Fehler zu erhalten [2].

2 Automatische Emotionserkennung

Mimiken entstehen durch die Kontraktion verschiedener Gesichtsmuskeln, welche zu den Änderungen der jeweiligen Gesichtsmerkmale führen. So kommt es zum Beispiel zum Zusammenkneifen der Augenlider, Heben der Augenbrauen und Lippen oder Hochziehen der Nase. Durch das Auftreten von Grübchen und Fältchen kann sich auch die Textur der Haut verändern. Verschiedene Emotionen führen zu unterschiedlichen charakteristischen Änderungen im Gesicht, wobei manche Emotionen schwieriger zu erkennen sind als andere. Wir beschäftigen uns hier mit den sechs Basis Emotionen: Freude, Trauer, Angst, Ekel, Überraschung und Wut [3]. Wichtig um diese Emotionen zu erkennen sind die Lage, Intensität und Dynamik der entscheidenden Merkmale [3]. Um Emotionen automatisch erkennen zu können müssen drei grundlegende Phasen durchlaufen werden. Als erstes muss das Gesicht und dessen Komponenten erkannt werden, dann müssen die Merkmale extrahiert und letztendlich als ein Ausdruck klassifiziert werden [4].

2.1 Gesichtsdetektion

Der erste Schritt für eine funktionierende Emotionserkennung ist die richtige Detektion des Gesichts. Im besten Fall enthält ein Emotionserkennungsprogramm einen Gesichtsdetektor, der es erlaubt ein Gesicht in komplexen Szenen mit einem schwierigen Hintergrund zu erkennen. Manche Methoden zur Emotionserkennung brauchen die exakte Position des Gesichts, bei anderen, wie beim Active Appearance Modell reicht die grobe Lokalisierung [3].
Bei der Detektion des Gesichts kann man zwischen Merkmal und Bild basierten Methoden unterscheiden. Ersteres eignet sich gut für Echtzeitsysteme bei denen Farbe und Bewegung beteiligt ist, letzteres ist eine stabile Technik für statische Grauwertbilder [5].
Die Entwicklung des merkmalsbasierten Ansatzes kann man weiter in drei Bereiche unterteilen: die Low-Level Analyse, Merkmalsanalyse und der Verwendung aktiver Formmodelle. Die Low-Level Analyse befasst sich zunächst mit der Segmentierung visueller

Merkmale unter Verwendung von Pixeleigenschaften wie Graustufen und Farben. Allerdings sind aufgrund der niedrigen Ebene die aus dieser Analyse generierten Merkmale nicht eindeutig. Deshalb macht man als nächstes eine Merkmalsanalyse, bei der visuelle Merkmale, unter Verwendung von Informationen zur Gesichtsgeometrie, zu einem globaleren Konzept des Gesichts und der Gesichtsmerkmale organisiert. Dadurch wird die Mehrdeutigkeit von Merkmalen reduziert und die Position von Gesicht beziehungsweise Gesichtsmerkmalen bestimmt. Der letzte Schritt umfasst die Verwendung aktiver Formmodelle von Schlangen bis hin zu neueren Punktverteilten Modellen (PDM) [5].
Beim bildbasierten Verfahren wird eine Fensterabtasttechnik zum Erfassen von Gesichtern angewendet [5]. Der Fensterabtastalgorithmus ist im Wesentlichen nur eine Suche auf dem Eingabebild nach in allen Maßstäben möglichen Gesichtsorten, wobei es Unterschiede bei der Implementierung dieses Algorithmus für fast alle bildbasierten Systeme gibt. Typischerweise variieren die Größe des Abtastfensters, die Unterabtastrate, die Schrittgröße und die Anzahl der Iterationen in Abhängigkeit von dem vorgeschlagenen Verfahren und der Notwendigkeit eines rechnerisch effizienten Systems [5].

2.2 Merkmal Extrahierung

Hat man das Gesicht einmal detektiert, geht es daran die Merkmale zu extrahieren. Man kann sich hierbei auf das gesamte Gesicht oder nur bestimmte Teilbereiche beziehen [3]. Um Gesichtsausdrücke aus einem frontal aufgenommenen Bild zu erkennen, muss eine Reihe von Schlüsselparametern, die die Gesichtsausdrücke am besten beschreiben, aus dem Bild erarbeitet werden. Diese Parameter, auch Merkmalsvektor genannt, werden genutzt, um später zwischen den verschiedenen Emotionen unterscheiden zu können. Der wichtigste Aspekt einer erfolgreichen Merkmalerkennung ist die Menge an Informationen, die aus einem Bild in den Merkmalsvektor extrahiert werden kann [6]. Wenn der Merkmalsvektor des einen Gesichts mit dem eines anderen Gesichts übereinstimmt, können die beiden Gesichter mittels Merkmalsklassifikation nicht mehr unterschieden werden. Dieser Zustand heißt Merkmal Überlappung und sollte bei einer idealen Merkmal Extrahierung niemals auftreten [6].
Für automatische Emotionserkennungssysteme wurden verschiedene Arten herkömmlicher Ansätze untersucht. Die Gemeinsamkeit dieser Ansätze besteht darin, den Gesichtsbereich zu erfassen und geometrische Merkmale, Erscheinungsmerkmale oder eine Mischung aus beidem aus dem Zielgesicht zu extrahieren [4]. Für die geometrischen Merkmale wird die Beziehung zwischen Gesichtskomponenten verwendet, um einen Merkmalsvektor für das Training zu konstruieren. Die Erscheinungsmerkmale werden

normalerweise aus dem globalen Gesichtsbereich extrahiert oder aus Regionen mit unterschiedlichen Arten von Informationen [4].

2.2.1 Gabor Wavelets

Mehrere Untersuchungen in der Bildverarbeitung haben ergeben, dass die Feature Extraktion für Gesichtserkennung- und Tracking mit Gabor Filtern gute Ergebnisse liefert, weshalb sie auch eine vielversprechende Methode für die Emotionserkennung darstellt [6].

Der allgemeine Rahmen zur Musterdarstellung in Bildern basiert hierbei auf topographisch geordneten, räumlich lokalisierten Filtern, die aus einer mehrfach auflösenden und orientierten Sammlung von Gabor Wavelet Funktionen bestehen [7]. Eine 2D-Gabor-Funktion ist eine ebene Welle mit dem Wellenfaktor k, die durch eine Gauß'sche Hüllkurvenfunktion mit der relativen Breite σ begrenzt wird [6]. Um Informationen über Gesichtsausdrücke zu extrahieren, wird jedes Bild mit einer mehrfach auflösenden und orientierten Sammlung von Gabor Wavelets Gk_+ und Gk_- gefaltet [7]. Das Vorzeichen stellt gerade (+) oder ungerade (-) Phasen dar, während k, der Filterwellenvektor, die räumliche Frequenz und Ausrichtung des Filters bestimmt [7]. Die Antworten der Filter auf das Bild werden zu einem Vektor zusammengefügt, der folgende Komponenten besitzt:

$$R_{\vec{k},\pm}(\vec{r}_0) = \int G_{\vec{k},\pm}(\vec{r}_0, \vec{r}) I(\vec{r}) d\vec{r}, \text{ mit}$$

$$G_{\vec{k},+}(\vec{r}) = \frac{k^2}{\sigma^2} e^{-k^2 \|\vec{r}-\vec{r}_0\|2/2\sigma^2} cos(\vec{k} \cdot (\vec{r} - \vec{r}_0)) - e^{-\sigma^2/2}, \text{ und}$$

$$G_{\vec{k},-}(\vec{r}) = \frac{k^2}{\sigma^2} e^{-k^2 \|\vec{r}-\vec{r}_0\|2/2\sigma^2} sin(\vec{k} \cdot (\vec{r} - \vec{r}_0))$$

[7]

Um das System gegenüber der absoluten Beleuchtungsstärke unempfindlich zu machen, wird das Integral des Cosinus Gaborfilters vom Filter subtrahiert. Beim Sinusfilter ist dies nicht notwendig, da dieser nicht von der absoluten Beleuchtungsstärke abhängt [7].
Es werden drei räumliche Frequenzen mit folgenden Wellennummern benötigt: k = $\pi/2$, $\pi/4$, $\pi/8$ [7]. In allen Kalkulationen wird die Bandbreite auf $\sigma = \pi$ gesetzt [7]. Es werden sechs Wellenvektor Orientierungen genutzt, mit Winkeln von 0 bis π, die im gleichmäßigen Abstand von $\pi/6$ angeordnet sind [7].

Die Komponenten des Gabor Vektors R_k werden definiert als die Amplitude der kombinierten geraden und ungeraden Filterantworten $R_{\vec{k}} = \sqrt{R^2_{\vec{k},+} + R^2_{\vec{k},-}}$ [7]. Im Gegensatz zu linearen Filterantworten reagiert hier die Antwortamplitude weniger empfindlich auf Positionsänderungen [7].
Um den Ähnlichkeitsraum von Gabor kodierten Gesichtsbildern zu untersuchen, können Filterantworten mit der gleichen räumlichen Frequenz und Orientierung in übereinstimmenden Punkten zweier Gesichtsbilder verglichen werden [7]. Das normalisierte Skalarprodukt wird verwendet, um die Ähnlichkeit zweier Gabor Antwortvektoren zu quantifizieren [7]. Die Ähnlichkeit wird dabei als Durchschnitt der Gaborvektorähnlichkeit über alle korrespondierenden Gesichtspunkte berechnet. Da Gaborvektoren bei Nachbar Pixel stark korreliert und redundant sind, reicht es aus, den Durchschnitt auf einem relativ spärlichen Gitter, welches das Gesicht bedeckt, zu berechnen [7].
Das hinzufügen mehrerer Gitterpunkte könnte die Leistung des geometrischen Maßes erhöhen, was allerdings mit einem stark erhöhten Rechenaufwand verbunden wäre. Die Lokalisierung der Gitterpunkte ist der teuerste Teil eines voll automatischen Systems [7]. Eine Verbesserung, allerdings nur eine geringfügige, könnte die Kombination von Gabor und Geometrie Systemen bieten[7].

2.2.2 Active Appearance Modell

Active Appearance Modelle (AAM) haben sich als eine gute Methode bewährt, um vordefinierte lineare Formmodelle zu einem Quellbild, welches das zu untersuchende Objekt enthält, auszurichten [8]. Das Prinzip der AAMs besteht darin, sich ihren Form- und Erscheinungsbildkomponenten durch eine Gradientenabstammungssuche anzupassen [8].
Die Form s eines AAMs wird durch ein 2D trianguliertes Netz beschrieben, wobei insbesondere die Koordinaten der n Gitter Eckpunkte die Form $s = [x_1, y_1, x_2, y_2, \ldots, x_n, y_n]$ definieren [8]. Diese Eckpunkte korrespondieren mit einem Quellbild, von dem aus die Form ausgerichtet wurde [8]. Da AAMs lineare Form Variation erlaubt, kann die Form s als Basisform s_0 ausgedrückt werden, plus einer linearen Kombination von m Formvektoren s_i:
$s = s_0 + \sum_{i=1}^{m} p_i s_i$ (1) [8]
wobei der Koeffizient $p = (p_1, , p_m)^T$ die Formparameter bildet. Diese Formparameter können typischerweise in rigide ähnliche Parameter p_s und nicht rigide Objektdeformationsparameter p_0 unterteilt werden, sodass sich $p_T = [p_s^T, p_0^T]$ ergibt [8]. Ähnlichkeitsparameter werden assoziiert mit einer geometrischen Ähnlichkeitstransformation (zum Beispiel Translation, Rotation, Skalierung) [8]. Die objektspezifischen Parameter

sind die Restparameter, die nicht starre geometrische Variationen darstellen, die der bestimmenden Objektform zugeordnet sind (zum Beispiel öffnen des Mundes, Augen schließen, etc.) [8]. Prokrustes Anpassung wird verwendet, um die Grundform s_0 abzuschätzen [8].
Wenn das Gesicht des Probanden durch Abschätzen der Form und Erscheinung der AAM Parameter einmal erkannt wurde, kann man die Information nutzen um folgende Eigenschaften abzuleiten:

SPTS Die normalisierte Form s_n bezieht sich auf die n Eckpunkte für die x- und y-Koordinaten, was in einem rohen 2n-dimensionalen Merkmalsvektor resultiert [8]. Diese Punkte sind die Scheitelpunkte, nachdem alle starren geometrischen Abweichungen (Translation, Rotation und Skalierung) in Bezug auf die Grundform entfernt wurden [8]. Die normalisierte Form s_n kann erhalten werden durch synthetisieren einer Form Instanz von s, durch das Nutzen der Gleichung (1), welche die Ähnlichkeitsparameter p ignoriert [8].

CAPP Das kanonisch normalisierte Erscheinungsbild a_0 bezieht sich auf den Bereich, in dem alle nicht rigiden Formvariationen normalisiert wurden, wobei die Basisform s_0 berücksichtigt wird [8]. Dies wird erreicht durch Anwenden einer stückweise affinen Neigung auf jede dreieckige Form im Quellbild, sodass es mit der Basisgesichtsform einhergeht [8]. Wenn man die starren Formvariationen weglassen würde, würde dies zu einer schlechteren Leistung führen [8].

2.3 Klassifizierung

2.3.1 Neuronale Netze

Ein guter Klassifikator, der für viele Mustererkennungsanwendungen verwendet wird, ist das künstliche neuronale Netzwerk (ANN). Es ist effektiv bei der Modellierung nichtlinearer Abbildungen und hat eine gute Klassifizierungsleistung bei einer relativ geringen Anzahl von Trainingsbeispielen. Fast alle ANNs können in drei Grundtypen eingeteilt werden: Multilayer Perceptron Klassifikator (MLP), wiederkehrende neuronale Netze (RNN) und Netze für radiale Basisfunktionen (RBF) [9].

Sobald die Struktur von ANN vollständig spezifiziert ist, sind MLP relativ einfach zu Implementieren und besitzen einen gut definierten Trainingsalgorithmus, weshalb sie

auch bei der Emotionserkennung eingesetzt werden [7]. ANN-Klassifikatoren weisen jedoch im Allgemeinen viele Entwurfsparameter auf, die normalerweise ad hoc festgelegt werden, wie zum Beispiel die Form der Neuronenaktivierungsfunktion, die Anzahl der verborgenen Schichten und die Anzahl der Neuronen in jeder Schicht [9]. Die Leistung von ANN hängt stark von diesen Parametern ab.

2.3.2 Vector Machines

Ein weiterer Klassifikator ist die Support Vector Machine (SVM), welche ein wichtiges Beispiel für die Diskriminanzklassifikatoren darstellt [9]. In einer Reihe von Mustererkennungsaufgaben, die das Gesicht und die Gesichtsaktionserkennung betreffen, haben sich SVMs als eine nützliche Methode zur Klassifizierung erwiesen [8] und übertreffen nachweislich andere bekannte Klassifizierungsmethoden [9], da die globale Optimalität des Trainingsalgorithmus und die Existenz hervorragender datenabhängiger Verallgemeinerungsgrenzen große Vorteile bieten. SVM-Klassifikatoren basieren hauptsächlich auf der Verwendung von Kernelfunktionen, um die ursprünglichen Features nichtlinear auf einen hochdimensionalen Raum abzubilden, in dem Daten mithilfe eines linearen Klassifikators gut klassifiziert werden können [9]. Da es keine systematische Möglichkeit gibt, die Kernelfunktion auszuwählen, ist die Trennbarkeit der transformierten Merkmale nicht garantiert. Dies stellt bei Mustererkennungsanwendungen allerdings kein Problem dar, weil dort bei vielen keine perfekte Trennung der Trennungsdaten angestrebt wird, um eine Überanpassung zu vermeiden [9].

2.4 Facial Action Coding System

Nachdem die verschiedenen Merkmale des Gesichts extrahiert und klassifiziert wurden, können sie mit Hilfe des Facial Action Coding Systems (FACS) nach Ekman und Friesen in visuelle Klassen kodiert werden.

FACS ist ein umfassendes anatomisch basiertes System zur Erfassung aller visuell erkennbaren Gesichtsbewegungen, welches auf 44 sogenannten Action Units (AUs), sowie verschiedenen Kategorien von Augen und Kopf Positionen beziehungsweise Bewegungen basiert [10]. Kombiniert man diese AUs miteinander entsteht eine große Variation verschiedener Gesichtsausdrücke. Die Verbindung der AUs 12 und 13, welche das Hochziehen der Mundwinkel und Füllen der Backen beschreiben, mit den AUs 25 und 27, die das leichte öffnen des Mundes darstellen, und dem AU 10, dem anheben der obe-

ren Lippe, ist nur eine Kombination von vielen, um ein Lachen zu beschreiben. Nach diesem Prinzip lassen sich alle Emotionen aus den einzelnen AUs zusammensetzen.

Trotz seiner Einschränkungen, das Fehlen einer zeitlichen und detaillierten räumlichen Information, ist diese Methode die weitverbreitetste für das messen menschlicher Gesichtsbewegungen [11].

3 Einsatzgebiete

Automatisierte Emotionserkennung kann in vielen Bereichen zum Einsatz kommen. Besonders in der Marktforschung und im Marketing kann sie von großem Nutzen sein. Viele Firmen haben bereits erkannt, dass es eine gute Werbestrategie ist, die Emotionen des Kunden anzusprechen. Lässt sich der Kunde emotional auf eine Marke ein, steigt nicht nur die Wahrscheinlichkeit, dass er das Produkt kauft, sondern auch dass er es weiterempfiehlt und einem Konkurrenzprodukt vorzieht [12]. Um nun diesen Effekt zu überprüfen, wird bereits heute automatisierte Emotionserkennung angewendet. Auch in der Software Entwicklung kann eine solche Technik während Usability-Tests zum Einsatz kommen um festzustellen, ob der Kunde mit der Nutzung zufrieden ist [13]. In der Medizin könnte sie Rückschlüsse auf das Befinden eines Patienten geben und, eingebaut in ein Smart Home System, im Ernstfall einen Notdienst kontaktieren. Im Bereich der Mensch-Maschinen-Interaktion können mittels Emotionserkennung große Fortschritte erzielt werden. Die menschliche Interaktion ist von Feinheiten geprägt. In der Regel passt man sein eigenes Verhalten der Stimmung seines Gesprächspartners an. So geht man mit einem aggressiven oder verängstigten Menschen anders um als mit jemandem der gut gelaunt ist. Eine Maschine muss diese Emotionen erst erkennen, um dann ihr Verhalten anpassen zu können. Hierfür kann neben dem Gesichtsausdruck auch die Stimme eine große Rolle spielen [14].

Konkretes Anwendungsbeispiel Die durch das Fraunhofer Institut für Integrierte Schaltungen entwickelte Bildanalyse Software "SHORE" (Sophisticated Highspeed Object Recognition Engine) ist in der Lage, neben einer Bestimmung des Geschlechts und Abschätzungen des Alters, Emotionen zu erkennen und zu unterscheiden. Theoretisch könnten 100 Emotionen erfasst werden, bei SHORE wurde sich auf die vier am einfachsten zu unterscheidenden Emotionen beschränkt: glücklich, traurig, überrascht und verärgert. Die Software wurde plattformunabhängig entwickelt, sodass sie in vielen Bereichen Anwendung finden kann. Beworben wird die Software für drei Anwendungsbereiche: Werbung und Marktforschung, Fahrassistenz-Systeme und Medizintechnik. Während bei der Marktforschung hier vor allem die Zielgruppenanalyse als Vorteil

genannt wird, soll bei Fahrassistenz-Systemen auch die Stimmungslage des Fahrers analysiert werden, um so Unfallrisiken zu reduzieren. In der Medizintechnik soll die Software zum einen zur Schmerzerkennung von Patienten, die sich nicht mehr verständigen können, angewendet werden können. Zum anderen kann sie, eingebettet in ein Ambient-Assistent-Living-System, verwendet werden um älteren und erkrankten Menschen so lang wie möglich die Möglichkeit geben selbstständig aber dennoch sicher zu leben. Eine weitere Möglichkeit der Verwendung von SHORE besteht darin, in einer Datenbrille eingebunden beispielsweise autistischen Menschen das Leben zu erleichtern. Erkrankte haben dabei in den meisten Fällen große Probleme, die Gefühlslage ihres Gesprächspartners und den Subtext so mancher Aussagen zu erkennen. Durch die über die Brille erhaltenen Informationen kann so der Alltag stark erleichtert werden[15] [16].

4 Material und Methoden

4.1 Neuronale Netze

4.2 Benutzeroberfläche

4.3 Schnittstelle

????

5 Ergebnisse

6 Diskussion

7 Zusammenfassung

Literatur

[1] I. Goodfellow, Y. Bengio und A. Courville. Deep Learning. http://www.deeplearningbook.org. MIT Press, 2016.

[2] Y. LeCun, Y. Bengio und G. Hinton. Deep learning. Nature 521, 436 EP -, 2015.

[3] B. Fasel und J. Lüttin. Automatic Facial Expression Analysis: A survey. Pattern Recognition 36, 259–275, 2002.

[4] B. C. Ko. A Brief Review of Facial Emotion Recognition Based on Visual Information. Sensors 18(2), 401, 2018.

[5] E. Hjelmås und B. K. Low. Face Detection: A Survey. Computer Vision and Image Understanding 83(3), 236–274, 2001.

[6] S. Bashyal und G. K. Venayagamoorthy. Recognition of facial expressions using Gabor wavelets and learning vector quantization. Engineering Applications of Artificial Intelligence 21(7), 1056–1064, 2008.

[7] M. Lyons, S. Akamatsu, M. Kamachi und J. Gyoba. Coding facial expressions with Gabor wavelets. In: Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, 200–205, Apr. 1998.

[8] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar und I. Matthews. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, 94–101, Juni 2010.

[9] M. E. Ayadi, M. S. Kamel und F. Karray. Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern Recognition 44(3), 572–587, 2011.

[10] P. Ekman und E. L. Rosenberg. What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS). Oxford University Press Inc., 1997.

[11] I. A. Essa und A. P. Pentland. Coding, analysis, interpretation, and recognition of facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 757–763, Juli 1997.

[12] D. Consoli. A new concept of marketing: the emotional marketing. BRAND. Broad Research in Accounting, Negotiation, and Distribution 1(1), 52–59, 2010.

[13] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch und M. R. Wróbel. "Emotion Recognition and Its Applications". In: Human-Computer Systems Interaction: Backgrounds and Applications 3. Hrsg. von Z. S. Hippe, J. L. Kulikowski, T. Mroczek und J. Wtorek. Springer International Publishing, S. 51–62. URL: https://doi.org/10.1007/978-3-319-08491-6_5.

[14] M. Brand, F. Klompmaker, P. Schleining und F. Weiß. Automatische Emotionserkennung – Technologien, Deutung und Anwendungen. Informatik-Spektrum 35(6), 424–432, Dez. 2012.

[15] T. Ruf, A. Ernst und C. Küblbeck. "Face Detection with the Sophisticated High-speed Object Recognition Engine (SHORE)". In: Microelectronic Systems: Circuits, Systems and Applications. Hrsg. von A. Heuberger, G. Elst und R. Hanke. Springer Berlin Heidelberg, S. 243–252. URL: https://doi.org/10.1007/978-3-642-23071-4_23.

[16] F.-I. für Integrierte Schaltungen. Mit der Bildanalysesoftware SHORE® zur erfolgreichen Erkennung und Auswertung von Gesichtern. URL: https://www.iis.fraunhofer.de/de/ff/sse/ils/tech/shore-facedetection.html (besucht am 21. 06. 2019).

A Anhang