Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Hauptseminar

Harald Kirschenmann

Personenerkennung

Fakultät Technik und Informatik
Department Informatik

Faculty of Engineering and Computer Science
Department of Computer Science

Harald Kirschenmann

Personenerkennung

Abgegeben am 31. August 2018

Inhaltsverzeichnis

1 Einleitung

Foto- und Videokameras finden seit etlichen Jahrzehnten immer weitere Verbreitung. Mit der Entwicklung von Digitalkameras und Camcordern fanden Kameras Einzug in private Haushalte. Zur Überwachung von öffentlichen Verkehrsmitteln, sicherheitskritischen Bereichen und öffentlichen Plätzen wurden Kameras intalliert. Mit Smartphones sind Kameras heutzutage ein alltäglicher Begleiter.

Sei es ein *Smartphone* oder eine *Smartwatch*, wir umgeben uns im Alltag mit Computern. Computer werden kleiner, leistungsfähiger und sind ständig vernetzt. Sind Berechnungen zu aufwändig, so können diese an leistungsfähigere IT-Systeme ausgelagert werden.

Für den Menschen stellt das Sehen und Erkennen von Gegenständen und Personen eine leichte Aufgabe dar, für Maschinen handelt es sich noch immer um eine große Herausforderung. Mit dieser Herausforderung beschäftigt sich ein Teilgebiet der Informatik, die *Computer Vision*. Mit der Verbreitung von Kameras und Computern wächst das Potenzial der Computer Vision. Aufwändigere Verfahren können genutzt werden und immer mehr Bildmaterial steht zur Verfügung, um die Verfahren zu verbessern und testen.

Teilbereiche in der *Computer Vision* sind die *Face Detection*, das Erkennen von Gesichtern, und die *Face Recognition*, das Zuordnen von Identitäten zu den Gesichtern. Der Bedarf, Menschen zu erkennen und zu identifizieren, steigt weiter an. Ausschlaggebend sind hierzu die Menge an Daten und der damit einhergehende Bedarf, diese Datenmengen zu ordnen und klassifizieren, aber auch ein steigendes Sicherheitsbedürfnis unserer Gesellschaft.

Der Zugang zu sicherheitskritischen Bereichen soll mit Hilfe der *Gesichtserkennung* kontrolliert werden. Zur Überwachung wird die Gesichtserkennung bereits ausführlich getestet und verwendet. Am Berliner Bahnhof Südkreuz findet eine Testphase statt, in der die Gesichtserkennung ausführlich erprobt wird, um den öffentlichen Bahnhof zu überwachen. Die Polizei in Hamburg plant, die Gesichtserkennung dauerhaft in Hamburg zu nutzen. Der Einsatz und die Entwicklung von Software zur Gesichtserkennung erzeugt aber auch viel Kritik und Bedenken im Bereich des Datenschutzes. Beispielsweise werden gegen die Pläne der Hamburger Polizei vom Hamburger Datenschutzbeauftragten Bedenken geäußert.

Ein weiteres Anwendungsfeld stellt die *personalisierte Kommunikation* dar. Hierbei soll die Interaktion zwischen dem Menschen und der Maschine durch Wissen über den Menschen angereichert werden. So können die Systeme den Menschen mit Hilfe von Kamerasensoren identifizieren und die Kommunikation anhand des Wissens über die Individuen anpassen.

Die Qualität der unterschiedlichen Verfahren zur Gesichtserkennung unterscheidet sich. Je nach Rahmenbedingungen sind einige Verfahren geeigneter als andere. Im Rahmen dieser Arbeit soll die Möglichkeit erörtert werden, vorhandene Gesichtserkennungsalgorithmen zu evaluieren und zu vergleichen. Hierzu werden Rahmenbedingungen herausgestellt, die die Qualität der Algorithmen beeinflussen können. Mit Hilfe vorhandener Datensätze soll das Verhalten der Verfahren bei diesen Bedingungen gemessen werden.

2 Grundlagen

In diesem Abschnitt wird ein Einblick in das in Kapitel 1 angesprochene Themenfeld der *Personalisierten Kommunikation* gegeben. Im Anschluss wird die *Gesichtserkennung* näher betrachtet. Abschließend wird das *Deep Learning* dargestellt. Viele moderne Verfahren nutzen *Deep Learning* zur Lösung komplexer Aufgaben. Auch im Bereich der *Gesichtserkennung* werden *Deep Learning*-Verfahren verwendet.

2.1 HCI und Personalisierte Kommunikation

Für die Benutzung und Steuerung verschiedenster Computersysteme interagiert der Mensch mit einer Maschine. Die Kommunikation zwischen Mensch und Maschine wird als *Human-Computer Interaction* (HCI) bezeichnet. Die ACM[1] SIGCHI-Arbeitsgruppe stellte folgende Definition für die HCI vor:

> *"Human-computer interaction is a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them." [1]*

Die HCI beschreibt eine Disziplin, die sich mit dem Design, der Evaluation und der Implementation von interaktiven Computer Systemen für die menschliche Nutzung und der Analyse von bedeutenden Phänomenen, die die HCI umgeben, beschäftigt.

Bei der Interaktion zwischen dem Menschen und der Maschine werden Schnittstellen geschaffen, die als *Interface* bezeichnet werden. Die Interfaces sollen möglichst natürlich und unscheinbar sein, so dass diese vom Menschen intuitiv verwendet werden können. [2]

[1] Association for Computing Machinery - http://www.acm.org

2.1.1 Sprache und Gesten

Seit der Entstehung von Computer Systemen entwickelt sich auch die HCI weiter. Die Maus und die Tastatur sind lange bekannte und noch immer weit verbreitete Eingabegeräte zur Kommunikation des Menschen mit der Maschine. Diese Geräte entsprechen keiner natürlichen oder intuitiven Form der Kommunikation.

Natürliche Arten der Kommunikation wären beispielsweise die Sprache oder Gesten. Bei der Sprache handelt es sich um aneinander gereihte Phoneme[2], die vom Menschen oder einer Maschine erkannt und interpretiert werden können [3]. Gesten sind Bewegungen, die ausgeführt werden, um eine Nachricht zu kommunizieren. Im 2-dimensionalen Raum können die Gesten dem „Zoomen“ oder „Wischen“ auf Touch-Displays entsprechen [4]. Im 3-dimensionalen Raum ist die Bandbreite der möglichen Gesten größer. Viele der Körper- und Handbewegungen sind intuitiv anwendbar.

Weitere Gesten können mit der Augenpartie ausgeführt werden. Dies können der Blick in unterschiedliche Richtungen, das Blinzeln oder Augenbrauenbewegungen sein. Die Blickrichtung kann als Zeigegeste interpretiert werden [5][6].

Eine Kombination der verschiedenen Gesten erweitert die Möglichkeiten sich auszudrücken und führt zu einer intuitiveren Interaktion. In der Kommunikation zwischen Menschen werden Gesten häufig kombiniert. So kann mit einer Zeigegeste ein Objekt bestimmt werden und eine weitere Geste speziell für den Kontext dieses Objekts ausgeführt werden [7].

Mit Hilfe von Gesten haben sich eigenständige Sprachen entwickelt, die sich aus einer Kombination von Mimik, Handgesten und Körperhaltung bilden. Neben Gesten für einzelne Wörter, gibt es in den Gebärdensprachen einzelne Handgesten, die Buchstaben in einem Fingeralphabet[3] repräsentieren [8].

2.1.2 Personalisierte Kommunikation

In vielen Anwendungsszenarien eines Computersystems ist es sinnvoll, wenn das System die Nutzer, mit denen interagiert wird, unterscheiden und auf Hintergrundwissen zugreifen kann. Sei es der Fahrkartenautomat, der einem die bevorzugten Tickets vorschlägt, oder das Smart Home, das sich an die Individuellen Bedürfnisse anpasst. Viele Systeme kommunizieren mit allen Nutzern auf die gleiche Art und Weise, eine individuelle Kommunikation findet selten statt.

[2] Phoneme – Laute, die aneinander gereiht Worte ergeben
[3] Fingeralphabet - www.sign-lang.uni-hamburg.de/fa/

Mit dem Wissen über das Individuum kann die Kommunikation zwischen Mensch und Maschine verbessert werden. Das Interesse und die Verbindlichkeit werden durch einfache Namensnennung gesteigert. Der Nutzer wird direkt angesprochen, hierdurch wird die Aufmerksamkeit auf die Kommunikation gelenkt. Dies wird in dem Szenario 2.1.3 *Peter – the smart plant* anschaulich.

In der Interpretation sind viele intuitiv ausgeführte Gesten nicht eindeutig. Das Verständnis von Gesten kann sich bei zwei Menschen drastisch unterscheiden. So kann eine positiv gemeinte Geste, wie der nach oben gerichtete Daumen, von anderen Menschen als Beleidigung interpretiert werden [9]. Auch das Kopfnicken wird nicht universell als eine Geste für das „Ja“ verwendet. In verschiedenen Teilen der Welt wird das Kopfnicken als „Nein“ interpretiert oder es werden andere Gesten verwendet. Bei der Interaktion mit einer Maschine verwenden Nutzer intuitiv unterschiedliche Gesten für die gleichen Befehle. In Untersuchungen haben nur 43,5 % der Nutzer zuvor definierte Gesten verwendet [10].

Auch die Gebärdensprache ist nicht eindeutig. Wie die gesprochene Sprache, haben sich die Gebärdensprachen unabhängig voneinander entwickelt, so dass unterschiedliche Sprachen und Dialekte historisch entstanden sind.

Für die personalisierte Kommunikation muss der Nutzer vom System unterschieden werden können. Ohne vorherige Authentifizierung über eine Eingabemaske ist der Maschine häufig nicht bekannt, mit welchem Individuum sie kommuniziert. Eine Unterscheidung mit dem Handy oder weiteren Geräten, wie den Ubitags des UbiSense-Systems [11] ist möglich. Hierzu wird das Mitführen der Geräte benötigt. Das Mitführen von Geräten zur Identifizierung erschwert die Nutzung der Systeme und führt nicht zu dem Ziel der HCI eines unscheinbar und intuitiv zu bedienenden Interfaces.[2] Visuelle Identifizierungsverfahren, wie die Gesichtserkennung, kommen ohne mitgeführte Geräte aus.

Für die personalisierte Kommunikation müssen die Nutzer zuverlässig erkannt und identifiziert werden. Des Weiteren soll die Kommunikation individuell und personenbezogen gestaltet werden.

2.1.3 Peter - the smart plant

Im Rahmen eines Bachelorprojekts wurde die Intelligente Pflanze „Peter“ im *Creative Space for Technical Innocations*[4] (CSTI) entwickelt. Die Pflanze ist mit verschiedenen Sensoren, wie einem Erdfeuchte-Sensor und einer Kinect[5]-Kamera, ausgestattet. Darüber hinaus verfügt die Pflanze über Lautsprecher, mit denen die Pflanze mit Menschen, die sich in der Nähe befinden, kommunizieren kann.

[4] CSTI - csti.haw-hamburg.de/
[5] Kinect - www.xbox.com/de-DE/xbox-one/accessories/kinect

Pflanzen benötigen eine ausreichende Menge an Wasser, um leben und wachsen zu können. Da sich Peter im Inneren eines Gebäudes befindet, kann die natürliche Wasserversorgung, wie Regen oder Grundwasser, den Wasserbedarf nicht decken. Peter ist darauf angewiesen, von Menschen gegossen zu werden. Die Pflanze darf nicht vertrocknen, aber auch nicht zu viel Wasser erhalten. Der Erdfeuchte-Sensor misst beständig die richtige Menge an Wasser.

Mit der Kinect-Kamera kann die Pflanze die Umgebung wahrnehmen und erkennen, ob eine Person anwesend ist. Über die Lautsprecher kann Peter mit den Personen in Kontakt treten und ihr Bedürfnis nach Wasser mitteilen. Die Pflanze hat kein Hintergrundwissen zu den Personen, da diese nicht unterschieden werden. Es ist den Personen überlassen, ob diese Peter mit Wasser gießen.

Wird Peter mit der Fähigkeit ausgestattet, die Personen zu erkennen und bekannten Individuen zuzuordnen, erweitern sich die Handlungsmöglichkeiten der intelligenten Pflanze. Es kann auf bestehende Hintergrundinformationen zu den Personen zugegriffen und daran die Handlungsweise angepasst werden.

Wenn bekannt ist, dass die Person die Pflanze nicht gießen wird oder nicht angesprochen werden möchte, kann sich die Pflanze dazu entscheiden, die Kommunikation zu unterlassen. Befindet sich eine Personengruppe in der Nähe, unter der sich Personen befinden, die die Pflanze gießen, so können diese direkt mit Namensnennung angesprochen werden. Mit der Namensnennung wird das Interesse auf die Kommunikation mit der Pflanze gelenkt und aufgrund der persönlichen Ebene eine höhere Verbindlichkeit geschaffen, dass die Pflanze gegossen wird. Durch die Auswahl bestimmter Person, erhöht sich die Wahrscheinlichkeit, dass die Pflanze mit Wasser versorgt wird, zusätzlich.

2.2 Gesichtserkennung

Die Gesichtserkennung ist ein Teilgebiet der *Computer Vision* und ermöglicht die Identifizierung von Menschen. Die Identifizierung der Personen wird mit Bildern der Gesichter durchgeführt. Hierzu werden Bilder aus Kamera-Sensoren verwendet, die zu identifizierenden Personen benötigen keine Geräte, die sie mit sich zu tragen haben. Im Gegensatz zu Fingerabdruckscannern oder vergleichbaren Verfahren kann die Identifizierung kontaktlos und in einiger Entfernung durchgeführt werden.

Das Erkennen und Zuordnen von Gesichtern zu einzelnen Personen ist für den Menschen alltäglich. In den ersten Lebensmonaten erlernt der Mensch die Fähigkeit der Gesichtserkennung.

Anwendung findet die Gesichtserkennung in der Zutrittskontrolle, Identifikation, Überwachung, dem Aufdecken von Identitätsdiebstahl und bei der Überprüfung von

Pässen. Im Besonderen aufgrund der Möglichkeiten, die sich mit der Gesichtserkennung im Bereich der Überwachung ergeben, wird die Gesichtserkennung kontrovers diskutiert und von vielen Seiten kritisiert.

Die Bandbreite der Anwendungsszenarien steigt stetig an. Zur Zeit wird an Systemen gearbeitet, die blinde Menschen im Alltag unterstützen sollen und auf die Gesichtserkennung zurückgreifen. [11][12]

In klassischen Verfahren erfolgt die Identifizierung der Personen anhand biometrischer Merkmale eines Gesichtes. Die biometrischen Merkmale lassen sich nicht einfach, wie die Mimik, ändern. Hier werden die Abstände der Augen, Nase, Mund und des Kinns geometrisch genutzt. Ein weiterer Ansatz, der in klassischen Verfahren genutzt wird, ist das *Template Matching*. Hier wird versucht, bestimmte Bereiche eines Bildes zu finden, die mit einem zuvor erstellten Template übereinstimmen. [14] Moderne Verfahren setzen auf den Ansatz des Deep Learnings (siehe Kapitel 2.3) und erzielen hiermit gute Ergebnisse.

2.2.1 Anforderungen an die Gesichtserkennung

Systeme zur Gesichtserkennung haben besondere Anforderungen und Herausforderungen zu erfüllen. Im Hinblick auf die möglichen Anwendungsfelder, wie der Zugangskontrolle und der Überwachung, ist eine zuverlässige und korrekte Erkennung von realen Personen besonders wichtig und falsche Treffer sind auszuschließen. Menschen dürfen nicht verwechselt werden oder als anwesend erkannt werden, wenn diese nicht anwesend sind.

Befindet sich eine Personengruppe im Bereich des Kamerabildes, so kann es zu Verdeckungen kommen. Die Seitenansicht von Personen stellt eine weitere Herausforderung an die Verfahren dar. Viele Verfahren arbeiten mit der frontalen Perspektive auf das Gesicht, um die Menschen zu identifizieren. Es gibt verschiedene Lösungsansätze, damit die Gesichter auch aus einer seitlichen Perspektive erkannt werden können [15][16]. Ein vielversprechender Ansatz ist die Berechnung der frontalen Perspektive aus einer seitlichen Perspektive mit Hilfe von *Generative Adversarial Networks* [17].

Die Erkennung von Menschen soll sich auf real anwesende Menschen beschränken. Abgebildete Gesichter auf Postern, Masken oder digitalen Geräten sollen nicht beachtet werden. Hierzu ist der Einsatz weiterer Sensoren, wie eine Wärmebildkamera oder eine 3-D Kamera, denkbar.

Die Qualität der Bilder und die Aufnahmebedingungen, wie die Belichtung, können sich erheblich unterscheiden. Gute Verfahren für die Gesichtserkennung müssen mit den unterschiedlichen Rahmenbedingungen zurechtkommen.

Viele Verfahren arbeiten im 2-dimensionalen Raum. Im 3-dimensionalen Raum stehen den Verfahren mehr Merkmale zur Verfügung, mit denen die Menschen unterschieden werden können. [18][19]

2.2.2 Klassischer Ablauf der Gesichtserkennung

Klassische Verfahren der Gesichtserkennung unterteilen den Prozess der Identifikation von Personen in drei Schritte, der *Face Detection*, *Feature Extraction* und der *Face Recognition*.

In der ersten Phase, der *Face Detection*, wird ein gegebenes Bild nach Gesichtern durchsucht und diese lokalisiert. In dieser Phase kann eine Vorverarbeitung der gefundenen Gesichter stattfinden. Zur Normalisierung werden die Gesichter in Bezug auf den Aufnahmewinkel und den Abstand vereinheitlicht. Die Gesichter werden auf die benötigten Bildinformationen reduziert.

Bei der *Feature Extraction* werden die benötigten Merkmale des gefundenen Gesichtes extrahiert. Je nach Verfahren können als Merkmale die Augen, Nase, Mund, Kinnpartie und deren geometrischen Zusammenhänge verwendet werden. Wird auf das Verfahren des *Template Matching* gesetzt, so wird in dieser Phase ein *Template* des Gesichtes extrahiert.

In der Phase der *Face Recognition* werden die extrahierten Merkmale bzw. das extrahierte Template mit vorhandenen Templates verglichen. Für die *Verifikation* eines Gesichtes wird bestimmt, ob es sich bei dem gefundenen Gesicht um das gesuchte Gesicht handelt. Bei der Identifizierung werden die Merkmale bzw. das Template mit den Daten einer Datenbank verglichen. Werden übereinstimmende Merkmale bzw. ein übereinstimmendes Template gefunden, entspricht das gefundene Gesicht der Person aus der Datenbank und konnte identifiziert werden.

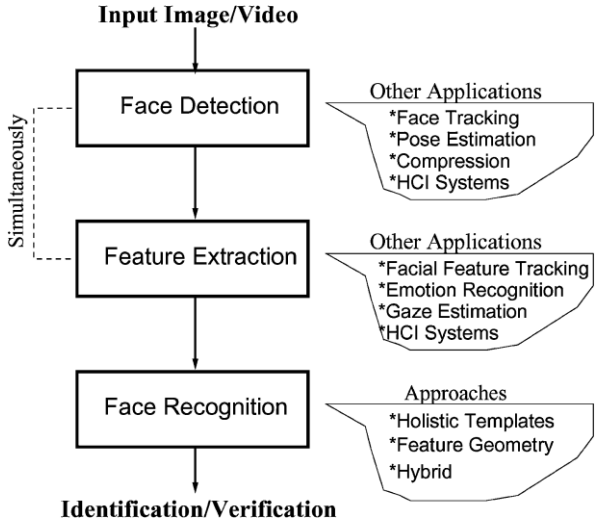

Fig. 1: Ablauf der Gesichtserkennung und mögliche Anwendungen [14]

Für das Lernen von Gesichtern, werden die Merkmale oder Templates des Gesichts in einer Datenbank gespeichert und einer Person zugeordnet. Für das spätere Identifizieren oder Verifizieren stehen diese als Referenz zur Verfügung.

Werden die drei Phasen ausgeführt, können Personen anhand ihres Gesichts identifiziert oder verifiziert werden. Wird nur die Phase der *Face Detection* genutzt, so lassen sich beispielsweise Gesichter in Videos verfolgen und die Ausrichtung des Gesichts bestimmen. Werden die erste und die zweite Phase, die *Feature Extraction* ausgeführt, können die gezeigten Emotionen erkannt und die Blickrichtung des Gesichts ermittelt werden.

2.3 Deep Learning

In modernen Ansätzen zur Lösung komplexer Problemstellungen, wie im Bereich der Computer Vision oder im Speziellen der Gesichtserkennung, kommen häufig Verfahren im Bereich des *Machine Learning* oder des *Deep Learning* zum Einsatz. Verfahren des *Deep Learning* basieren auf *Künstlichen Neuronalen Netzen* und sind in der Lage, eigenständig Lösungswege für bestimmte Aufgaben zu erlernen.

Neuronale Netze nehmen Informationen auf, verarbeiten diese Informationen, können das Netz selbst modifizieren und geben eine Antwort auf Basis des vorherigen Inputs aus. Ein Neuronales Netz besteht aus vielen *Neuronen*, die in sogenannten *Layer* angeordnet sind. Die Neuronen sind mit Kanten aus einem Layer in den Kanten des darauffolgenden *Layer* verbunden.

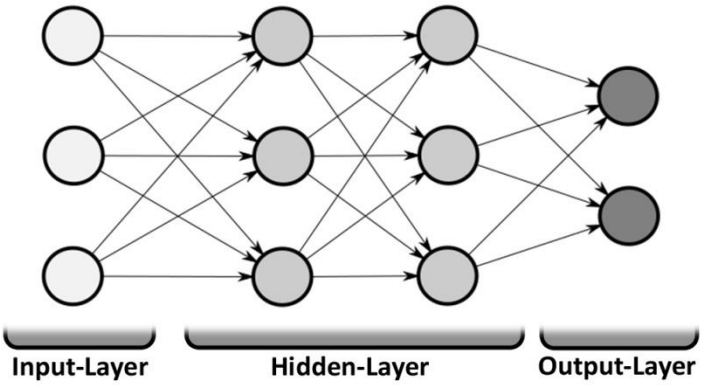

Fig. 2. Schematische Darstellung eines Neuronalen Netzes

In der schematischen Darstellung Fig. 2 befindet sich auf der linken Seite der Input-Layer. Hier erhält das Netz von außen den Input, verarbeitet diesen und gibt den berechneten Output der Neuronen im *Input-Layer* an die Neuronen des nächsten Layers weiter. Auf der rechten Seite befindet sich der *Output-Layer*, hier wird die Ausgabe des Neuronalen Netzes ermittelt. Zwischen dem *Input-* und dem *Output-Layer* können sich mehrere Schichten weiterer Layer, den *Hidden-Layer*, befinden. *Deep Learning*-Netze enthalten komplexere Strukturen, die sich in den Hidden-Layern befinden.

Die Verbindungen zwischen den einzelnen Neuronen sind gewichtet. Je größer die Gewichtung einer Verbindung, desto mehr Einfluss hat das Neuron auf die nachfolgenden Neuronen. In vielen Fällen repräsentiert die Kantengewichtung der einzelnen Verbindungen zwischen den Neuronen das gespeicherte Wissen eines Netzes. [21]

Damit Neuronale Netze bestimmte Aufgaben erlernen, müssen die Netze trainiert werden. Für den Lern- bzw. Trainingsprozess werden viele Daten benötigt. Die Daten werden in Trainings- und Testdaten aufgeteilt. Mit den Trainingsdaten wird das Netz trainiert und mit den Testdaten wird überprüft, wie gut das Netz die gewünschte Aufgabe erfüllt. Während des Trainings wird das Netz so modifiziert, dass sich bei einer Wiederholung der Eingabe das Ergebnis verbessert. Bei der Modifizierung werden die Kantengewichtungen verändert, jedoch sind Änderungen des Schwellenwertes, der Aktivitätsfunktion in den Neuronen und ein Hinzufügen oder Entfernen von Neuronen ebenfalls möglich. Für die konkrete Herangehensweise an das Training von Netzen gibt es mehrere Möglichkeiten [22]

Für die Neuronalen Netze gibt es unterschiedliche Netztypen, die sich für verschiedene Zwecke eignen. Im Bereich des Computer Vision haben sich die *Convolutional Neural Networks* (CNN) bewährt. CNNs wurden frühzeitig von LeCunn et al. zur Texterkennung verwendet [23]. Heutzutage werden CNNs vielfach für die *Face Recognition* und die *Face Detection* eingesetzt [24][25][26].

3 Benchmark

Für die Problemstellung der Gesichtserkennung existieren viele verschiedene Verfahren. Alle haben sie kleinere oder größere Unterschiede. Die Genauigkeit und die Geschwindigkeit unterscheiden sich oder die Verfahren sind unter bestimmten Rahmenbedingungen besonders akkurat. Um zu untersuchen, was heutige Gesichtserkennungsverfahren leisten können, muss eine Vergleichbarkeit zwischen diesen Verfahren geschaffen werden. Diese lässt sich erreichen, wenn die Bedingungen, unter denen Messungen durchgeführt werden, für alle Verfahren gleich sind. Dies bedeutet, dass die Verfahren auf Basis der gleichen Daten Gesichter bestimmten Identitäten zuordnen müssen und auch, dass für alle Verfahren die gleichen Messungen durchgeführt werden. Für die Analyse der Gesichtserkennungsverfahren sind die Auswahl der Daten und die Messergebnisse entscheidend. Auf die verschiedenen Messungen wird in 3.1 näher eingegangen. Eine Übersicht zu den verfügbaren Datensätzen wird in 3.2 dargestellt. Viele der vorgestellten Datensätze sind gemeinsam mit dazugehörigen Benchmarks veröffentlicht worden.[27] Weitere Benchmarks in wissenschaftlichen Veröffentlichung zu finden.[28] Für das Benchmark von NIST[6] werden weiterhin Messungen durchgeführt.[29]

3.1 Untersuchungen

Für das Benchmark und eine anschließende Analyse der Ergebnisse sollen viele Rahmenbedingungen beachtet werden. Wichtige Aspekte, bei denen die Verfahren überprüft werden sollen, sind die Qualität der Bilder, die Eigenschaften der Testpersonen, die Anzahl der Vergleichsbilder, die Emotionen der Testpersonen und weitere Rahmenbedingungen. Hierbei sollen einige verschiedene Messwerte ermittelt werden, mit denen anschließend die Verfahren analysiert werden können.

Die Qualität der Bilder ist ein wichtiger Aspekt in allen Bereichen der Computer Vision. Die Auflösung und die vorhandenen Bildinformationen beeinflussen direkt, welche Informationen die Daten enthalten und von den Verfahren genutzt werden können. Ist die Auflösung zu gering, dann können die Gesichter nicht mehr unterschieden werden. Der Farbraum der Bilder spielt ebenfalls eine Rolle. Bilder können in Schwarz/Weiß bzw. Graustufen, mit Farbwerten oder gar mit Thermalwerten vorliegen.

Neben der Auflösung und dem Farbraum sind Rahmenbedingungen, wie die Belichtung der Gesichter, die Perspektive und die Entfernung zum Aufnahmeobjekt weitere Punkte, die

[6] NIST: National Institute of Standards and Technology - https://www.nist.gov/

vom Benchmark abgedeckt werden sollen. Das ISO-Rauschen bei Aufnahmen in dunklen Umgebungen oder verschwommene Aufnahmen, wenn sich das Objekt bewegt, sind zu beachten.

Ein weiterer Punkt ist die Auswahl der Testpersonen. Eine homogene Zusammensetzung der Geschlechter, des Alters und der Hautfarbe der Testpersonen ist wünschenswert. Viele Datensätze beachten eine ausgeglichene Zusammensetzung der Geschlechter, jedoch nicht des Alters oder der Hautfarbe.

Die Mimik der Testpersonen sollte bei Gesichtserkennungsverfahren keine Rolle spielen. Personen mit lachenden, weinenden oder neutralen Emotionen müssen zuverlässig erkannt werden können. Die Verfahren sollten ebenso robust gegenüber Verdeckungen einzelner Gesichtspartien sein.

Mit den unterschiedlichen Rahmenbedingungen sollen möglichst viele Varianten abgedeckt werden, auf die die Verfahren im realen Einsatz treffen können, seien es die Urlaubsbilder, Bilder aus einer Überwachungskamera oder Videobilder.

Für aussagekräftige Analysen des Benchmarks, müssen Messungen durchgeführt werden. Auf Basis der Messergebnisse können Aussagen über die Qualität der Verfahren getroffen werden. In vielen Anwendungsfällen kann die Geschwindigkeit der Verfahren einen Faktor darstellen, daher wird eine Zeitmessung durchgeführt. Die Qualität der Verfahren wird entscheidend durch die Genauigkeit beeinflusst. Daher ist es wichtig zu messen, wie gut die Erkennungsrate ist und die falschen Treffer näher zu betrachten: Wie hoch ist die Rate an Personen, die erkannt wurden, obwohl sich diese nicht auf dem Bild befinden (False Positive) und wie hoch ist die Rate an Personen, die sich auf dem Bild befinden, allerdings nicht erkannt wurden (False Negative)?

3.2 Daten

In diesem Abschnitt wird eine Übersicht über einige Datenbanken gegeben, die für wissenschaftliche Zwecke zur Verfügung stehen. Eine Übersicht über die Größe der Datenbanken ist in Fig. 1 zu betrachten. Die Datenbank MS-Celeb-1M ist in der Auflistung nicht enthalten. Diese Datenbank enthält mit 10 Millionen Bildern deutlich mehr als die restlichen Datenbanken.[30] Im Nachfolgenden werden exemplarisch einige der Datensätze beschrieben.

In einigen Datensätzen, wie der *Disguised Faces in the Wild*-Datensatz, wurden die Bilder in Trainings- und Testdaten unterteilt. [31] Viele Datensätze richten sich speziell auf einzelne Aspekte, die in 3.1 benannt wurden. In der *Large Age-Gap*-Datenbank befinden sich Bilder von erwachsenen Menschen und deren Kinderbilder. Mit diesen Bildern lässt sich überprüfen, wie sich die Algorithmen bei großen Altersunterschieden verhalten. [32]

Die *IIITD In and Beyond Visible Spectrum Disguise*-Datenbank enthält Thermalbilder. Des Weiteren sind viele Gesichter z.B. mit Brillen, Perücken, Bärten und Masken verdeckt. [33][34]

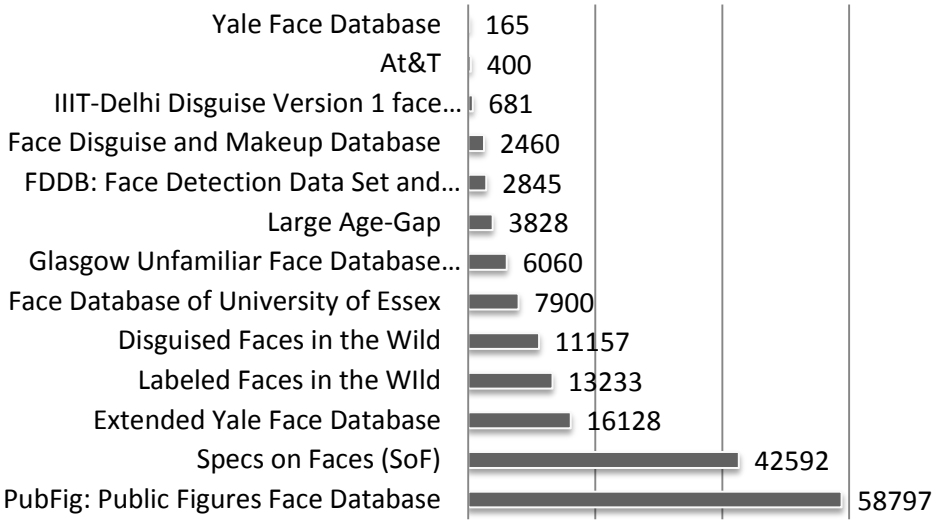

Fig. 3: Anzahl der Bilder in den Datenbanken

4 Ausblick

Mit dieser Ausarbeitung wurden Methoden und Verfahren einer Gesichtserkennung vorgestellt und eine Übersicht über bestehende Bild-Datenbanken gegeben. Mit verschiedenen Messungen und Rahmenbedingungen, die überprüft werden sollen, kann so ein umfangreiches Benchmark entwickelt werden. Und somit zu einer Übersicht über die Leistung der Verfahren in den einzelnen Aspekten beitragen.

jeder Anforderung muss eine passende Datenbank zugeordnet werden, damit zu diesen Aspekten im Benchmark Messungen für die unterschiedlichen Verfahren durchgeführt und diese evaluiert werden können.

Mit weiteren Messungen lässt sich ermitteln, ob die Vorverarbeitungen der Bilder sinnvoll sind. Denkbar sind hier das passende Zuschneiden der Gesichter, die Rekonstruktion der frontalen Ansicht [17] oder die Rekonstruktion verdeckter Bereiche [35].

5 Literatur

[1] HEWETT, T. T.; BAECKER, R.; CARD, S.; CAREY, T.; GASEN, J.; MANTEI, M.; PERLMAN, G.; STRONG, G. & VERPLANK, W.: ACM SIGCHI Curricula for Human-Computer Interaction, New York, NY, USA: *ACM*., 1992.

[2] WACHS, J. P.; KÖLSCH, M.; STERN, H. & EDAN, Y.: Vision-based Hand-gesture Applications. In: *Commun. ACM* 54 (2011), Nr. 2, S. 60—71

[3] SARMA, M. & SARMA, K. K.: Recognition of Assamese Phonemes Using Three Different ANN Structures. In: *ACM*. : *Proceedings of the CUBE International Information Technology Conference*., 2012, S. 299—302.

[4] WOBBROCK, J. O.; MORRIS, M. R. & WILSON, A. D.: User-defined Gestures for Surface Computing. In:: *ACM*. : *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*., 2009, S. 1083—1092.

[5] MAJARANTA, P. & BULLING, A.: Eye Tracking and Eye-Based Human–Computer Interaction. In: FAIRCLOUGH, S. H. & GILLEADE, K. (Hrsg.), London: *Springer London*. : *Advances in Physiological Computing*., 2014, S. 39—65.

[6] JACOB, R. J. K.: The Use of Eye Movements in Human-computer Interaction Techniques: What You Look at is What You Get. In: *ACM Trans. Inf. Syst.* 9 (1991), Nr. 2, S. 152—169.

[7] KOELSCH, M.; BANE, R.; HOELLERER, T. & TURK, M.: Multimodal Interaction with a Wearable Augmented Reality System. In: *IEEE Comput. Graph. Appl.* 26 (2006), Nr. 3, S. 62—71.

[8] ZAFRULLA, Z.; BRASHEAR, H.; STARNER, T.; HAMILTON, H. & PRESTI, P.: American Sign Language Recognition with the Kinect. In:: *ACM*. : *Proceedings of the 13th International Conference on Multimodal Interfaces*., 2011, S. 279—286

[9] MORRIS, D.; COLLETT, P.; MARSH, P. & O'SHAUGHNESSY, M..: Gestures: Their Origin and Meanings. London: Cape, 1978.

[10] WOBBROCK, J. O.; MORRIS, M. R. & WILSON, A. D.: User-defined Gestures for Surface Computing. In:: *ACM*. : *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*., 2009, S. 1083—1092.

[11] ABDOLRAHMANI, A.: Examining Facial Recognition Technology to Augment the Indoor Navigation Experience of Individuals Who Are Blind. In: *SIGACCESS Access. Comput.* (2017), Nr. 117, S. 35--38

[12] BRANHAM, S. M.; ABDOLRAHMANI, A.; EASLEY, W.; SCHEUERMAN, M.; RONQUILLO, E. & HURST, A.: "Is Someone There? Do They Have a Gun": How Visual Information About Others Can Improve Personal Safety Management for Blind Individuals. In:: *ACM*. : *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*., 2017, S. 260--269

[13] STEGGLES, P.; GSCHWIND S.: *The Ubisense Smart Space Platform*. In: Adjunct Proceedings of the Third International Conference on Pervasive Computing, Vol. 191*, Mai 2005,* S. 73 – 76.

[14] BRUNELLI, R.; POGGIO, T.: Face Recognition – Features versus Templates. In:: *IEEE* Transactions on pattern analysis and machine intelligence, Vol. 15, No10, 1993, S. 1042 - 1052.

[15]CAMENT, L. A.; GALDAMES, F. J.; BOWYER, K. W. & PEREZ, C. A.: Face recognition under pose variation with active shape model to adjust Gabor filter kernels and to correct feature extraction location. In:: . 1 : *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*., 2015, S. 1-6

[16]VU, N. S. & CAPLIER, A.: Efficient statistical face recognition across pose using Local Binary Patterns and Gabor wavelets. In:: . : *2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*., 2009, S. 1-5

[17]HUANG, R., ZHANG, S., LI, T. & HE, R.: Beyond Face Rotation: Global and Local Perception GAN for Photorealistic and Identity Preserving Frontal View Synthesis. In: *2017 IEEE International Conference on Computer Vision,* S. 2458-2467

[18]MEDIONI, G. & WAUPOTITSCH, R.: Face modeling and recognition in 3-D. In:: . : *2003 IEEE International SOI Conference. Proceedings (Cat. No.03CH37443)*., 2003, S. 232-233.

[19]WANG, Y.; LIU, J. & TANG, X.: Robust 3D Face Recognition by Local Shape Difference Boosting. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (2010), Nr. 10, S. 1858—1870.

[20] Zhao, W.; Chellappa, R.; Phillips, P. J. & Rosenfeld, A.: Face Recognition: A Literature Survey. In: *ACM Comput. Surv.* 35 (2003), Nr. 4, S. 399—458.

[21] REY, G. D. & WENDER, K. F.: Neuronale Netze - Eine Einführung in die Grundlagen, Anwendungen und Datenauswertung: *Hans Huber Verlag*. 2. Auflage, 2017

[22]BISHOP, C. M.: Pattern Recognition and Machine Learning (Information Science and Statistics), Berlin, Heidelberg: *Springer-Verlag*., 2006

[23]Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition.Proceedings of the IEEE 86, 11 (1998), 2278–2324.

[24]TRIANTAFYLLIDOU, D.; NOUSI, P. & TEFAS, A.: Fast Deep Convolutional Face Detection in the Wild Exploiting Hard Sample Mining. In: *Big Data Research* 11 (2018), S. 65 – 76

[25]SUNDARARAJAN, K. & WOODARD, D. L.: Deep Learning for Biometrics: A Survey. In: *ACM Comput. Surv.* 51 (2018), Nr. 3, S. 65:1--65:34

[26]WU, Y.; LI, J.; KONG, Y. & FU, Y.: Deep Convolutional Neural Network with Independent Softmax for Large Scale Face Recognition. In:: *ACM*. : *Proceedings of the 2016 ACM on Multimedia Conference*., 2016, S. 1063—1067

[27]HUANG, Z.; MEMBER, S.; SHAN, S.; MEMBER, S.; WANG, R.; ZHANG, H.; LAO, S.; KUERBAN, A.; CHEN, X. & MEMBER, S.: A Benchmark and Comparative Study of Video-Based Face Recognition on COX Face Database

[28]FERRARA, M.; FRANCO, A.; MAIO, D. & MALTONI, D.: Face Image Conformance to ISO/ICAO Standards in Machine Readable Travel Documents. In: *IEEE Transactions on Information Forensics and Security* 7 (2012), Nr. 4, S. 1204-1213

[29] GROTHER, P.; NGAN, M.; HANAOKA, K.: Ongoing Face Recognition Vendor Test (FRVT). Report. URI: https://www.nist.gov/sites/default/files/documents/2018/06/21/frvt_report_2018_06_21.pdf

[30]GUO, Y.; ZHANG, L.; HU, Y.; HE, X. & GAO, J.: MS-Celeb-1M: A Dataset and Benchmark for Large Scale Face Recognition. In:: . : *European Conference on Computer Vision*., 2016

[31]KUSHWAHA, V.; SINGH, M.; SINGH, R.; VATSA, M; RATHA, N; CHELLAPPA, R.: Disguised Faces in the Wild, IEEE International Conference on Computer Vision and Pattern Recognition Workshop on Disguised Faces in the Wild, 2018

[32]BIANCO, S.: Large Age-Gap Face Verification by Feature Injection in Deep Networks. In: *Pattern Recognition Letters* 90 (2017), S. 36-42

[33]T. I. Dhamecha, R. Singh, M. Vatsa, and A. Kumar, Recognizing Disguised Faces: Human and Machine Evaluation, *PLoS ONE*, 9(7): e99212, 2014.

[34] T. I. Dhamecha, A. Nigam, R. Singh, and M. Vatsa Disguise Detection and Face Recognition in Visible and Thermal Spectrums, *In proceedings of International Conference on Biometrics, 2013*

[35]LIU, G.; REDA, F. A.; SHIH, K. J.; WANG, T.-C.; TAO, A. & CATANZARO, B.: Image Inpainting for Irregular Holes Using Partial Convolutions. (2018)