

Institut für Visualisierung und Interaktive Systeme

Abteilung Mensch-Computer-Interaktion (MCI)

Universität Stuttgart

Pfaffenwaldring 5a

70569 Stuttgart

Studienarbeit Nr. 2450

**Assistenzsysteme mit Emotionserkennung.
Prototypische Realisierung mit Betrachtung
der ethischen Dimension**

Sandra, Boffo

Studiengang: Informatik

Prüfer: Prof. Dr. Albrecht Schmidt

Betreuer: Oliver Korn M.A.

begonnen am: 09.01.2014

beendet am: 28.07.2014

CR-Klassifikation: H.5.2, H.5.1, K.3.1

Inhaltsverzeichnis

1	Einleitung	8
2	Theoretische und philosophische Grundlagen	9
2.1	Der Begriff des Glücks	9
2.2	Motivation und Arbeit	11
2.2.1	Maslows Motivationstheorie	12
2.2.2	Herzbergs Zwei-Faktoren-Theorie	13
2.2.3	Fazit	15
2.3	Emotionen	15
2.3.1	Facial Action Coding System (FACS)	18
2.4	Praktische Umsetzung mit Gamification	19
2.4.1	Zielgruppen	20
2.4.2	Notwendigkeit eines Feedback	21
2.4.3	Mögliche Auswirkungen	21
3	Stand der Wissenschaft - Techniken der Emotionsklassifizierung	23
3.1	Affektive MCI	23
3.2	Herzschlagfrequenz	28
3.3	Methoden der Emotionsklassifikation	32
3.3.1	Viola-Jones-Detektor (<i>face detection</i>)	33
3.3.2	Extraktion relevanter Regionen - SURF	39
3.3.3	Klassifikation der Merkmale (BoW + SVM)	46
4	Emotionserkennungssystem	55
4.1	OpenCV : Viola-Jones-Algorithmus	55
4.2	Accord.NET : BoW, inklusive SURF und k-means-Clustering	57
4.3	Accord.NET: Support Vektor Maschine	58
5	Studie	60
5.1	Technisches System (Wizard-of-Oz)	60
5.1.1	Bestandteile des Systems	61
5.1.2	Struktur und Ablauf	64
6	Auswertung	66
6.1	Daten	66

6.2	Interpretation der Daten	70
6.2.1	Statistische Interpretation - Signifikanztest	70
6.2.2	Statistische Interpretation - Diagramme	72
6.2.3	Beobachtungen	74
7	Ethisch-philosophische Aspekte	77
8	Fazit	78
9	Literaturverzeichnis	80
9.1	Literaturnachweis	80
9.2	Onlinequellen	83
9.2.1	Bibliotheken	85
A	Anhang	86
	Eidesstattliche Erklärung	89

Abbildungsverzeichnis

Abb. 2.1:	Motivationszyklus	12
Abb. 2.2:	Bedürfnishierarchie nach Maslow,1972	12
Abb. 2.3:	Vergleich der “satisfiers” mit “dissatisfiers” nach Herzberg.	14
Abb. 2.4:	Zusammenspiel von Motivations- und Hygienefaktoren	14
Abb. 2.5:	Basisemotionen nach EKMAN	16
Abb. 2.6:	Übersicht über die zugeordneten AUs	18
Abb. 2.7:	Übersicht über die AUs der EMFACS nach EKMAN	19
Abb. 3.1:	Überblick über die Struktur des Demokoffers des Fraunhofer Institut	27
Abb. 3.2:	Überblick über die Anwendung der Eulersche Videoverstärkung	29
Abb. 3.3:	MIT - Eulersche Videoverstärkung	29
Abb. 3.4:	Verschiedene Lösungen zur Herzschlagmessung	31
Abb. 3.5:	Lage des cosinuss-Sensor im Gehörgang	32
Abb. 3.6:	Basisemotionen	32
Abb. 3.7:	Umwandlung eines Grauwertbildes in ein Integralbild	34
Abb. 3.8:	Berechnung der nächsten Pixelsumme	35
Abb. 3.9:	Berechnung der Pixelsumme eines beliebigen Rechtecks	35
Abb. 3.10:	Haar-Merkmale eines Gesichts	36
Abb. 3.11:	Haar-Merkmale der Viola-Jones-Methode	36
Abb. 3.12:	Oktaven von SIFT	40
Abb. 3.13:	Bestimmung der Orientierung eines Gradienten	41
Abb. 3.14:	Haar-Filter für die Bestimmung der Orientierung des Keypoints in seiner Umgebung	44
Abb. 3.15:	Berechnung des Vektors mit dem größten Wert anhand der Haar-Filter-Ergebnisse	45
Abb. 3.16:	Berechnung des Deskriptors mit ausgerichtetem Quadrat	45
Abb. 3.17:	Histogramm als Wortwolke zum Märchen “Schneewittchen” dargestellt. Die Wörter, die am Häufigsten vorkommen, sind größer dargestellt.	46
Abb. 3.18:	Charakteristische Textone in einer Textur und das entsprechende Histogramm	47
Abb. 3.19:	Merkmale sechs gleicher Objekte anschaulich für nur 3 Merkmale dargestellt .	49
Abb. 3.20:	Übersicht über mögliche Clusterverfahren	49
Abb. 3.21:	Dendrogramm als Streudiagramm	50
Abb. 3.22:	Von der Menge der Merkmalsvektoren mit k-means Clustering zum visuellen Wörterbuch	52
Abb. 3.23:	Trennung der Objekte durch eine Hyperebene H	53

Abb. 3.24: Struktur der 1-vs-all-SVM	54
Abb. 4.1: Charakteristische Punkte im Gesicht	57
Abb. 4.2: UML-Diagramm der SVM-Klassen in Accord.NET	59
Abb. 5.1: Anzeige des LEGO Digital Designer für die Studie	61
Abb. 5.2: Unterschiedliche Anzeigen des LEGO Digital Designer beim Aufbau	62
Abb. 5.3: System für den Durchgang ohne Gamification	62
Abb. 5.4: Emotionserkennung mit SHORE TM	63
Abb. 5.5: Bedieneroberfläche ohne Gamification	64
Abb. 5.6: Bedieneroberfläche mit Gamification	64
Abb. 6.1: Entwicklung der Emotionen ohne Gamification	67
Abb. 6.2: Entwicklung der Emotionen mit Gamification	67
Abb. 6.3: Entwicklung "Glücklich"	68
Abb. 6.4: Entwicklung "Genervt"	68
Abb. 6.5: Entwicklung "Nicht glücklich"	69
Abb. 6.6: Entwicklung der Fehleranzahl	69
Abb. 6.7: Entwicklung der emotionalen Selbsteinschätzung	73
Abb. 6.8: Vergleich der Emotionen pro Durchlauf	74
Abb. 6.9: Bewertung der Gamification durch den Probanden	76
Abb. 6.10: Bewertung der User Experience über AttrakDiff	77

Tabellenverzeichnis

Tab. 1:	Übersicht Punktevergabe	65
Tab. 2:	Übersicht der durchschnittlichen Selbsteinschätzung	66
Tab. 3:	Übersicht der durchschnittlichen Zeit pro Durchlauf	70
Tab. 4:	Zwei-Stichproben-t-Test für abhängige Stichproben - Emotionen	71
Tab. 5:	Zwei-Stichproben-t-Test für abhängige Stichproben - Fragebogen	72
Tab. 6:	Durchschnittliche Bewertung - Fragebogen	72
Tab. 7:	Übersicht der Daten von Proband 9	76

Zusammenfassung

Lange Zeit wurden die Emotionen als wichtiger Aspekt im Bereich der Motivation ignoriert, sie waren sogar unerwünscht. In den letzten Jahrzehnten jedoch wurde die Verbindung zwischen Emotionen, Zufriedenheit und Motivation, gerade in Arbeitskontexten, stärker gewichtet und in den Mittelpunkt mehrerer Studien gestellt.

Mihaly Csikszentmihalyi verfolgt mit dem Konzept des „Flow-Erlebnis“ einen anderen Ansatz. Er untersuchte, warum Freizeitaktivitäten, die gefährlich oder kostspielig sind, ausgeübt werden. Er legt den Begriff der „intrinsischen Belohnung“ zugrunde, was bedeutet, dass Menschen in ihrer Freizeit Aktivitäten unternehmen, die in sich belohnend sind, während die Berufstätigkeit extrinsisch motiviert ist. Gamification könnte einen Weg darstellen, diese intrinsischen Motivationsmechanismen in die Arbeitswelt zu übertragen.

Um die Auswirkungen von Gamification auf die Emotionen zu bewerten, müssen diese zunächst zuverlässig erkannt werden. Der emotionale Zustand ist an der Mimik erkennbar. Daher wird das technische System durch eine Gesichtserkennung die Mimik erfassen. Aufgrund eines hinterlegten neutralen Gesichtsausdrucks des Benutzers werden die relevanten Basisemotionen Freude, Trauer und Wut mit Hilfe des Facial Action Coding System (FACS) nach Paul Ekman erkannt. In einer Studie wird untersucht, ob sich durch Gamification der emotionale Zustand verändert. Unterstützt wird diese Analyse durch einen Fragebogen für die Probanden vor und nach der Studie.

Da Emotionen in den intimen Persönlichkeitsbereich des Menschen fallen, wird darüber hinaus die Legitimität eines solchen Systems unter ethischen Aspekten untersucht.

Abstract

For a long time emotions were ignored as an important aspect within the field of motivation. But in the last decades the link between emotions, satisfaction and motivation, especially in context of employee motivation, became more important. Many studies consider this relation between motivation and emotions.

Mihaly Csikszentmihalyi chose another approach for his theory of flow. In many studies he analysed the reasons for humans to carry out activities in their leisure time which normally are known as expensive and dangerous. Csikszentmihalyi figured out that people are motivated to do such activities in their leisure time because they're falling into a state of complete immersion in that activities. For him people are looking for intrinsic reward which they don't get during their working life. Gamification within work can be a way to transfer this intrinsic reward into the world of employment because work often causes extrinsic reward based on salary and other tangibles more than intrinsic reward .

Therefore it's necessary to detect emotions reliable in order to evaluate the effect of gamification on emotions. The best way to recognize emotions is the facial expression. Thus a technical system has to use a facial expression recognition to evaluate in which emotional state a person is. Based on Paul Ekman's Facial Action Coding System(FACS) our emotion recognition system will recognize basic emotions like happy, sad and angry. Therefore we analyse within a study if gamification will have an effect on emotional states. A questionnaire which a test person complete before and after the study will support our results.

Emotions belong to human's privacy and therefore it's important to think about possible consequences of our system. As a result we look at the system from ethical aspects and try to evaluate how tolerable our system will be with regard to self-determination.

1 Einleitung

Die Teilüberschrift eines Artikels der Zeitung “Die Zeit” lautet: “Junge Beschäftigte verlangen eine neue Arbeitswelt.[...] Ihre Ansprüche verändern die gesamte Wirtschaft.”¹. Von der ‘Generation Y’, die alles in Frage stellt, ist zu lesen. Eine Generation junger Menschen, die ganz andere Ansprüche an ihre Berufsausübung und damit auch an Unternehmen hat, als die Generation zuvor. Anders PARMENT erläutert in seinem Buch “Die Generation Y - Mitarbeiter der Zukunft motivieren, integrieren, führen” die Unterschiede der neuen Generation zur vorhergehenden Babyboomer-Generation, für die “Funktionalismus wichtiger als Ästhetik und Emotionen”² ist. Für die “Babyboomer” war Arbeit eine Pflicht, der man nachging, um die Familie zu ernähren. Aber die Generation Y, im Gegensatz zur vernunftsorientierten Babyboomer-Generation, “legt mehr Wert auf emotionale Aspekte des Arbeitgeberangebots”³, spricht auf Spaß an der Arbeit, Zufriedenheit und Individualität. Diese Aspekte sollen die Motivation fördern.

Überhaupt nimmt der Begriff der Motivation viel Raum im Kontext der Leistungsförderung von Menschen ein. Man nimmt an, dass zufriedene Mitarbeiter motivierter arbeiten und dies positive Auswirkungen auf das Unternehmen und dessen wirtschaftlichen Erfolg haben. Mit verschiedenen Studien wird versucht, diese Annahme zu untermauern, wie z.B. mit der Studie der DUW zur “Mitarbeitermotivation durch Weiterbildungsangebote”⁴ oder auch die Studie der Firma Stepstone in sieben europäischen Ländern zur Frage, ob glückliche Mitarbeiter erfolgreiche Unternehmen bedingen⁵. Man stellt dabei fest, dass es Arbeitnehmern zwar nicht unwichtig ist, wie viel sie verdienen, aber dass - im Gegensatz zur Babyboomer-Generation - andere Aspekte eine ebenso bedeutende Rolle spielen, die auf emotionaler Ebene stattfinden, wie bspw. Anerkennung durch den Vorgesetzten.

Diese Arbeit möchte den Motivationsansatz in diesem Zusammenhang näher untersuchen. Da Motivation ein Gefühl ist, wird sich diese Arbeit näher mit Emotionen befassen. Inwieweit ist es möglich, emotionale Zustände zuverlässig zu erkennen? Können mit darauf abgestimmter Gamification diese Zustände so verändert werden, dass sich Motivation einstellt? Und wenn ja, welche Auswirkungen und Folgen kann das haben? Sehr wichtig ist hierbei die Berücksichtigung ethischer Aspekte, da Emotionen in den intimen Persönlichkeitsbereich des Menschen fallen.

¹Die Zeit, 2013

²Parment, 2013

³Brinks, 2005

⁴DUW Studie

⁵Stepstone Studie

2 Theoretische und philosophische Grundlagen

2.1 Der Begriff des Glücks

Die Frage “Was ist Glück” ist nicht so einfach zu beantworten. Es kommt darauf an, in welchem Kontext man den Begriff ‘Glück’ anwenden will. So zeigt Philipp MAYRING ⁶ auf, dass es wichtig ist, zu unterscheiden, welche semantische Bedeutung man benutzen möchte, da in der deutschen Sprache ‘Glück’ entweder in Bezug auf Zufall oder auf Erfüllung benutzt werden kann, während viele andere Sprachen zwei Begriffe zur Unterscheidung haben, wie bspw. im Lateinischen fortuna (Glück, Schicksal, Zufall, aber auch Unglück) und beatitudo (Glück, Glückseligkeit). Ich folge in dieser Arbeit der gängigen Festlegung für ‘Glück’ als eines erfüllenden Gefühls ⁷, als ein Ziel, welches der Mensch anstrebt und von ihm individuell ausgelegt wird ⁸. Für die Philosophie ist der Begriff ‘Glück’ die “zentrale Zielgröße vieler philosophischer Systeme”⁹. Viele große Denker versuchten, eine allgemeingültige Antwort auf Fragen wie: ‘Was ist Glück?’ und: ‘Braucht man Glück?’, zu finden. Heraus kam dabei eine Vielzahl von Glückstheorien. So war für PLATON tugendhaftes und gutes Leben eine wichtige Voraussetzung für Glückseligkeit, aber “wahres Glück [blieb] dem vorbildhaftesten Philosophen vorbehalten” ¹⁰. Sein Schüler ARISTOTELES legte den Fokus ebenfalls auf die Tugendhaftigkeit, aber Eudaimonia ¹¹ war für ihn ein “Gut, dass um seiner selbst angestrebt wird, also in sich das Ziel ist”¹². Es kann von jedem erreicht werden, wenn er über “moralische Vollkommenheit und ein Vollmaß des Lebens”¹³ verfügt. Glücklich ist der, “der moralisch handelt, zugleich aber aller äußeren und leiblichen Güter [...] teilhaftig ist”¹⁴, die er benötigt, um “ein vollkommenes Leben zu führen”¹⁵. Für Aristoteles ist Glück “Inbegriff eines [...] gelungenen, sinnerfüllten Lebens”¹⁶. Diese Aussage bildet die Grundlage für die eudämonische Ethik.

Die hedonistische¹⁷ Ethik ist ein weiterer Ansatz, die Glück ebenso als das vom Menschen angestrebte höchste Ziel ansieht, jedoch stellt sie Glück mit Freude oder Lust gleich. Glück kann der Mensch nur erlangen, wenn er das tut, “was ihm Freude macht [...], dem es gelingt, seine Triebe und Bedürfnisse [...] zu harmonisieren (Ataraxie, Anm. d. Verf.)”¹⁸. Durch diesen Zustand erst

⁶Mayring, 1991

⁷vgl. ebda. S.12

⁸ebda.S.14, aber auch Pieper,2000

⁹ebda.S.18

¹⁰ebda.S.20

¹¹griech. Glück, bzw. Glückseligkeit

¹²ebda.S.20

¹³Aristoteles, Nikomachische Ethik I,10.1100a

¹⁴Pieper,2000,S.267

¹⁵ebda.

¹⁶ebda.S.266

¹⁷griech. hedone=Lust

¹⁸ebda.S.267

erreicht man das “eigentlich, wahre Glück, die Lust am Gleichmaß, an der Ausgewogenheit der Interessen”¹⁹. Die utilitaristische²⁰ Ethik wiederum sieht den größtmöglichen Nutzen als Mittel zum Erreichen von Glück. Dieser Nutzen bildet die Grundlage “aller moralisch guten Handlungen[...], durch die das Glück aller von diesen Handlungen Betroffenen gefördert wird”²¹. Der materialistische Ansatz, der sich in der marxistischen Ethik dahingehend äußert, dass die “ökonomische Basis [...] die Grundlage aller menschlichen Handlungen und Handlungsnormen”²² ist, definiert Glück als etwas, das man “unter bestimmten gesellschaftlichen Bedingungen”²³ erreicht. Wenn sich der Unterbau aus der “Unterdrückung durch die Herrschenden”²⁴ befreit hat, dann wird sich Glück einstellen. Immanuel KANT sieht, basierend auf dem kategorischen Imperativ, Glück als etwas an, dessen man sich durch ein vernünftiges, tugendhaftes Leben als würdig erweisen muss²⁵.

Keiner dieser Ansätze wurde der Allgemeingültige. Glück ist zu individuell und zu sehr von der momentanen Lebenssituation abhängig, als dass man eine allgemeingültige Antwort geben kann. Auch die Methodik zum Erreichen von ‘Glück’ hängt von mehreren Faktoren ab. Max WEBER, der Anfang des 20. Jahrhunderts den Zusammenhang zwischen dem modernen Kapitalismus und protestantischer Ethik untersuchte, vermutete, dass der religiöse Hintergrund, in dem wir aufwachsen und erzogen werden, unsere Methoden maßgeblich prägt. Auslöser war für WEBER die Annahme, dass “die Puritaner als Geburtshelfer des modernen Kapitalismus”²⁶ fungierten und sah den Kapitalismus als eine Gewalt, die “alle traditionellen gesellschaftlichen Ordnungen unwiderrufflich zerstörte”²⁷ und damit dem Individuum die persönliche Freiheit zu nehmen droht²⁸.

Heute ist das Thema Glück, bzw. Zufriedenheit - gerade in der Wirtschaft und der Soziologie - sehr aktuell, wie zahlreiche Werke²⁹, oder auch seit Jahren die Glücksdatenbank von Ruut VEENHOVEN³⁰ zeigen. Der Begriff ‘Lebensqualität’, der betonen soll, dass zum

¹⁹ebda.S.267

²⁰lat. utilis = nützlich

²¹ebda.S.271

²²ebda.S.279

²³Mayring,1991,S.22

²⁴Pieper,2000,S.280

²⁵vgl. Pieper,2000,S.162

²⁶Guttandin,1998,S.9

²⁷ebda.,S.8

²⁸vgl. ebda. S.7

²⁹Kleinbeck/ Kleinbeck,2009; Scherke,2009; Holodynski,2006; Wrase,2010

³⁰emeritierter Professor für Soziologie, Fachbereich Bedingungen des Glücklichseins, Direktor der World Database of Happiness

Glücklichsein “nicht nur materielle Güter, sondern auch subjektives Wohlbefinden gehört”³¹, rückt dabei in den Mittelpunkt. Zufriedenheit und Glück gehen Hand in Hand. VEENHOVEN definiert “Happiness [...] as the degree to which an individual judges the overall quality of his/her life-as-a-whole favorably”³² und analysiert internationale Publikationen zum Thema Glück, um damit ein weltweites Glücksranking erstellen zu können³³. VEENHOVEN erläutert, dass sich Glück auf zwei Ebenen aufteilt: dem hedonischen Gefühlsniveau und dem Zufriedenheitsniveau, die das Glücksniveau bilden.³⁴ Während ersteres die Stimmung widerspiegelt, in der man sich befindet, bezeichnet zweiteres “den Grad, in dem ein Individuum seine Bestrebungen als erfüllt betrachtet”³⁵, basierend auf der Annahme, dass man “bestimmte bewusste Ziele und bestimmte Vorstellungen zu deren Verwirklichung entwickelt hat”³⁶.

2.2 Motivation und Arbeit

Motive sind bestimmte und bewusste Ziele, die man erreichen möchte. Laut KLEINBECK entscheidet sich in einem Prozess über vier Phasen, wie motiviert ein Mensch sein Ziel erreichen will. In den ersten beiden Phasen entstehen die “Motivationszustände”³⁷ und das Planen des “zielorientierten Handlungsverlaufs”³⁸, um das Handlungsziel zu spezifizieren. Phase drei und vier sind dann Ausführung und Modifikation des Plans. Je motivierter der Handelnde ist, desto besser kann er mit Erfolg und Misserfolg umgehen und mit Hilfe seiner Erwartungen den Prozess steuern. Das ist aber nicht universell für jeden gleich, sondern hängt stark von den Motiven ab, die “in jeweils unterschiedlichen individuellen Ausprägungen und Kombinationen”³⁹ jedem eigen sind. Im Kontext der Arbeitsmotivation kann man durch “äußere Einflussgrößen wie Arbeitsaufgaben und -bedingungen”⁴⁰ den Grad der Arbeitsmotivation, zusammen mit der individuellen Zielsetzung, bestimmen.

Bisher gibt es einige Theorien, die den Zusammenhang zwischen Arbeitszufriedenheit und Motivation untersuchen. Vereinfacht folgen diese einem allgemeinen zyklischen Schema⁴¹

³¹Mayring,1991,S.76

³²http://www1.eur.nl/fsw/happiness/hap_quer/hqi_fp.htm[Zugriff am 17.11.2013]

³³Für weitere Details und die genaue Erklärung der Methode sei der Leser auf <http://worlddatabaseofhappiness.eur.nl/index.html> verwiesen

³⁴<http://www2.eur.nl/fsw/research/veenhoven/Pub2010s/2011d-fulld.pdf>

³⁵ebda.,S.2

³⁶ebda., S.2

³⁷Kleinbeck/ Kleinbeck, 2009, S.10

³⁸ebda.

³⁹ebda.

⁴⁰ebda.

⁴¹vgl. Fischer,1989, S.28, Abb.5

widmen kann, weil “a want that is satisfied is no longer a want.”⁴³ [...]If the physiological needs are relatively well gratified, there then emerges a new set of needs.⁴⁴ Sobald die Bedürfnisse befriedigt sind, haben sie “keinerlei motivationale Wirkung mehr”⁴⁵. Während der Drang nach Erfüllung der ersten vier Stufen der Bedürfnisse aus einem Defizit heraus entsteht, ist die letzte Stufe nach MASLOW das Streben nach Selbstverwirklichung, als der Wunsch des Menschen “to become actualized in what he is potentially”⁴⁶.

Auch wenn laut FISCHER einige Studien durchgeführt wurden, um die Theorie zu widerlegen, so bildet MASLOWS Bedürfnishierarchie eine Grundlage, die es ermöglicht, die menschlichen Bedürfnisse breiter gefächert in den Fokus der Aufmerksamkeit zu rücken.

2.2.2 Herzbergs Zwei-Faktoren-Theorie

Einen Versuch, einen ähnlichen Ansatz zu finden, unternahm HERZBERG mit seinen Mitarbeitern, als sie untersuchten, wie Arbeitnehmer zu ihrer Arbeit stehen. Dazu befragte er Mitarbeiter und bat sie, “[to] think of a time in the past when you felt especially good or bad about your job.[...] Can you think of such a high or low point in your feelings about your job?”⁴⁷. Nach HERZBERG gibt es nur bestimmte Faktoren, die “zur Herstellung wirklich befriedigender Arbeitsverhältnisse”⁴⁸ geeignet sind. Abb. 2.3 zeigt HERZBERGS Vergleich von Motivatoren (first-level-Faktoren) mit Hygienefaktoren (second-level-Faktoren).

⁴³Maslow, Abraham H., 1972,S.38

⁴⁴Maslow, 1972,S.39

⁴⁵Fischer, 1989, S.30

⁴⁶MASLOW, 1972,S.46

⁴⁷Herzberg et.al.,1959,S.20

⁴⁸Fischer, 1989, S.34

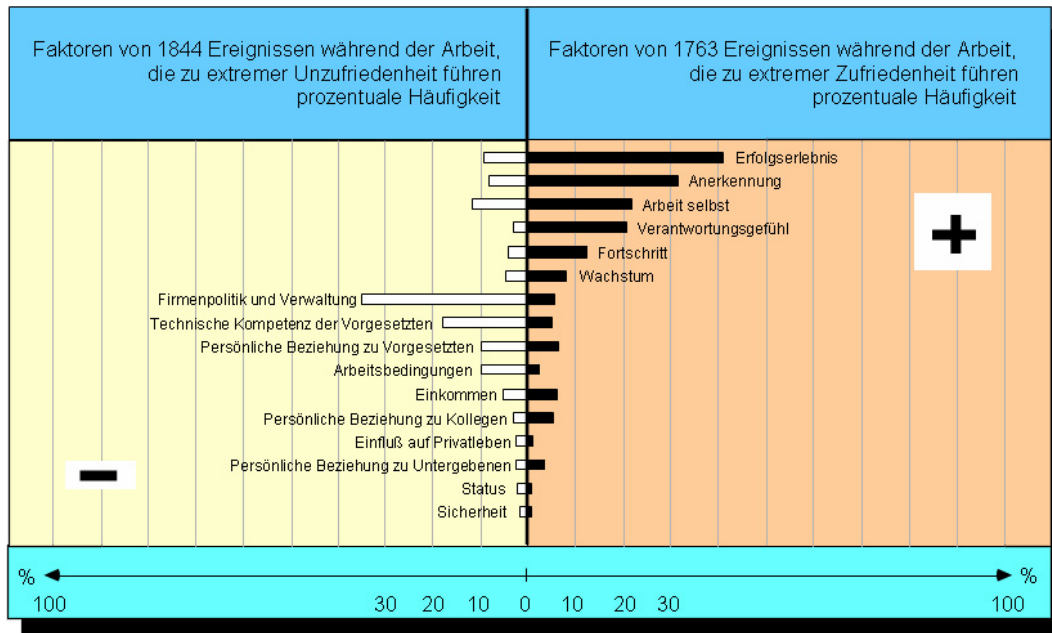


Abbildung 2.3: Vergleich der “satisfiers” mit “dissatisfiers” nach Herzberg.

Bildquelle: wikipedia.de

Motivatoren(satisfier) sind u.a. Leistung und Erfolg, Anerkennung, Beförderung, die entscheidend auf die Arbeitsmotivation einwirken, aber deren Fehlen nicht zwangsläufig unzufrieden macht. Hygienefaktoren(dissatisfier) sind u.a. Gehalt, Firmenpolitik, Sicherheit am Arbeitsplatz, die unzufrieden machen können, aber nicht unbedingt zufrieden. Die verschiedenen Breiten der Balken verdeutlichen, ob ein Faktor langfristig (“long-range”) oder kurzfristig (“short-range”) auf die Einstellung des Mitarbeiters einwirkt. Kombiniert man die Motivatoren mit den Hygienefaktoren, ergeben sich vier mögliche Situationen für einen Mitarbeiter:

Motivatoren	Hygienefaktoren	Arbeitsmotivation	Beschwerdengrad
Hoch	Hoch	Hoch	Niedrig
Niedrig	Hoch	Niedrig	Niedrig
Hoch	Niedrig	Hoch	Hoch
Niedrig	Niedrig	Niedrig	Hoch

Abbildung 2.4: Zusammenspiel von Motivations- und Hygienefaktoren

Quelle: eigener Entwurf

Die erste Zeile stellt die beste Konstellation, in der ein Mitarbeiter sich befinden kann, dar. Er ist hoch motiviert und hat wenig in Bezug auf die Hygienefaktoren auszusetzen. Die letzte Zeile verdeutlicht die schlechteste Konstellation, da der Mitarbeiter unmotiviert und mit den

Bedingungen unzufrieden ist.

2.2.3 Fazit

Es existieren weitere Theorien, wie bspw. die Instrumentalitätstheorie von Vroom, die hier nicht weiter erläutert werden sollen⁴⁹. Für uns ist der Ansatz von MASLOW und HERZBERG ausreichend, um zu klären, wie Motivation, Zufriedenheit und vielleicht auch Glück erreicht werden können. Auch wenn weder MASLOWs Theorie, noch HERZBERGs 2-Faktoren-Theorie ein universales Modell darstellen, bietet jede von ihnen eine lohnende Grundlage für weitere Untersuchungen.

KLEINBECK bildet einen weiterführenden Ansatz, indem er annimmt, dass Zufriedenheit eine Art der Motivation freisetzt, die auf die Produktivität einwirkt und damit dem Unternehmen hilft, langfristig eine starke Position auf dem Markt einzunehmen⁵⁰. Damit das Unternehmen seine Ziele erreichen kann, ist es nicht nur wichtig, welche Fähigkeiten, bzw. Fertigkeiten die Mitarbeiter besitzen und welche Ressourcen das Unternehmen bereitstellt, sondern es hängt wesentlich von der Arbeitsmotivation der Mitarbeiter ab⁵¹. Ein wesentlicher Aspekt hierbei ist die Anerkennung und das damit einhergehende "Gefühl des Stolzes"⁵², etwas vollbracht zu haben. Dieses Gefühl steigert das Selbstwertgefühl und die Zufriedenheit. Ein zufriedener Mitarbeiter ist motivierter und dadurch zielstrebig, ausdauernder und bemüht, bzw. kreativer bei der Erledigung seiner Aufgaben⁵³.

2.3 Emotionen

Ob ein Mensch glücklich und zufrieden ist, erkennt man an seinem Gesichtsausdruck genauso, wie wenn er genervt oder wütend ist. Wir sind ständig von Emotionen umgeben und Emotionen lenken unser Handeln. Mit Hilfe von Emotionen sind wir in der Lage, auf Situationen zu reagieren, im Falle von Gefahr sogar reflexartig und ohne darüber nachzudenken⁵⁴. EKMAN erforscht Emotionen seit über 30 Jahren und stellte bei Studien in Völkern, bzw. Stämmen, die weder durch moderne Medien, noch durch den Einfluss der modernen Welt, geprägt waren, fest, dass es Basisemotionen gibt, die angeboren und nicht durch Kultur oder Gesellschaft angezogen sind. Auch DARWIN untersuchte schon 1872 die Gesichtsausdrücke von Menschen (und von Tieren), um

⁴⁹für weitere Informationen siehe u.a Fischer, 1989, S.44 ff.

⁵⁰vgl.hierzu Kleinbeck/ Kleinbeck,2009,S.12

⁵¹ebda., S.18

⁵²ebda.S.20

⁵³vgl.ebda.,S.21

⁵⁴vgl. Ekman,2010, S.27

nachzuweisen, dass es Emotionen gibt, die, unabhängig von Kultur und Herkunft, allen gleich sind⁵⁵. Nach EKMAN gibt es sieben allgemeingültige Gesichtsausdrücke, die, unabhängig von kultureller und ethischer Herkunft, die gleichen Emotionen ausdrücken. Diese sind Freude, Wut, Ekel, Abscheu, Trauer, Angst, Verachtung und Überraschung⁵⁶



Abbildung 2.5: Basisemotionen nach EKMAN

Bildquelle: rueetschli.net

Emotionen, so vermutet EKMAN, entstanden auf vielfältige Art und Weise. Grundsätzlich helfen Emotionen, eine Situation automatisch zu bewerten. Verspürt man Angst, macht sich der Körper bereit zur Flucht. Während der Evolution entwickelten sich Emotionen als eine Variable der Selektion, die denen, die z.B. aus Angst einen weiten Bogen um giftige Tiere machten, das Überleben, die Fortpflanzung und damit die Vererbung dieser Eigenschaft sicherten. DARWIN selbst beschreibt einen Selbstversuch, in dem er sich vor ein Schlangenterrarium im Zoo stellte, mit dem festen Entschluss, keine Reaktion zu zeigen, falls die Schlange angreift. Als dieser Fall aber eintrat, sprang er vor Schreck zurück und hatte es nicht geschafft, sich mit bloßem Willen und rational “gegen die Einbildung einer Gefahr, welche niemals direkt erfahren worden war”⁵⁷ zu widersetzen. Interessant ist EKMANs Erläuterung, dass wir von der Evolution her noch nicht in der Lage sind, entsprechende Reaktionen auf Situationen zu zeigen, die “nicht zum

⁵⁵Darwin,2000

⁵⁶Manchmal ist nur von sechs Basisemotionen die Rede, wenn Überraschung nicht dazu gezählt wird. Nach EKMAN ist Überraschung irgendwo mittendrin, weil es ein negativer, aber auch ein positiver Zustand sein kann(vgl. Ekman, 2012,S14 ff.

⁵⁷Darwin, 2000, S.47

Erfahrungsbereich unserer Urahnen”⁵⁸ gehörten und daher von uns nicht als gefährlich gedeutet werden können. Gemeint sind Situationen, die durch technische Geräte (Autos, Maschinen,...) erzeugt werden und im Vergleich zu anderen Gefahrensituationen (Höhe, Tiere, etc.) noch nicht lang genug existieren, als dass der Mensch dafür schon eine geeignete emotionale Reaktion hätte entwickeln können.

Emotionen kann man selbst erzeugen, indem man sich an Vergangenes erinnert, wobei es Erinnerungen gibt, die man freiwillig erlebt und welche, die sich einem durch einen Auslöser aufdrängen. Ersteres wurde von Walter BENJAMIN, letzteres von Marcel PROUST aufgegriffen, als sie von Erinnerungen schrieben, die durch Artefakte ausgelöst wurden. So war es bei PROUST der Genuß von Madeleines (frz. Kleingebäck)⁵⁹, während es bei BENJAMIN das bewusste Durchstöbern einer Sammlung war⁶⁰. Erinnerungen, so zeigen beide auf, lösen immer Emotionen aus. Emotionen können aber auch durch das Aufsetzen eines bestimmten Gesichtsausdrucks ausgelöst werden, wie EKMAN in Studien entdeckte. Das klappte nicht bei “beliebigen Ausdrücken, sondern ausschließlich bei jenen, [...] [die] universal allen Menschen eigen sind”⁶¹. Durch Experimente fanden EKMAN und seine Kollegen heraus, dass dies auch “physiologische Veränderungen auslöst” und im Gehirn messbar ist⁶².

Im Allgemeinen teilt man die verschiedenen Emotionen in zwei Gruppen auf: in negative Emotionen (Trauer, Wut, Ekel, Abscheu, Angst, Verachtung) und in positive Emotionen (Freude), Überraschung kann sowohl negativ als auch positiv sein. Eine Eigenschaft von Emotionen ist, dass sie “unsere Sicht der Welt [ändern] und unsere Interpretation des Handelns anderer”⁶³. Wenn wir uns in einer bestimmten Emotion befinden, versuchen wir sie zu erhalten, indem wir uns die Bestätigung dafür aus der Umgebung holen und das Wissen ignorieren, “das nicht mit ihm übereinstimmt”⁶⁴. Wir lassen uns nicht so leicht davon überzeugen, dass das Zurechtweisen durch einen anderen berechtigt war, solange wir deswegen wütend deswegen, sondern ignorieren wohlgemeinte Argumente. EKMAN nennt dies einen “Refraktärzustand”⁶⁵, einen Zustand, in dem wir die Informationen ignorieren, “die zu dem uns beherrschenden Gefühl nicht passen, es nicht nähren und rechtfertigen”⁶⁶.

⁵⁸Ekman, 2010, S.63

⁵⁹Proust, Marcel; Auf der Suche nach der verlorenen Zeit

⁶⁰Benjamin, 1972–1999)

⁶¹Ekman, 2010, S.51

⁶²ebda.

⁶³ebda., S.55

⁶⁴ebda.

⁶⁵ebda. S.56

⁶⁶ebda.

Auch Stimmungslagen bilden den Nährboden für verschiedene Emotionen. Was eine Emotion auslöst, können wir immer ganz gut erklären, warum wir aber in einer bestimmten Stimmung sind, nicht ⁶⁷. Das Problem an schlechten Stimmungen ist, dass man dazu neigt, in der Umgebung Gründe für das Ablassen der eigenen Gereiztheit zu suchen, als bei sich selbst. In Bezug auf die Motivationstheorien von MASLOW und HERZBERG, ist der Aspekt, dass auch die Umgebung Auslöser für entsprechende Emotionen sein kann, ein wichtiger. Ein demotivierter Mitarbeiter geht schon in entsprechender Stimmung zur Arbeit und es ist für einen Mitarbeiter, der traurig oder wütend ist, nicht so einfach, alleine aus diesem Zustand herauszufinden, wenn er die Situation, bzw. den Ort nicht verlassen kann. Er sucht in seiner Umgebung Gründe dafür, dass es ihm so geht. Deshalb wäre es wünschenswert, ihm eine Hilfestellung zu geben, den Zustand verlassen zu können.

2.3.1 Facial Action Coding System (FACS)

EKMAN erstellte aufgrund seiner jahrzehntelangen Untersuchungen ein Kodiersystem, da die Basisemotionen immer durch die gleichen Gesichtsmuskeln in der Mimik ausgedrückt werden. Das FACS beschreibt, welche Muskeln (Actions Units (AU)) um Mund und Augen durch eine Bewegung zu einem bestimmten Gesichtsausdruck führen. Auch DARWIN untersuchte schon, welche Muskelgruppen bspw. sich beim Lachen zusammenziehen und welche sich entspannen ⁶⁸.

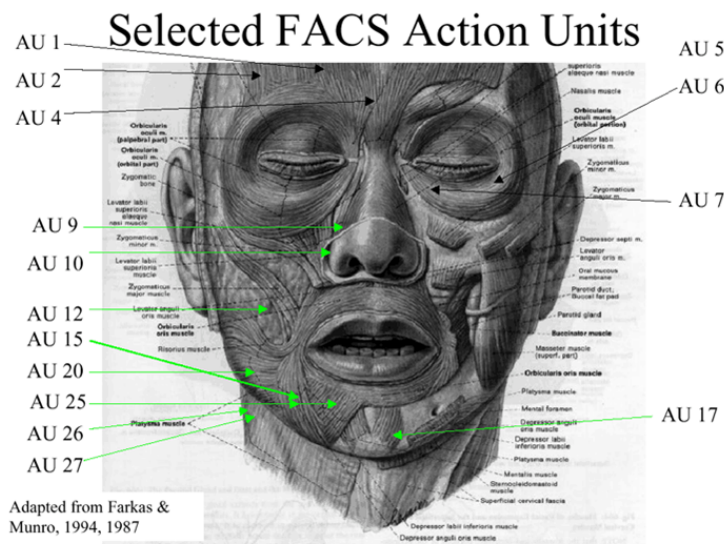


Abbildung 2.6: Übersicht über die zugeordneten AUs
 Bildquelle: www.wedernoch.de/

⁶⁷ vgl. ebda. S.72
⁶⁸ vgl. Darwin, 2000, S.222 ff.

Welche AUs in vielfältigen Kombinationen miteinander welchen Gesichtsausdruck verursachen, untersuchten EKMAN und seine Kollegen durch Studien mit anderen, aber auch durch Studien an sich selbst. Für Emotionen sind nur bestimmte AUs beteiligt, die die Basisemotionen erzeugen können. Diese werden auch gern das “Emotion Facial Action Coding System” (EMFACS) genannt und basieren auf den FACS.

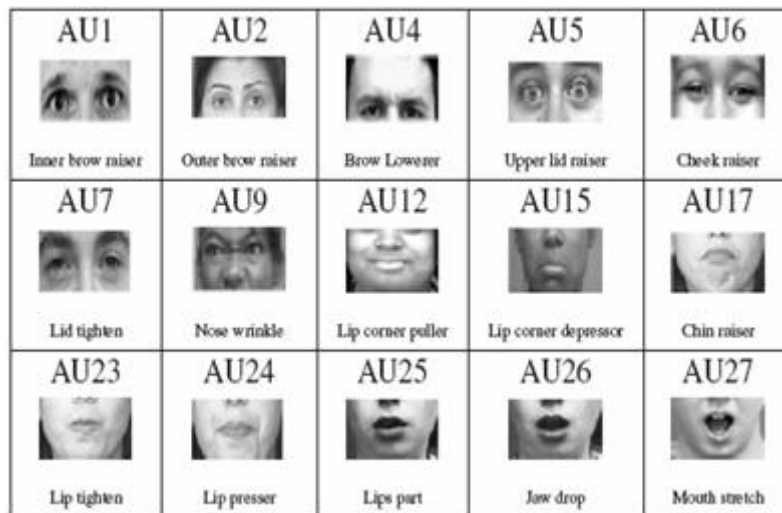


Abbildung 2.7: Übersicht über die AUs der EMFACS nach EKMAN

Bildquelle: www.ecse.rpi.edu

An der Bildung von Emotionen sind in der Regel die AUs von Mund und Augen relevant (Abb.2.7). Der Abstand der Mundwinkel zueinander verändert sich, je nach Emotion. Bei einem Lächeln gehen sie auseinander, während sie sich bei Zorn zusammenziehen. Auch die Nasenwurzel und die Augenbrauen geben Aufschluss über die Emotion. Nur bei Zorn ziehen sie sich stark zusammen, bei einem Lächeln entspannen sie sich deutlich und bei Überraschung heben sie sich.

2.4 Praktische Umsetzung mit Gamification

Warum Gamification, bzw. das Spiel als Mittel der Wahl? Das hängt mit der Erkenntnis zusammen, dass der Mensch nicht nur ein 'homo oeconomicus' ist, sondern auch ein 'homo ludens'. Der Mensch spielt gerne und verlässt damit “das Joch der Fron, [...]den]Druck der unentwegten Strebungen”⁶⁹. Es hilft dem Menschen, ganz er selbst zu sein. Ebenso ist ein Spiel intrinsisch orientiert, es ist “potentiell erfreulich”⁷⁰. Das Spiel wird “weder für Handlungen benutzt, die als unkontrollierbar bekannt sind, noch für Handlungen, die als Resultat eines Plans bekannt sind”⁷¹, d.h. wenn man einen Spieler beobachtet, verfolgt er weder ein “vernünftiges Ergebnis

⁶⁹Fink, Eugen, Oase des Glücks, Alber, 1957, S.12

⁷⁰Csikszentmihalyi, 2005, S.12

⁷¹Millar, 1973, S.236

eines Planes”⁷² noch verliert er die Kontrolle über sein Handeln. Spielen ist für die meisten Lebewesen eine Möglichkeit, “Vertrautes [zu] erforschen, bereits Beherrschtes [zu] üben”⁷³, sowie Motorik und Kognition zu trainieren.

Im Bereich der Gamification geht es darum, dieses Potential zu nutzen, um den Benutzer zu motivieren und - in Bezug auf ältere oder leistungsgeminderte Benutzer - motorische und kognitive Fähigkeiten zu fördern. Das Spiel ist eine Methode, die “in sich selber befriedigend sein”⁷⁴ und uns Möglichkeiten liefern kann, eine andere Motivationsform zu finden, da es immer ein direktes Feedback liefert. Darüber hinaus kann durch Gamification der Spaß am Tun wesentlich erhöht werden⁷⁵. Der Spaß, der sich beim Spielen einstellt, motiviert wiederum, am Ball zu bleiben und bei einer Niederlage nicht gleich aufzugeben. Durch die positiven Erlebnisse ist es leichter, Fähigkeiten im Spiel zu trainieren, da der Spieler nicht unter Leistungsdruck steht und auf Biegen und Brechen versucht, etwas zu erlernen. Auf die Arbeitswelt bezogen wäre es ein möglicher Ansatz, einem gestressten oder erkennbar schlecht gelauntem Mitarbeiter ein passendes Spiel anzubieten, um ihm mittels Spaß und Erfolgserlebnisse zu helfen, seine Arbeit konzentriert und motiviert mit Erfolg fortzusetzen. Dadurch würde nicht nur die Fehlerrate gesenkt werden, sondern auch der Frustpegel.

2.4.1 Zielgruppen

Gamification ist eine Methode, die sich bereits in vielfältiger Ausprägung in verschiedenen Zweigen der Wirtschaft wiederfindet. Einen Mitarbeiter kann man mit spielerischen Elementen in seiner Arbeit durchaus motivieren, Aufgaben gleich zu erledigen, die er sonst gerne vor sich her schiebt. Eine Vorreiterrolle besitzt hierbei die Firma SAP, die durch Gamification ihre Mitarbeiter ermuntert, Aufgaben gleich zu erledigen⁷⁶. Unser System soll mehrere Personengruppen ansprechen:

Jüngere Menschen Gesunde Menschen zwischen 20 und 55 sind in der Regel motorisch und kognitiv in der Lage, auch kniffligere Aufgaben zu lösen. Jedoch muss auch hier der Fokus auf dem Spaßfaktor liegen, weil ein gesunder Mensch irgendwann überfordert oder unterfordert sein kann. Die emotionale Mimik - so werden wir es voraussetzen - ist bei einem gesunden Menschen

⁷²ebda.

⁷³ebda.S.247

⁷⁴ebda. S.19

⁷⁵Eine gute Darstellung findet sich in einem Vortrag von Mario Herger, Gamification-Designer bei SAP

⁷⁶So bot die Firma SAP ihren Mitarbeitern ein Golfspiel an, in dem die Mitarbeiter Kundenanfragen durch Einlöcher an den Kollegen weiterleiten konnten

nicht durch körperliche Defizite eingeschränkt.

Ältere Menschen Ab einem gewissen Alter ⁷⁷ verändern sich Reaktionsfähigkeit und motorische bzw. visuelle Fähigkeiten. Darüber hinaus ist bei älteren Menschen darauf zu achten, ob sie mit der Steuerung digitaler Geräte z.B. mittels Touchscreen oder Pen vertraut sind oder schnell damit überfordert sind.

Leistungsgeminderte Menschen Im Umgang mit leistungsgeminderten Menschen ist es schon etwas schwieriger, da die körperlichen Einschränkungen nicht gleich, sondern individuell ausgeprägt sind. Auch ist die Erkennung von Emotionen unter Umständen anders anzugehen, wenn körperliche Einschränkungen es deutlich erschweren, die Mimik oder Gestik zu analysieren. Der Umgang mit ihnen ist auf einer anderen Ebene anzusiedeln, als bei gesunden Menschen, gerade in Bezug auf Unlust oder möglicher anderer Gefahren(Sucht).

2.4.2 Notwendigkeit eines Feedback

Um den Benutzer bestmöglich zu motivieren, ist es nötig, ihm ein klares Ziel und Feedback zu geben. Durch das zeitnahe Feedback, ob das Ziel erreicht wurde, baut sich ein Erfolgsgefühl auf, das aufrecht erhalten werden muss. Es ist nicht Sinn und Zweck der Sache, ihm keinerlei Hürden anzubieten. Das System sollte so konzipiert sein, dass sich der Benutzer auch schweren Aufgaben motiviert widmet. Dafür ist ein entsprechendes Feedback wichtig, um ihn zu ermuntern, d.h. ein positives Feedback bei Erfolg und ein aufmunterndes bei Misserfolg. Wichtig ist hierbei eine hohe User Experience, denn wir wollen den Benutzer nicht überfordern und nicht unterfordern. Es ist wichtig, dem Benutzer auch einen "Notausgang" anzubieten, um eine Verschlechterung des emotionalen Zustands zu verhindern. Ziel ist es, den Benutzer positiv zu motivieren, nicht ihn zu demotivieren. Das bedeutet auch, dass wir den Benutzer da abholen, wo er von seinen Fähigkeiten und körperlichen Voraussetzungen steht.

2.4.3 Mögliche Auswirkungen

Diese Arbeit folgt den Ansätzen von CSIKSZENTMIHALYI, der der Frage nachging, warum man Aktivitäten in seiner Freizeit unternimmt, die materiell keinerlei Anreize bieten. Dazu entwarf er eine Studie, die versucht, Freude zu verstehen und zwar "als laufenden Prozess, welcher in der Gegenwart eine lohnende Erfahrung darstellt"⁷⁸. Freude, so sein Ziel, muss man als eine "autonome Realität betrachten, die aus sich heraus verstanden werden muss"⁷⁹. Er wählte dafür

⁷⁷In unserem Fall also ab 56

⁷⁸Csikszentmihalyi,2005,S.29

⁷⁹ebda.

Probanden aus, die verschiedene autotelische⁸⁰ Aktivitäten wie z.B. Schach, Klettern, Tanzen, Basketball oder Musizieren, betreiben. Diese Menschen hatten in ihrer Tätigkeit eines gemeinsam: Sie wenden viel Energie für eine Aktivität auf, die materiell nur “minimale Belohnungen”⁸¹ mit sich bringen.

Warum das Konzept des flow-Erlebnisses so hoch einzuschätzen ist, gründet auf der Tatsache, dass wir arbeiten gehen müssen, um Geld zu verdienen und das, “was man tun muss, kann nicht befriedigend sein”⁸². Wir schleppen uns jeden Tag gegen unseren Willen zur Arbeit. Auch Anders Parment weist darauf hin, dass dies ein Generationenproblem war⁸³ und die Gesellschaft zur Vernunft hin erzogen wurde. In den letzten Jahrhunderten wurde durch “die Gesellschaften ein rationales und universales Motivationssystem geschaffen, mit dessen Hilfe sie erwünschtes Verhalten hervorrufen”⁸⁴ und kontrollieren konnte. Geld wird zwar benötigt, um seinen Lebensunterhalt zu bestreiten, aber dient von der Motivation her eher “zur Entschädigung für eine sinnlose Lebens-Plackerei”⁸⁵. Und dafür hat der Mensch die Freizeit geschaffen, denn da unternehmen wir Dinge, die uns Spaß machen und für die wir kein Geld bekommen, sondern eher noch dafür bezahlen.

Einen Ansatz liefert uns das Konzept mit dem Gedanken, dass man die Arbeit “befriedigender gestalten”⁸⁶ kann und damit extrinsische und intrinsische Motivation miteinander verknüpft. Spielen eignet sich hervorragend dafür, weil man jederzeit damit anfangen und aufhören kann. Die Testpersonen in Csikszentmihalyi Studie gaben auf die Frage nach den Gründen, warum sie eine Aktivität erfreulich finden, als wichtigsten Punkt an, dass die Aktivität Lust bringe und sie ihr Können anwenden können, dicht gefolgt von der Aussage, dass die Aktivität selber es ist, die in ihr befindliche Welt, die sich beim Ausüben eröffnet⁸⁷. Auch in den weiteren Befragungen steht der Zustand des Eintauchens in eine andere Welt und des damit verbundenen Selbstvergessens an vorderer Stelle. Diesen Zustand nennt Csikszentmihalyi ‘flow’, “das holistische Gefühl bei völligem Aufgehen in einer Tätigkeit”⁸⁸. Dieses erreicht man nur dann, wenn die Anforderungen und die Fähigkeiten im Gleichgewicht stehen. Das bedeutet konkret, wenn die Anforderungen höher sind als das Können des Benutzers, wird er Angst aufbauen. Wenn es genau andersrum ist stellt sich Langeweile ein. Jedoch sagt Csikszentmihalyi selbst, dass dieses Gleichgewicht “von

⁸⁰Eine Aktivität ist autotelisch, wenn sie ihrer selbst willen ausgeübt wird und nicht wegen extrinsischen Belohnungen, vgl. Csikszentmihalyi,2005, S.30

⁸¹ebda. S.34

⁸²Csikszentmihalyi,2005,S.21

⁸³Parment,2013, S.8 ff

⁸⁴Csikszentmihalyi,2005,S.20

⁸⁵ebda.,S.22

⁸⁶ebda.,S.22

⁸⁷vgl. Csikszentmihalyi,2005, Tbl.1,S.35

⁸⁸ebda.,S.59

der subjektiven Wahrnehmung dieser beiden Gegebenheiten”⁸⁹ abhängt und es ist nicht möglich sei, vorherzusagen, ob wir bei einer Person in einer konkreten Situation Angst, Langeweile oder flow erzeugen.

Unsere Methode der Gamification soll jedem Benutzer die Möglichkeit bieten, eine eigene Welt zu betreten und im besten Fall sogar ein flow-Erlebnis zu erfahren. Durch den Anreiz intrinsischer Belohnung wollen wir die Motivation erhalten und fördern, aber auch dabei helfen, eine negative Emotion in eine positive umzuwandeln. Nach Csikszentmihalyi müsste dies eine Auswirkung sein, wenn wir es schaffen, eine Balance zwischen Können, Herausforderung und Erreichbarkeit eines Zieles zu erhalten. Wie bei Kindern auch, könnten geeignete Spiele Auswirkungen auf die Fähigkeiten älterer oder leistungsgeminderter Menschen haben, die durch Wiederholung gleicher Bewegungen und durch eine stufenweise Steigerung des Schwierigkeitsgrades bestimmte Fähigkeiten im Kognitiven (z.B. Merkfähigkeit) und Motorischen (z.B. gezielte Bewegungen, Reaktion) trainieren und verbessern können.

Eine Auswirkung, die wir gesondert beachten müssen, ist die Gefahr einer Sucht. Ein Spiel birgt durchaus ein gewisses Suchtpotential. Daher muss man sich im Vorfeld Gedanken machen, ob man das Spiel so gestalten muss, dass es dieses Potential minimiert oder ob es durch den Einsatz im Arbeitsbereich schon von sich aus Beschränkungen in der Attraktivität mitbringt. Es wird also notwendig sein, sich Beschränkungen im Realitätsgrad, in der Levelanzahl und dem Spaßfaktor zu überlegen.

3 Stand der Wissenschaft - Techniken der Emotionsklassifizierung

3.1 Affektive MCI

Die Idee eines Wesens, welches erschaffen wurde und intelligent ist, aber dem Menschen untertan, findet sich seit Jahrhunderten in Religion, Mystik und Literatur. Die Entwicklung der Technik seit der Industrialisierung beflügelte die Umsetzung dieser Idee mit technischen Mitteln. Hier sei besonders die mechanische Ente aus dem Jahr 1738 von Jaques des Vaucanson , sowie der Schachtürke, den Wolfgang von Kempelen 1769 präsentierte, erwähnt. 1921 tauchte in Karel Čapeks Theaterstück R.U.R. die Bezeichnung des Roboters auf, in Anlehnung an „robota“, was im tschechischen Frondienst oder Zwangsarbeit bedeutet, als Idee einer intelligenten Maschine, die dem Menschen zu Diensten und sonst emotionslos sein soll. Dass sie das aber oft nicht ist,

⁸⁹ebda.,S.76

zeigt sich wiederum in diversen Buch- und Filmfiguren⁹⁰. Die Erschaffung eines intelligenten, künstlichen Wesens, welches dem Ebenbild des Menschen in vielerlei Hinsicht entspricht, ist der Kern der Forschung der Künstlichen Intelligenz(KI).

Die Idee aber, dass eine Maschine auch emotional dem Menschen ebenbürtig ist, genauer gesagt, erkennt, in welchem Gemütszustand sich der Mensch befindet, ist relativ jung. Da es im Bereich “affektiver Benutzerschnittstelle” mehrere Ansätze gibt, werden kurz die richtungsweisenden Ideen und Systeme vorgestellt.

Mittlerweile ist der Mensch überall von Technik umgeben, die ihm Arbeit abnehmen und erleichtern, im Endeffekt fehlende Fähigkeiten kompensieren soll, um ein vorgegebenes Ziel zu erreichen. Sprach man einst von “Technik”⁹¹ als (sinngemäß) einer Vorgehensweise, wie man mit Planung und praktischem Wissen dieses Ziel erreichen kann, ist der Begriff heute eher mit den Technologien verknüpft, die wir tagtäglich benutzen. Die Bedienung und der Umgang mit diesen Technologien sollen im Idealfall dem Menschen Freude machen und von den eigenen Fähigkeiten her zugänglich sein, was nicht immer gelingt. Schnell unterstellt man dem Computer oder anderen Geräten Absichten, wenn das Ergebnis nicht das Erwartete ist. Oder man fühlt sich als unfähig oder “zu dumm”, damit umzugehen. Der Mensch neigt dazu, die technischen Geräte, die er täglich benutzt, zu “vermenschlichen”. Er baut eine emotionale Beziehung dazu auf und passt seine Reaktionen und seine Art des Umgangs darauf an. Die Soziologin Marleen Brinks⁹² führte im Bereich Arbeits- und Organisationspsychologie an der Fernuni Hagen eine empirische Studie mit 340 Probanden durch, die täglich mehr als 30 Stunden mit dem PC arbeiten. Sie wollte das oftmals aggressive Verhalten gegenüber dem Computer, was man regelmäßig in Büros erlebt, untersuchen. Die Auswertung der Fragebögen zeigte, dass über $\frac{2}{3}$ schon mal Aggressionen gegenüber dem Computer zeigten, indem sie selbigen anschrien. Über 30% schlugen schon mal die Maus und ein kleiner Anteil von 1,5% beschädigte Teile des Computers. Sie vermutet, dass Frust, Unverständnis bei unerwartet auftretenden Fehlern und Vermenschlichung des Computers dieses Verhalten auslösen. Durch diese Eigenschaft des Menschen, Technologien auf die gleiche Stufe der Handlungsabsichten zu stellen wie der Mensch, resultiert eine bestimmte Form der Interaktion mit dem Auto, dem Computer oder dem Smartphone, die noch wenig erforscht ist, aber Unbeteiligte immer wieder erstaunt.

Rosalind Picard,Gründerin und Direktorin der Affective Computing Forschungsgruppe am

⁹⁰An dieser Stelle sei auf die Kurzgeschichten Isaak Asimovs, insbesondere auf “Der Zweihundertjährige” verwiesen, sowie auf den Androiden Data aus Star Trek - das nächste Jahrhundert

⁹¹altgriech. *technē*

⁹²vgl. Brinks, 2005. Leider ist die Studie online nicht mehr abrufbar.

Medienlabor des MIT, brachte 1997 mit ihrem Artikel ⁹³ erstmals den Begriff “Affective Computing” in die Informatik, mit dem Ziel, aufzuzeigen, wie sehr das Arbeiten mit dem Computer die Emotionen des Benutzers beeinflusst. Grundlage ihrer Überlegungen sind diverse Ergebnisse der Hirnforschung, unter anderem der Studie Cytowics, einem Neurologen, der synästhetische Phänomene untersuchte und feststellte, dass der limbische Bereich unseres Gehirns, welcher für Gedächtnis, Aufmerksamkeit und Emotionen zuständig ist, eine größere Rolle im Zusammenspiel mit dem Bereich, der für rationale Entscheidungen zuständig ist, spielt, als man bisher annahm. Bisher dachte man, dass der Cortex, der für die sensorischen Eindrücke zuständig ist, den größeren Einfluss hat. Die Untersuchungen LeDoux’ unterstreichen ebenfalls, dass Emotionen unsere Entscheidungen und Handlungen stark beeinflussen⁹⁴. Die Intensität der empfundenen Emotion entscheidet, wie rational (oder auch nicht) die Handlungen daraufhin ausfallen. Auch der Forschungsbereich der Künstlichen Intelligenz hat mittlerweile erkannt, dass die menschliche Intelligenz nicht allein durch logische Konklusionen und scheinbar intelligente Algorithmen erreicht werden kann. Die Annahme, dass der Computer die perfekte Umsetzung des Kartesischen Dualismus ⁹⁵ sei, kann so nicht gestützt werden.

Picard sah die Notwendigkeit, Computer mit mehr Einfühlungsvermögen zu entwerfen und forscht seither in diesem Bereich. Ein Computer, der erkennt, dass der Benutzer gestresst ist, vermag mehr zu bewirken als ein gefühlloser Kasten, “it might enhance computers’ abilities to make decisions”⁹⁶. Die Zusammenarbeit Mensch-Computer würde dabei stark verbessert und auf eine Ebene gebracht werden, die die Form der Zusammenarbeit unter Menschen am Besten simuliert. Menschen untereinander haben verschiedene Formen der Kommunikation und Interaktion. Der Mensch orientiert sich vor allem an den nonverbalen Signalen, die der Partner aussendet und reagiert entsprechend emotional darauf. Wenn man also die User Experience (UX) im Bereich Mensch-Computer-Interaktion so gestaltet, dass Frust und Ärger abnehmen, kann man die Effizienz und den Erfolg des Mitarbeiters steigern. Bisher versucht man, die Fehlerrate und die Bedienzeit so niedrig wie angemessen zu halten, sowie ein Interface zu gestalten, welches den Benutzer unterstützt und nicht verwirrt, genauso wie eine leichte Erlernbarkeit neuer Funktionen zu ermöglichen. Ziel hierbei ist es, den Benutzer zu einem Glücksgefühl und maximaler Zufriedenheit zu verhelfen. Wäre das durch die Fähigkeit des Computers, kontextsensitiv auf die Emotionen des Benutzers zu reagieren, verknüpfbar, würde man die natürliche Form der

⁹³Picard, 1997

⁹⁴LeDoux, 1995

⁹⁵Descartes teilt den Menschen in zwei Teile: *res cogitans* und *res extensa*. Der Verstand ist seiner Auffassung nach klar vom Körper getrennt und kann daher nicht als etwas Körperliches gesehen werden.

⁹⁶Picard, 1997, S.1

menschlichen Kommunikation nachstellen können und so dem Benutzer zu einem zufriedenen Umfeld verhelfen.

Picard wurde Mitbegründerin des MIT-Spin-Off *Affectiva*, die mit **Affdex** ein Programm zur Emotionserkennung über den Gesichtsausdruck entwickelten, welches gerade im Bereich des Marketing Einsatz beim Entwickeln geeigneter Werbung und Verkaufsgesprächsführungen helfen soll⁹⁷. Mit *Affectiva* entstand, im Kontext der Problematik der Emotionserkennung bei autistischen Menschen, die Idee, Veränderungen, die durch Stress oder starke Emotionen (vgl. hierzu Kap. 3.2) in der Leitfähigkeit der Haut durch Sensoren am Handgelenk aufzuzeichnen. Auslöser hierfür war die Problematik autistischer Menschen, ihre Emotionen mittels ihres Gesichtsausdrucks auch nach außen hin zeigen zu können. Daraus wurde der Q-Sensor entwickelt, der im medizinischen Bereich Betreuern und Ärzten bei der Analyse nach möglichen Auslösern für Stress bei Autisten helfen soll und im Alltag nicht stört. Um sich verstärkter der Sensorik im medizinischen Bereich, die auch im Alltag einsetzbar sind ⁹⁸, zu widmen, wurde Picard 2014 Mitbegründerin der Firma *Empatica*.

Ein weiterer spezieller Bereich für affektive Benutzerschnittstellen sind Lernprogramme als Beispiel möglicher Anwendungsbereiche. Hierbei geht es um die Maximierung des Lernerfolgs von Benutzern, zum Beispiel Schüler, Studenten oder Mitarbeiter, durch angemessene Reaktion des Programms auf den momentanen Gemütszustand des Benutzers. Der Lernerfolg würde, nach Picard, deutlich größer sein, wenn der Computer entsprechend sensibel auf die Reaktion der Benutzer reagiert und erkennt, ob sich bspw. Frust durch andauernden Misserfolg einstellt, der Benutzer eher überfordert oder unterfordert und deshalb gelangweilt ist. Dies führte zur Entwicklung und Untersuchung konkreter Systeme, die das Lernen durch Emotionserkennung und entsprechender Reaktion erleichtern sollen⁹⁹. In Hinblick auf das Thema dieser Arbeit sei besonders die spielbasierte Lernumgebung *Crystal Island* für Studenten von Jennifer L. Sabourin und James C. Lester¹⁰⁰ erwähnt, welche als Grundlage für die Analyse, wie stark kann man die Einstellung und Motivation von Studenten beim Lernen ändern und positiv beeinflussen kann, dient.

Das Fraunhofer Institut wiederum beschäftigte sich in dem Projekt (D)Emokoffer (Abb.3.1) damit, wie man Emotionen messen und gezielt bestimmte Emotionen beim Benutzer, während dessen Interaktion mit einem Computer, hervorrufen kann. Dabei kam unter anderem das Spiel

⁹⁷<http://www.affdex.com/>

⁹⁸Stichwort "wearable sensor"

⁹⁹vgl. Moridis, Economides, 2012. Whitehill et. al., 2014

¹⁰⁰Sabourin, Lester, 2014

EmoTetris als Applikation zum Einsatz. Während das Spiel lief, wurden die Emotionen des Benutzers mit Hilfe von Sensoren überwacht. Das Spiel reagierte auf die gemessenen Emotionen, indem es durch bestimmte Spielereignisse versuchte, bestimmte Emotionen wie Zorn oder Freude absichtlich hervorzurufen. Es kamen dafür bspw. ein guter und ein böser Zauberer zum Einsatz, die dem Benutzer Gutes tun, indem sie ihn in ein leichteres Level halfen, oder Schlechtes, indem sie ihn in ein schlechteres Level versetzten. Auch unterschiedliche Steine mit unterschiedlichen Eigenschaften stellten den Benutzer zeitweise auf eine Geduldprobe, indem die Steine sich nicht so verhielten, wie das sonst bei Tetris möglich ist.

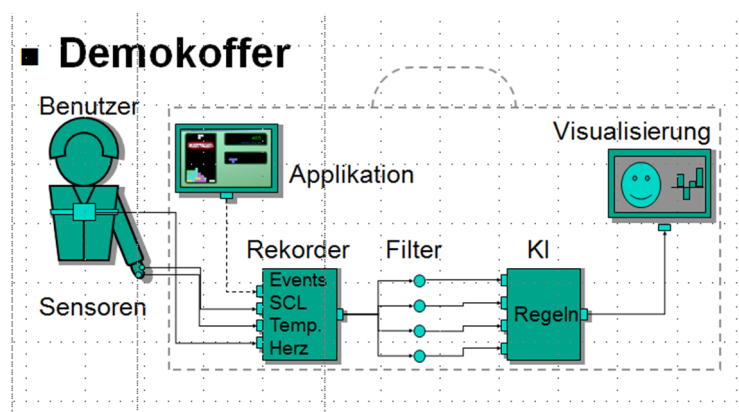


Abbildung 3.1: Überblick über die Struktur des Demokoffers des Fraunhofer Institut

Bildquelle: Vorlesungsfolien Fraunhofer Institut Graphische Datenverarbeitung

Die Computer - und Smartphone-Hersteller erforschen ebenfalls seit Jahren, wie man Emotionen des Benutzers analysieren und das Verhalten von Computer oder Smartphone entsprechend personalisieren kann, um den Benutzer ein vertrautes Gefühl beim Umgang mit den Technologien zu ermöglichen. Samsung ließ sich in diesem Zuge bereits im Oktober 2011 ein System zur Emotionserkennung patentieren, welches jedoch bis heute noch nicht einsatzbereit ist. Bisher basiert die prototypische Emotionserkennung, die im Januar 2012 auf der Consumer Communications and Networking Conference” (CCNC) vorgestellt wurde, auf der Analyse der Tippgeschwindigkeit und der Druckstärke des Benutzers, um festzustellen, ob er traurig oder wütend oder fröhlich ist. Wenn das System vom Benutzer ausreichend trainiert wurde, klassifiziert das Smartphone mit einer Trefferquote von 67,5% die Emotionen richtig¹⁰¹.

Auch Google ist aktiv am Forschen, um Emotionserkennung in ihrem Produkt “Google Glass”¹⁰² zu realisieren. Bisher ist die Brille nicht im Handel verfügbar, der Verkaufsstart ist für

¹⁰¹vgl. hierzu diverse Pressemitteilungen von Januar 2012, zb. <https://www.presetext.com/news/20120112001?likes=like>

¹⁰²vgl. hierzu <http://www.mygoogleglass.de/googleglass/>

Ende 2014 geplant, aber sollte das den Entwicklern gelingen, dann kann man, neben anderen beeindruckenden Features, die Emotionen des Gegenüber sich auf der Brille anzeigen lassen und entsprechend reagieren.

Abschließend sei Intel erwähnt, die auf der CES im Januar 2014 ihre “RealSense 3D Kamera” vorstellte, die auf der Kinect-Technik basiert und ab Mitte 2014 in Standardgeräten mit Microsoft Windows 8 und 8.1 zum Einsatz kommen. Das wäre zum Beispiel in UltrabooksTM, Computern mit abnehmbarer Tastatur, Notebooks, 2in1-Geräten und All-in-One-PCs¹⁰³. Diese Kamera bietet nicht nur eine gestengesteuerte Bedienung des Computers, sondern auch eine Emotions- und Gesichtserkennung. Wie jedoch die Emotionserkennung auf die Interaktion mit dem Benutzer konkret einwirkt, dazu äußert sich Intel leider nicht. Jedoch zeigen alle drei Beispiele, wie wichtig der Bereich “affective computing” wurde und dass es sich lohnt, sich mit den technischen Möglichkeiten der Emotionserkennung auseinander zu setzen.

3.2 Herzschlagfrequenz

Ein möglicher Ansatz für die Emotionserkennung wäre die Messung des Herzschlags, da sich dieser, durch die Ausschüttung von Hormonen¹⁰⁴, bei starken Emotionen wie Ärger und Freude erhöht. Jedoch ist - im Gegensatz zur Mimikerkennung - damit allein nicht eindeutig entscheidbar, welche Emotion den Herzschlag verursachte. Dazu muss der Herzschlag mit dem Gesichtsausdruck kombiniert betrachtet werden, weil man sonst die Emotion nicht oder falsch klassifizieren kann. War es jedoch bisher so, dass die Herzfrequenzmessung oftmals störende Kabel und wenig praktische Einsatzmöglichkeit lieferte, bieten technische Entwicklungen inzwischen Lösungen, die durchaus einsetzbar wären.

Das Computer Science and Artificial Intelligence Laboratory (CSAIL) des Massachusetts Institute of Technology (MIT) stellte im Juni 2012 auf der Siggraph (Special Interest Group on Graphics and Interactive Techniques)¹⁰⁵ eine neue Software vor, die aus aufeinanderfolgenden Videobildern Veränderungen verstärkt, die man mit dem bloßen Auge nicht wahrnimmt. Daraus extrahieren sie entweder die Herz- oder die Atemfrequenz. Ihre Software arbeitet vergleichsweise wie ein Equalizer, der manche Frequenzen verstärkt und andere unterdrückt.¹⁰⁶

¹⁰³<http://www.intel.de/content/www/de/de/architecture-and-technology/realsense-depth-technologies.html>

¹⁰⁴Vgl. Rost, 2005, S.69 ff.

¹⁰⁵international Conference and Exhibition on Computer Graphics and Interactive Techniques, siehe unter: <http://s2012.siggraph.org/>

¹⁰⁶vgl. web.mit.edu/newsoffice/2012/amplifying-invisible-video-0622.html

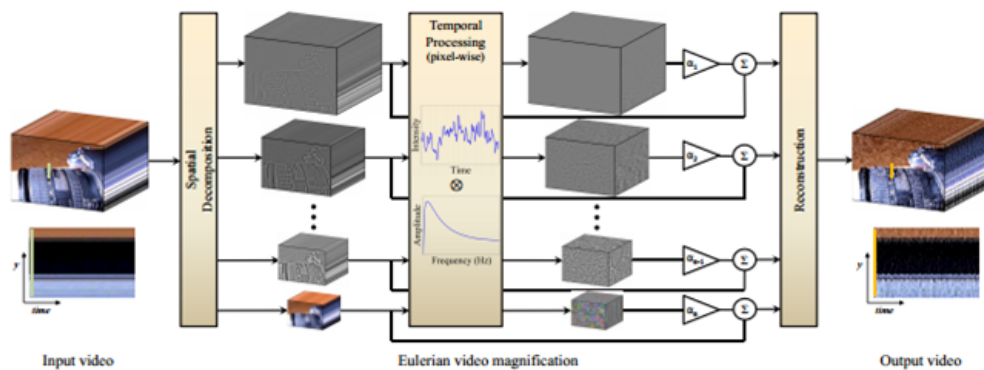


Abbildung 3.2: Überblick über die Anwendung der Eulersche Videoverstärkung

Bildquelle: people.csail.mit.edu/mrub/vidmag/

RUBINSTEIN et.al. nennen ihre Methode Eulersche Videoverstärkung (Eulerian Video Magnification). Ziel ist es, Variationen mit schwacher räumlicher Amplitude so zu verstärken, dass unsichtbare Signale in Videos erkennbar werden¹⁰⁷ (Abb.3.2). Dazu wird jedes Pixel der Videoeingangssequenz auf verschiedene Farbvariationen hin untersucht. Diese Werte ergeben einen Graphen in einem Koordinatensystem, aus welchem mit Hilfe der Fouriertransformation die enthaltenen Frequenzen ermittelt werden. Die verschiedenen Frequenzbereiche werden durch die Berechnung der Laplace-Pyramide, welche ähnlich wie ein Bandpassfilter wirkt, bearbeitet. Dabei werden nur die Frequenzbereiche, die interessieren, verstärkt und die anderen abgeschwächt. So sprechen RUBINSTEIN et. al. von einem Frequenzbereich von 0,4 bis 4 Hz, der einem Herzschlag von 24 bis 240 Schlägen/ Minute entsprechen, der erhalten wird, wenn man die Herzfrequenz extrahieren möchte¹⁰⁸. Die gefilterten Signale werden schlussendlich mit einem Faktor a verstärkt, zum Originalsignal addiert und zu einem Video zusammengefügt.

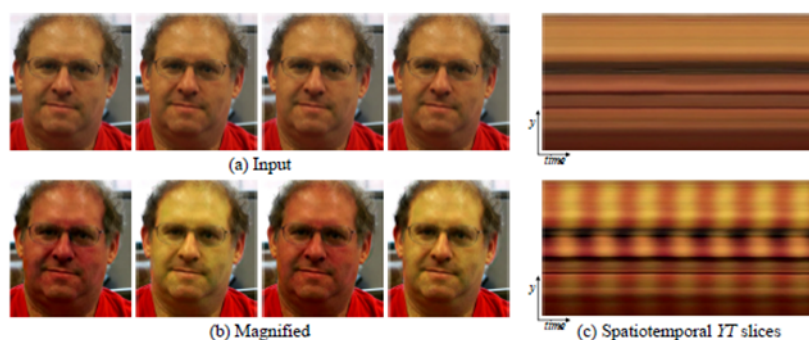


Abbildung 3.3: MIT - Eulersche Videoverstärkung

Bildquelle: people.csail.mit.edu/mrub/vidmag/

Abb. 3.3 zeigt die Erkennung des Herzschlags mit Hilfe des RGB-Farbraums. Dazu bedienen

¹⁰⁷<http://people.csail.mit.edu/mrub/papers/vidmag.pdf>

¹⁰⁸<http://people.csail.mit.edu/mrub/papers/vidmag.pdf>

sich die Forscher des MIT der Tatsache, dass durch den Herzschlag regelmäßig Blut in Gefäße gepumpt wird, was eine Rötung der Haut verursacht. Gerade im Gesicht, in dem feine Äderchen großflächig und oberflächlich die Haut versorgen, ist diese Färbung am Besten zu erkennen, da dem Betrachter Adern, die im menschlichen Körper relativ weit oben unter der Haut liegen, Rot erscheinen. Diese Blutgefäße absorbieren, im Gegensatz zu tiefer gelegenen Gefäßen, die anderen Spektralfarben und reflektieren das langwellige Rot ¹⁰⁹. Durch die verschiedenen Frequenzbereiche der Spektralfarben kann mittels Zerlegung und Filtern der Rot- bzw. Grünanteil in der Gesichtsfarbe analysiert werden. Die Häufigkeit der sich wiederholenden Rotfärbung in einem Zeitraum (temporal variations) ergibt die Herzfrequenz. Das bloße Auge ist nicht in der Lage, diese Verfärbungen zu erkennen, die Software deckt dies mittels Verstärkung auf.

Interessant ist die beiläufige Entdeckung, dass man nicht nur Farbänderungen verstärken kann, sondern auch Bewegungen. Dies könnte für die Überwachung im medizinischen Bereich sehr von Nutzen sein, wie RUBINSTEIN et. al. anhand der möglichen Überwachung von Frühgeborenen auf einer Intensivstation zeigen. Dabei kann die Software das Heben und Senken des Brustkorbs, welches in der Regel bei Frühchen mit dem bloßen Auge schwer erkennbar ist, verstärken und auf einem Monitor darstellen. Aber nicht nur bei Frühchen, auch in anderen medizinischen Bereichen, in denen ein ständiges Untersuchen den Patienten sehr belastet, könnte diese Technik, vor allem durch die Messung in Echtzeit, hilfreich sein.

Auch der freie Markt bietet mittlerweile mehrere Möglichkeiten, den Herzschlag zu messen (Blutdruckmessgeräte, Pulsuhren mit Brustgurt,...). ADIDAS brachte 2013 eine Pulsuhr auf den Markt, die optisch den Herzschlag misst (Abb.3.4 a). Dazu wird mittels Licht an der Innenseite der Uhr das Handgelenk beleuchtet und durch Sensoren der Blutfluss der Kapillaren am Handgelenk erfasst. Der Takt der Intensität der Kapillareneinfärbung ergibt dann die Herzfrequenz. Die Uhr kam im November 2013 auf den deutschen Markt und gibt Sportlern die Hoffnung, für die Herzfrequenzmessung eine Uhr ohne den lästigen Brustgurt nutzen zu können und genauso zuverlässige Werte zu erhalten. Erste Tests fallen entsprechend positiv aus. ¹¹⁰. Eine ähnlich gelagerte Möglichkeit, den Herzschlag zu erfassen, gibt es als App namens "Cardiograph" für das Smartphone (Abb.3.4 b). Auch diese ermittelt die Einfärbung mittels Licht. Dazu legt man den Finger, nach dem Starten der App, auf die Kamera des Smartphones. Die App ermittelt mit Hilfe des Blitzlichts die Intensität der Einfärbung der Fingerspitze ¹¹¹. Durch

¹⁰⁹<http://www.farbimpulse.de/424.pdf>

¹¹⁰<http://www.runnersworld.de/ausruestung/adidas-micoach-smart-run-im-test.301986.htm>

¹¹¹http://www.aerztezeitung.de/praxis_wirtschaft/w_specials/gesundheitsapps2011/article/806569/herzfrequenzmesser-unterwegs.html

die Häufigkeit der Farbänderung errechnet sich dann die Herzfrequenz.

Die Firma Cardio.Inc.¹¹², ein Spin-Off des CSAIL, entwickelte aus den Forschungsergebnissen für Apple eine App namens Cardio, die den Ansatz RUBINSTEINs et.al. auch im privaten Bereich anwendbar macht (Abb.3.4 c). Dafür wird das Gesicht des Nutzers mittels der Frontkamera seines iPhones oder iPads für eine bestimmte Zeit ebenso daraufhin untersucht, wie viel Licht das Gesicht absorbiert.

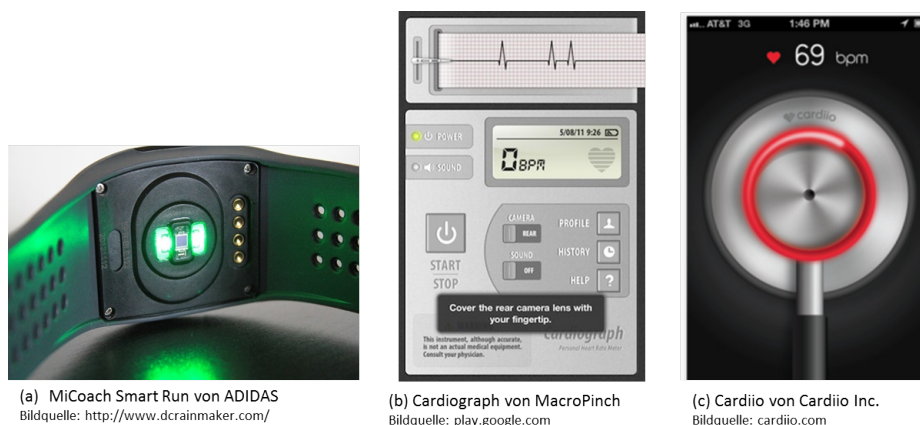


Abbildung 3.4: Verschiedene Lösungen zur Herzschlagmessung

Abschließend sei noch der “cosinuss - Sensor”¹¹³, eine Erfindung Johannes Kreuzers, erwähnt, der 2011 im Rahmen seiner Doktorarbeit an der Technischen Universität München am Lehrstuhl für Realzeit-Computersysteme ein Gerät zur Messung von Temperatur und Puls im Ohr entwickelte. Kreuzers Ziel war es, einen drahtlosen Sensor zu entwickeln, der in der Bewegung nicht stört und vielseitig, vor allem im Sport, einsetzbar ist. Auslöser war das Problem, wie man an Probanden, die Funktionskleidung im Sportbereich testen, Temperatur und Puls am Angenehmsten messen kann. Bisherige Möglichkeiten über Sensoren in der Speiseröhre oder rektal waren nicht sehr attraktiv, auch wenn sie die beste Möglichkeit waren, kontinuierlich die Körpertemperatur zu erfassen. Auch der Brustgurt für die Erfassung der Herzfrequenz wird von vielen Sportlern als unangenehm empfunden. So kam Kreuzer die Idee, den Gehörgang als geeignete Stelle auszusuchen, da diese Stelle, gerade durch drahtlosen Datentransfer, den Menschen am Wenigsten beeinträchtigt, und entwickelte einen Sensor, der so groß ist wie ein Hörgerät und eng im Gehörgang anliegt (Abb.3.5). Er misst die Temperatur auf 0,01 Grad Celsius genau und gleichzeitig den Puls mit der gleichen Technik wie die zuvor beschriebenen (Prinzip der Pulso-

¹¹²<http://www.cardio.com/>

¹¹³<http://cosinuss.com/>

ximetrie¹¹⁴). Die erfassten Daten werden hinter dem Ohr im Gerät gespeichert und können an ein Handy, an eine kleine Box, die man bequem in der Hosentasche tragen kann, oder an den PC gesendet werden.

Inzwischen hat Kreuzer mit anderen Absolventen ein Unternehmen gegründet und entwickelt den Sensor, vor allem im Bereich der Herzfrequenz Erfassung, weiter, damit die Messungen noch exakter werden.

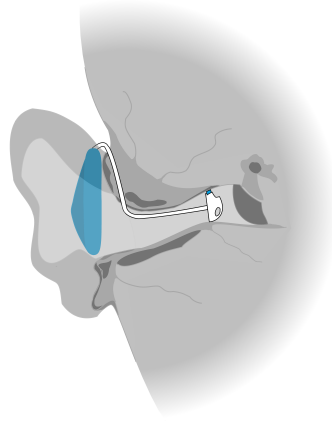


Abbildung 3.5: Lage des cosinuss-Sensor im Gehörgang

Bildquelle: <http://mediatum.ub.tum.de/>

3.3 Methoden der Emotionsklassifikation

Seit vielen Jahren erforschen Wissenschaftler das Problem, Emotionen anhand von Bildern oder dem Stream einer Webcam zu erkennen. Was für den Menschen, wie EKMAN in seinen jahrzehntelangen Studien zeigte, schon nicht leicht ist, ist schon gar nicht so leicht zu implementieren. Auch wenn die AUs maßgeblich zur Klassifizierung des Gesichtsausdrucks sind, so müssen die Informationen technisch erst gewonnen und sinnvoll verarbeitet werden. Ein weiteres Problem ist, dass die Analyse von Bilddaten sehr rechen- und ressourcenintensiv ist.



Abbildung 3.6: Basisemotionen

Bildquelle: Cohn-Kanade-Datenbank

Jede Emotion äußert sich durch bestimmte Merkmale, die charakteristisch sind (Abb.??). So ist bei Zorn die Nasenwurzel anders gekräuselt als bei Verachtung oder Freude. Bereits bestehen-

¹¹⁴Das bedeutet, dass der Puls nicht-invasiv erfasst wird, indem die Sauerstoffsättigung mittels Messung der Lichtabsorption bestimmt wird. Mehr unter: <http://de.wikipedia.org/wiki/Pulsoxymetrie>

de Arbeiten¹¹⁵ zeigen, dass mit vorhandenen Methoden der Objekterkennung eine zuverlässige Emotionsklassifizierung möglich ist und die Hinzunahme der Herzfrequenz überflüssig macht. Gängige Emotionserkennungssysteme setzen sich im Allgemeinen aus drei Kernmethoden zusammen:

- i) Gesicht entdecken (*Viola-Jones-Detektor*)
- ii) Extraktion der Emotionsmerkmale (*z.B. SIFT oder SURF*)
- iii) Klassifikation des Gesichtsausdrucks (*z.B. Bag of Words + Support Vector Machine*)

Ergänzend sei erwähnt, dass es weitere Möglichkeiten der Merkmalsextraktion statt SURF gibt, die auch vereinzelt eingesetzt werden, wie zum Beispiel mit Gaborfilter¹¹⁶ oder Active-Apperance-Model (AAM)¹¹⁷. Auf diese soll hier jedoch nur verwiesen werden, da die Kombination von SIFT bzw. SURF mit einer SVM der aktuelle Stand der Wissenschaft ist.

3.3.1 Viola-Jones-Detektor (*face detection*)

Wie bereits erwähnt, ist die Analyse von Bilddaten sehr rechenintensiv. Um Ressourcen und damit Zeit zu sparen, ist es sinnvoll, den Bildbereich auf einen Bildausschnitt einzuschränken. Zur Klassifikation von Emotionen benötigen wir nur das Gesicht, daher reicht es aus, das Bild nach einem Gesicht zu durchsuchen (region of interest(ROI)) und dadurch den Bereich einzuschränken. Hierfür hat sich eine Methode als besonders effizient erwiesen: die Viola-Jones-Methode¹¹⁸. Diese ist nicht nur für Gesichter einsetzbar, sondern generell für eine Vielzahl von Objekten, wenn sie entsprechend trainiert wird. Im Falle der *face detection* werden die Klassifikatoren auf Gesichtsmerkmale trainiert. Diese Methode gibt die Position des Gesichtes im Bild wieder, wodurch die restlichen Bereiche des Bildes verworfen werden können und nicht im weiteren Prozess berücksichtigt werden müssen.

VIOLA und JONES kombinieren in ihrem Algorithmus drei unterschiedliche Konzepte miteinander und bilden damit ein System, welches auf den Merkmalen eines gegebenen Bildes basiert und dadurch schneller operiert als ein pixelbasiertes System. Es setzt sich zusammen aus einem *integral image*, welches eine sehr schnelle Bewertung von Merkmalen ermöglicht, einer *modifizierten AdaBoost Prozedur* und einer *cascade of classifiers* zur Beschleunigung des Detektors.

¹¹⁵Azcarate et.al.,Amsterdam,2005; Su et. al., Taiwan,2007; Huang, Hung-Fu, Tai,Shen-Chuan, Taiwan,2012; Dumas, Melanie, San Diego

¹¹⁶Chia, John, University of British Columba, Vancouver; Krishna, Tarun et. al., The LNM Institute of Information Technology, Jaipur, Indien

¹¹⁷Cootes,Edwards,Taylor,University of Manchester

¹¹⁸Viola P., Jones M.,Rapid Object Detection using a Boosted Cascade of Simple Features,2001

- **Integral Image**

Die Methode basiert auf dem Wert einfacher Merkmale eines Grauwertbildes, d.h. das Eingabebild wird, wie im Bereich der Bildverarbeitung oft üblich, in ein Graustufenbild konvertiert. Jeder Grauwert errechnet sich durch die gewichtete Transformation der Farbinformationen des Pixels¹¹⁹ und gibt den Helligkeitswert (bzw. den Intensitätswert) des Pixels wieder. Um die Berechnung zu beschleunigen, führten VIOLA und JONES erstmals das Konzept des Integralbildes in der Bildverarbeitung ein, welches bis dahin von CROW im Bereich des *texture mapping* benutzt wurde¹²⁰.

Der Wert des Pixels (x,y) im Integralbild stellt die Summe aller Pixel innerhalb eines Rechtecks dar, welches zwischen dem aktuellen Pixel (x,y) und dem Ursprung $(0,0)$ aufgespannt wird.

$$I_{\Sigma}(x, y) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j) \quad (1)$$

So errechnet sich die Summe in Abb. 3.7 aus den Werten der Pixel, die zwischen dem Ursprung und dem markierten Pixel (x,y) liegen

$$I_{\Sigma}(x, y) = 255 + 3 \cdot 200 = 855 \quad (2)$$

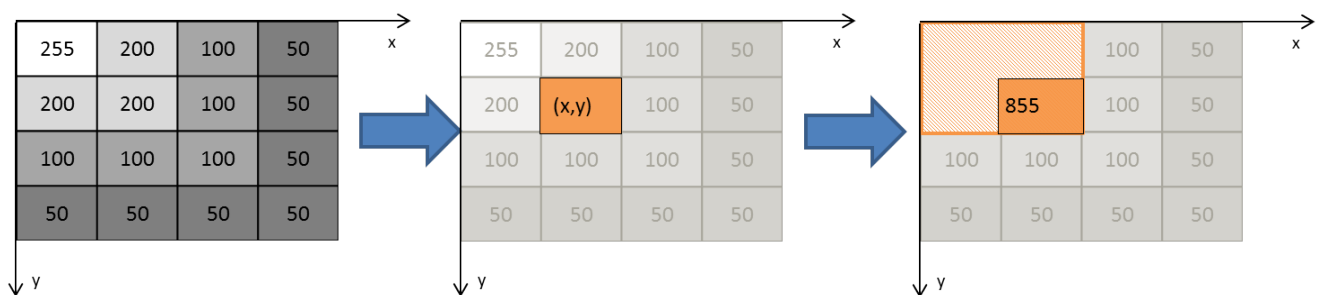


Abbildung 3.7: Umwandlung eines Grauwertbildes in ein Integralbild

Bildquelle: eigener Entwurf

Die Verwendung des Integralbildes als Grundlage für weitere Berechnungen reduziert die Rechenzeit deutlich, da die Ermittlung der Pixelsumme auf Elementaroperationen basiert,

¹¹⁹Die Berechnung erfolgt im RGB für ein Pixel mit der Formel: Grauwert = 0,299*Rot + 0,587*Grün + 0,144*Blau. Die Gewichtungen ergeben sich aus der unterschiedlichen Farbempfindlichkeit des menschlichen Auges ggü. den Grundfarben

¹²⁰Crow, 1984

für die der Rechner wenig Speicher und Zeit benötigt. Darüber hinaus bietet die Struktur des Integralbildes eine einfache Berechnung mit Hilfe von bereits errechneten Werten (Abb.3.8).

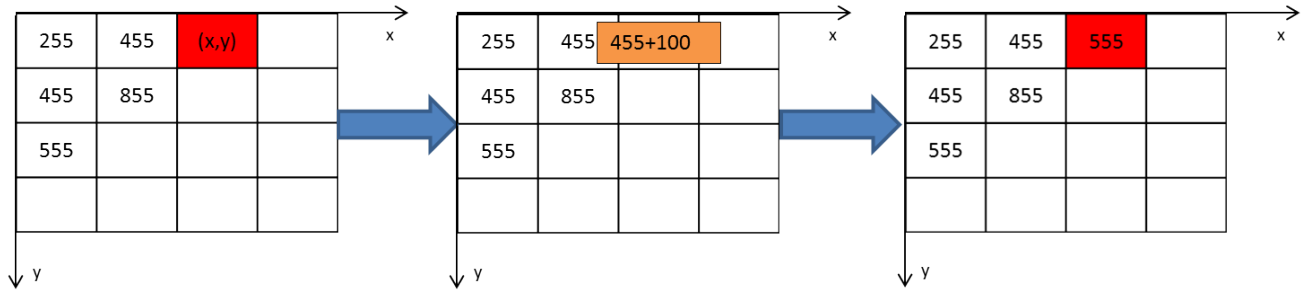


Abbildung 3.8: Berechnung der nächsten Pixelsumme

Bildquelle: eigener Entwurf

Der Rechenaufwand ändert sich auch nicht, wenn man eine beliebige Fläche, die nicht bündig am Ursprung anliegt, mit Hilfe des Integralbildes ermitteln will. Das wiederum ermöglicht den weitergehenden Prozess auf beliebigen Bildausschnitten (Abb.3.9)

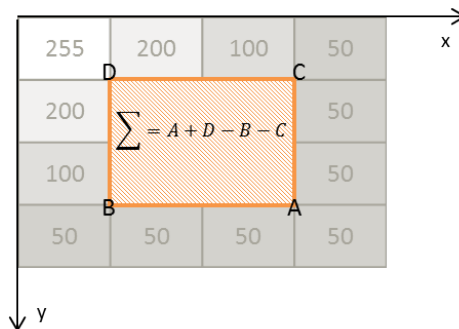


Abbildung 3.9: Berechnung der Pixelsumme eines beliebigen Rechtecks

Bildquelle: eigener Entwurf

Da die Werte des Integralbildes sich aus den Grauwerten in der Umgebung des Pixels im Originalbild errechnen, geben sie über die Wichtigkeit und die Größe des Pixels in seiner Umgebung Aufschluss. Je heller der Bereich ist, desto höher ist die Summe. Die Information, ob ein Bereich hell oder dunkel ist, ist eine wichtige Grundlage für den weiteren Prozess. Wenn man sich nämlich ein Gesicht genauer betrachtet, entdeckt man typische Helligkeitsunterschiede. So sind zum Beispiel Augenbrauen in der Regel dunkler als die Augen selbst, bzw. die Nasenwurzel heller als die Augen (Abb. 3.10).



Abbildung 3.10: Haar-Merkmale eines Gesichts

Bildquelle: <https://github.com/kylemcdonald/AppropriatingNewTechnologies>

Aus diesen Erkenntnissen heraus entwickelten VIOLA und JONES für die Detektion von Gesichtern fünf verschiedene Rechtecksmerkmale(siehe die Merkmale A3,A4,B2,B5 und C in Abb.3.11), die durch Rotation möglichst all die Bereiche auffinden lassen, die durch den Helligkeitsunterschied ein Gesicht beschreiben. Diese Merkmale haben durch ihre Rechtecksform eine starke Ähnlichkeit mit Haar-basierten Wavelet-Funktionen, weshalb in verschiedenen Ausarbeitungen auch immer wieder von *Haarlike features* die Rede ist¹²¹.

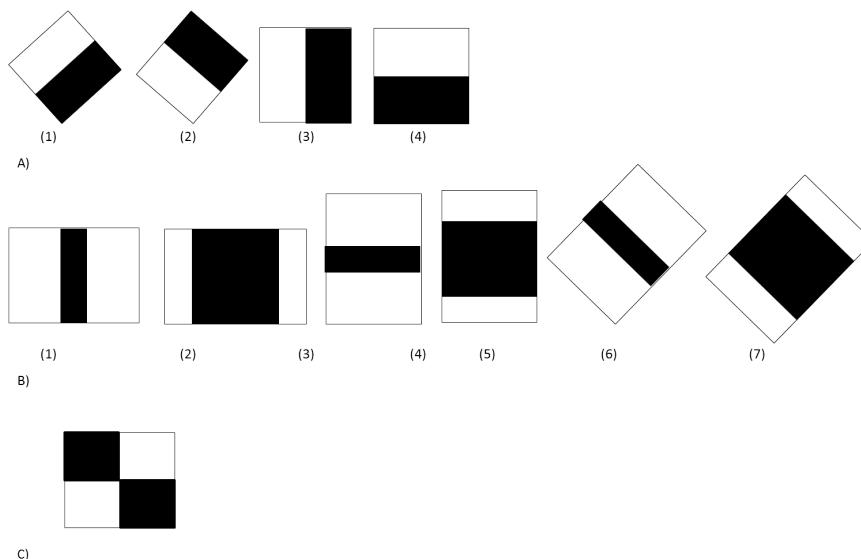


Abbildung 3.11: Haar-Merkmale der Viola-Jones-Methode

Bildquelle: eigener Entwurf

¹²¹vgl. hierzu u.a. AZCARATE et.al.,2005

Um den Wert eines Haar-like Merkmals zu berechnen, wird die Differenz zwischen der Pixelsumme des hellen und der Pixelsumme des dunklen Rechtecks über die zuvor ermittelten Pixelwerte im Integralbild berechnet. So ermittelt man den Wert der Kantenmerkmale (*two-rectangle features*¹²²) (A) anhand rechteckiger Regionen, die in der Regel gleich groß sind, die gleiche Form besitzen und entweder horizontal oder vertikal angeordnet sind. Diese Merkmale dienen beispielsweise der Detektion der Augenpartie.

Das Linienmerkmal *three-rectangle-feature*¹²³(B) besteht aus zwei außenliegenden Regionen, die eine dritte einschließen. Die Summe zwischen den außenliegenden Regionen wird um die Summe in der mittig gelegenen Region verringert, um den Wert zu bestimmen. Dieses Merkmal wird unter anderem für das Auffinden des Nasenrückens benutzt.

Im letzten relevanten Merkmal *four-rectangle feature* (C) sind vier Regionen diagonal angeordnet. Hier wird die Summe der jeweils diagonal gelegenen Regionen gebildet und voneinander abgezogen.

Auch wenn Rechtecksmerkmale nur begrenzt flexibel und durch die einzig möglichen Ausrichtungen in vertikaler, horizontaler und diagonaler Richtung eher grobkörnig sind, so lässt sich aus der Menge der Merkmale eine reichhaltige Bildrepräsentation gewinnen, die ein effektives Lernen unterstützt. Die Verbindung mit dem *integral image* gleicht die eingeschränkte Flexibilität genügend aus.

- **Modifizierte AdaBoost Prozedur**

Der AdaBoost Lernalgorithmus wird generell benutzt, um die Leistung bei der Klassifikation eines einfachen (“weak”)Lernalgorithmus zu verstärken (“boosten”). Dieser gehört in den Bereich der Methoden des *machine learning*, die in der Regel mit Hilfe einer gegebenen Menge an Merkmalen und einer Trainingsmenge von positiven und negativen Bildern eine Funktion zur Klassifikation erlernen. Da die Fehlerrate, die ein verstärkter *classifier* in der Trainingsmenge schafft, exponentiell mit der Anzahl der Durchläufe gegen Null konvergiert, ist AdaBoost sehr gut für das Aufspüren von Objekten geeignet.

Das Unterfenster, mit welchem das Bild nach Merkmalen durchsucht wird, hat eine Auflösung von 24 x 24 Pixel. Dadurch ist die vollständige Menge der Rechtecksmerkmale mit über 180 000 Elementen sehr mächtig, die jedem Unterfenster zugeordnet sind. Man bekommt zwar mehr Merkmale in dem Unterfenster als Pixel, aber die Berechnung dieser Menge

¹²²VIOLA/JONES,2001,S.2

¹²³ebda.

verursacht zu hohe Kosten. VIOLA und JONES hatten durch Beobachtungen während durchgeführter Experimente die Vermutung, dass auch eine sehr kleine Merkmalsmenge ausreichen müsste, wenn man die Elemente optimal miteinander kombinieren könnte, um einen effektiven *classifier* zu konstruieren. Daher wandelten sie den AdaBoost Algorithmus so ab, dass dieser eine kleine Menge Merkmale nutzt **und** einen *classifier* trainiert.

Im modifizierten AdaBoost bestimmt der Lernalgorithmus für jeden schwachen *classifier*, welchen Fehler ϵ_j dieser für ein Merkmal f_j in einem Trainingsbild macht. Dazu werden am Anfang die y -Werte der Eingangsbilder (x_n, y_n) , die dem Training dienen, auf 0 gesetzt, die kein Gesicht enthalten, ansonsten auf 1. Jedes Bild wird zusätzlich gewichtet mit $w_n = 1/2m$ (m = Anzahl positiver Beispiele) bzw. $w_n = 1/2l$ (l =Anzahl negativer Beispiele). Für jedes Merkmal berechnet der Lerner zuerst einen optimalen Schwellwert θ . Dieser dient dem *classifier* $h_j(x)$ zur Klassifikation derart, dass er den Wert 1 zurückgibt, wenn das Merkmal f_j im Trainingsbild größer als der Schwellwert θ_j ist, ansonsten 0. Das Ziel ist, dass die Anzahl der falsch klassifizierten Beispiele minimal ist. Wenn nun also der Fehler ϵ_j jedes *classifiers* $h_j(x)$ berechnet wurde, wird der mit dem kleinsten Fehler ausgewählt und die Gewichte der Trainingsbilder, die von diesem *classifier* als falsch klassifiziert wurden, erhöht. Am Ende der Methode wurde ein starker *classifier* konstruiert, als Linearkombinationen der in dem Durchlauf durch alle Bilder ausgewählten *classifier*.

- **Cascade of Classifiers**

Das letzte Konzept der Viola-Jones-Methode ist ein Algorithmus zur Konstruktion einer Klassifikationskaskade, der gleichzeitig die Detektionsleistung erhöht und die Rechenleistung massiv verkürzt. Die Grundidee dahinter ist, dass kleinere und dafür effizientere, verstärkte *classifier* konstruiert werden können, welche zahlreiche negative Unterfenster aussortieren, dabei aber fast alle positiven Instanzen finden. Die Optimierung von Leistung und Ressourcen wird also durch einfachere *classifier* erreicht, die die Masse von Unterfenster vorsortieren, bevor komplexere *classifier* angewandt werden. Dadurch wird die Rate der fälschlich ausgemusterten Positive minimiert. Da die Form dieses Prozesses wie ein entarteter Entscheidungsbaum aussieht, nennen VIOLA und JONES ihn *Kaskade*. Der Begriff drückt dabei aus, dass jeder vorhergehende *classifier* die Berechnung des Nächsten durch ein positives Ergebnis auslöst. Kommt ein falsches Ergebnis heraus, wird das Unterfenster entfernt

3.3.2 Extraktion relevanter Regionen - SURF

Menschen können einen Bildinhalt erst dann interpretieren, wenn Objekte deutlich erkennbar sind. Jedes Objekt wird durch Kanten oder Ecken begrenzt und hebt sich dadurch von einem Hintergrund ab. Je homogener der Hintergrund und je größer der Farbunterschied von Hintergrund und Objekt ist, desto deutlicher hebt sich die Kontur hervor. Daher sucht man in der Regel in der Bildverarbeitung nach charakteristischen Punkten bzw. Merkmalen, die Kanten und Ecken beschreiben, weil ein Objekt anhand dieser eindeutig erkannt werden kann.

Als Herbert BAY et. al. ihren Merkmalsextraktor entwickelten, gab es in der Objekterkennung zwar schon bestehende Systeme aus Merkmalsdetektor und - deskriptor, aber sie waren entweder zu rechenintensiv oder zu empfindlich gegenüber Änderungen der räumlichen Orientierung, Lichtverhältnisse oder Skalierung. Einer der meistgenutzten Detektoren war der *Harris Corner Detector*¹²⁴, der auf dem Ecken-Detektor von MORAVEC¹²⁵ basiert. Während MORAVECs Operator mittels einem lokalen Fenster, welches er in alle vier Richtungen verschob, die Intensitätsveränderungen mit Hilfe der aufsummierten Intensitätswerte der Pixel ermittelt, um Ecken aufzufinden, verbesserten HARRIS et. al. diese Idee, indem sie den Gradienten, d.h. die Änderungsrate der Pixelwerte und nicht die absoluten Pixelwerte zu Grunde legten. Die Leistung des Harris-Detektors ist zwar wesentlich besser als der Detektor von MORAVEC, und im Gegensatz dazu auch rotationsinvariant, aber er benötigt deutlich mehr Rechenaufwand. Und beide Detektoren sind nicht skalierungsinvariant, was gerade im Bereich der Objekterkennung sehr nachteilig ist, da die Objekte nicht immer in der gleichen Größe vorliegen.

Jahre später entwickelte LOWE einen Algorithmus, der Merkmale detektiert und jedes durch einen Merkmalsvektor eindeutig beschreibt. Ein Merkmal ist ein Bereich, der charakteristisch und einzigartig und daher zum Vergleich sehr gut geeignet ist. Da der Algorithmus skalierungs- und rotationsinvariant ist, nannte LOWE ihn *Scale-invariant Feature Transform* (SIFT). Da sich die Funktionsweise von BAYs Merkmalsextraktor stark am Konzept des SIFT orientiert, soll die Funktionsweise von SIFT kurz erläutert werden.

- **Scale-invariant Feature Transform(SIFT)**

Die Arbeitsweise des Detektors basiert auf einer approximierten Laplacepyramide. Die

¹²⁴C. Harris, M. Stephens, 1988

¹²⁵Moravec, Hans P., 1977

einzelnen Ebenen (Oktaven) dieser Pyramide werden aber nicht, wie sonst bei der Laplacepyramide üblich, durch Größenveränderung des Bildes erzeugt, sondern durch die Anwendung von Gaußfilter und die Differenz zweier Ergebnisse daraus. Die einzelne Oktave entsteht durch das Anwenden von fünf Gaußfilter auf das Bild und enthält am Ende vier *Difference-of-Gaussians* (DoG)-Intervalle (Abb.3.12). Durch die Differenz zweier Gaußfilter-Ergebnisse erreicht man eine Approximation der Gauß-Laplace-Funktion. Da der Gaußfilter ein Tiefpassfilter ist, stärkt er die Bereiche des Bildes, die niedrige Frequenzen aufweisen und verwischt Störungen (hohe Frequenzen), erhält aber Kanten und Ecken. Durch die Differenz wird anschließend der hohe Frequenzbereich der Kanten und Ecken im Bild betont (Laplace). Jede weitere Oktave bekommt dann das letzte Bild der vorhergehenden Oktave als Eingabe, welches in x- und y-Richtung halbiert wird. Dann wird der Vorgang wiederholt und die neue Oktave mit wiederum vier DoG-Intervallen gefüllt.

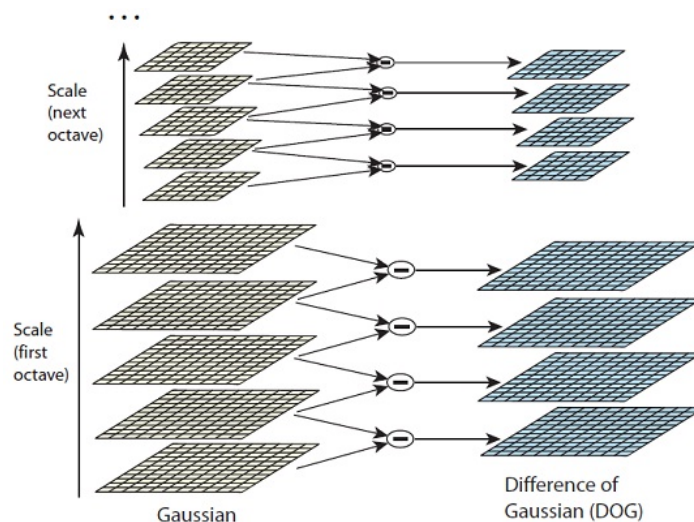


Abbildung 3.12: Oktaven von SIFT

Bildquelle: Lowe, 2004

Nach der Filterung lokalisiert der Algorithmus die Position eines *interest point*, indem er die Extremwerte innerhalb der Oktave durch Vergleich mit den erzeugten DoG-Ergebnissen, also mit den Nachbarn in der Ebene darunter und darüber, ermittelt. Der Betrag des Gradienten an einem ermittelten Punkt dient anschließend dazu, die Punkte auszusortieren, deren Betrag unter einem Schwellwert liegen, um Punkte auszusortieren, die Störungen verursachen. Die anschließende Bestimmung der dominanten Orientierung des Punktes in Bezug auf seine Umgebung gewährleistet die Rotationsinvarianz. Dazu wird ein 16 x 16 (entspricht 256 Pixel) Fenster zentral auf den Punkt gelegt und in 4x4 große Unterfenster unterteilt (ergibt 16 Unterfenster mit je 16 Pixel). In jedem Unterfenster werden nun die

Längen und die Richtungen der Gradienten berechnet. Die Richtung eines Vektors wird in der Regel durch den Winkel, der von dem Vektor mit einer horizontalen Achse eingeschlossen wird, bestimmt. Bei SIFT wird jedoch die Orientierung durch den Winkel α bestimmt, den der Vektor in einem Umkreis verstrichen hätte (Abb.3.13).

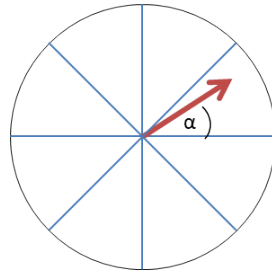


Abbildung 3.13: Bestimmung der Orientierung eines Gradienten

Bildquelle: eigener Entwurf

Der Innenkreiswinkel (360°) des Umkreises wird in acht gleich große Winkelklassen aufgeteilt, die ihrerseits die *bins* eines Histogramm bilden, d.h. $\alpha = 0^\circ - 44^\circ$ bildet die erste Klasse, $\alpha = 45^\circ - 89^\circ$ die Zweite, usw. . Nun wird jeder der 16 Gradienten im Unterfenster entsprechend seines Betrags gewichtet und in Abhängigkeit seiner Richtung in eine der acht Klassen eingeteilt, woraus sich nachher ein Merkmalsvektor mit $16 \cdot 8 = 128$ Werten zusammensetzt, der eine einzigartige Signatur für das Merkmal darstellt.

- **Speeded-Up Robust Features (SURF)**

BAY et. al. benutzen, im Gegensatz zu SIFT, für ihre Methode das Integralbild als Grundlage und erreichen damit, sowie durch die Kombination mit einem Mittelwertfilter, eine deutliche Beschleunigung der Merkmalsextraktion. SURF teilt sich ebenfalls in einen Detektor und einen Extraktor auf.

Der Detektor arbeitet mit der Hesse-Matrix und der Faltung des Bildes mit Filter, die auf der 2. Ableitung der Gaußfunktion basieren. Die Anwendung des Filters, d.h. die Faltung des Bildes, in einem Pixel ermöglicht zu bestimmen, wie wichtig das Pixel ist. Die Wahl der Gaußfunktion begründet sich in der Eigenschaft, dass charakteristische Punkte in der Regel in einer *blob*¹²⁶-ähnlichen Struktur zu finden sind. Ein Blob ist eine Bildbereich, in dem die Grauwerte benachbarter Pixel ähnlich groß oder sogar gleich sind. Ein Objekt, das sich deutlich von einem Hintergrund heraushebt, stellt solch einen Blob dar. Daher kann das Konzept der Normalverteilung, die der Gaußfunktion zu Grunde liegt, hier zum Tragen kommen. Die Gaußfunktion gewichtet, gemäß der Normalverteilung, das Zentral-

¹²⁶engl. Klecks, Tropfen, in der Bildverarbeitung bedeutet es Binary Large Object (Blob)

pixel stärker als die benachbarten Pixel. Dadurch haben weiter entfernte Pixel weniger Einfluss auf das Ergebnis als die unmittelbaren Nachbarn. Darüber hinaus verwischt die Gaußfunktion vorhandene Kanten und Ecken nicht so stark wie ein Mittelwertfilter und entfernt Störungen. Mit der Variation des Parameters σ lässt sich die Reichweite des Filters einstellen, d.h. wie viel der Umgebung wird mit einbezogen.

Um zuverlässig Kanten und Ecken zu ermitteln, wird gern, wie bei HARRIS schon, der Gradient verwendet, der aus den partiellen ersten Ableitungen an einem Punkt besteht. Kanten und Ecken sind Bereiche in einem Bild, an denen sich die Grauwertintensität auf kleinem Raum stark ändert (Extremum des Gradienten). Der Nachteil ist aber, dass Kanten oder Ecken, die sich nicht so stark vom Hintergrund hervorheben, nicht zuverlässig genug erkannt werden. Daher nutzt man gerne statt der ersten Ableitung die zweite Ableitung. Diese gibt im Graph der Gaußfunktion die Wendepunkte mittels dem Nulldurchgang an, also die Stellen, an denen sich der Grauwert am Stärksten ändert.

Fast-Hessian Methode

Der Filter L , den BAY et. al. benutzen, setzt sich aus drei Filter zusammen. Wenn eine Kante vorhanden ist, ändert sich der Grauwert in zwei Richtungen. In einer Ecke sogar in vier Richtungen. Daher kommen drei Filter zum Einsatz, die den Wert der zweiten Ableitung für ein Pixel in horizontaler (L_{xx}), vertikaler (L_{yy}) und diagonaler (L_{xy}) Richtung berechnen. Das Ergebnis der jeweiligen Faltung wird für jedes Pixel $\vec{x} = (x, y)$ in einer Hesse-Matrix¹²⁷ gespeichert.

$$H(\vec{x}, \sigma) = \begin{pmatrix} L_{xx}(\vec{x}, \sigma) & L_{xy}(\vec{x}, \sigma) \\ L_{yx}(\vec{x}, \sigma) & L_{yy}(\vec{x}, \sigma) \end{pmatrix} \quad (3)$$

Der Nachteil bei der Verwendung der Gaußfunktion ist aber, dass diese kontinuierlich ist und bei der Anwendung auf einen diskreten Bildbereich entsprechend diskretisiert und abgeschnitten werden muss, was Einschränkungen bei der Wiederholbarkeit auf gedrehte Bilder bedeutet. Daher approximieren BAY et. al. die Hesse-Matrix H durch einfache Boxfilter D , die sich sehr gut auf ein Integralbild anwenden lassen und die Berechnung beschleunigen, denn diese ist dann unabhängig von der Größe des Filters und des Bildes. Mit den neuen Filtern wird die approximierte Hesse-Matrix $H_{approx}(\vec{x}, s)$ mit s als Seitenlänge

¹²⁷Die Hesse-Matrix enthält immer die partiellen zweiten Ableitungen

des Boxfilters zu:

$$H_{approx}(\vec{x}, s) = \begin{pmatrix} D_{xx}(\vec{x}, s) & D_{xy}(\vec{x}, s) \\ D_{yx}(\vec{x}, s) & D_{yy}(\vec{x}, s) \end{pmatrix} \quad (4)$$

Der Wert der Determinanten $detH_{approx}$ der approximierten Hesse-Matrix gibt nun Auskunft darüber, welche Bedeutung das Pixel hat, und wird berechnet durch

$$detH_{approx} = D_{xx}D_{yy} - (\omega \cdot D_{xy})^2 \quad (5)$$

Um die Approximation zu verbessern, gewichten BAY et. al. den zweiten Summanden mit dem Faktor $\omega \approx 0.912$, um den Unterschied zwischen den Gaußkernel in H und den approximierten Gaußkernel in H_{approx} auszugleichen. Die Determinante einer Hesse-Matrix gibt in der Regel durch ihren Wert Auskunft über die Eigenschaften eines Punktes. So liegt ein Extremwert vor, wenn die Determinante positiv ist, der mit Hilfe des Wertes f_{xx} in Maximum oder Minimum unterschieden werden kann. Ist die Determinante negativ, so liegt ein Sattelpunkt vor. Ergibt sie null, kann keine Aussage getroffen werden. In unserem Fall ist der erste Fall wichtig, d.h. $detH_{approx} > 0$, denn dann haben wir einen potentiellen markanten Punkt (“*Keypoint*”) gefunden.

Statt nun die Bildgröße wie bei SIFT zu verändern, verändern BAY et. al. die Größe der Filtermasken innerhalb einer Oktave, um eine Vergrößerung des Bildes zu simulieren. Die kleinste Filtergröße in der ersten Oktave ist 9x9. Da alle Masken einen diskreten Mittelpunkt haben sollen, muss die Filtermaske auf jeder Seite um zwei Pixel vergrößert werden. Daher ist die nächste Größe 15x15, nachfolgend 21x21 und 27x27, die in der ersten Oktave zum Einsatz kommen. In der zweiten Oktave startet der zweitkleinste Filter, der in der ersten Oktave eingesetzt wurde, also 15x15. Da mit jeder weiteren Oktave eine Verdopplung der Bildgröße simuliert werden soll, wird daher der nächste Filter um 4 Pixel pro Seite, also um insgesamt jeweils 12 Pixel vergrößert. In der dritten Oktave beträgt die Vergrößerung insgesamt 24 Pixel. In der Regel genügen drei Oktaven, um alle interessanten Punkte zu finden.

Wie LOWE berücksichtigen nun BAY et. al. ebenfalls die Nachbarn in den benachbarten Intervallen und in der Umgebung selbst. Daher sind das erste und das letzte Intervall in einer Oktave nur zum Vergleich geeignet, da sie keine zwei Nachbarintervalle besitzen, im

Gegensatz zu Intervall 2 und 3. Die herausgefundenen potentiellen Keypoints werden nun dahingehend untersucht, ob sie das Maximum in ihrer Umgebung sind. Dazu werden die Werte der Determinanten der Nachbarn in der $3 \times 3 \times 3$ Umgebung miteinander verglichen. Wenn ein Maximum vorliegt, dann wurde ein Keypoint gefunden, dessen Position und Größe interpoliert wird. Man kann, um es stärker einzugrenzen, auch einen Schwellwert (*Hessian threshold*) vorab festlegen, um die Anzahl der Keypoints zu reduzieren. Dann werden nur die Punkte berücksichtigt, die diesen Schwellwert überschreiten.

Bestimmung der Orientierung und Bildung des Merkmalsvektors

Analog zu SIFT bestimmt nun auch SURF die Orientierung des Keypoints, sowie die Intensitätsverteilung in Abhängigkeit seiner Umgebung. Es kommen dafür, anders als bei SIFT, zwei Haar-Filter zum Einsatz, welche die Orientierung des Keypoints in x- und y-Richtung berechnen. Dazu wird die Umgebung in einem Umkreis mit einem Radius von $6s$ um den Keypoint mit einer Haar-Filtergröße von $4s$ in horizontaler (d_x) und vertikaler (d_y) mit einer Abtastrate von $1s$ untersucht (Abb. 3.14), wobei s die Seitenlänge des Intervalls ist, in dem der Keypoint gefunden wurde.

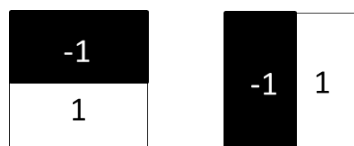


Abbildung 3.14: Haar-Filter für die Bestimmung der Orientierung des Keypoints in seiner Umgebung

Bildquelle: eigener Entwurf

Nun werden die Ergebnisse der Haar-Filter innerhalb des Kreises berechnet und dabei mit $\sigma = 2s$ gewichtet, wodurch die Punkte im Kreis, die näher am Keypoint liegen, wieder stärker gewichtet werden. Damit sind alle weiteren Ergebnisse allein vom Maßstab s abhängig.

Die Haar-Filterergebnisse für d_x und d_y für jeden Punkt im Kreis werden abschließend in einem entsprechenden Koordinatensystem eingetragen und zu einem Vektor (d_x, d_y) zusammengefasst, indem ein Fenster der Größe $\pi/3$ um den Ursprung des Koordinatensystems rotiert und alle Werte innerhalb dieses Fensters addiert werden. Der größte so ermittelte Wert nach einer Rotation von 360° für d_x und d_y beschreibt dann die Orientierung des Keypoints (Abb. 3.15) und bildet einen Vektor (d_x, d_y) .

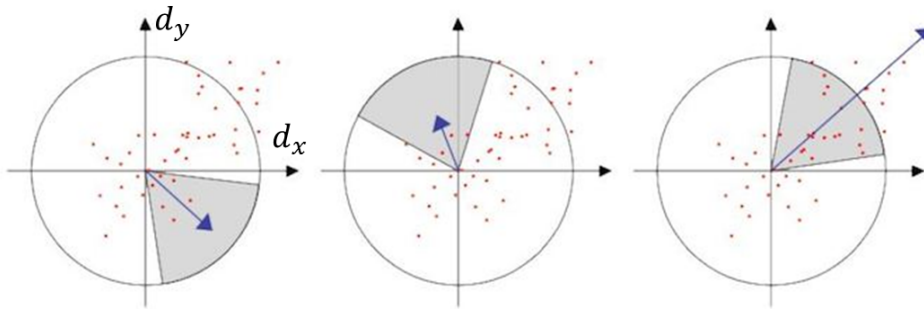


Abbildung 3.15: Berechnung des Vektors mit dem größten Wert anhand der Haar-Filter-Ergebnisse

Bildquelle: <http://eric-yuan.me/surf/>, modifiziert vom Verfasser

Dieser wird im Extraktionsschritt als Ausrichtung eines Quadrats benutzt, welches mit der Kantenlänge $a = 20s$ zentriert auf den Keypoint gelegt und mittels des Vektors ausgerichtet wird (Abb.3.16). Dieses Quadrat wird dann in 16 gleich große Bereiche aufgeteilt. In jeder dieser Regionen werden dann 5×5 Punkte gleichmäßig verteilt. Erneut kommen die Haar-Filter zum Einsatz, diesmal mit einer Filterlänge von $2s$, und werden auf jedem Punkt in den 16 Bereichen mit dem Gewicht $\sigma = 2s$ angewendet.

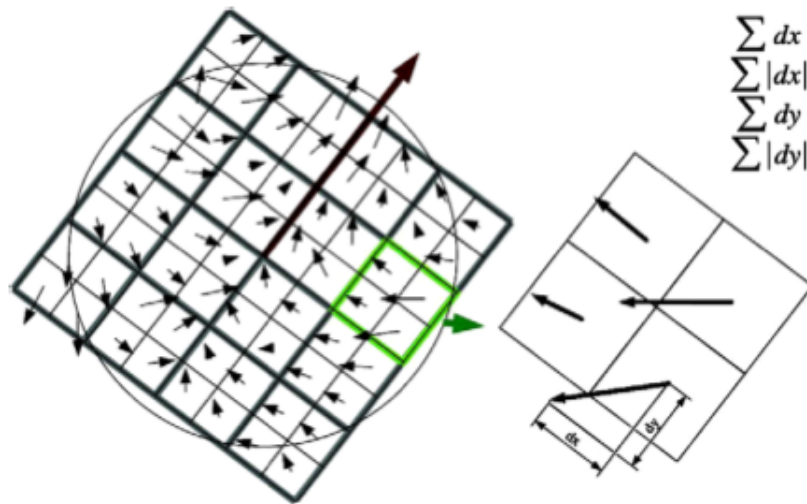


Abbildung 3.16: Berechnung des Deskriptors mit ausgerichtetem Quadrat

Bildquelle: <http://pixhawk.ethz.ch>

Die dabei entstehenden Ergebnisse werden für jeden Bereich addiert und bilden einen 4D-Merkmalvektor

$$\vec{v} = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|) \quad (6)$$

Daraus ergibt sich am Ende des Vorgangs für einen Keypoint ein Merkmalsvektor von $4 \cdot 16 = 64$ Einträgen, der eine charakteristische Beschreibung des Merkmals enthält. Für ein Objekt mit n Kexpoints erhält man dann eine Liste von n Merkmalsvektoren, die das Objekt beschreiben.

3.3.3 Klassifikation der Merkmale (BoW + SVM)

BoW Das *Bag of Words* Verfahren kommt ursprünglich aus dem Bereich der Textverarbeitung, bzw. des Text-Mining. Der Ansatz ist, dass jeder Text durch die Anzahl vorkommender charakteristischer Wörter eindeutig einem bestimmen Themengebiet zugeordnet werden kann. So finden sich in der Regel in Märchen bestimmte charakteristische Wörter (Königin, Prinzessin, Schloss) eher als in einem Fachartikel.

Für die Analyse werden daher Wörter aus dem Text extrahiert und die, die keine Informationen beinhalten (un-/bestimmte Artikel, etc.) entfernt. Die verbleibenden charakteristischen Wörter (*Passphrasen*) werden dann ihrem Wortstamm entsprechend (z.B. schläft, Schlaf, schlief) zusammengefasst und anschließend gezählt. Die Stärke der einzelnen Wortstämme wird dann in einem Histogramm (*Bag of Words*) dargestellt, um darauf basierend den Text zu klassifizieren (vgl. Abb. 3.17) Königin, Spiegel, Schneewittchen).



Abbildung 3.17: Histogramm als Wortwolke zum Märchen “Schneewittchen” dargestellt. Die Wörter, die am Häufigsten vorkommen, sind größer dargestellt.

Bildquelle: <http://www.wordle.net/>

Ein anderes Anwendungsbeispiel ist das Vergleichen zweier Texte, die in unterschiedlicher Sprache, jedoch mit gleichem Inhalt und gleicher Satzfolge verfasst wurden. Dann

kann man aus den Sätzen und Wörtern das Vokabular ermitteln und ein Wörterbuch zur Übersetzung von neuen Texten erstellen.

Auch in der Bildverarbeitung gehört dieser Ansatz mittlerweile zur Analyse von Bildern dazu. Im Bereich der Texturerkennung wird dieser Ansatz analog auf Texturen umgesetzt. Jede Textur wird durch bestimmte visuelle Basiselemente, bzw. Textone (Mikrotexturen) definiert, die sich wiederholen. Daher kann man die Häufigkeit bestimmter Textone ermitteln und mit Hilfe des entstehenden Histogramms Texturen erkennen (Abb. 3.18).

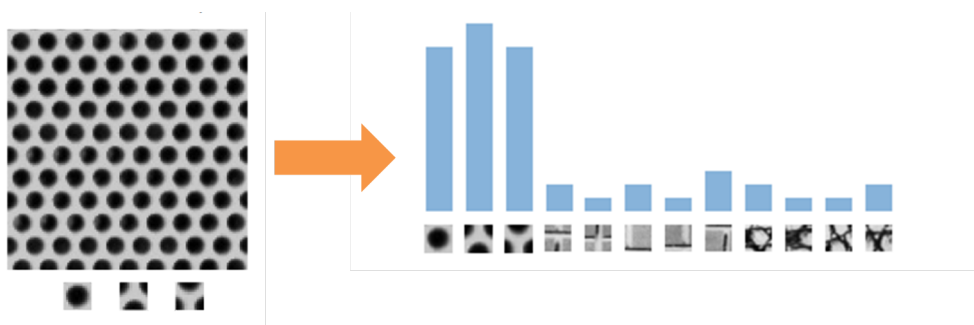


Abbildung 3.18: Charakteristische Textone in einer Textur und das entsprechende Histogramm

Bildquelle: BARTHEL, Vorlesungsfolie CV, HTW Berlin

Im Bereich der Objekterkennung basiert der Ansatz auf der Tatsache, dass jede Objektklasse über charakteristische Merkmale eindeutig bestimmt ist. So hat ein Tisch eine Tischplatte und vier Tischbeine, die individuell gestaltet sein können, aber doch eindeutig als solche erkennbar sind und entsprechende Keypoints darstellen. Die Idee ist, dass verschiedene Merkmale auch verschiedene Merkmalsvektoren liefern, während gleiche oder sehr ähnliche Merkmale wiederum in den **Einträgen** ihrer Merkmalsvektoren übereinstimmen. Die Merkmale eines Objekts stellen visuelle Wörter des Objekts dar und bilden das *Bag of Features*. Und da jedes Objekt über charakteristische Merkmale verfügt, die es eindeutig beschreiben, kann man überflüssige Merkmale mittels *Clustering* aussortieren.

k-Means Clustering Das Verfahren des *Clustering* ist die technische Umsetzung des typischen Sortierverhalten des Menschen von Objekten. Darauf basiert auch die Cantorsche Mengenlehre, der eine Menge definierte als

“jede Zusammenfassung M von bestimmten wohlunterschiedenen Objekten m

unserer Anschauung oder unseres Denkens (welche die „Elemente“ von M genannt werden) zu einem Ganzen.”¹²⁸

In der Regel wird nach gleichen Merkmalen, bzw. Eigenschaften gruppiert und anhand der verschiedenen Merkmale unterschieden. So kann man beispielsweise Tassen und Gläser in verschiedener Variation entweder in eine Gruppe “Trinkgefäß” zusammenfassen oder in die Gruppe “Tasse” und in die Gruppe “Glas” aufteilen. Man kann sie auch farblich gruppieren. Es kommt darauf an, was man damit erreichen möchte, aber in der Regel versucht man eine möglichst große Homogenität innerhalb der Gruppe und eine möglichst große Heterogenität zwischen den Gruppen zu erreichen.

Im Falle der Bildverarbeitung ist *Clustering* eines der Verfahren, die helfen, Objekte auf Bildern anhand ihrer Merkmalsvektoren zu beschreiben. So werden die Merkmalsvektoren der Klasse “Tasse” sehr ähnliche Merkmale (Gefäßrand, Boden, Form) und sehr unterschiedliche Merkmale (Tasse hat einen Henkel) im Vergleich zur Klasse “Glas” liefern. Legt man also die Merkmale der Klasse “Trinkgefäß” mit Gefäßrand, Boden, Form fest, kann man eine Trainingsmenge, bestehend aus Tassen und Gläser, zusammenstellen und die extrahierten Merkmalsvektoren danach gruppieren, d.h. z.B. alle Merkmalsvektoren, die den Gefäßrand beschreiben, kommen in eine Gruppe, die Merkmale, die den Boden beschreiben, in eine andere. Wenn man allgemein eine Trainingsmenge aus beliebigen Objekten mit gleichen Merkmalen bildet und die N Merkmalsvektoren extrahiert, ergibt das im Falle von SURF eine $N \times 64$ Matrix. Man benötigt jedoch nicht alle N Merkmale, um eine Tasse eindeutig zu beschreiben und kann die Anzahl durch das Clustering auf die reduzieren, die das Objekt eindeutig beschreiben, indem man ein *visuelles Vokabular* erstellt.

Ziel hierbei ist es, extrahierte Merkmale in k Gruppen zusammenzufassen und damit eine globale Beschreibung für das Objekt zu bekommen. Hintergrund ist, dass jeder der N Merkmalsvektoren - im Fall von SURF - einen 64-dimensionalen Merkmalsraum aufspannen, jeder Eintrag des Merkmalvektors einer Achse des Merkmalsraums zugeordnet wird und insgesamt einen Punkt im 64D bilden. In Abb. 3.19 ist es anschaulich dargestellt. Nehmen wir an, wir haben drei Merkmale von sechs Tassen extrahiert, die in jeder Tasse zu finden sind, zB. Rand, Henkel, Boden. Dann erkennen wir: Sind sich zwei Objekte sehr ähnlich, so werden die Merkmalsvektoren auf nahe beieinander liegende Punkte abbilden, also eine sehr ähnliche Orientierung und eine geringe Distanz zueinander besitzen; im anderen Fall ändert sich die Orientierung der Vektoren und die Punkte liegen weit auseinander.

¹²⁸Cantor, Georg, 1895

Um nun für eine Objektklasse eine globale Beschreibung zu erhalten, müssen die einzelnen Einträge gruppiert werden.

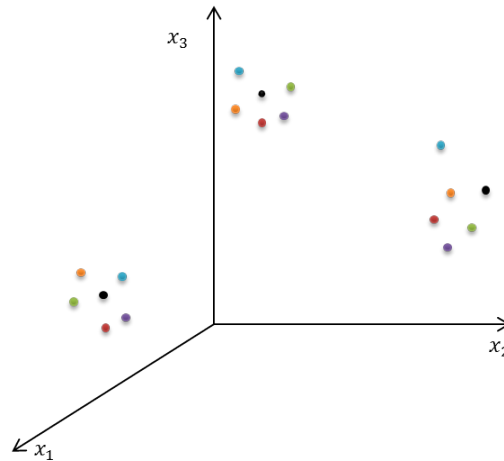


Abbildung 3.19: Merkmale sechs gleicher Objekte anschaulich für nur 3 Merkmale dargestellt
Bildquelle: eigener Entwurf

Für das Clustern wird nun eine Trainingsmenge benötigt, die Objekte einer Klasse, aber in k verschiedenen Ausführungen, enthält. Die entsprechenden Merkmalsvektoren werden berechnet. Um nun aus einer Menge von Merkmalsvektoren eines Objekts ein visuelles Vokabular zu erstellen, kann man sich verschiedener Verfahren, wie zum Beispiel dem Verfahren von MACQUEENs *k-means Clustering-Algorithmus*¹²⁹, bedienen. Abb. 3.20 gibt einen groben Überblick über die bestehenden Verfahren, es soll hier jedoch nur der Unterschied zwischen hierarchischen und partitionierenden Verfahren erläutert werden. Beide Verfahren basieren auf der Berechnung der Ähnlichkeit, bzw. Distanz zwischen Vektoren, jedoch die Vorgehensweise ist unterschiedlich.

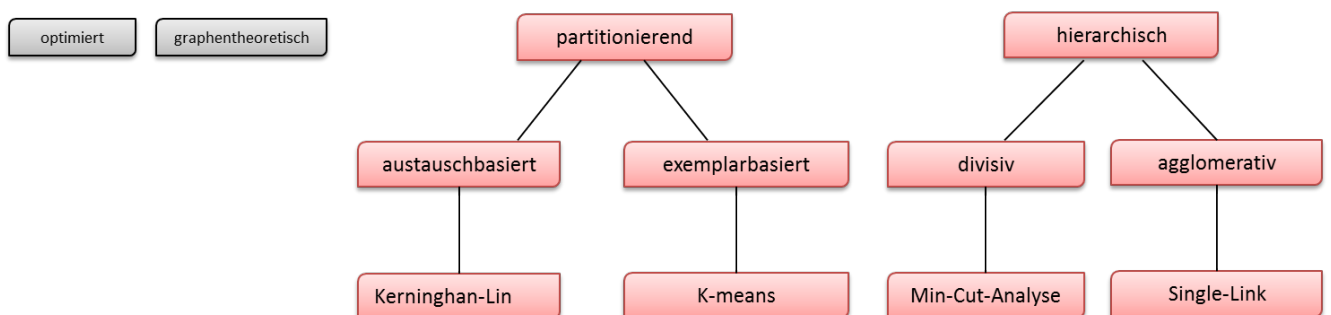


Abbildung 3.20: Übersicht über mögliche Clusterverfahren
Bildquelle: eigener Entwurf

¹²⁹MacQueen, J.,1967

Hierarchische Clusteranalyse Das hierarchische Verfahren unterscheidet sich in *divisive* (verfeinern) und *agglomerative* (vergrößern) Berechnung. Während beim divisiven Vorgehen ein “*Top-Down-Verfahren*” zur Erstellung der Cluster angewandt wird, entspricht das Vorgehen des agglomerativen Verfahrens einem “*Bottom-Up-Verfahren*”. In ersterem Verfahren werden alle Merkmale einem Cluster zugeordnet und durch Aufteilung solange in neue Cluster aufgeteilt, bis jedes Element ein Cluster darstellt. In zweiterem Verfahren geht der Algorithmus genau anders herum vor. Alle Elemente bilden jeweils ein Cluster und werden dann solange zu mehreren Clustern zusammengefügt, bis ein Cluster übrig bleibt, in dem alle Teilcluster zusammengefügt wurden.

Visuell lässt sich der Vorgang in einem *Dendrogramm* darstellen, d.h. in einer Baumstruktur oder in einem Streudiagramm. So ist die Wurzel des Baums das Gesamtcluster, welches alle Elemente enthält. Das Blatt stellt die einzelnen Elemente dar und die Knoten im Baum zeigen die Zusammenführung der einzelnen Cluster an. Im Streudiagramm (Abb. 3.21) kann man erkennen, bei welchem Ähnlichkeitsmaß zwei Cluster zusammengelegt wurden. Diese Information erhält man jedoch nicht im Baum.

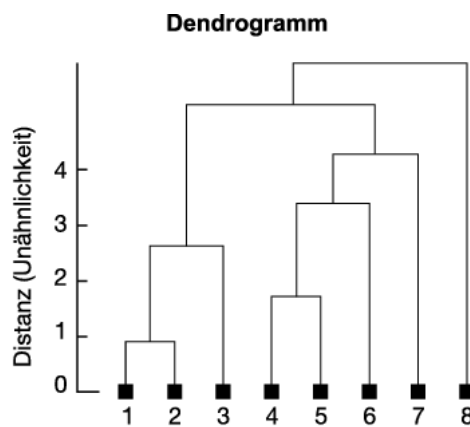


Abbildung 3.21: Dendrogramm als Streudiagramm

Bildquelle: <http://wirtschaftslexikon.gabler.de/Archiv/2564/dendrogramm-v9.html>

Partitionierende Clusteranalyse Der Hauptvertreter der partitionierenden Clusteranalyse ist der *k-means-Algorithmus* zur Vektorquantisierung. Er basiert darauf, dass man vorab eine feste Anzahl k an Clustern wählt und mit Hilfe von Ähnlichkeits- bzw. Distanzmaße die restlichen Elemente den Clustern zuordnet. Der Algorithmus teilt also die Gesamtmenge auf k Cluster auf und zwar solange, bis kein Vektor mehr verschoben werden kann.

Es gibt zwei weit verbreitete Verfahren des *k-means-Algorithmus*, einmal nach Lloyd und

einmal nach MacQueen. Für den Algorithmus von MacQueen wählt man zum Beispiel die ersten k Merkmalsvektoren aus. Diese legt man als Clusterzentren fest und ordnet jeden weiteren Merkmalsvektor der Trainingsmenge mit Hilfe eines *Distanzmaßes* dem jeweiligen Cluster zu. Lloyd ging in seinem Algorithmus einen etwas anderen Weg, indem er zufällig k Elemente auswählt und dann genauso arbeitet wie der Algorithmus von MacQueen.

Grundsätzlich läuft der *k-means-Algorithmus* in folgenden Schritten ab:

- Initialisierung: beginne mit k Elementen (zufällig oder festgelegt), lege diese als Startzentren fest
- ordne jedes Element mittels Distanzmaß dem Clusterzentrum zu, zu dem es den kleinsten Abstand hat
- lege die neuen Clusterzentren durch den Mittelwert der bisher vorhandenen Elemente und der bisheriges Clusterzentrum fest
- wiederhole Schritt 2 und 3 solange, bis kein Element mehr verschoben werden kann

Zur Berechnung der Distanzmaße wählt man gern zwischen der L_1 (City-Block-Metrik)- bzw. L_2 - (euklidischer (quadratischer) Abstand) Norm, aber auch andere, wie die *Minkowsky-Metrik* sind möglich. Das hängt von der Beschaffenheit der Daten ab und welche Metrik man der Berechnung zugrunde legt. In unserem Fall reicht der euklidische Raum und damit die L_2 -Norm.

Zum Beispiel lautet der quadratische euklidische Abstand zwischen den Merkmalen x_j und dem Zentrum m_i

$$d(X, M) = \sum_{i=1}^k \sum_{x_j \in S_i} (x_j - m_i)^2 \quad (7)$$

wobei k die Anzahl der Clusterzentren, S_i die Menge der Vektoren im Cluster i und m_i das Zentrum des i -ten Clusters ist.

Abb. 3.22 stellt diesen Vorgang an verschiedenen Ausführungen der Klasse Schuhe übersichtlich dar. Die extrahierten Merkmale werden in 3 Cluster eingeteilt. In jedem Cluster befinden sich am Ende die Merkmale, deren Vektoren sich sehr ähnlich sind¹³⁰. Im nächsten Schritt wird für jedes Cluster ein neues Zentrum ermittelt und die Vektoren erneut mit Hilfe des

¹³⁰Wobei zu beachten ist, dass TOMBASIK et.al. die drei Cluster erneut in drei Untercluster unterteilen, um einen Clusterbaum erzeugen zu können. Dies ist die effizientere Methode durch die Struktur des Baumes

minimalen Abstands den Clustern zugeordnet. Dies wird so lange iterativ durchgeführt, bis die Verschiebungen enden, d.h. jeder Merkmalsvektor dem entsprechenden Clusterzentrum am nächsten ist.

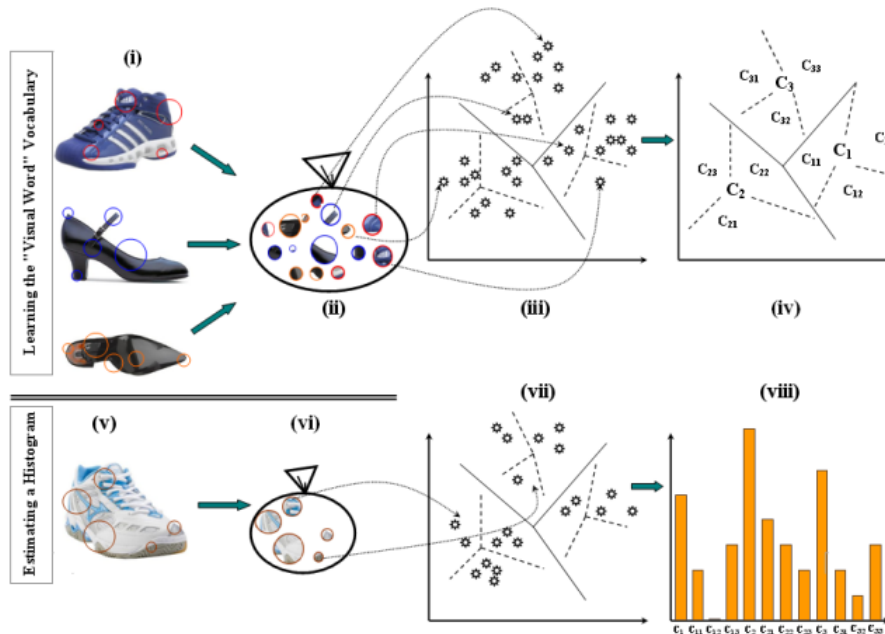


Abbildung 3.22: Von der Menge der Merkmalsvektoren mit k-means Clustering zum visuellen Wörterbuch

Bildquelle: TOMASIK et. al

Man erhält am Ende gruppierte Merkmalsvektoren, die durch die Zentren das visuelle Wörterbuch darstellen und die Merkmale global beschreiben. Mit Hilfe des Vokabulars werden abschließend die Merkmalsvektoren quantisiert, d.h. zu den extrahierten Merkmalen wird der jeweils nächste Nachbar im Wörterbuch ermittelt, um das Bild mittels der visuellen Wörter beschreiben zu können. Dieser Schritt ist wichtig, um ein Histogramm zu erstellen und eine Klassifikation zu erreichen.

Der Vorteil des hierarchischen Verfahrens ist, dass man vorab nicht festlegen muss, wie viele Cluster man bilden möchte. Der Nachteil jedoch ist die Laufzeit und den Aufwand, den der Nutzer bei der Analyse hinterher betreiben muss. MacQueens Algorithmus setzt die Gruppierung in einer kurzen Laufzeit um, was den Algorithmus sehr attraktiv macht. Jedoch liefert der Algorithmus nicht immer das optimale Ergebnis, sondern ist von der Menge der Punkte abhängig. Je mehr Merkmale man in die Berechnung einschließt, desto mehr nähert sich das Ergebnis dem optimalen Ergebnis an.

Eine mögliche Alternative wäre der Algorithmus von Lloyd, der das lokale Optimum der verteilten Punkte ermittelt, bei großen Datenmengen jedoch zu Kosten der Laufzeit. Da-

her haben sich im Laufe der Zeit weitere *k-means-Algorithmen* wie bspw. der *k-Median*- und der *k-means++*-Algorithmus entwickelt. Grundsätzlich aber ist das Finden einer optimalen Lösung NP-hart, daher reicht die heuristische Vorgehensweise von MacQueen und Lloyd zum Erstellen eines visuellen Vokabulars aus.

Support Vector Machine (SVM)

Die *Support Vektor Maschine* dient der Klassifikation bei der Objekterkennung und kommt aus dem Bereich des *machine learning*. Die grundsätzliche Aufgabe der SVM ist es, anhand der Daten eine Entscheidungsregel aufzustellen, die eine Klassifikation ermöglicht. Im einfachsten Fall soll sie für ein Objekt anhand seiner Merkmale entscheiden, ob es in Klasse 1 oder Klasse 2 gehört. In der Objekterkennung geschieht dies mit Hilfe der extrahierten Merkmale.

Wie in Abschnitt 3.2.3 erläutert, bildet jedes Merkmal auf einen Punkt (im Falle der Verwendung von SURF) im 64D-Merkmalraum ab. Durch das Clustering kann man genau die markanten Merkmale bestimmen, die das Objekt ausreichend genug beschreiben, welche wiederum Punkte im Merkmalsraum darstellen. Der Grundgedanke ist nun, dass sich die gelabelten Merkmale zweier Klassen so im Merkmalsraum verteilen, dass sie durch eine Hyperebene H eindeutig getrennt werden können. (Abb. 3.23)

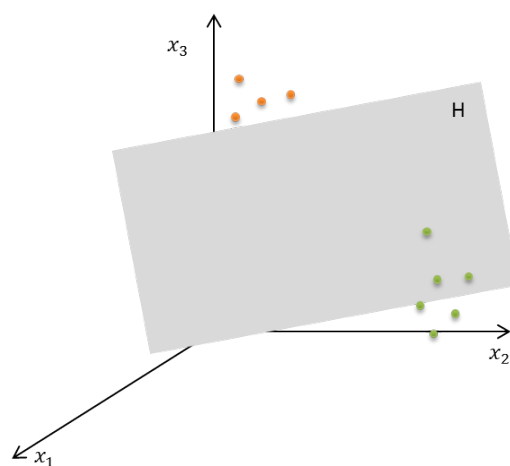


Abbildung 3.23: Trennung der Objekte durch eine Hyperebene H

Bildquelle: eigener Entwurf

Grundsätzliche Funktionsweise Eine SVM lernt auf einem Datensatz von Objekten, bzw. Punkte \vec{x}_i , die aufgrund ihrer Klassenzugehörigkeit mit einem Label $y_j \in \{-1; 1\}$ so

gekennzeichnet werden, dass sie in einem Merkmalsraum getrennt verteilt werden können. Das ergibt also eine Menge an Punkten $\{x_i, y_j\}$, die in einem geeigneten Koordinatensystem eingetragen werden können. Das Ziel der SVM ist es, mittels einer Entscheidungsfunktion $g(\vec{x}) = \vec{w}^T \cdot \vec{x}_i + b$ eine Hyperebene H derart zu bestimmen, dass der Abstand (*“margin”*) zwischen den Klassen maximal wird (*“hard margin”*). Der Vektor \vec{w} ist hierbei der Normaleneinheitsvektor¹³¹ der Ebene H und b ist der Abstand von der Ebene H zum Ursprung, d.h. die Verschiebung der Ebene. Dabei orientiert sich der Abstand an den Elementen, die der Hyperebene am nächsten liegen. Wie in Abb.3.24 ersichtlich, gibt es mehrere Möglichkeiten, eine Ebene zwischen zwei Datenwolken zu legen, in unserem Fall H_1 und H_2 . Die SVM hat die Aufgabe, die Werte für \vec{w} und b zu bestimmen und damit die Ebenengleichung $g(\vec{x})$ festzulegen, die diese Bedingung erfüllt. In unserem Beispiel wäre das die Gleichung, die H_1 beschreibt. Für jedes neue \vec{x}_i kann die SVM nun entscheiden, welcher der zwei Klassen es angehört, indem sie den Wert $g(\vec{x}_i)$ berechnet. In der Regel gilt für den Bereich oberhalb der Hyperebene H , dass $g(\vec{x}_i) \geq 1$ gilt, wenn $y_j = 1$ ist, denn $g(\vec{x}_i) = 1$ stellt in unserer Grafik die gestrichelte Grenzebene links von H_1 dar. Unterhalb von H gilt $g(\vec{x}_i) \leq -1$, wenn $y_j = -1$ ist, da analog $g(\vec{x}_i) = -1$ die Grenzebene unterhalb darstellt.

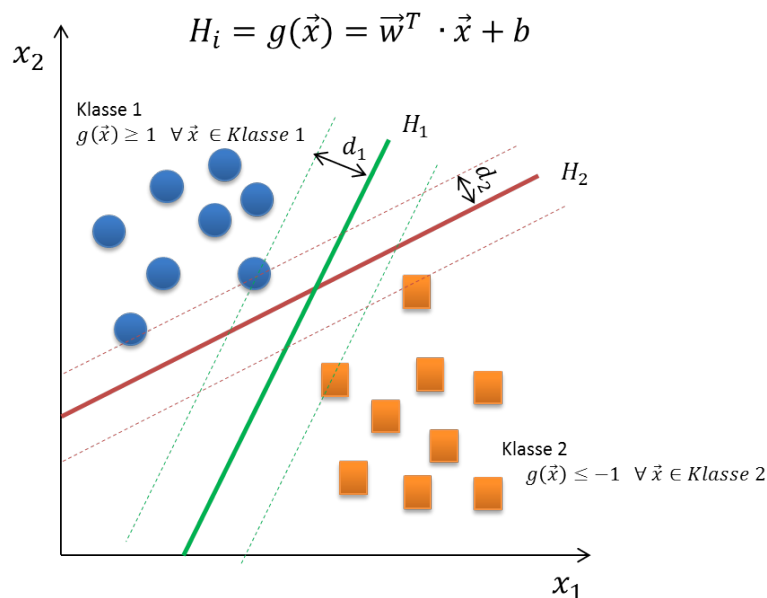


Abbildung 3.24: Struktur der 1-vs-all-SVM

Bildquelle: eigener Entwurf

¹³¹Der normierte Einheitsvektor ist notwendig, um auszuschließen, dass dieser unendlich lang werden kann. Dies ist nachteilig, weil der Abstand zum nächsten Element mit Hilfe des Normalenvektors berechnet wird

Die einfachste Art der Klassifikation ist die **1-vs.-all**. Dazu trainiert man k verschiedene SVMs, um eine Klasse von den anderen $k - 1$ Klassen zu unterscheiden. Das Problem an der 1-vs-All-SVM ist jedoch, dass sich in der Regel Objekte nicht linear voneinander trennen lassen. Daher benutzt man einen Kernel-Trick, um diese Daten in einen höherdimensionalen Raum K zu transformieren, in dem die Daten dann linear getrennt werden können. Die einzige Bedingung dabei ist, dass in K das Skalarprodukt definiert ist und dass die Dimension von K ausreichend groß ist. Denn nur dann findet man eine lineare Trennung, die korrekt ist. Im Regelfall erreicht man dies, indem die Dimension genauso groß ist wie die Anzahl der Punkte. Der Begriff Kernel-Trick kommt daher, dass die Skalarprodukte durch Kernfunktionen, wie zum Beispiel Polynome oder radiale Funktionen berechnet werden.

Die andere, komplexere Art der Klassifikation ist die **1-vs-1**. Dazu werden, um k Klassen zu unterscheiden, $\frac{k \cdot (k-1)}{2}$ SVMs trainiert, die durch Voting die Klassen paarweise unterscheiden. Dazu wird von jeder SVM ein Merkmal klassifiziert, was als Stimmabgabe gewertet wird. Der Klasse, dessen SVM am Häufigsten votet, wird das Objekt dann zugeordnet.

4 Emotionserkennungssystem

4.1 OpenCV : Viola-Jones-Algorithmus

Für die Implementierung in C# unseres Emotionserkennungssystems, basierend auf den Methoden, die in Kapitel 3 näher erläutert wurden, kommt EmguCV, ein Wrapper der quelloffene Programmierbibliothek OpenCV, welche den Entwickler mit Methoden aus dem Bereich Computervision versorgt, zum Einsatz, sowie AForge.NET, bzw. dessen Erweiterung Accord.NET, ebenfalls quelloffene Bibliotheken, die speziell Methoden der Bildverarbeitung bereitstellen. Während wir in EmguCV einen sehr gut funktionierenden und einfach zu implementierenden Viola-Jones-Algorithmus für die Gesichtsdetektion angeboten bekommen, stellt Accord.NET jeweils den SURF - und BoW-Algorithmus, sowie Methoden für das k-mean-Clustering und die SVM bereit. Es gäbe auch andere offen zugängliche Programmierbibliotheken für C# Entwicklungen, wie OpenSURF von Chris EVANS¹³² oder LibSVM¹³³, die man statt der Methoden von Accord.NET nutzen kann, die ich jedoch verworfen habe. Das System soll dann mit der Cohn-Kanade Datenbank trainiert werden, welche aus 486 Sequenzen von 97 Personen besteht. Jede Sequenz beginnt hierbei mit einem neutralen Gesichtsausdruck und weiteren 5 für die Ba-

¹³²<http://www.chrisevansdev.com/computer-vision-opensurf.html>.

¹³³<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

sisemotionen. Die Bilder wurden dem Gesichtsausdruck entsprechend vorab in entsprechende Ordner sortiert, um ein Trainings- und Testset zu erzeugen.

Der Viola-Jones-Algorithmus umfasst wenige Zeilen. Für den Gesichtsdetektor bietet OpenCV die Methode *DetectHaarCascade* an, deren Funktion und wichtigste Bestandteile anhand eines Code-Ausschnitts des eigenen Programms kurz erläutert werden sollen:

```
1 private void facedetector (object sender, EventArgs e)
2 {
3     [...]
4
5     var faces =
6         grayframe.DetectHaarCascade(haar, 1.4, 4,
7             HAAR_DETECTION_TYPE_DO_CANNY_PRUNING, new Size(25,25)) [0];
8     [...]
9 }
```

1. **haar**: Dies ist unser neues Objekt vom Typ *HaarCascade*, welches auf die XML-Datei *haarcascade_frontalface_alt_tree*, die von OpenCV zur Verfügung gestellt wird, verweist. Diese Datei enthält Trainingsdaten und erzielt gute Ergebnisse. OpenCV bietet diverse XML-Dateien zur Detektion an, so ermöglicht es auch die Detektion von Augen, Mund oder das Gesicht im Profil¹³⁴.
2. **4** : gibt die kleinste Anzahl nächster Nachbarn wieder. Wenn dieser Wert größer als 0 ist, dann gruppiert der Algorithmus die Rechtecke, die sich überschneiden und gibt nur die zurück, die mindestens genauso viele Überschneidungen mit anderen haben wie die kleinste Anzahl der nächsten Nachbarn. Setzt man diesen Parameter auf 0, so werden keine Rechtecke gruppiert und jedes Rechteck wird angezeigt (*k-Nearest-Neighbor-Algorithmus*). Wählt man einen beliebigen Wert *n*, so zeigt der Algorithmus nur dann ein Ergebnis an, wenn es eine Gruppe dieser Größe gibt. Je höher dieser Wert ist, desto strenger ist das Detektorkriterium.
3. **DO_CANNY_PRUNING**: Hier kann man zwischen verschiedenen HAAR-DETECTION-TYPES wählen. OpenCV bietet mehrere an, wie bspw. **FIND_BIGGEST_OBJECT** zum Auffinden des größten Objekts. In unserem Fall genügt uns entweder die *default*-Einstellung 0, die jedoch keine Optimierung darstellt, oder **CANNY_PRUNING**. Dies ist eine *Canny edge detection*, die - **bevor** der Gesichtsdetektor startet - die Bereiche aussortiert, die für gewöhnlich keine Gesichter beinhalten. Dadurch wird die Rechenzeit beschleunigt. Der

¹³⁴Hier sei auf die zahlreichen XML-Dateien im Ordner "OpenCV" von EmguCV verwiesen

Nachteil ist jedoch, dass Gesichter unter Umständen auch verloren gehen können.

4.2 Accord.NET : BoW, inklusive SURF und k-means-Clustering

Accord.NET ist die Erweiterung der AForge.NET- Bibliothek, einem C# Framework, welches dazu entwickelt wurde, Entwicklern den freien Zugang zu .NET Anwendungen, speziell für die Bereiche des *Computer Vision* und der *Künstlichen Intelligenz*, d.h. *machine learning*, *neuronale Netzwerke* und *Robotik* bereitzustellen. Accord.NET wurde entwickelt, um ein noch reichhaltiger ausgestattetes Methodenpaket anzubieten. Beide unterliegen der LGPL (*GNU Lesser General Public License*) und sind somit frei einsetzbar.

Accord.NET bietet für die Umsetzung der in Kapitel 3.3.2 und 3.3.3 beschriebene Verfahren passende Klassen an, deren Methoden in unserem System implementiert werden sollen und daher kurz vorgestellt werden. Ich wähle dieses Verfahren, da jede Emotion über ganz eigene Merkmale und entsprechende Keypoints verfügt (Abb.4.1) und daher die Klassifikation ermöglicht¹³⁵.



Abbildung 4.1: Charakteristische Punkte im Gesicht

Bildquelle: www.research-results.de

- BagOfVisualWords-Class:

Die BagOfVisualWords-Klasse bietet zwei grundlegende Methoden zum Erstellen eines BoW-Modells. Der Ablauf, ein Wörterbuch des Trainingssets zu erstellen, wird mit drei Methoden realisiert.

- compute(**Bitmap** [] images, **double** treshold):

Die Methode berechnet auf ein vorgegebenes Trainingsset das BoW-Modell unter Verwendung des SURF-Algorithmus⁷ und erstellt eine Liste der SURF-*Keypoints* des Typs **TPoint**. Der Schwellwert für das k-means-Clustering (Default-Wert $1E - 05$) kann hierbei individuell angepasst werden.

¹³⁵Hierbei sei der Leser auf Abb.12 in Kapitel 3.3 verwiesen

- `GetFeatureVector(List <TPoint> points)`:

GetFeatureVector berechnet für die Liste der SURF-*Keypoints* das BoW-Wörterbuch. Das Wörterbuch wird als **double**- Vektor zurückgegeben. Accord.NET bietet von *GetFeatureVector* auch Methoden zum Erfassen des Wörterbuchs für ein unbearbeitetes Bild, bzw. Bitmap, um unbekannte Bilder mit der Methode *equals()* mit dem vorhandenen Wörterbuch vergleichen zu können.

Der Vorteil von Accord.NET zeigt sich hier bei der Integrität des SURF-Algorithmus' und des *k-means*-Clustering in einer Methode, die dadurch nicht extra als Bibliothek eingebunden, bzw. implementiert werden müssen.

4.3 Accord.NET: Support Vektor Maschine

Für die Klassifikation von Emotionen gibt es zwei grundsätzliche Ansätze: Die Klassifikation Fröhlich vs. Nicht-Fröhlich oder die Klassifikation in vier Emotionsklassen, d.h. Glücklich, Wütend, Traurig und Überrascht, wobei die ersten drei Klassen für die Entscheidung, ob Gamification in einem Arbeitsprozess glücklich macht, ausreichend sind.

Accord.NET bietet mehrere Klassen an, je nach Bedarf(vgl. Abb.4.2).Für eine 1-vs.1-SVM bietet sich die Klasse *MulticlassSupportVectorMachine* an, für die 1-vs.-all-SVM die Klasse *MultilabelSupportVectorMachine*. Beide Klassen sind von der Klasse *ISupportVectorMachine* abgeleitet. Die *MultiClassSupportVectorMachine* bedient sich hierbei an Methoden der Klasse *KernelSupportVectorMachine*, welche die Klasse *ISupportVectorMachine* erweitert. Daher bieten diese zwei Klassen alle benötigten Methoden an, um eine lineare Klassifikation oder eine Multiklassen-Klassifikation zu realisieren. Hier sollen nur die grundlegenden Methoden vorgestellt werden.

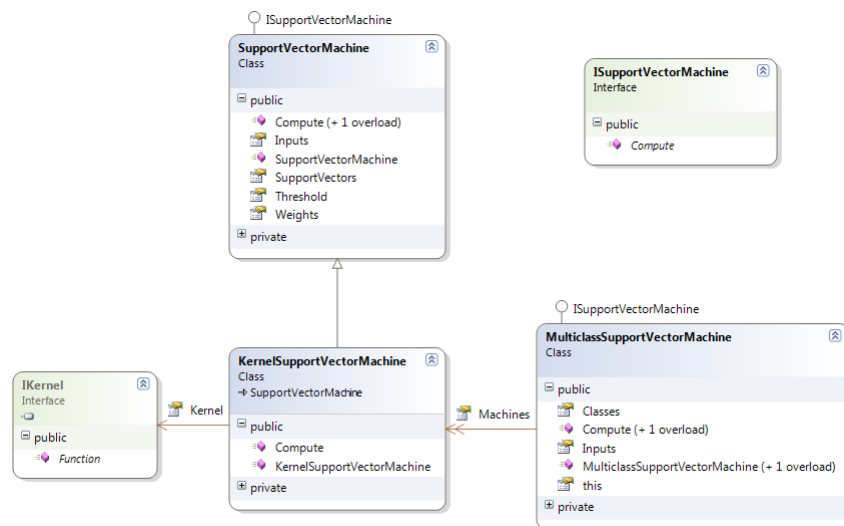


Abbildung 4.2: UML-Diagramm der SVM-Klassen in Accord.NET

Bildquelle: <http://www.codeproject.com/KB/recipes/handwriting-svm>

Zur Instanziierung eines neuen SVM-Objekts werden sich sehr ähnliche Konstruktoren verwendet, einzig der Bezeichner ist ein anderer:

- `MultilabelSupportVectorMachine(int inputs, IKernel kernel, int classes)`
- `MulticlassSupportVectorMachine(int inputs, IKernel kernel, int classes)`

Hierbei gibt der Parameter *inputs* die Länge der Merkmalsvektoren des Trainingssets, sowie der Parameter *classes* die Anzahl der Klassen an. Der bereitgestellte Kernel $k(x, y) = \langle S(x), S(y) \rangle$ berechnet das Skalarprodukt der Merkmalsvektoren, jedoch kann man auch entweder den linearen Kernel oder aus anderen angebotenen Kernels, wie zum Beispiel Gaußkernel, wählen. Für den `MulticlassSupportVectorMachine`-Konstruktor ist der lineare Kernel als Default-Wert voreingestellt.

Nach der Instanziierung muss die SVM trainiert werden. Auch dafür gibt es zwei Konstruktoren, die die gleiche Struktur besitzen, jedoch aus dem Namespace *VectorMachineLearning*.

- `MulticlassSupportVectorLearning(MulticlassSupportVectorMachine machine, double[][] inputs, int[] outputs)`
- `MultilabelSupportVectorLearning(MulticlassSupportVectorMachine machine, double[][] inputs, int[] outputs)`

Beide Methoden bekommen das zuvor erstellte SVM-Objekt, sowie die zuvor extrahierten Merkmalsvektoren des Trainingssets und die entsprechenden Labels. Die Labels sind notwendig, um jedes Objekt auf den Bildern seiner Klasse zuzuweisen, damit die SVM erlernt, welche Merkmale zu welcher Klasse gehören werden. Daher wird jedem Bild ein entsprechendes Label aus $0 \dots n$, wobei n die maximale Klassenanzahl ist, im Array *outputs* so hinzugefügt, dass die Indizes mit dem jeweiligen Merkmalsvektor in *inputs* übereinstimmt.

Die Klasse *VectorMachineLearning* stellt eine Auswahl an Lernalgorithmen zum Trainieren einer SVM bereit. Um jedoch einen Lernalgorithmus anwenden zu können, muss dieser konfiguriert werden. So bietet die Klasse zum Beispiel den **SequentialMinimalOptimization**(SMO)¹³⁶-Algorithmus an, d.h. der Margin wird mit der Methode der quadratischen Optimierung maximiert. Es stehen aber auch andere Algorithmen, wie die Methode der kleinsten Quadrate, zur Verfügung.

5 Studie

Sinn und Zweck der Studie soll sein, herauszufinden, ob Gamification in einem Arbeitsprozess Auswirkungen auf den emotionalen Zustand hat, im besten Fall sogar Glücksgefühle hervorruft. Dafür musste ein Arbeitsprozess nachgeahmt werden, der einer Montagetätigkeit am Nächsten kommt, sowie mit einer automatischen Emotionserkennung und einer Form der Gamification gekoppelt werden. Da das Thema dieser Studienarbeit im Projekt **motionEAP** bearbeitet wird, ist der Gedankengang, diese Idee an einem Arbeitstisch mit Kinect-Technik umzusetzen, um die Effizienz und Leistungsfähigkeit von leistungsgeminderten Arbeitnehmern zu gewinnen, bzw. zu steigern.

5.1 Technisches System (Wizard-of-Oz)

Da ich leider bei der Implementierung auf unerwartete und für mich unlösbare Schwierigkeiten bei der Verknüpfung der einzelnen Schritte zu einem ganzen Emotionserkennungssystem stieß, griff ich auf ein manuell zusammengesetztes System zurück. Dafür nutzte ich die Emotionserkennung, die im Programm **SHORE**TM (*Sophisticated High-speed Object Recognition Engine*)¹³⁷ des Fraunhofer Instituts enthalten ist, und ich für meine Studie verwenden durfte und imple-

¹³⁶Platt, John C., 1998

¹³⁷Kueblbeck, Ernst, 2006

mentierte eine Benutzeroberfläche zur Steuerung und Dokumentation der Studie. Ausgestattet mit diesen Programmen setzte ich eine *Wizard-of-Oz*-Studie um, d.h ich simulierte manuell das, was das System später alleine können soll.

5.1.1 Bestandteile des Systems

- Simulation des Arbeitsprozesses mit Lego

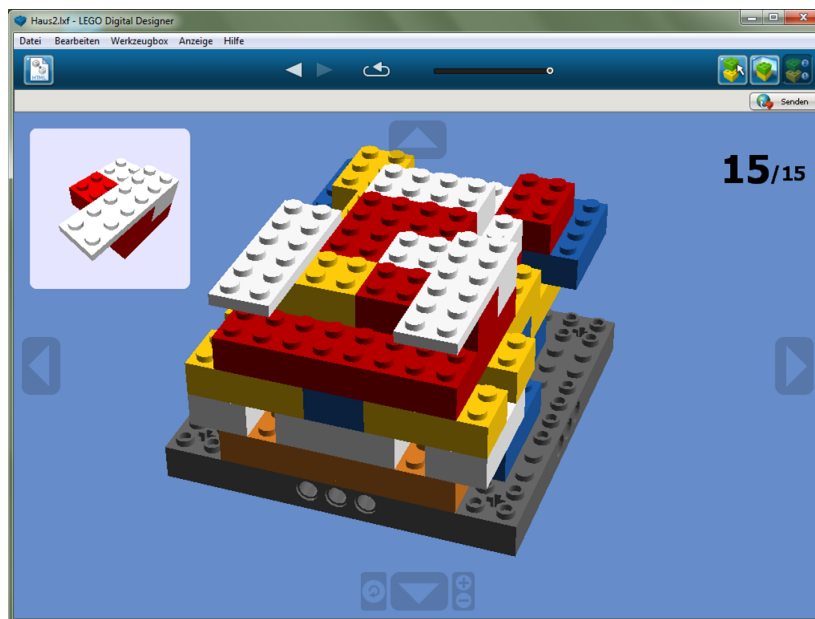


Abbildung 5.1: Anzeige des LEGO Digital Designer für die Studie

Bildquelle: eigener Entwurf

Zur Simulation des Montageprozesses suchte ich ein System, welches einen solchen Prozess äquivalent simuliert. Daher fiel meine Wahl auf das Baukastensystem von *Lego*[®]. Die Aufgabe der Probanden war der wiederholte Aufbau eines 'Hauses' aus Legosteinen (Abb.5.1). Das 'Haus' wurde so entworfen, dass es dem Probanden einen Lerneffekt ermöglichte, aber erst nach einigen Durchgängen. Der Proband musste also das 'Haus' einige Male neu aufbauen, um Symmetrien und Farbkombinationen erkennen und sich merken zu können. Darüber hinaus entwarf ich das 'Haus' so, dass ein vollständiger Aufbau nicht über Gebühr lange braucht. In Eigentests zuvor erreichte man ca. 3 Minuten für einen Durchgang.

Angeleitet wurde der Proband beim Aufbau durch das frei verfügbare Programm *Lego*[®] Digital Designer. Dieses bietet, nach der Erstellung des gewünschten Objektes im Programm, eine Schritt-für-Schritt Aufbauanleitung an, welche die Projektion der einzelnen

Montageschritte auf die Arbeitsfläche des Kinect-Arbeitstisches ersetzt. Für die Anzeige der Montageschritte bietet der Digital Designer zwei Möglichkeiten an: Entweder direkt im Programm als Bauanleitungsmodus oder als HTML-Version(vgl. Abb. 5.2).

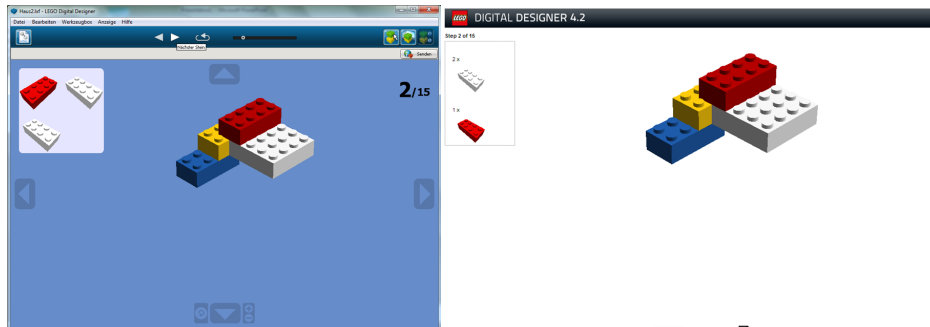


Abbildung 5.2: Unterschiedliche Anzeigen des LEGO Digital Designer beim Aufbau
Bildquelle: eigener Entwurf

Der Unterschied liegt in der Art der Anzeige. Während der Proband im Bauanleitungsmodus das 'Haus' in alle Richtungen drehen konnte und mit Hilfe einer Animation in jeder der 15 Stufen genau angezeigt bekam, welcher Stein auf welchen als Nächstes gesetzt wird, zeigte die HTML-Bauanleitung gleich das fertige Teilstück im jeweiligen Schritt an und bietet dem Probanden keinerlei Möglichkeit, das selbstgebaute Teilstück auf Korrektheit durch eine Drehung des Teilstücks im Programm oder durch Wiederholung des letzten Schritts zu überprüfen. Damit der Proband sich jedoch während des gesamten Prozesses orientieren konnte und nicht aufgrund dessen unnötigen Frust empfindet, wählte ich den Bauanleitungsmodus im *Lego*[®] Digital Designer.

- Räumlicher Aufbau



Abbildung 5.3: System für den Durchgang ohne Gamification
Bildquelle: eigener Entwurf

Der Proband hatte einen Bildschirm vor sich, auf dem der Digital Designer und das Feedback angezeigt wurden und von Lautsprecher umgeben war (Abb. 5.3). Da eine korrekte Emotionserkennung im Allgemeinen stark abhängig ist vom Licht, wurde die Studie bewusst in einen Raum verlegt, der gleichmäßig ausgeleuchtet war. Die Kamera musste individuell auf den Probanden eingestellt werden, da die Sitzhöhe durch die Körpergröße des Probanden variierte. Bevorzugt wurde die Kameraposition unter dem Bildschirm. Dokumentiert wurde auf einem Laptop, der sich weiter entfernt befand.

- Emotionserkennung

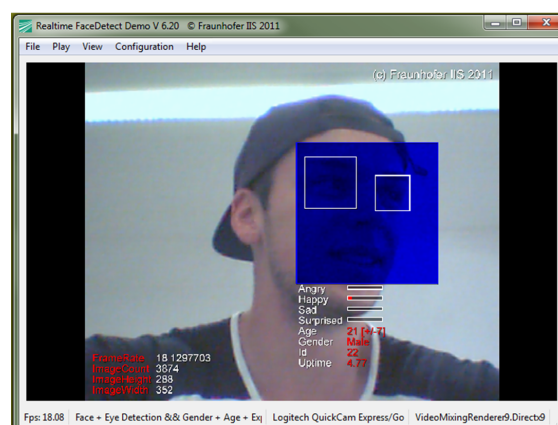


Abbildung 5.4: Emotionserkennung mit **SHORE**TM
Bildquelle: eigener Entwurf

Für die Dokumentation der Emotionserkennung, die im Laufe des Prozesses erkannt werden, wurde eine Logitech QuickCam Express mit einer Auflösung bis zu 640x480 Pixel eingesetzt, die den Probanden überwachte. Die von **SHORE**TM (5.4) angezeigten Emotionen wurden dann von Hand über die Bedienoberfläche in einer Textdatei gespeichert. Dabei achtete ich darauf, nur die Emotionen zu dokumentieren, die den höchsten Ausschlag hatten.

- Dokumentation - Emotionen, Zeit, Fehleranzahl, Feedback

Für die Dokumentation implementierte ich zwei Programme. Das Programm für den Durchlauf ohne Gamification speicherte die benötigte Gesamtzeit, die erkannten Emotionen und die Fehleranzahl tabellarisch in einer Textdatei (Abb. 5.5).

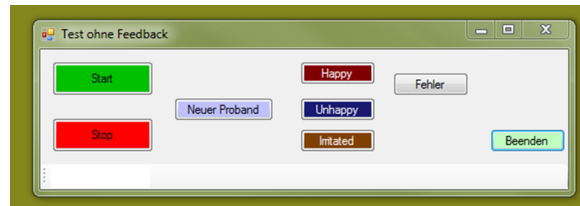


Abbildung 5.5: Bedieneroberfläche ohne Gamification

Bildquelle: eigener Entwurf

Für die Dokumentation im Durchlauf mit Gamification implementierte ich in der Bedieneroberfläche weitere Module, um das Feedback zu steuern. Dazu wurde beim Start des Programms ein zweites Fenster mit aufgerufen, welches dem Probanden den aktuellen Punktestand, den Fortschritt des Durchlaufs und am Ende die Gesamtpunktezahl anzeigte (Abb. 5.6). Wie in Durchlauf 1 speicherte auch dieses Programm die benötigte Zeit pro Durchgang und Durchlauf, sowie die Fehleranzahl und die Punktezahl ab.



Abbildung 5.6: Bedieneroberfläche mit Gamification

Bildquelle: eigener Entwurf

5.1.2 Struktur und Ablauf

Für die Studie standen mir 10 Probanden im Alter von 19 bis 38 Jahren zur Verfügung. Es wurden mit jedem Probanden je zwei Durchläufe durchgeführt, jeweils mit 10 Aufbaudurchgängen. Dabei wurden 10 ohne und 10 mit Gamification gestaltet.

Jeder Proband bekam, nachdem er sein Einverständnis zur Studie gab, einen Fragebogen, der seine allgemeine und seine momentane emotionale Einschätzung vor und nach der Studie erfasst. Ersteres ist für einen groben Überblick, wie sich der Proband im allgemeinen fühlt (gut/schlecht, glücklich/ unglücklich), wie er an neue Herausforderungen herangeht (zustimmend/ ablehnend) und wie er auf unterschiedliches Feedback reagiert. Die Einschätzung der

momentanen Gemütslage soll helfen, einen Unterschied zwischen der Laune vor und nach der Studie feststellen zu können. Zwischen den zwei Durchläufen wurde der Proband gebeten, einen Fragebogen zur Beurteilung des Systems auszufüllen, um hinterher bewerten zu können, wie benutzerfreundlich das jeweilige System war.

1. Durchlauf ohne Gamification

Der Proband wurde gebeten, das 'Haus' 10 Mal hintereinander neu aufzubauen. Dabei half ihm der Bauanleitungsmodus des *Lego*[®] Digital Designers, der vom Probanden gesteuert wurde. Während dieses Durchlaufs wurden für jeden Aufbau Zeit, Fehleranzahl und erkannte Emotionen protokolliert, der Proband bekam jedoch keinerlei Feedback über Zeit oder Fehleranzahl.

2. Durchlauf mit Gamification

Der Ablauf war identisch mit dem Durchlauf in 1.), jedoch bekam der Proband ein akustisches Feedback, gekoppelt mit der Vergabe von Punkten. Ein Aufbaudurchgang umfasste 15 Teilschritte. Jeder fehlerfreie Teilschritt wurde akustisch mit "Sehr gut", "Gut gemacht" und "Ausgezeichnet", welches das Programm zufällig ausgewählt abspielte, gelobt und mit 10 Punkten belohnt. Wurde ein Fehler in einem Teilschritt begangen, wurden 10 Punkte abgezogen und der Proband akustisch mit "Das kannst du besser", "Das stimmt leider nicht" und "Schau genau hin" darauf hingewiesen, dass er einen Fehler gemacht hat. Da nach Korrektur des Fehlers die 10 Punkte nicht zugeteilt wurden, verlor der Proband pro Fehler insgesamt 20 Punkte(vgl. Tab. 1).

Fehleranzahl	Punkte
0	150
1	130
2	110
3	90
4	70
5	50
6	30
7	10
> 8	0

Tabelle 1: Übersicht Punktevergabe

Auch in diesem Durchgang wurden Zeit, Fehleranzahl und Emotionen, sowie Punktzahl protokolliert. Der Proband bekam den Punktestand in einem kleinen Fenster auf dem gleichen Bildschirm angezeigt wie der Digital Designer(vgl. Abb. 5.6).

6 Auswertung

Die Frage, die ich klären möchte, ist, ob Gamification in einem Arbeitsprozess glücklicher macht. Daher sind Daten, die erhoben wurden, nötig, um eine Antwort finden zu können.

6.1 Daten

1. Selbsteinschätzung der Probanden

Jeder Proband wurde befragt, wie er sich allgemein selbst einschätzt in Hinblick auf Motivation und seiner grundsätzlichen Stimmungslage. Er konnte Punkte zwischen 1 und 5 vergeben, wobei 5 "trifft voll zu" darstellte und 1 entsprechend "trifft gar nicht zu". Dabei kam ein eindeutig positiver Selbsteindruck heraus (vgl. Tab. 2).

Frage	durchschnittlicher Zustimmungindex
Oft gut gelaunt	4.0
Positiv an Neues	4.3
Mag Herausforderungen	4.2
Positives Feedback ist wichtig	3.5
Schnell frustriert	2.4
Gibt schnell auf	1.8
Positives Feedback motiviert	4.0
Ist oft glücklich	4.1
Negatives Feedback demotiviert	2.4
Anerkennung steigert die Leistung	4.2

Tabelle 2: Übersicht der durchschnittlichen Selbsteinschätzung

2. Entwicklung der Emotionen ohne Gamification

Die Zahl der gemessenen "Genervt"-Gesichtsausdrücke war in den ersten Durchgängen insgesamt relativ hoch, im Vergleich zu den gemessenen "Glücklich"- Gesichtsausdrücken. Im Verlauf der Studie nahmen erstere ab, während die Anzahl für "Glücklich" gegen Ende sogar leicht zunahm. Den Gesichtsausdruck für "Nicht glücklich", welches zusammengefasst die Emotionen "Traurig" und "Überrascht" darstellten, erkannte **SHORE**TM kaum,

daher wurden insgesamt wenige dieser Emotionen protokolliert (Abb. 6.1)

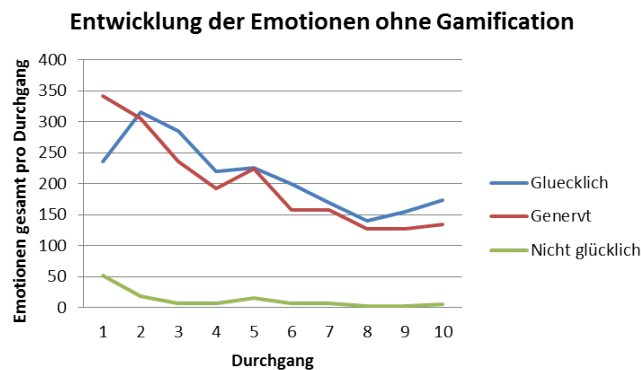


Abbildung 6.1: Entwicklung der Emotionen ohne Gamification

Bildquelle: eigener Entwurf

3. Entwicklung der Emotionen mit Gamification

Im Vergleich zu dem ersten Durchlauf (vgl. Abb. 6.1) veränderte sich die Anzahl der gemessenen "Glücklich"-Gesichtsausdrücke gegenüber den "Genervt"-Mimiken deutlich. Diesmal stiegen die Probanden scheinbar entspannter in den zweiten Durchlauf ein (Abb. 6.2).

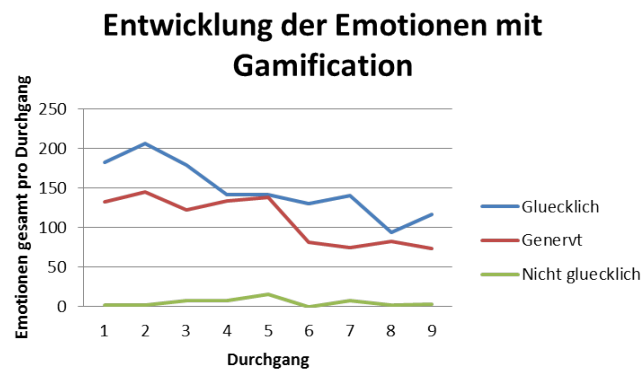


Abbildung 6.2: Entwicklung der Emotionen mit Gamification

Bildquelle: eigener Entwurf

4. Entwicklung der einzelnen Emotionen im Vergleich

Im einzelnen betrachtet, entwickelten sich die einzelnen Emotionen bei dem Durchlauf ohne Gamification im Vergleich zum Durchlauf mit Gamification wie im Folgenden:

- " Glücklich" wurde im Durchlauf ohne Gamification öfter gemessen als im Durchlauf

mit Gamification. Jedoch sank die Anzahl im ersten Durchlauf stetig und nahm zum Ende hin wieder deutlich zu. Im zweiten Durchlauf stieg die Anzahl im mittleren Bereich ein, sank jedoch flacher als im ersten Durchlauf und nahm in den letzten zwei Durchgängen deutlich ab.

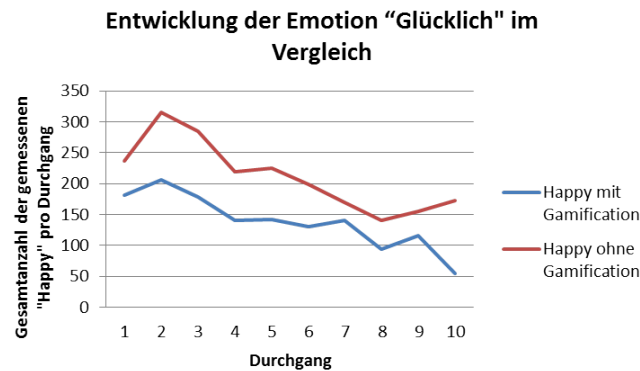


Abbildung 6.3: Entwicklung "Glücklich"

Bildquelle: eigener Entwurf

- " Genervt" stieg im ersten Durchlauf sehr hoch ein, fiel jedoch von Durchgang zu Durchgang stark und pendelte sich schließlich im unteren Drittel ein. Im zweiten Durchlauf nimmt die Entwicklung einen flacheren Verlauf. In beiden Durchläufen ist ein deutlicher Anstieg jeweils in der Mitte des Durchlaufs zu erkennen.

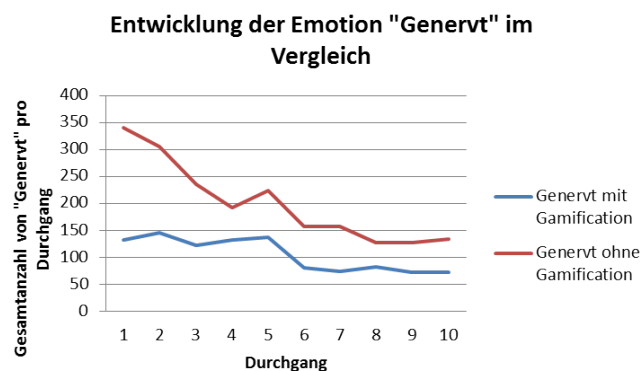


Abbildung 6.4: Entwicklung "Genervt"

Bildquelle: eigener Entwurf

- Der Verlauf von "Nicht glücklich" zeigt wiederum, dass die Probanden im ersten Durchlauf öfter überrascht oder traurig zu sein schienen, als im zweiten Durchlauf. Im ersten Durchlauf begannen die Probanden mit einer deutlich höheren Anzahl als im zweiten Durchlauf.

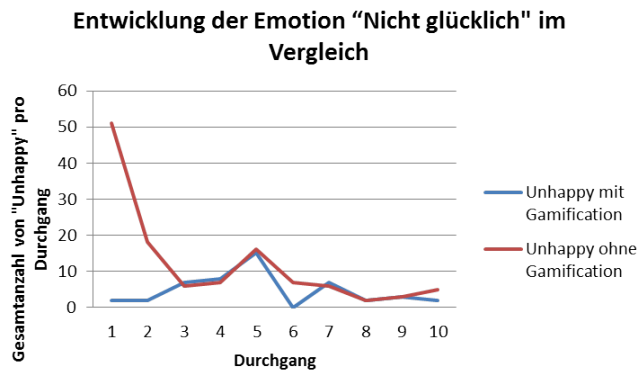


Abbildung 6.5: Entwicklung "Nicht glücklich"
 Bildquelle: eigener Entwurf

5. Fehleranzahl im Vergleich

Die Entwicklung der gesamten Fehleranzahl der zwei Durchläufe zeigt, dass die Probanden vor allem in den ersten Durchgängen ohne Gamification deutlich mehr Fehler machten als in den darauf folgenden Durchgängen. Die Fehleranzahl nahm jedoch mit jedem weiteren Durchgang deutlich ab, um gegen Ende erneut anzusteigen.

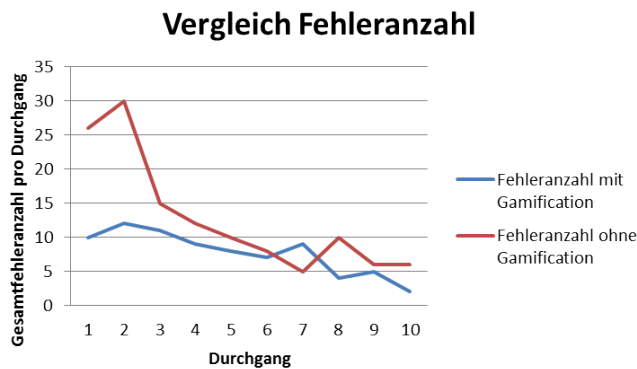


Abbildung 6.6: Entwicklung der Fehleranzahl
 Bildquelle: eigener Entwurf

6. Benötigte Zeit

Auch die benötigte Zeit nahm erwartungsgemäß deutlich ab. So dauerte der erste Durchgang im Durchlauf ohne Gamification im Schnitt 00:06:56, während der erste Durchgang im zweiten Durchlauf nur noch 00:03:37 benötigte (Tab. 3).

Nr.	Durchlauf 1	Durchlauf 2
1	0:06:56,89	0:03:37,53
2	0:04:59,53	0:03:21,88
3	0:03:47,90	0:03:30,76
4	0:03:43,58	0:02:52,90
5	0:03:37,29	0:02:58,22
6	0:03:12,14	0:02:24,38
7	0:02:54,49	0:02:30,48
8	0:02:24,88	0:02:22,11
9	0:02:28,00	0:02:34,08
10	0:02:37,44	0:02:08,26

Tabelle 3: Übersicht der durchschnittlichen Zeit pro Durchlauf

6.2 Interpretation der Daten

Bei der Durchführung der Studie traten teilweise Probleme bei der Emotionserkennung auf. So zeigte **SHORE**TM oft "Genervt" an, wenn sich der Proband stark konzentrierte und einen entsprechenden Gesichtsausdruck zeigte. Auf Nachfrage bestätigten alle Probanden, nicht genervt oder wütend gewesen zu sein, sondern sich auf den Bau des 'Hauses' zu konzentrieren. Daher kann die erkannte "Genervt"-Emotion mit "konzentriert" gleichgesetzt werden.

6.2.1 Statistische Interpretation - Signifikanztest

Um zu untersuchen, ob eine Veränderung aufgetreten ist, ziehe ich einen Signifikanztest heran - den Zwei-Stichproben-t-Test bei abhängigen Stichproben. Dafür stelle ich die Nullhypothese $H_0 =$ "Gamification ändert nichts am emotionalen Zustand" auf und untersuche anhand meiner Daten und einem zweiseitigen t-Test¹³⁸, wie sich die Wahrscheinlichkeit im Vergleich zum Signifikanzniveau $\alpha = 5\%$ darstellt und ob meine eigentliche Hypothese $H_1 =$ "Gamification verändert den emotionalen Zustand" zutreffen könnte (Abb. 4).

¹³⁸Zweiseitig, weil ich durch die Ausgangslage beide Richtungen offen lassen möchte

Analyse "Glücklich"		
Ermittelter Wert	ohne Gamification	mit Gamification
Mittelwert	211.6	138.5
Beobachtungen	10	10
t-Statistik	7.420948	
$P(T \leq t)$ einseitig	2.0067E-05	
Kritischer t-Wert eins.	1.83311293	
$P(T \leq t)$ zweiseitig	4.0134E-05	
Kritischer t-Wert zweiseitig	2.26215716	
Analyse "Genervt"		
Ermittelter Wert	ohne Gamification	mit Gamification
Mittelwert	199.9	105.4
Beobachtungen	10	10
t-Statistik	5.69046557	
$P(T \leq t)$ einseitig	0.00014894	
Kritischer t-Wert eins.	1.83311293	
$P(T \leq t)$ zweiseitig	0.00029789	
Kritischer t-Wert zweiseitig	2.26215716	
Analyse "Nicht-Glücklich"		
Ermittelter Wert	ohne Gamification	mit Gamification
Mittelwert	12.1	4.8
Beobachtungen	10	10
t-Statistik	1.48118431	
$P(T \leq t)$ einseitig	0.08634795	
Kritischer t-Wert eins.	1.83311293	
$P(T \leq t)$ zweiseitig	0.17269591	
Kritischer t-Wert zweiseitig	2.26215716	

Tabelle 4: Zwei-Stichproben-t-Test für abhängige Stichproben - Emotionen

Die ermittelten Werte t-Statistik und $P(T \leq t)$ sind die Größen, die zur Beurteilung wichtig sind. Ist $P(T \leq t) < \alpha$, dann wird H_0 verworfen, im anderen Fall beibehalten. Analog gilt, dass der Wert "t-Statistik" mit dem "kritischen t-Wert" vergleichbar ist, um H_0 dann zu verwerfen, wenn "t-Statistik" größer als der "kritische Wert" ist. In unserem Fall trifft das für die Emotionen "Glücklich" und "Genervt" zu, da bei beiden $P(T \leq t)$ deutlich kleiner als $\alpha = 5\%$ ist. In diesem Fall ergibt sich ein signifikanter Unterschied im Vorher-Nachher-Vergleich und H_1 trifft zu. Gamification scheint deutliche Veränderungen zu bewirken. Bei der Emotion "Nicht-glücklich" ist $P(T \leq t) > \alpha$ und daher trifft H_0 zu, was den Schluss zulässt, dass Gamification hier nicht signifikante Veränderungen bewirkte.

Analyse "emotionale Selbsteinschätzung"		
Ermittelter Wert	vor der Studie	nach der Studie
Mittelwert	2,1	2,143333333
Beobachtungen	10	10
t-Statistik	-3,08129551	
$P(T \leq t)$ einseitig	0,013714613	
Kritischer t-Wert eins.	2,015048373	
$P(T \leq t)$ zweiseitig	0,027429225	
Kritischer t-Wert zweiseitig	2,570581836	

Tabelle 5: Zwei-Stichproben-t-Test für abhängige Stichproben - Fragebogen

Die Auswertung mittels t-Test der emotionalen Selbsteinschätzung liefert uns ebenfalls eine signifikante Änderung, da $P(T \leq t) < \alpha$.

6.2.2 Statistische Interpretation - Diagramme

Die Interpretation der Daten anhand der erstellten Diagramme hilft uns, die Ergebnisse der t-Tests richtig zu deuten. Anhand der emotionalen Selbsteinschätzung (Abb. 6.7) wird durch eine leichte Steigerung zumindest bestätigt, dass die Probanden ihre anfänglich gute Laune nicht verloren haben, sondern lediglich etwas müder wurden. Diese Steigerung wirkt aber in Tab. 6 nicht so signifikant. Einzig die Beurteilung der Müdigkeit zeigt eine deutliche Differenz.

Analyse "emotionale Selbsteinschätzung"	
vor der Studie	nach der Studie
4,20	4,22
1,50	1,55
3,10	3,21
1,20	1,22
1,30	1,33
1,30	1,33

Tabelle 6: Durchschnittliche Bewertung - Fragebogen

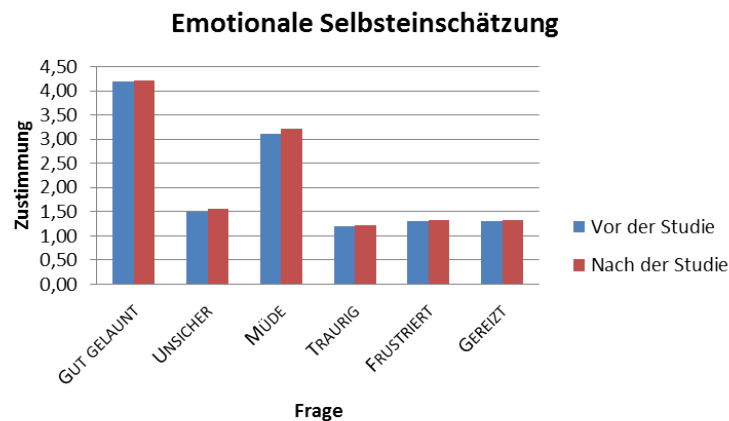


Abbildung 6.7: Entwicklung der emotionalen Selbsteinschätzung

Bildquelle: eigener Entwurf

Vergleicht man Abb. 6.3 mit Abb. 6.4, dann wären die Probanden in den zweiten Durchlauf unglücklicher als in den ersten Durchlauf gestartet, da im ersten Durchlauf 236 gemessene "Glücklich"-Emotionen zu 182 gemessene im zweiten Durchlauf stehen. Daraus könnte folgen, dass Gamification unglücklich macht (**Annahme I**). Betrachtet man aber nur die Emotionen "Glücklich" und "Genervt" innerhalb des Durchlaufs (Abb. 6.8), könnte man anhand dieser Zahlen wiederum vermuten, dass Gamification glücklich macht (**Annahme II**). Denn im Durchlauf mit Gamification blieb die Anzahl der "Glücklich"-Emotionen von Beginn an über der Anzahl der "Genervt"-Emotionen, während im Durchlauf ohne Gamification dies erst ab Durchgang 2 der Fall war. Aber auch das überrascht nicht, da die Konzentration beim allerersten Aufbau des 'Hauses' natürlich sehr hoch und zu Beginn des zweiten Durchlaufs (und nach 10 Aufbau-durchgängen bis dahin) erwartungsgemäß deutlich niedriger war. Da die "Genervt"-Emotionen hauptsächlich dann verzeichnet wurden, wenn der Proband sehr konzentriert auf den Bildschirm oder das 'Haus' blickte, wäre auch die Interpretation, dass die Aufgabe leichter zu lösen war und dadurch glücklicher machte, möglich, was im Grunde Annahme II trifft .

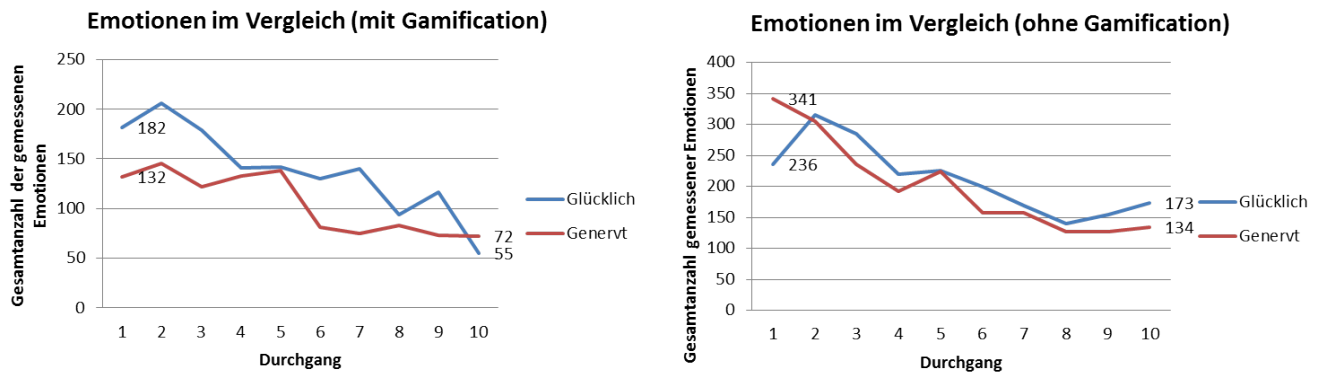


Abbildung 6.8: Vergleich der Emotionen pro Durchlauf

Bildquelle: eigener Entwurf

Es ist ebenso deutlich erkennbar, dass die Probanden mit jedem weiteren Durchlauf weniger konzentriert an den Aufbau gingen (Abb. 6.4). Im Gegenzug sank die Fehleranzahl durch den eintretenden Lerneffekt stetig (Abb. 6.6). Gerade bei der Fehleranzahl ist Durchgang 7 in beiden Durchläufen auffällig. Im Durchlauf ohne Gamification stellt sich dort die minimalste Fehleranzahl dar, während im Durchlauf mit Gamification ein zweites Maximum erkennbar ist, nachdem die Fehleranzahl in beiden Durchläufen bis dahin kontinuierlich sank. Während die Fehleranzahl im zweiten Durchlauf danach wieder sinkt, steigt sie im ersten Durchlauf deutlich an. Ich vermute, dass sich hier die sinkende Konzentration zeigt, die im zweiten Durchlauf durch die Gamification eine andere Ausprägung erfährt als im ersten Durchlauf ohne Gamification.

Zusammengefasst ist es schwer, anhand der gemessenen Emotionen und mit dem t-Test eine eindeutige Antwort auf die Eingangsfrage zu geben. Vielleicht wäre das Ergebnis eindeutiger, wenn ich mit **SHORE**TM eine feinere Emotionserkennung gehabt hätte. Trotz guter Ausrichtung auf das Gesicht des Probanden der Kamera selbst und einem Emotionserkennungstest mit jedem Probanden erkannte **SHORE**TM auch mit einem deutlich ausgeprägten traurigen Gesichtsausdruck die Emotion "Sad" nicht.

6.2.3 Beobachtungen

Sehr interessant waren die Reaktionen der Probanden auf die hinzugenommene Gamification. Während das akustische Feedback entweder irgendwann als Hintergrundgeräusch fungierte oder als nervig empfunden wurde, wurde der Punktstand als leichter Motivator erkennbar. Am Anfang des Durchlaufs 2 wurden die Probanden im Schnitt eine Minute langsamer als im letzten Durchlauf von Durchlauf 1. Dabei war erkennbar, dass die Probanden in den neuen Durchlauf

konzentrierter und auf Fehlervermeidung bedacht starteten. Dies erklärt den Anstieg der Zeit, sowie den leichten Anstieg der gemessenen "Genervt"-Emotionen in Abb. 6.4. Ebenso äußerte der überwiegende Teil der Probanden, dass die Gamification half, konzentriert weiter zu machen. Äußerte man dem Probanden gegenüber, ob er schneller oder langsamer als im vorhergehenden Durchgang den Aufbau erledigte, konnte man bei 6 von 10 Probanden beobachten, wie sie von sich aus versuchten, im nächsten Durchgang schneller zu werden, ohne Fehler zu machen. Ebenfalls entscheidend, wie sehr Gamification den Probanden ansprach, war die Haltung gegenüber dem Baukastensystem von *Lego*[®]. An der Studie nahmen 4 Frauen und 6 Männer teil. Während die Frauen zwar *Lego*[®]-Steine aus ihrer Kindheit her kannten und auch damit gebaut hatten, hatten 3 der 6 Männer noch nie wirklich mit *Lego*[®] gebaut. 2 von ihnen hatten sogar eine ausgeprägte Aversion gegenüber *Lego*[®] und erwähnten das mit Beginn der Studie und während der Durchläufe regelmäßig. Gerade bei diesen zwei Probanden war es geradezu auffällig, dass Gamification sie mehr demotivierte als motivierte. Ich vermute, dass der Verlust von Punkten durch Fehler diese Aversion bestätigte und verstärkte und eher provozierend wirkte.

Interessant war die Herangehensweise der anderen 8 Probanden. Auch wenn das nicht Bestandteil der Studie war und deshalb nicht gemessen wurde, fiel auf, dass die Frauen mit einem anderen persönlichen Ehrgeiz an die Aufgabe gingen als die Männer. Sie wirkten bei der Vermeidung ihrer Fehler weniger von der Gamification als von eigener Zielsetzung beeinflusst. Ebenso äußerten sie sich nicht, ob sie *Lego*[®] mögen oder nicht. Für sie schien das keine Rolle bei der Lösung der Aufgabe zu spielen. Die restlichen 4 Männer, die keine ausgeprägte Abneigung gegenüber *Lego*[®] hatten, teilten sich auf in einen Probanden, der keinerlei Erfahrung mit dem Baukastensystem hatte und in drei Probanden, die sehr leidenschaftliche *Lego*[®]-Bauer in ihrer Kindheit und Jugend waren. Alle 4 versuchten nicht nur, in jedem Durchgang es noch geschickter als im vorhergehenden zu bauen und schon relativ früh Muster zu erkennen, sondern auch ihre benötigte Zeit in jedem weiteren Durchgang zu verkürzen. Auffällig war, dass diese 4 Probanden nicht einmal fragten, wie oft sie das 'Haus' noch aufbauen müssten, im Gegensatz zu den anderen. Bei 2 von ihnen war deutlich zu sehen, aber leider nicht messbar, wie sehr sie in ihrem Tun aufgingen. Je weniger Fehler, je geschickter sie gebaut haben und je schneller sie wurden, desto enthusiastischer wurden sie. Gerade der vorletzte Proband fiel hierbei stark auf. Für ihn war die Mischung aus Punktzahl, Zeit und immer wieder anderer innovativer Vorgehensweise so ansprechend, dass er von Durchgang zu Durchgang immer enthusiastischer wurde und nach dem letzten Durchlauf selbst erwähnte, er hätte noch weitere Durchgänge bauen können. Von den Daten her war das bei ihm tatsächlich messbar (Tab. 7)

Durchgang	Zeit	Glücklich	Genervt
1	0:01:24,59	10	0
2	0:01:23,54	13	8
3	0:01:21,15	11	1
4	0:01:09,46	12	3
5	0:01:20,74	10	1
6	0:01:17,17	17	2
7	0:01:09,06	13	2
8	0:01:03,66	10	5
9	0:01:14,44	12	1
10	0:01:04,67	12	1

Tabelle 7: Übersicht der Daten von Proband 9

Dennoch bewerteten die Probanden die Gamification insgesamt eher negativ ¹³⁹(Abb. 6.9). Auch die Beurteilung der User Experience über den Fragebogen von *AttrakDiff*¹⁴⁰ zeigt, dass die ermittelte pragmatische und hedonische Qualität des Systems ohne Gamification höher sei als des Systems mit Gamification. Beide Systeme jedoch wirkten auf die Probanden im Durchschnitt eher handlungsorientiert und mittelmäßig attraktiv(vgl. Abb. 6.10).

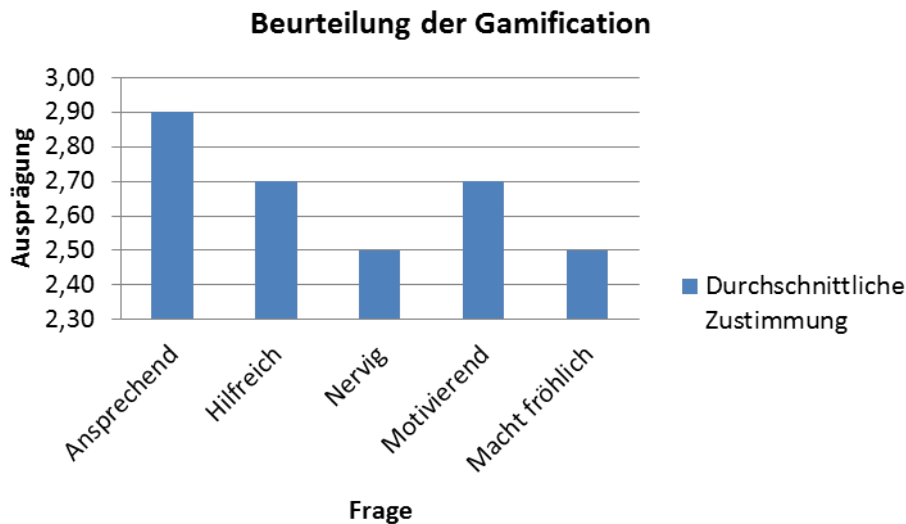


Abbildung 6.9: Bewertung der Gamification durch den Probanden

Bildquelle: eigener Entwurf

¹³⁹Wenn man bedenkt, dass der höchste Zustimmungswert im Fragebogen bei 5 lag und der Mittelwert bei 3

¹⁴⁰Webdienst zur Beurteilung und Auswertung von unterschiedlichen Produkten

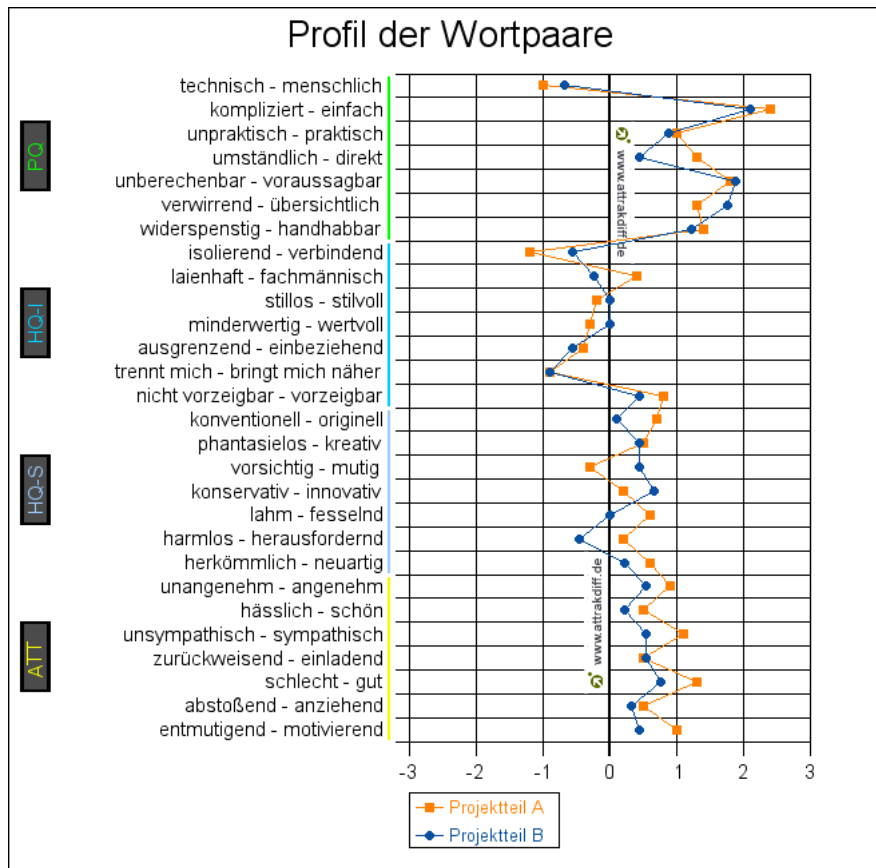


Abbildung 6.10: Bewertung der User Experience über AttrakDiff

Bildquelle: www.attrakdiff.de

7 Ethisch-philosophische Aspekte

Wichtige Aspekte werden Datenschutz und das Recht auf Selbstbestimmung, bzw. persönlicher Freiheit sein, die man während der ganzen Erarbeitung und Testphase nicht vergessen darf. Da wir es mit Menschen zu tun haben werden, die aufgrund ihrer körperlichen Besonderheiten teilweise nicht einschätzen können, welche Auswirkungen das System haben kann, müssen wir eine besondere Sensibilität aufbringen. Natürlich soll es, der “UN Konventionen über die Rechte von Menschen mit Behinderung” folgend, gewahrt sein, dass die ”individuelle Autonomie, einschließlich der Freiheit, eigene Entscheidungen zu treffen, sowie [der] Unabhängigkeit”¹⁴¹ geachtet werden. Jedoch wird es auch um Menschen gehen, die durch eine geistig-kognitive Behinderung nicht in der Lage sind, dieses Recht völlig in Anspruch zu nehmen. Es werden dann andere übernehmen müssen, da man sie ein Stückweit vor sich selbst beschützen muss, weil sie die Folgen nicht vollständig abschätzen können. Das erfordert wiederum die Überlegung, ob man sagen kann, dass im Interesse des Betroffenen gehandelt wird. Einig dürfte man sich aber darin

¹⁴¹Übereinkommen über die Rechte von Menschen mit Behinderungen vom 13. Dezember 2006 aus: Bundesgesetzblatt (BGBl) 2008 II, S. 1419, Art.3a

sein, wenn man alles tut, um jedem Benutzer die Selbstbestimmung und persönliche Freiheit so gut wie es ihm möglich zu ermöglichen.

Ein anderer Aspekt ist die Frage, ob das überhaupt erlaubt sein sollte, emotionale Zustände zu verändern. Auch hier kann man davon ausgehen, dass der Benutzer keine Marionette sein soll, dessen Emotionen an Fäden hängen und von uns gesteuert werden. Sinn und Zweck des Systems soll es sein, Menschen darin zu unterstützen, positive flow-Erlebnisse erfahren und so ihre Motivation steigern zu können. Im Allgemeinen soll erreicht werden, dass Benutzer zufrieden sind. Natürlich stellt sich dann die Frage daran, ob dieses Vorhaben gut ist. Was aber wiederum ist daran verkehrt, eine Methode zu entwickeln, die einem Freude am Tun vermittelt und im besten Fall sogar helfen kann, Fähigkeiten spielerisch zu fördern? Wenn man mit einer angebrachten Sensibilität im Umgang mit den Wünschen und Erwartungen der Benutzer umgeht, sollte nichts dagegen sprechen.

Das bringt mich zum letzten Aspekt und zwar dem der Zielsetzung. Wichtig ist es, die Ziele und Wünsche der Benutzer in den Vordergrund der Entwicklung zu stellen und nicht den eigenen Anspruch, ein besonders schönes System zu entwerfen. Da wir trotz allem mit der Emotionserkennung in einen Bereich schauen, der strikt der persönlichen Freiheit unterliegt, ist es wichtig, dass man dieses Recht nicht untergräbt. Es ist wichtig, darauf zu achten, Ablehnung und Widerwillen zu erkennen, auch wenn der Benutzer es nicht ausdrücken kann. Das klingt nicht einfach und ist es auch nicht. Wenn aber zu erkennen ist, dass der Benutzer dieses System nicht nutzen will, muss man es respektieren und ihn aussteigen lassen. Bei einem gesunden Menschen ist das nicht schwer zu erkennen. Diese Problematik betrifft eher Behinderte oder ältere Menschen, die mit der Situation überfordert, aber vielleicht auch nicht in der Lage sind, ihren Unwillen klar zu äußern.

8 Fazit

Die Studie ist aufgrund ihres Stichprobenumfangs nicht repräsentativ genug, aber sie gibt eine ungefähre Ahnung, was Gamification bewirken kann (oder auch nicht). Meine Beobachtungen waren, dass Gamification im Arbeitsprozess als eine nette Abwechslung und von dem einen oder anderen auch als ein Mittel betrachtet wurde, "was bei der Stange hält". Ein Argument für Gamification ist die Reduktion der Fehleranzahl, gerade hinsichtlich wirtschaftlicher Aspekte betrachtet. Ein anderes Argument ist die Steigerung der Aufmerksamkeit, wenn man sich durch Gamification motivieren lässt.

Ein Problem, was offensichtlich war, ist die zuverlässige Emotionserkennung. Es ist vermutlich schwer, eine geeignete Position der Kamera so zu finden, dass das Gesicht des Monteurs tatsächlich gut eingefangen wird. Da man sich nicht nur bei einem Arbeitsprozess ständig hin und her bewegt, sondern auch Körpergrößen und Haltungen individuell verschieden sind, wird der Einsatz mehrerer Kameras mehr erbringen, als der Einsatz einer Kamera, die zentral platziert ist. Auch die Lichtverhältnisse werden eine große Rolle spielen, da Schattenlinien im Gesicht eine falsche Identifizierung verursachen. Sollte man eine emotionsbasierte Gamification anstreben, ist die Gefahr groß, dass das System die Art der Gamification falsch wählt.

Ein weiterer Aspekt ist die Individualität potentieller Arbeiter. Bei der Studie konnte man gut beobachten, dass es Probanden gab, die sich davon motivieren ließen, welche, die es nicht berührte, aber auch welche, die davon weniger unterstützt als genervt wurden. Beide Durchläufe eines Probanden mussten nach je 5 Durchgängen abgebrochen werden, weil er immer entnervter wurde. Es sei erwähnt, dass dieser Proband einer von den zwei war, die das Baukastensystem geradezu ablehnten.

Trotz allem finde ich, wenn man das System weiterentwickelt und vor allem das Problem der Emotionserkennung löst, also deutlich über dieses einfache Prinzip meines Systems hinaus und weitere Arten der Gamification untersucht, es ein zukunftssträchtiges System wird, was den meisten bei der Durchführung ihrer Arbeit helfen kann. Natürlich wird es auch manche geben, die dem System nichts abgewinnen können, aber ich glaube, dass diese in der Minderheit sein werden. Der überwiegende Teil wird sich davon ansprechen und motivieren lassen.

Aus ethischer Sicht denke ich, wenn gewisse Rahmenbedingungen erfüllt sind, spricht nichts dagegen, ein System für den Einsatz bei älteren oder beeinträchtigten Menschen zu entwerfen und auszuprobieren. Mein einfaches System erzeugte schon Emotionen, die zu verschiedenen Verhaltensmuster führte. Dann wird sich das auch bei älteren oder beeinträchtigten Menschen in einer erkennbaren Form zeigen. Wir manipulieren ja nicht in dem Sinne, dass wir die Benutzer zu 'emotionalen Sklaven' machen, sondern unser Ziel ist es, eine Möglichkeit zu finden, um ihnen zu helfen, ein - im Rahmen ihrer Möglichkeiten - normales Leben führen zu können, indem wir versuchen, Barrieren zu beseitigen, die das verhindern, bzw. den Zugang dazu erschweren. Bei älteren Menschen ist das Ziel, ihre Fähigkeiten zu erhalten, die für gewöhnlich im Alter durch fehlende Stimuli nachlassen. Durch den Einsatz von Gamification könnte man Motivation steigern und Fähigkeiten trainieren.

9 Literaturverzeichnis

9.1 Literaturnachweis

- Azcarate, Aitor; Hageloh, Felix; Van de Sande, Koen; Valenti, Roberto: Automatic facial emotion recognition. University of Amsterdam, 2005. Unter : <http://damienleroux.com/wp-content/uploads/2011/06/Automatic-Facial-Emotion-Recognition.pdf> [Abruf 12.01.2014]
- Aristoteles: Nikomachische Ethik I. Unter : <http://www.textlog.de/33420.html> [Abruf 16.11.2013]
- Bay, Herbert; Ess, Andreas; Tuytelaars, Tinne; Van Gool, Luc: Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding, 110(2008), S.346-359, 2006
- Benjamin, Walter; „Ich packe meine Bibliothek aus. Eine Rede über das Sammeln“ in: Gesammelte Schriften, Bd. IV, Unter Mitwirkung von Theodor W. Adorno und Gershom Scholem, hrsg. von Rolf Tiedemann und Hermann Schweppenhäuser. Bände I–VII, Suppl. I–III (in 17 Bänden gebunden). 1. Auflage, Suhrkamp, Frankfurt am Main 1972–1999)
- Brinks, Marleen: Aggression gegen Computer - Eine wissenschaftliche Untersuchung eines alltäglichen Phänomens. Ibidem-Verlag, Stuttgart, 2005
- Cantor, Georg: 9. Beiträge zur Begründung der transfiniten Mengenlehre, S.481-512(1895). In: Gesammelte Abhandlungen mathematischen und philosophischen Inhalts. Hrsg. Zermelo, Ernst Friedrich Ferdinand, 1932
- Chia, John: Emotion Recognition through Facial Expressions. Department of Computer Science, University of British Columbia, Vancouver, BC
- Cootes, T.F.; Edwards, G.J.; Taylor, C.J.: Active Appearance Models. Department of Medical Biophysics, University of Manchester
- Csikszentmihalyi, Mihaly, Das flow-Erlebnis, Klett-Cotta, 2005
- Crow, Franklin C.: Summed-Area Tables for Texture Mapping. Computer Graphics, Vol. 18, No. 3, 1984
- Darwin, Charles; Der Ausdruck der Gemütsbewegungen bei dem Menschen und den Tieren, Eichborn AG, Frankfurt a.M., 2000
- Ekman, Paul; Gefühle lesen, Spektrum Verlag, Heidelberg, 2010

- Fink, Eugen: Oase des Glücks - Gedanken zu einer Ontologie des Spiels. Karl Alber Verlag, 1957
- Fischer, Lorenz: Strukturen der Arbeitszufriedenheit. Verlag für Psychologie, Göttingen, 1989
- Guttandin, Friedhelm: Einführung in die "Protestantische Ethik" Max Webers. Westdeutscher Verlag, 1998
- Harris, Chris; Stephens, Mike: A combined corner and edge detector. Proceedings of the Alvey Vision Conference, S.147 - 151, 1988
- Herzberg, Frederick et. al.: The Motivation of Work. John Wiley and Sons, Inc., New York, 1959
- Holodynski, Manfred: Emotionen - Entwicklung und Regulation. Springer Medizin Verlag, Heidelberg, 2006
- Kanungo, Tapas; Mount, David M.; Netanyahu, Nathan S.; Piatko, Christine D.; Silverman, Ruht; Wu, Angela Y.: An Efficient k-Means Clustering Algorithm: Analysis and Implementation. In: IEEE Transactions on pattern analysis and machine intelligence, Vol.24, No. 7, Juli 2002
- Kleinbeck/ Kleinbeck: Arbeitsmotivation. Pabst Science Publishers, 2009
- Krishna, Tarun; Rai, Ayush; Bansal, Shubman; Khandelwal, Shubman; Gupta, Shubman; Goyal, Dushyant: Emotion Recognition using Facial and Audio Features. The LNM Institute of Information Technology, Jaipur, Rajasthan, Indien
- Kueblbeck, Christian; Ernst, Andreas: Face detection and tracking in video sequences using the modified census transformation. Journal on Image and Vision Computing, Vol.24, Issue 6, S.564-572, 2006
- LeDoux, Joseph E.: Emotion - Clues from the brain. Annual Rev. Psychol. 1995, S.209-235
- MacQueen, J.: Some methods for Classification and Analysis of Multivariate Observations. Proceeding of the fifth Berkely Comp. on math. Statistics and Probability, 1967
- Maslow, Abraham H.: Motivation and Personality. Harper and Row Publishers Inc, 1972
- Mayring, Philipp: Psychologie des Glücks. Kohlhammer-Verlag, 1991
- Millar, Susanna: Psychologie, des Spiels. Otto Maier Verlag, Ravensburg, 1973

- Moravec, Hans P.: Towards automatic visual obstacle avoidance. Stanford University, 1977. Unter: <http://www.frc.ri.cmu.edu/~hpm/project.archive/robot.papers/1977/aips.txt> [Abruf 04.05.2014]
- Morgan, Eric Lease: Use and understand: A DPLA beta-Sprint proposal. University of Norte Dame, 2011
- Moridis, Christos N.; Economides, Anastasios A.: Affective Learning: Empathetic Agents with Emotional Facial and Tone of Voice Expressions. IEEE Transactions on affective Computing, Vol.3,No.3, Juli - September 2012
- Parment, Anders: Die Generation Y - Mitarbeiter der Zukunft motivieren, integrieren, führen. Springer Verlag, 2013
- Picard, Rosalind W.: Affective Computing. M.I.T. Media Laboratory Perceptual Computing Section, Technical Report No. 321, Cambridge, 1997
- Picard, Rosalind W.: Affective Computing: From Laughter to IEEE. IEEE Transactions on affective Computing, Vol.1,No.1, Januar - Juni 2010
- Pieper, Annemarie: Einführung in die Ethik. Francke-Verlag, 2000
- Pieper, Annemarie; Thurnherr, Urs (Hrsg.): Angewandte Ehtik. Beck, München, 1998
- Platt, John C.: Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines
- Rost, Wolfgang, Emotionen -Elixiere des Lebens, Springer Medizin Verlag, Heidelberg, 2005
- Scherke, Katharian: Emotionen als Forschungsgegenstand der deutschsprachigen Soziologie. Verlag für Sozialwissenschaften/ GWV Fachverlage GmbH, Wiesbaden, 2009
- Sabourin, Jennifer L.; Lester, James C.: Affect and Engagement in Game-Based Learning Environments. IEEE Transactions on affective computing, Vol.5,No.1, Januar - März 2014
- Schlag, Bernhard: Lern - und Leistungsmotivation. Leske + Budrich, 1995
- Su, Mu-Chun; Hsieh, Yi-Jwu; Huang, De-Yuan: A simple Approach to Facial Expression Recognition. National Central university, Taiwan, 2007. Unter: <http://damienleroux.com/wp-content/uploads/2011/06/simple-approach-to-facial-expression.pdf>[Abruf 12.01.2014]

- Tomasik, Brian; Thiha, Phyo; Turnbull, Douglas: Tagging Products using Image Classification. Dept. of Computer Science, Swarthmore College, Swarthmore
- Viola, Paul; Jones, Michael: Rapid Object Detection using a Boosted Cascade of Simple Features. Cambridge, 2001. Unter: <http://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/viola-cvpr-01.pdf>[Abruf 12.01.2014]
- Whitehill, Jacob; Serpell, ZewelANJI; Lin, Yi-Ching; Foster, Aysha; Movellan, Javier R.: The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions. IEEE Transactions on affective computing, Vol.5,No.1, Januar - März 2014
- Wrase, Isabelle: Mitarbeitermotivation im Outsourcing unter besonderer Berücksichtigung des Facility Managements. Gabler Verlag/ Springer Fachmedien, Wiesbaden, 2010
- Übereinkommen über die Rechte von Menschen mit Behinderungen vom 13.Dezember 2006. Aus: Bundesgesetzblatt(BGBl)2008 II

9.2 Onlinequellen

- Adidas Sportuhr - Erfahrungsbericht unter: <http://www.runnersworld.de/ausruestung/adidas-micoach-smart-run-im-test.301986.htm>
- Affdex. Unter: www.affdex.com [Abruf 30.06.2014]
- AttrakDiff zur Beurteilung der UX eines Produkts. Unter: www.attrakdiff.de [Abruf 07.07.2014]
- AttrakDiff: Hassenzahl, Marc; Burmester, Michael; Koller, Franz: Der User Experience(UX) auf der Spur: Zum Einsatz von www.attrakdiff.de. Unter : www.attrakdiff.de[Abruf 07.07.2014]
- Barthelt, Kai Uwe: Bag of (Visual) Features. Vorlesungsfolien zur Vorlesung Computer Vision, HTW Berlin[Abruf 06.03.2014]
- Cardio unter: <http://www.cardio.com/>[Abruf 08.12.2013]
- Cardiograph- App unter : http://www.aerztezeitung.de/praxis_wirtschaft/w_specials/gesundheitsapps2011/article/806569/herzfrequenzmesser-unterwegs.html [Abruf 07.01.2014]
- Cosinuss unter: <http://cosinuss.com/>[Abruf 31.01.2014]

- DUW Studie unter: www.duw-berlin.de/fileadmin/user_upload/content/presse/DUW-Studien/DUW_Motivation_klein.pdf [Abruf 09.11.2013]
- Ekman, Paul: Einiversals and cultural differences in facial expressions of emotion. In: J.Cole(Ed.), Nebraska Symposium on Motivation, Vol. 19, 1971, Lincoln University of Nebraska Press, 1972. Unter: <http://www.paulekman.com/wp-content/uploads/2013/07/Universals-And-Cultural-Differences-In-Facial-Expressions-Of.pdf> [Abruf 12.11.2013]
- EmoTetris. Unter: <http://www.igd-r.fraunhofer.de/emotetris/> [Abruf 30.06.2014]
- Empatica E3. Unter: <https://www.empatica.com> und: <http://www.technologyreview.com/news/421316/sensor-detects-emotions-through-the-skin/> [Abruf 08.07.2014]
- Eulersche Videoverstärkung unter: <http://people.csail.mit.edu/mrub/papers/vidmag.pdf>[Abruf 12.11.2013]
- Fraunhofer (D)Emokoffer. In: Voskamp, Jörg; Blech, Michael; Schultz, Randolph: Vorlesungsfolien. Institut für Graphische Datenverarbeitung, Abteilung AR2 - Human Centered Interaction Technologies. Unter: <http://slideplayer.de/slide/1546/#>. [Abruf 10.07.2014]
- Verfärbung der Haut unter: <http://www.farbimpulse.de/424.pdf> [Abruf 12.11.2013]
- Gestik-Mimik. Unter: rueetschli.net/psychologie/mimik-gestik/mimik-7-basis-emotionen-217.html [Abruf 12.11.2013]
- Herger, Mario. Vortrag zum Thema Gamification. Unter: http://www.mid.de/fileadmin/mid/Bilder/Fotos/Insight12/Vortrage_2012 [Abruf 12.11.2013]
- Projekt MotionEAP. Unter: www.motioneap.de
- Q-Sensor. Unter: <http://affect.media.mit.edu/projectpages/iCalm/iCalm-2-Q.html> [Abruf 12.07.2014]
- Samsung Advanced Institute of Technology. Unter: <http://www.sait.samsung.co.kr/> [Abruf 07.07.2014]
- Samsung Prototyp Smartphone mit Emotionserkennung. Näheres unter: <https://www.presstext.com/news/20120112001?likes=like> [Abruf 07.07.2014]
- Stepstone Studie unter: http://www.stepstone.de/b2b/stellenanbieter/jobboerse-stepstone/upload/studie_gluck_am_arbeitsplatz.pdf?cid=B2C_CLC_SYS19 [Abruf 12.11.2013]

- Weber, Max: Die protestantische Ethik und der Geist des Kapitalismus. Unter: <http://www.zeno.org/Soziologie/M/Weber,+May/Schriften+zur+Religionssoziologie/Die+protestantische+Ethik+und+der+Geist+des+Kapitalismus> [Abruf am 08.11.2013]
- Veenhoven, Ruut: Glück als subjektives Wohlbefinden: Lehren aus der empirischen Forschung. Unter: <http://www2.eur.nl/fsw/research/veenhoven/Pub2010s/2011d-fulld.pdf> [Abruf 16.11.2013]
- "Wollen die auch arbeiten?" - aus: Die Zeit, 07.03.2013, Nr. 11. Unter: www.zeit.de/2013/11/Generation-Y-Arbeitswelt [Abruf 16.11.2013]
- Mimik und FACS. Unter www.wedernoch.de/mimik/facs.html [Abruf 30.11.2013]

9.2.1 Bibliotheken

- Accord.NET. Unter: www.accord-framework.net
- AForge.NET. Unter: www.aforgenet.com
- Cohn-Kanade AU-Coded Expression Database. Unter: <http://www.pitt.edu/~emotion/ck-spread.htm>
- OpenSURF. Unter: <http://www.chrisevansdev.com/computer-vision-opensurf.html>
- EmguCV. Unter: <http://www.emgu.com>
- LibSVM. Unter: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- SHORE (Fraunhofer Institut). Siehe auch: Kueblbeck, Ernst, 2006 und <http://www.iis.fraunhofer.de/de/bf/bsy/fue/isyst.html>

A Anhang

ERFASSUNG DES EMOTIONALEN ZUSTANDS

Proband - Nr:	
Alter:	
Geschlecht:	Datum:

Kreisen Sie für jedes nachfolgend genannte Element die Zahl rechts ein, die Ihrer Meinung nach am besten Ihre Emotion zum Ausdruck bringt. Wählen Sie die Zahl entsprechend der oben angegebenen Skala aus.

Allgemeine Einschätzung	Skala				
	trifft nicht zu				trifft voll zu
1. Ich bin oft gut gelaunt	1	2	3	4	5
2. Ich gehe an neue Aufgaben positiv heran	1	2	3	4	5
3. Ich mag neue Herausforderungen	1	2	3	4	5
4. Positives Feedback ist mir wichtig	1	2	3	4	5
5. Ich bin schnell frustriert	1	2	3	4	5
6. Ich gebe schnell auf	1	2	3	4	5
7. Positives Feedback motiviert mich	1	2	3	4	5
8. Ich bin oft glücklich	1	2	3	4	5
9. Negatives Feedback demotiviert mich	1	2	3	4	5
10. Anerkennung meiner Arbeit steigert meine Leistung	1	2	3	4	5

Emotionale Zustände vor der Studie	Skala				
	trifft nicht zu				trifft voll zu
11. Ich bin gut gelaunt	1	2	3	4	5
12. Ich fühle mich unsicher	1	2	3	4	5
13. Ich bin müde	1	2	3	4	5
14. Ich bin traurig	1	2	3	4	5
15. Ich bin frustriert	1	2	3	4	5
16. Ich bin gereizt	1	2	3	4	5
17. Ich bin motiviert	1	2	3	4	5
18. Ich probiere gern Neues	1	2	3	4	5

Prototypische Untersuchungen im Kontext von Gamification in Arbeitsprozessen

Emotionale Zustände <u>nach der Studie</u>	trifft nicht zu	Skala			trifft voll zu
		1	2	3	
19. Ich bin gut gelaunt	1	2	3	4	5
20. Ich fühle mich unsicher	1	2	3	4	5
21. Ich bin müde	1	2	3	4	5
22. Ich bin traurig	1	2	3	4	5
23. Ich bin frustriert	1	2	3	4	5
24. Ich bin gereizt	1	2	3	4	5
25. Ich fühle mich motiviert	1	2	3	4	5

Bitte beurteilen Sie hier den zweiten Durchgang und Ihre Empfindungen beim Einsatz von Gamification

Beurteilung Gamification <u>nach der Studie</u>	trifft nicht zu	Skala			trifft voll zu
		1	2	3	
26. Die angewandte Art der Gamification finde ich ansprechend	1	2	3	4	5
27. Die angewandte Art der Gamification ist hilfreich beim Lösen der Aufgabe gewesen	1	2	3	4	5
28. Die angewandte Art der Gamification nervte mich	1	2	3	4	5
29. Die angewandte Art der Gamification motivierte mich, besser zu werden	1	2	3	4	5
30. Die angewandte Art der Gamification machte mich fröhlicher	1	2	3	4	5
31. Ich würde eine andere Form der Gamification bevorzugen (Wenn ja, bitte noch 32. beantworten)	1	2	3	4	5
32. a) Was würden Sie ändern?					
b) Warum?					

Erklärung

Ich versichere, diese Arbeit selbstständig verfasst zu haben. Ich habe keine anderen als die angegebenen Quellen benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommene Aussagen als solche gekennzeichnet. Weder diese Arbeit noch wesentliche Teile daraus waren bisher Gegenstand eines anderen Prüfungsverfahrens. Ich habe diese Arbeit bisher weder teilweise noch vollständig veröffentlicht. Das elektronische Exemplar stimmt mit allen eingereichten Exemplaren überein.

Filderstadt, 28.07.2014

Sandra Boffo

Declaration

I hereby declare that the work presented in this thesis is entirely my own. I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted copies.

Filderstadt, 28.07.2014

Sandra Boffo