



Hochschule für Angewandte Wissenschaften Hamburg  
*Hamburg University of Applied Sciences*

# Wissenschaftliche Arbeit

**Kai Bielenberg**

**Companion Technologie - Emotionen erkennen, verstehen und  
verarbeiten**

*Fakultät Technik und Informatik  
Studiendepartment Informatik*

*Faculty of Engineering and Computer Science  
Department of Computer Science*

Kai Bielenberg

**Companion Technologie - Emotionen erkennen, verstehen und  
verarbeiten**

Wissenschaftliche Arbeit eingereicht im Rahmen der AW2 Prüfungsleistung

im Studiengang Master of Science Angewandte Informatik  
am Department Informatik  
der Fakultät Technik und Informatik  
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Prof. Dr. Kai von Luck  
Zweitgutachter: Prof. Dr. Bettina Buth

Eingereicht am: 31. August 2014

**Kai Bielenberg**

**Thema der Arbeit**

Companion Technologie - Emotionen erkennen, verstehen und verarbeiten

**Stichworte**

Companion Technologie, Emotionserkennung, Klassifikation, EmoBike

**Kurzzusammenfassung**

Dieses Dokument stellt mehrere Ansätze aus dem Bereich der Companion Technologie vor und überprüft die Anwendungsmöglichkeiten im EmoBike Projekt der HAW Hamburg

**Kai Bielenberg**

**Title of the paper**

Companion Technologie - emotion detection, recognition and processing

**Keywords**

Companion Technologie, emotion detection, classification, EmoBike

**Abstract**

This documents describes several possibilities to use companion technologie in the EmoBike Projekt from the HAW Hamburg

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Gliederung</b>	<b>1</b>
<b>3</b>	<b>FACS - Facial Action Coding System[7]</b>	<b>2</b>
<b>4</b>	<b>Head Pose and Movement Analysis as an Indicator of Depression [4]</b>	<b>3</b>
4.1	System zur automatischen Depressionserkennung . . . . .	5
4.2	Versuchsaufbau . . . . .	6
4.3	Ergebnis . . . . .	6
4.4	Fazit . . . . .	6
<b>5</b>	<b>Facial Expression Analysis using Eye Gaze Information [10]</b>	<b>6</b>
5.1	Vorgehen . . . . .	7
5.2	Ergebnis . . . . .	8
5.3	Fazit . . . . .	8
<b>6</b>	<b>Analysis of Emotion Recognition using Facial Expressions, Speech and Multi-modal Information [6]</b>	<b>8</b>
6.1	Vorgehen . . . . .	9
6.2	Ergebnis . . . . .	10
6.3	Fazit . . . . .	11
<b>7</b>	<b>Fazit/Ausblick</b>	<b>11</b>

# 1 Einleitung

Diese Ausarbeitung befasst sich mit Companion Technologie und insbesondere mit dem Teilbereich der Emotionserkennung und Verarbeitung. Der Begriff Companion Technologie [9] wurde von Yorick Wilks geprägt. Er beschreibt Produkte, die den emotionalen Zustand des Nutzers erfassen und nutzen können, um diesem zu helfen. Ein Companion<sup>1</sup> ist ein emotionaler vertrauenswürdiger Partner in einer hoch technischen Welt.

Nach dem FSB Transregio 62[3] setzt sich die Companion Technologie aus folgenden Teilbereichen zusammen:

1. Planung und Entscheidung
2. Interaktion und Verfügbarkeit
3. Situation und Emotion
4. Transferprojekte
5. Zentralbereich

Diese Ausarbeitung wird genauer auf den Teilbereich 3 *Situation und Emotion* eingehen. Insbesondere wird die Emotionserkennung betrachtet.

Es wird untersucht inwieweit sich Erkenntnisse aus diesen Bereichen auf das EmoBike Projekt(siehe Abbildung 1.1) der HAW Hamburg übertragen lassen, um dort eine bessere Erkennung des emotionalen Zustandes des Nutzers zu erreichen.

Dieses Projekt behandelt das Erfassen und Regulieren des emotionalen Zustandes eines Probanden innerhalb einer Rehabilitationsmaßnahme. Der Patient soll bei der Nutzung des Reha-Fahrrads unterstützt und überwacht werden. Hierzu ist eine umfassende Klassifikation des emotionalen Zustandes des Probanden unabdingbar.

---

<sup>1</sup>Für weitere Informationen siehe: [5]



Abbildung 1.1: EmoBike [8]

## 2 Gliederung

Nachdem im Kapitel 1 der Kontext der Arbeit umrissen wurde, wird die Arbeit in Kapitel 3 das Facial Action Coding System und dessen Anwendungsmöglichkeiten beschreiben. Dieses bildet die Grundlage vieler Emotionserkennungssysteme.

Die nächsten Kapitel beschreiben verschiedene Ansätze im Bereich der Emotionserkennung. Kapitel 4 und 5 zeigen, dass auch simple Klassifikatoren gute Erkennungsraten aufweisen können. Das Kapitel 6 befasst sich mit Kombinationsmöglichkeiten verschiedener Klassifikatoren. Diese Ansätze zeigen, welche Möglichkeiten die Companion Technologie bietet und geben einen Überblick über den aktuellen Stand der Forschung in diesem Bereich. In Kapitel 7 werden die Erkenntnisse aus den vorgestellten Ansätzen zusammengefasst dargestellt und es wird ein Ausblick auf die weiterführende Arbeit gegeben.

### The Seven Universal Facial Expressions of Emotion

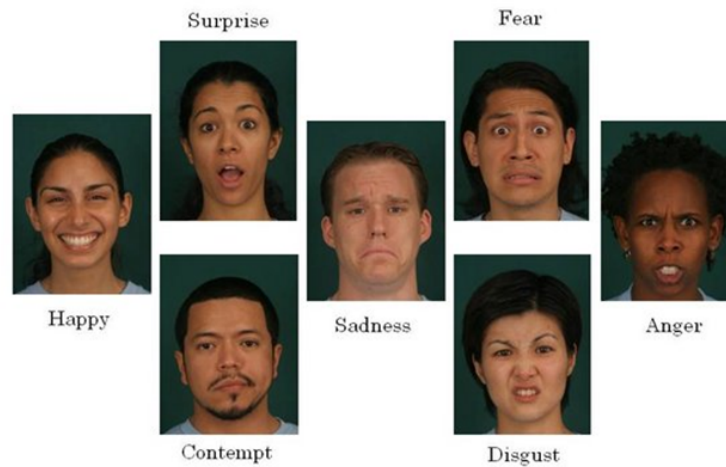


Abbildung 3.1: Sieben Universale Gesichtsausdrücke für Emotionen [1]

## 3 FACS - Facial Action Coding System[7]

Das Facial Action Coding System(FACS) wurde unter anderem von Ekman und Friesen[7] verwendet. Ekman und Friesen haben festgestellt, dass es universelle Gesichtsausdrücke für sieben Emotionen [2] gibt. Diese Gesichtsausdrücke sind in Abbildung 3.1 dargestellt.

Diese Gesichtsausdrücke lassen sich in so genannte Action Units(AU) zerlegen. Dies sind kleinst mögliche Bewegungen im Gesicht, die auf die Gesichtsmuskulatur zurückzuführen sind. Jeder Ausdruck kann in eine Summe von AUs zerlegt werden. Jede AU kann zusätzlich in unterschiedliche Intensitäten unterteilt werden.

Abbildung 3.2 zeigt eine Beispielhafte FACS Kodierung der Emotion Angst. Die Nummer referenziert die entsprechende AU und die Buchstaben zeigen die Intensität der AU an.

Mithilfe des FACS können so Emotionen in Bildern oder Videos erkannt werden. Das FACS kann ebenfalls verwendet werden um Emotionen darzustellen.

### Sample FACS Coding of a Fear Expression

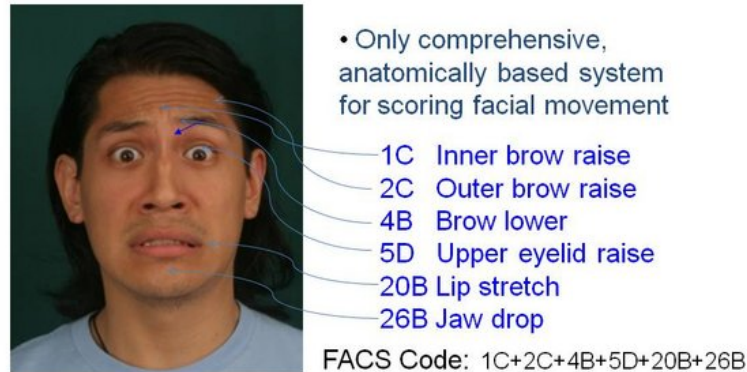


Abbildung 3.2: FACS Kodierungsbeispiel [1]

## 4 Head Pose and Movement Analysis as an Indicator of Depression [4]

Dieser Abschnitt befasst sich mit der Möglichkeit die Erkennung von Depressionen zu verbessern. Hierfür wird ein aufgezeichnetes Patientengespräch mit Mitteln der Informatik automatisiert ausgewertet. Die Ergebnisse zeigen, dass die Depressionserkennung mit der Hilfe der Informatik verbessert werden kann.

**Alghowinem u. a.** haben festgestellt das viele Studien zu Depressionserkrankungen die Gesten und Gesichtsausdrücke der Patienten auswerten, aber sehr wenige Studien die Kopfbewegung berücksichtigen. Daher wird untersucht wie gut sich die Kopfbewegung als Erkennungsmerkmal einer Depression eignet. Laut **Alghowinem u. a.** hat eine Studie gezeigt, dass depressive Patienten bei einem Gespräch weniger Kopfbewegung und weniger Augenkontakt zeigen als ein gesunder Patient.



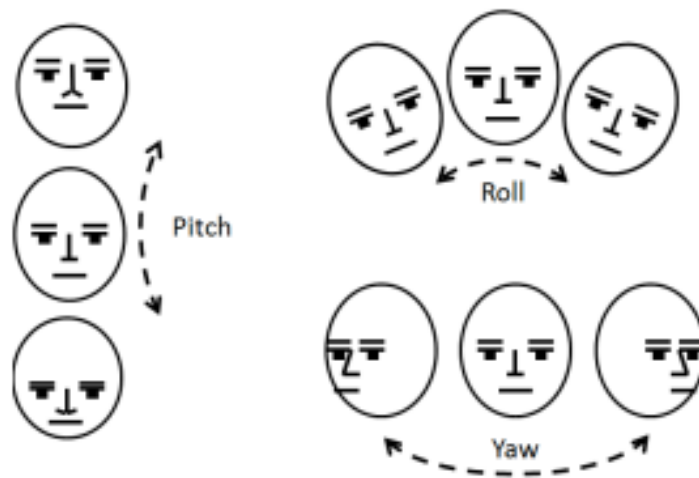


Abbildung 4.1: Kopfbewegungsrichtungen [4]

## 4.1 System zur automatischen Depressionserkennung

Für diesen Ansatz wurde ein System zur automatischen Kopfbewegungsanalyse in Videostreamen entwickelt. Folgende Klassifikatoren wurden aus den Videos extrahiert und untersucht:

- **Kopfbewegung:** Insgesamt werden drei Bewegungsrichtungen ausgewertet. Dies sind die Kopfneigung(pitch), die Kopfdrehung(yaw) und die Ausrichtung(roll). Die Abbildung 4.1 zeigt dies noch einmal.
- **Bewegungsgeschwindigkeit/-beschleunigung:** Es wird gemessen wie schnell und plötzlich Kopfbewegungen von statten gehen.
- **Funktionale Werte:** Verschiedene statistische Werte der oben genannten Messwerte werden mit einbezogen. Hierzu gehören die Maxima, die Minima, die Durchschnittswerte, die Varianz und die Standardabweichung.

Bei jeder Überschreitung eines Schwellenwertes, werden die Daten zur Auswertung genutzt. Die Schwellenwerte setzen sich aus der Summe des Durchschnitts und der Standardabweichung der jeweiligen Messwerte zusammen.

Für die Klassifikation der Messwerte wurde ein hybrider Ansatz gewählt. Es wurden Gaussian Mixture Models(GMM) und Support Vector Maschines(SVM) genutzt. Die GMMs dienen der Dimensionsreduzierung und die SVMs werden anschließend zur Klassifikation genutzt.

## 4.2 Versuchsaufbau

Es wurden Videos von 30 gesunden und 30 depressiven Patienten untersucht. Beide Geschlechter sind gleichermaßen vertreten.

Diese Videos beinhalten ein Interview mit 8 Fragen an den Patienten. Diese Fragen zielen darauf ab, positive und negative Emotionen zu erzeugen. Die Fragen werden nach folgendem Muster aufgebaut: "Denken sie an ein positives/negatives Ereignis. Wie haben sie sich in dieser Situation gefühlt?" So ist es möglich gesteuert positive und negative Emotionen in einem Patienten zu erzeugen und diese dann Messdaten zuzuordnen.

Die Videos wurden manuell annotiert um die Fragen und die gewünschten Emotionszustände zu trennen.

Es wurde getestet wie zuverlässig das System Depressionen erkennen kann. Des Weiteren wurden die Erkennungsraten der Low-Level-Features(Kopfbewegung,-geschwindigkeit,-beschleunigung) und der High-Level-Features(Statistische Werte) verglichen.

## 4.3 Ergebnis

Das System hat im Durchschnitt eine Erkennungsrate von 72% erreicht. Es wurde festgestellt, dass die Erkennungsrate verbessert wird, wenn nur die positiven Emotionszustände der Patienten ausgewertet werden. [Alghowinem u. a.](#) erklären dies damit, dass im positiven Emotionsbereich die Unterschiede der Messwerte von gesunden und depressiven Patienten größer sind.

Die Erkennungsraten konnten durch die zusätzliche Verwendung der High-Level-Features von 68% auf 78% gesteigert werden.

Das System alleine ist nicht zuverlässig genug. Es kann als Hilfe für die Diagnosestellung dienen und den entsprechenden Nutzer unterstützen.

## 4.4 Fazit

Es wurde gezeigt, dass auch simple Klassifikatoren, wie die Kopfbewegung, effektiv genutzt werden können um mehr über den emotionalen Zustand einer Person zu erfahren. Es ist daher Lohnenswert nicht nur den Gesichtsausdruck oder die Körperhaltung einer Person zu erfassen, um einen Einblick in den emotionalen Zustand zu erhalten. Im Kontext des EmoBikes bedeutet dies, dass die Betrachtung weiterer Klassifikatoren sinnvoll sein kann. Es gilt zu untersuchen wie gut sich verschiedene Erkenntnisse aus verschiedenen Klassifikatoren zu einem Zustand zusammenfassen lassen.

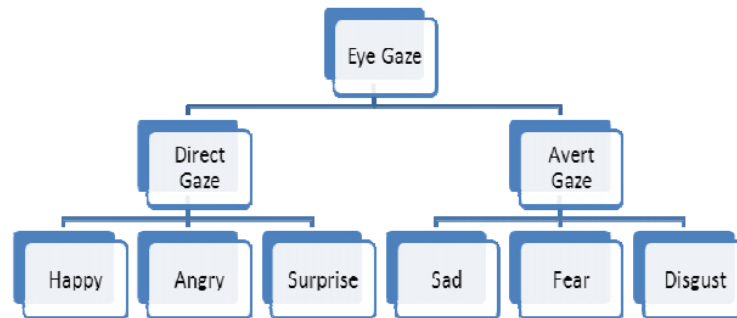


Abbildung 5.1: Unterteilung der Emotionen [10]

## 5 Facial Expression Analysis using Eye Gaze Information [10]

Dieses Kapitel zeigt eine Möglichkeit zur Verbesserung von Emotionserkennung in Videoströmen. *Zhao u. a.* haben festgestellt, dass bestimmte Emotionspaare oftmals schlechter klassifiziert werden. Hierbei handelt es sich zum Beispiel um Ärger und Trauer.

Es wird ein Ansatz vorgestellt indem sechs Emotionen in zwei Gruppen geteilt werden. Die Emotionen werden in *gerichtet* und *ungerichtet* unterteilt. Es wird untersucht wie diese Erkenntnis die Erkennungsraten verbessern kann.

### 5.1 Vorgehen

Die Abbildung 5.1 zeigt die Unterteilung der Emotionsgruppen. Emotionen werden in gerichtete und ungerichtete Emotionen aufgeteilt. *Freude(Happy)*, *Ärger(Angry)* und *Überraschung(Surprise)* gehören zu den gerichteten Emotionen. Diese werden mit direkter Blickrichtung assoziiert. *Trauer(sad)*, *Angst(fear)* und *Ekel(disgust)* gehören zu den ungerichteten Emotionen, diese werden mit einer abgewendeten Blickrichtung assoziiert.

Aus einem Videostrom werden Bilder extrahiert und ausgewertet. Zuerst wird die Blickrichtung untersucht und danach wird die entsprechende Emotion klassifiziert.

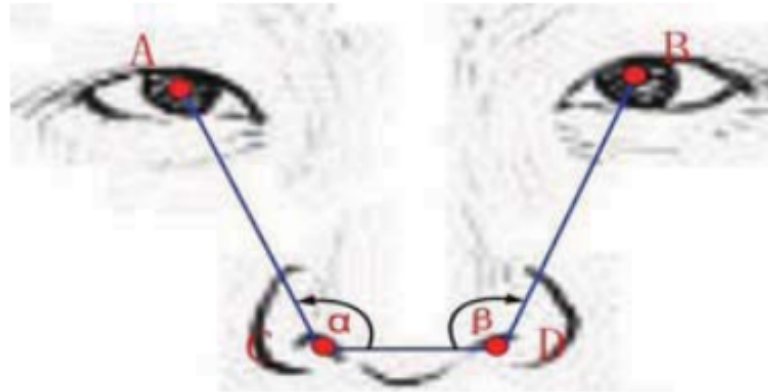


Abbildung 5.2: Blickrichtungserkennung [10]

Die Blickrichtung wird anhand der Winkel zwischen Nasenlöchern und Pupillen bestimmt. Nasenlöcher und Pupillen sind sehr dunkel im Vergleich zum Rest des Gesichtes und lassen sich sehr zuverlässig automatisch in den Bildern finden (Erkennungsrate: 96,7%).

Nach der Blickrichtungserkennung wird der emotionale Zustand der Person in dem Bild klassifiziert.

## 5.2 Ergebnis

Durch die Aufteilung in gerichtete und ungerichtete Emotionen konnte die Erkennungsrate der einzelnen Emotionen deutlich verbessert werden. Die Erkennungsraten von *Wut*, *Trauer* und *Angst* konnten von durchschnittlich 83% auf über 90% gesteigert werden.

## 5.3 Fazit

Es wurde gezeigt, dass kleine Verbesserungen an der Klassifikation deutliche Auswirkungen auf den Erfolg haben können. Eine genaue Analyse des Anwendungsfalles ist erforderlich um den geeignetsten Klassifikator zu wählen. Je nach Anwendungsfall können so deutlich genauere Ergebnisse erzielt werden.

Da im EmoBike Projekt ein Eye-Tracker zur Verfügung steht, kann diese Methode gut integriert werden um die Qualität der Emotionserkennung zu verbessern.

## 6 Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information [6]

Dieses Kapitel befasst sich mit der Möglichkeit verschiedene Sensordaten zur Klassifikation des emotionalen Zustandes zu verwenden. Es wird untersucht zu welchem Zeitpunkt des Klassifikationsprozesses eine Zusammenführung der Daten sinnvoller ist.

### 6.1 Vorgehen

Es werden vier unterschiedliche Emotionen klassifiziert. Dies sind *Trauer*, *Freude*, *Ärger* und *neutral*.

Zur Klassifikation wird die Mimik und die Sprache des Probanden analysiert.

- **Sprachanalyse:** Insgesamt werden 11 Features aus dem Audiostream extrahiert. Der Featurevektor setzt sich aus der *Tonhöhe*, der *Intensität*, der *Länge der Wortpausen* und *statistischen Werten*, wie zum Beispiel der Standardabweichung und dem Durchschnitt, zusammen. Für die Klassifikation wurde eine SVM verwendet.
- **Mimikanalyse:** Das Gesicht des Probanden wird aus dem Videostream extrahiert und in fünf Regionen aufgeteilt. Dies sind die *Stirn*, die *Augenbrauen*, die *Nase* und die *rechte und linke Wange*. Der Mund wird ausgelassen, da dieser durch das Sprechen nicht aussagekräftig ist. Pro Region wird ein 10-Dimensionaler Featurevektor extrahiert. Dies sind markierte Punkte in jeder Region. Diese Vektoren werden jeweils mittels eines K-Nearest-Neighbour Klassifikators(k=3) zu einem 4-Dimensionalen Featurevektor zusammengefasst. Die Dauer der Emotion wird mit berücksichtigt. Pro Region wird eine eigene SVM zur Klassifizierung eingesetzt. So kann untersucht werden, welche Regionen besonders geeignet sind um eine bestimmte Emotion zu klassifizieren.

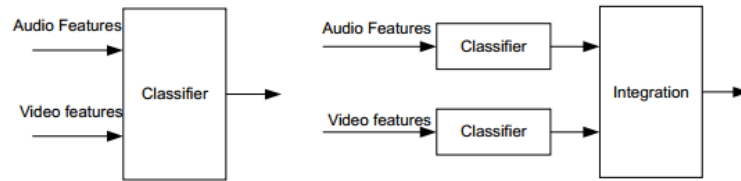


Abbildung 6.1: Unterschied von *feature level(links)* und *decision level(rechts)* [6]

Neben den einzelnen Klassifikatoren wird ein **Bimodales System** entwickelt, dass beide Analysen zusammen zur Klassifikation verwendet. Untersucht wird die Zusammenführung auf *decision level* oder *feature level*. Der Unterschied wird in Abbildung 6.1 verdeutlicht.

Die **feature level** Zusammenführung wird mittels eines gemeinsamen Klassifikators erzielt, der die oben beschriebenen Sensordaten als Input erhält. Die **decision level** Zusammenführung erfolgt durch unterschiedliche Kombinationsmöglichkeiten. Es wird das *Maximum*, der *Durchschnitt*, das *Produkt* und ein *gewichteter Durchschnitt* der ausgewerteten Daten verwendet.

## 6.2 Ergebnis

Der **Sprachklassifikator** erzielt Erkennungsraten von ca. 64%. Probleme wurden bei der Unterscheidung von *Freude/Ärger* und *Trauer/neutral* festgestellt. Diese Paare wurden in ca. 15% für das Gegenstück gehalten.

Der **Mimikklassifikator** erreicht eine Erkennungsrate von ca. 85% wenn alle Regionen kombiniert betrachtet werden. Die Klassifikation durch einzelne Regionen ist schlechter als die Kombination der Regionen. Da sich nicht jede Emotion in jeder Region eindeutig widerspiegelt, ist dieses Ergebnis nicht unerwartet.

Das **bimodale System** erreicht Erkennungsraten von 89% sowohl bei *feature level* als auch bei *decision level* Zusammenführung. Wobei der Unterschied von *Freude/Ärger* deutlicher bei der *decision level* Zusammenführung erkannt wird. Gleichzeitig wird das *Trauer/neutral* Paar besser vom *feature level* Klassifikator erkannt.

Die Zusammenführung im *decision level* Bereich war mit der Produkt Methode am erfolgreichsten.

## 6.3 Fazit

Der Audioklassifikator erreicht die schlechtesten Erkennungsraten. Trotzdem wird gezeigt, dass die Zusammenführung der Ergebnisse trotzdem von den schlechteren Audiodaten profitiert. Es ist sinnvoll schlechtere Sensordaten trotzdem in dem Entscheidungsprozess zu berücksichtigen, da diese zu einem genaueren Endergebnis beitragen können.

Für den EmoBike Kontext bedeutet dies, dass eine Zusammenführung der einzelnen Sensoren vielversprechend ist. Die Art der Zusammenführung muss genau untersucht werden um die besten Ergebnisse für den Anwendungsfall zu erzielen.

Wie das Kapitel 6 zeigt, gibt es keine generelle Lösung für die Art der Fusion der Sensordaten um das beste Ergebnis zu erzielen.

## 7 Fazit/Ausblick

Die unterschiedlichen Ansätze zeigen, dass verschiedene Klassifikatoren je nach Anwendungsfall besser oder schlechter geeignet sind. Des Weiteren ist eine Zusammenführung mehrerer Klassifikatoren sinnvoll um bessere Erkennungsraten zu erreichen.

Zu beachten ist jedoch, dass sich die Psychologen, auf deren Theorien die verschiedene Software entwickelt wurde, sich nicht immer einig sind. Anhand der vorgestellten Ansätze ist zum Beispiel zu sehen, dass Teilweise sechs oder sieben Grundemotionen beschrieben werden. Daher ist eine kritische Betrachtung der einzelnen Theorien wünschenswert, oft aber nicht umfassend möglich.

Im Kontext des EmoBikes und der vorgestellten Ansätze ist es erfolgversprechend unterschiedliche Klassifikatoren je nach Anwendungsfall zu nutzen. Eine Zusammenführung verschiedener Ansätze kann Helfen eine robustere Emotionserkennung zu schaffen.

In Zukunft muss untersucht werden, wie die einzelnen Komponenten besser zusammenarbeiten können.

## Literaturverzeichnis

- [1] : *Facial expression analysis*. – URL [http://www.scholarpedia.org/article/Facial\\_expression\\_analysis](http://www.scholarpedia.org/article/Facial_expression_analysis)
- [2] : *Micro Expressions*. – URL <http://www.paulekman.com/micro-expressions/>
- [3] : *SFB Transregio 62: Eine Companion-Technologie für kognitive technische Systeme*. – URL <http://www.sfb-trr-62.de/>
- [4] ALGHOWINEM, S. ; GOECKE, R. ; WAGNER, M. ; PARKERX, G. ; BREAKSPEAR, M.: Head Pose and Movement Analysis as an Indicator of Depression. In: *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, Sept 2013, S. 283–288. – ISSN 2156-8103
- [5] BIELENBERG, Kai: *Companion Technology*
- [6] BUSO, Carlos ; DENG, Zhigang ; YILDIRIM, Serdar ; BULUT, Murtaza ; LEE, Chul M. ; KAZEMZADEH, Abe ; LEE, Sungbok ; NEUMANN, Ulrich ; NARAYANAN, Shrikanth: Analysis of Emotion Recognition Using Facial Expressions, Speech and Multimodal Information. In: *Proceedings of the 6th International Conference on Multimodal Interfaces*. New York, NY, USA : ACM, 2004 (ICMI '04), S. 205–211. – URL <http://doi.acm.org/10.1145/1027933.1027968>. – ISBN 1-58113-995-0
- [7] EKMAN, P. ; FRIESEN, W.: *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto : Consulting Psychologists Press, 1978
- [8] ROSSEBURG, Kai: *EmoBike Bild*
- [9] WILKS, Yorick: *Close engagements with artificial companions: key social, psychological, ethical and design issues*. Bd. 8. John Benjamins Publishing, 2010
- [10] ZHAO, Yisu ; WANG, Xin ; PETRIU, E.M.: Facial expression anlysis using eye gaze information. In: *Computational Intelligence for Measurement Systems and Applications (CIMSA), 2011 IEEE International Conference on*, Sept 2011, S. 1–4. – ISSN 2159-1547