Masud Fazal-Baqaie et al. (Hrsg.): Projektmanagement und Vorgehensmodelle,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022

Emotionserkennungssoftware auf Basis von Gesichtsausdrücken zur Unterstützung der virtuellen Zusammenarbeit

Bärbel Bissinger[1], Dana Hohmann[2], Michael Fellmann[3] und Christian Märtin[4]

Abstract: Durch die fortschreitende Digitalisierung und vor allem bedingt durch die Covid-19 Pandemie findet eine Vielzahl von Besprechungen derzeit nur noch online statt. Dadurch verändern sich Kommunikation und Zusammenarbeit. Bei virtuellen Interaktionen ist das Erkennen und Zeigen von Emotionen, was für die menschliche Kommunikation essenziell ist, erschwert. Die softwarebasierte Emotionserkennung durch Auswertung von Gesichtsausdrücken ist ein zunehmend wichtiges Gebiet in der Mensch-Computer-Interaktions-Forschung und es gibt mehr und mehr Produkte, die dies ermöglichen sollen. Unser Beitrag gibt eine Übersicht über Aktivitäten in der Forschungs- und Produktwelt und skizziert die Reife der Technologien. Weiter stellen wir potenzielle Anwendungsfälle für Individuen und Kleingruppen vor. Dadurch soll der Frage nachgegangen werden können, ob diese Technologien als ein Werkzeug dienen können, um dem Menschen bei der Erkennung von Emotionen in Videokonferenzen zu assistieren.

Keywords: Computerbasierte Emotionserkennung, Gesichtsausdrücke, Mimik, Anwendungsfälle, Unternehmen, Virtuelle Interaktion

1 Einleitung

1.1 Wichtigkeit von Emotionen und Gesichtsausdrücken

Emotionen spielen bei allen menschlichen Handlungen und Interkationen eine bedeutende Rolle und sind daher für unsere Wahrnehmung, unsere Kommunikation und Entscheidungsfindung essenziell [Pi00]. Am emotionalsten und am direktesten kommunizieren Menschen in persönlichen Gesprächen von Angesicht zu Angesicht. Dabei werden Emotionen unbewusst und über die Mimik mitgeteilt. Innerhalb einer Gruppe wird diese Mimik verwendet, um ein bestimmtes Verhalten auszulösen oder die Gruppe zu organisieren. Durch das Gesicht ausgedrückte Emotionen können daher eine sehr starke Wirkung haben und andere Menschen zu Handlungen anregen [SS14].

1.2 Veränderung der Zusammenarbeit in der virtuellen Welt

Bedingt durch die Covid-19 Pandemie finden eine Vielzahl an Besprechungen derzeit virtuell statt und dieser Trend wird sich voraussichtlich auch in Zukunft fortsetzten [Th21]. Ein durch diesen Trend aufgetretenes Phänomen ist die sogenannte *Zoom Fatigue,* was allgemein für Erschöpfungserscheinungen, die durch Videokonferenzen auftreten, steht [NW21].

[1] Hochschule Augsburg, Fakultät Informatik, 86161 Augsburg, baerbel.bissinger@hs-augsburg.de
[2] Universität Rostock, Institut für Informatik, 18059 Rostock, dana.hohmann@uni-rostock.de
[3] Universität Rostock, Institut für Informatik, 18059 Rostock, michaeel.fellmann@uni-rostock.de
[4] Hochschule Augsburg, Fakultät Informatik, 86161 Augsburg, christian.maertin@hs-augsburg.de

Wenn es um die menschliche Kommunikation geht, gibt es zwischen virtuellen und Vor-Ort-Veranstaltungen einen großen Unterschied, insbesondere beim Zeigen und Erkennen von Emotionen. Bei Videokonferenzen kann nur ein Teil der menschlichen Ausdrucksweise, wie die Mimik und die Stimme, transportiert werden. Selbst diese wenigen Merkmale können durch Umstände wie kleine Kachelbilder oder schlechte Bildqualität nicht immer eindeutig übermittelt werden [ASO18]. Daher ist in virtuellen Umgebungen die Kommunikation von nonverbalen Signalen stark eingeschränkt und das Zeigen oder Erkennen dieser nonverbalen Botschaften kann sowohl für den Sender als auch den Empfänger der Nachrichten erschwert oder nicht möglich sein [OSB20]. Zudem können unbeabsichtigt Botschaften gesendete werden, wenn etwa der angestrengte oder ärgerliche Blick auf eine Fehlermeldung am Rechner beim Gegenüber ungewollt als Skepsis oder Ablehnung gedeutet werden. Derartige Phänomene können bei virtuellen Teams beispielsweise zu Planungs- und Koordinationsschwierigkeiten, Unklarheiten bei Rollen- und Aufgabenverteilungen, verminderte Entscheidungsqualität und zu Herausforderungen beim Vertrauensaufbau im Team führen [ASO18].

1.3 Bedarf computerbasierter Emotionserkennung über Gesichtsausdrücke

Im Kontext der virtuellen Zusammenarbeit sind in Bezug auf Emotionserkennung besonders Gesichtsausdrücke und Prosodie, die Erscheinungen der Stimme, relevant. Auch eine Analyse des Gesagten oder von Chat-Nachrichten wäre durch eine semantische Analyse beispielsweise auf Basis von Natural Language Processing (NLP) denkbar, wie Ansätze dazu zeigen [BLR19]. Weitere Möglichkeiten, um Rückschlüsse auf emotionale Zustände ziehen zu können, stellen zum Beispiel die Messung von körperlichen Signalen wie Herz-Kreislauf-Aktivitäten, die Hautleitfähigkeit oder elektrischen Aktivitäten des Gehirns dar [Pi00, BKM22]. Wir fokussieren uns aber auf die in virtuellen Arbeitsumgebungen verfügbaren Informationen, im Speziellen auf Gesichtsausdrücke, da diese auch dann übertragen und analysiert werden können, wenn die Person nicht spricht bzw. stummgeschaltet ist.

Ein Gesichtsausdruck sagt „mehr als tausend Worte“ und war schon bei Darwin Kernbestandteil der Emotionsforschung [Ne13, Ek09]. Doch obwohl ein Großteil der Anzeichen für Emotionen durch die Mimik vermittelt wird, sind wir nicht daran gewöhnt, diese lediglich über kleine Bildausschnitte des Gesichts zu erkennen. Teilnehmer von Videokonferenzen berichten beispielsweise, dass sie Schwierigkeiten haben, diese Ausdrücke in Videoübertragungen zu lesen und ebenso diese anderen per Video mitzuteilen [NW21]. Eine Möglichkeit das Erkennen von Emotionen bei Videoübertragungen zu unterstützen oder zu trainieren können Tools zur automatischen Emotionserkennung über Gesichtsausdrücke (Facial Expression Recognition – FER) sein [BMF22].

2 Computerbasierte Emotionserkennung über Gesichtsausdrücke

Im Jahr 1990 identifizierte Duchenne verschiedene Gesichtsmuskeln, die unterschiedliche Ausdrücke darstellen. Auf der Grundlage dieser Muskeln und der universell gültigen Grundemotionen von Ekman entwickelten dieser und Kollegen das *Facial Action Coding*

System (FACS), eine Beschreibung von Gesichtsausdrücken. Darüber hinaus entwickelten sie das *Emotions-FACS* (EmFACS), wodurch Korrelationen zwischen Muskelbewegungen im Gesicht und Emotionen ermittelt werden können. Mit Hilfe des FACS und des EmFACS können Gesichtsausdrücke auf standardisierte Weise bewertet werden. Viele der heutigen computerbasierten Ansätze zur Gesichtsemotionserkennung basieren auf diesem System [Pi00].

Generell sieht der Prozess von FER-Systemen folgende drei Schritte vor. Der Prozess beginnt nach der Eingabe des Bildmaterials mit dem Preprocessing [Jo20] In diesem Schritt erfolgen die Gesichtserkennung, die Einteilung der Gesichter in Bereiche (Regions of Interest – ROIs) für die spätere Klassifikation sowie eventuell nötige Bildkorrekturen. Im zweiten Schritt werden aus den ROIs die nötigen Gesichtsmerkmale extrahiert. Abschließend folgt die Klassifikation bzw. Regression mittels verschiedener Methoden wie Neuronalen Netzen, dem Hidden-Markov-Modell oder dem k-Nearest-Neighbour-Algorithmus [CN19].

Um einzuschätzen, ob Emotionen bereits zuverlässig von FER-Software erkannt werden können, geben wir im Folgenden einen Einblick über aktuelle Forschungsaktivitäten und stellen eine Auswahl von FER-Produkten und deren jeweiligen Reifegrad vor.

2.1 Forschungsaktivitäten

Die Anzahl der Publikationen ist in den vergangenen Jahren stark angestiegen. Dies trifft sowohl auf Arbeiten zu, die sich mit verschiedenen Vorgehensweisen zur Emotionserkennung auseinandersetzen als auch auf Veröffentlichungen, die mögliche Anwendungsgebiete aufzeigen. Das zeigt eine Literaturanalyse zum Thema FER, bei der vier Datenbanken (ACM, IEEE, BASE, Springer) von 2006 bis 2019 durchsucht wurden. Insgesamt wurden ohne Duplikate 783 Paper zum Thema „Emotion Recognition“ oder „Facial Expression Recognition“ gefunden. Davon passen 112 Paper zu den Selektionskriterien FER und Computervision [CN19].

Um einen Überblick über die Forschungsaktivitäten zu computerbasierter FER, den Forschungsansätzen und den verwendeten Tools zu bekommen, haben wir eine strukturierte Literaturanalyse mit Scopus durchgeführt. Dabei haben wir Publikationen ab dem Jahr 2012 betrachtet und uns auf FER-Tools für die Emotionserkennung von bewegten Bildern fokussiert. Abbildung 1 zeigt die Auswertung unserer Analyse mit den finalen Suchqueries.

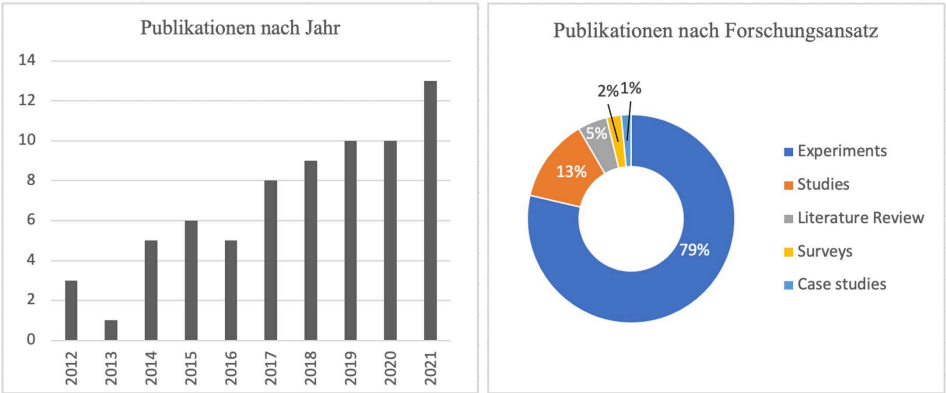

Abb. 1: Publikationen zur softwarebasierten FER nach Jahr und Forschungsansatz

Obwohl die Publikationen und Forschungsaktivitäten zunehmen und das Interesse an FER wächst, bewegen sich die Publikationen insgesamt noch auf einem niedrigen Niveau.

In der rechten Hälfte der Abbildung wird die Ergebnismenge anteilig nach dem jeweiligen Forschungsansatz veranschaulicht. Eine Vielzahl der betrachteten Arbeiten zum Thema FER beschreibt Experimente bzw. Entwicklungen von Algorithmen, mit denen verschiedenen Methoden zur softwarebasierten Emotionserkennung untersucht werden. In Publikationen zu potenziellen Anwendungsgebieten wurden zudem vermehrt Studien durchgeführt, die neue Einsatzmöglichkeiten aufzeigen bzw. evaluieren.

2.2 Produkte und Reifegrad

Auch das Angebot an FER-Produkten ist in den letzten Jahren gestiegen. Die vier am häufigsten genutzten Produkte in den betrachteten Publikationen sind in der untenstehenden Tabelle aufgeführt. Zu jedem dieser Produkte ist zudem die *True Positive Rate* (TPR) für die Erkennung von gestellten bzw. spontanen Emotionen in Videos angegeben, welche in unabhängigen Versuchen von Dupré et al. ermittelt wurde. Für diese Studie wurden zwei Datenbanken mit dynamischen Gesichtsausdrücken verwendet, die gestellte und spontane Grundemotionen enthalten. Daraus wurden über 900 Videostimuli von acht verschiedenen FER-Systemen sowie menschlichen Beobachtern für die Emotionsbestimmung herangezogen. Tabelle 1 zeigt die TPR der Untersuchungen von ausgewählten Produkten. Im Vergleich zum Menschen ist die Akkuratesse der Produkte noch geringer, wie aus dieser Studie hervorgeht [Du20].

	Mensch		Facet (Emotient)		FaceReader (Noldus)		MS Azure Face API		Affdex (Affectiva)	
Emotionen	Gestellt	Spontan	Gestellt	Spontan	Gestellt	Spontan	Gestellt	Spontan	Gestellt	Spontan
Ärger	0,68	0,08	0,72	0,02	**0,60**	**0,01**	0,56	0,03	0,60	0,00
Abscheu	0,64	0,89	0,56	0,83	**0,59**	**0,82**	0,65	1,00	0,29	0,39
Angst	0,73	0,11	0,95	0,22	**0,76**	**0,05**	1,00		0,41	0,09
Freude	0,92	0,97	0,61	0,70	**0,73**	**0,75**	0,47	0,63	0,68	0,75
Traurigkeit	0,90	0,54	0,74	0,32	**0,62**	**0,27**	0,53	0,21	0,47	0,19
Überraschung	0,84	0,81	0,77	0,59	**0,84**	**0,69**	0,72	0,50	0,57	0,77

Tab. 1: TPR der Erkennung von gestellten und spontanen Grundemotionen [Du20]

Die Ergebnisse zeigen einerseits, dass die Genauigkeit der Emotionserkennung stark zwischen den Grundemotionen variiert. Andererseits zeigen sie auch, dass die Produkte Emotionen über Gesichtsausdrücke noch nicht exakt ermitteln können. Allerdings wird auch ersichtlich, dass selbst Menschen spontane Emotionen über Gesichtsausdrücke nicht immer eindeutig erkennen (wie in den Tests z. B. bei Ärger, Angst und Traurigkeit) und es hier Potenziale zur Verbesserung gibt.

Die Tatsachen, dass Emotionen nicht direkt kommuniziert werden, sondern nur indirekt über Signale, durch welche auf verschiedene emotionale Zustände geschlossen werden kann, sowie die Unterschiede in körperlichen Anzeichen je nach Person, Kultur und Kontext, erschwert die Emotionserkennung sowohl für den Menschen als auch für die computerbasierte Erkennung [Pi00].

Maschinelles Lernen sowie die Ansätze und Algorithmen von FER-Systemen entwickeln sich stetig weiter, wodurch jedes Jahr exaktere FER-Systeme entstehen [CN19]. Dennoch ist die Emotionserkennung sehr komplex und bringt viele Herausforderungen mit sich. Der sinnvolle Einsatz der Technologien im organisationalen Kontext hängt also stark von der zukünftigen Weiterentwicklung der Systeme ab.

3 Potenzielle Anwendungsfälle

Nachdem ein erster Überblick über FER in der Wissenschaft und in der Produktwelt gegeben wurde, werden in diesem Abschnitt mögliche Anwendungsfälle kurz aufgezeigt. Diese haben wir in zwei Anwendungsbereiche, auf Individualebene und auf Gruppenebene, unterteilt. Für jede Kategorie werden exemplarisch zwei Anwendungsfälle beschrieben.

3.1 Individualebene

Auf Individualebene kann ein System mit FER zum Beispiel eingesetzt werden, um das *Erkennen von eigenen und fremden emotionalen Zuständen* auf Basis von Gesichtsausdrücken zu trainieren und die Aufmerksamkeit dahingehend zu schulen.

In einem Anwendungsfall auf dieser Ebene schaut sich der Teilnehmende Videosequenzen von Gesichtern mit unterschiedlichen emotionalen Zuständen an und notiert sich die Emotionen. In einem nächsten Schritt werden die Videosequenzen nochmals angeschaut, zusätzlich werden nun auch die vom System erkannten Emotionen angezeigt. Dies soll beim Betrachter die Aufmerksamkeit für Veränderungen der Gesichtsmuskeln und das Lesen von Emotionen nur mit Hilfe von Gesichtsausdrücken trainieren.

Außerhalb von Trainings könnte dies auch im beruflichen Alltag Anwendung finden, indem sich die Person zum Beispiel bei Online-Besprechungen selbst filmt, die gezeigten Gesichtsausdrücke analysieren lässt und ihre Einschätzung der empfundenen Emotionen mit den Ergebnissen der Analyse vergleicht. So kann die eigene Aufmerksamkeit auf Veränderungen in emotionalen Zuständen trainiert werden. Ebenso kann die Genauigkeit der Analyse mit den Einschätzungen der Personen getestet und verglichen werden.

3.2 Gruppenebene

Auf Gruppenebene könnte der Einsatz eines Systems mit FER die *Interaktionsqualität verbessern, indem emotionale Zustände aller Teilnehmer und weitere Interaktionsdaten, wie zum Beispiel die Redeanteile der Teilnehmer, visualisiert werden*. Dies kann insbesondere dem Moderierendem als Hilfestellung dienen, um gezielt auf die gemessenen Daten eingehen und reagieren zu können.

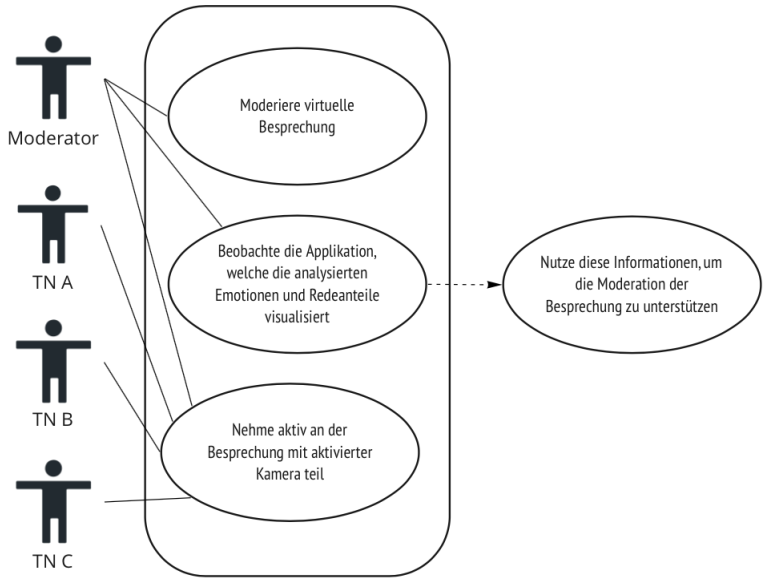

Abb. 2: Beispiel eines Anwendungsfalls auf Gruppenebene

4 Fazit und zukünftige Arbeiten

Dieser Beitrag im *Future Track* soll einen ersten Einblick über die computerbasierte Emotionserkennung mit Hilfe von Gesichtsausdrücken, deren Reife und mögliche Anwendungsfälle geben und als Anregung zur Diskussion dieser Technologie dienen.

Die Forschungsaktivitäten und Produkte zur FER nehmen zu, allerdings ist es fraglich, ob der Reifegrad der Produkte für die Zielerreichung der beschriebenen Anwendungsfälle schon ausreichend ist. Dies soll daher in zukünftigen Forschungsarbeiten durch erste Tests geprüft werden. Im Business-Kontext sehen wir diverse Anwendungsfälle auf *Individual-* und *Gruppenebene*, von denen wir exemplarisch zwei beschrieben haben. Dabei soll FER unterstützen, Emotionen von sich selbst und anderen leichter erkennen zu können und die Aufmerksamkeit dafür zu trainieren, um virtuelle Interaktionen verbessern zu können.

Neben den Möglichkeiten durch diese Technologien sehen wir aber auch Risiken durch deren Einsatz. Risiken wie ein Missbrauch der Daten, beispielsweise durch Überwachung müssen durch zukünftige ethische Standards oder wertebasiertes Design [FH19] minimiert

werden. Ebenso müssen Richtlinien für die systematische Bewertung und Verringerung der Risiken, die von Emotionserkennungssystemen ausgehen, berücksichtigt werden [He21].

Ein weiteres mögliches Risiko könnte zum Beispiel eine kognitive Überlastung oder eine unnatürliche Kommunikation durch den Einsatz von FER sein. Andererseits könnte der Einsatz von FER-Systemen auch dazu genutzt werden, um die generelle kognitive Last in Videokon1enzen (sog. *Zoom-Fatigue*) zu senken, indem Emotionen beispielsweise nicht durch die Übertragung des eigenen Videos, sondern über Emojis oder Avatare, welche die vom System erkannten Emotionen repräsentieren, an andere vermittelt werden. Diese Einsatzmöglichkeiten und Herausforderungen von FER müssen durch Studien untersucht und evaluiert werden.

Literaturverzeichnis

[ASO18] A. Acai, R. R. Sonnadara, and T. A. O'Neill: Getting with the times: a narrative review of the literature on group decision making in virtual environments and implications for promotions committees, Perspect Med Educ, vol. 7, no. 3, pp. 147–155, 2018.

[BKM22] Balic, S., Kleybolte, L., Märtin, C.: A Swarm Intelligence Approach: Combination of Different EEG-Channel Optimization Techniques to Enhance Emotion Recognition. In: Kurosu, M. (eds) Human-Computer Interaction. Technological Innovation, HCII, 2022.

[BLR19] E. Batbaatar, M. Li and K. H. Ryu: Semantic-Emotion Neural Network for Emotion Recognition From Text, in IEEE Access, vol. 7, pp. 111866-111878, 2019.

[BMF22] Bissinger, B., Märtin, C., Fellmann, M.: Support of Virtual Human Interactions Based on Facial Emotion Recognition Software. In: Kurosu, M. (eds) Human-Computer Interaction. Technological Innovation, HCII, 2022.

[CN19] Canedo, D; Neves, A J. R.: Facial Expression Recognition Using Computer Vision: A Systematic Review. Applied Sciences, Bd. 9, Nr. 21, 2019.

[Du20] Dupré, D. et. al: A performance comparison of eight commercially available automatic classifiers for facial affect recognition. *PLOS ONE*, Bd. 15, Nr. 4, S. 1–17, 2020.

[Ek09] Ekman P.: Darwin's contributions to our understanding of emotional expressions, Philos Trans R Soc Lond B Biol Sci, 2009.

[FH19] Friedman, B., Hendry, D.G.: Value Sensitive Design: Shaping Technology with Moral Imagination, MIT Press, Cambridge Massachusetts, 2019.

[He21] J. Hernandez et al., "Guidelines for Assessing and Minimizing Risks of Emotion Recognition Applications," 2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII), 2021.

[Jo20] John, A. et.al: Real-time facial emotion recognition system with improved preprocessing and feature extraction. In: Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT, 2020.

[Ne13] Newmark, C.: Charles Darwin: The Expression of the Emotions in Man and Animals. In: Senge, K., Schützeichel, R. (eds) Hauptwerke der Emotionssoziologie. Springer VS, Wiesbaden, 2013.

[NW21] H. Nesher Shoshan and W. Wehrt, “Understanding ‘Zoom fatigue’: A mixed-method approach,” Applied Psychology, p. apps.12360, 2021.

[OSB20] R. S. Oeppen, G. Shaw, and P. A. Brennan: Human factors recognition at virtual meetings and video conferencing: how to get the best performance from yourself and others, British Journal of Oral and Maxillofacial Surgery, vol. 58, no. 6, pp. 643–646, 2020.

[Pi00] R. W. Picard: Affective computing (1st ed.). Cambridge, Mass.: MIT Press., 2000.

[SS14] R. Stürmer and J. Schmidt, “Erfolgreiches Marketing durch Emotionsforschung,” Haufe, Freiburg, 2014.

[Th21] McKinsey Global Institute: The future of work after COVID-19, 2021.