

Tracking von Gesichtsmimik mit Hilfe von Gitterstrukturen zur Klassifikation von schmerzrelevanten Action Units

Christine Barthold¹, Anton Papst¹, Thomas Wittenberg¹
Christian Küblbeck¹, Stefan Lautenbacher², Ute Schmid², Sven Friedl^{1,3}

¹Fraunhofer-Institut für Integrierte Schaltungen IIS, Erlangen,
²Otto-Friedrich-Universität Bamberg, ³Universitätsklinikum Erlangen
sven.friedl@iis.fraunhofer.de

Kurzfassung. In der Schmerzforschung werden schmerzrelevante Mimikbewegungen von Probanden mittels des Facial Action Coding System klassifiziert. Die manuelle Klassifikation hierbei ist aufwändig und eine automatische (Vor-)klassifikation könnte den diagnostischen Wert dieser Analysen erhöhen sowie den klinischen Workflow unterstützen. Der hier vorgestellte regelbasierte Ansatz ermöglicht eine automatische Klassifikation ohne große Trainingsmengen vorklassifizierter Daten. Das Verfahren erkennt und verfolgt Mimikbewegungen, unterstützt durch ein Gitter, und ordnet diese Bewegungen bestimmten Gesichtsarealen zu. Mit diesem Wissen kann aus den Bewegungen auf die zugehörigen Action Units geschlossen werden.

1 Einleitung

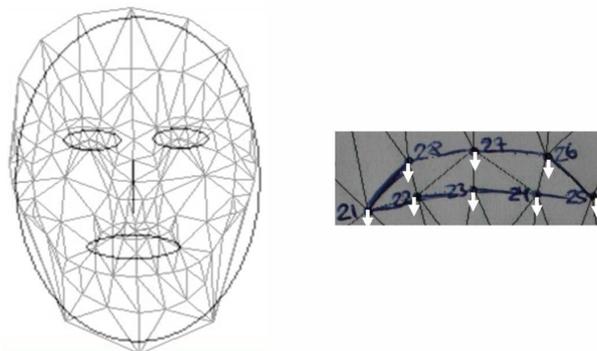
Menschliche Empfindungen wie Emotionen oder Schmerz lösen spezifische Muster von Kontraktionen der Gesichtsmuskulatur aus, die Grundlage dessen sind, was wir Mimik nennen. Aus der Beobachtung der Mimik kann wiederum auf menschliche Empfindungen rückgeschlossen werden. Im Rahmen der Schmerzforschung werden Videoaufnahmen von Probanden hinsichtlich des mimischen Schmerzausdrucks analysiert. Zur Beschreibung des mimischen Ausdrucks und dessen Veränderungen wird das Facial Action Coding System (FACS) [1] verwendet, das anatomisch begründet, kleinste sichtbare Muskelbewegungen im Gesicht beschreibt und als einzelne Action Units (AUs) kategorisiert. Eine Vielzahl von Untersuchungen hat gezeigt, dass spezifische Muster von Action Units auftreten, wenn Probanden Schmerzen angeben [2]. Die manuelle Klassifikation und Markierung der Action Units von Probanden in Videosequenzen bedarf einer langwierigen Beobachtung durch ausgebildete FACS-Coder. Eine automatische (Vor-)klassifikation kann hierbei den klinischen Workflow unterstützen und dieses Verfahren zum brauchbaren diagnostischen Instrument machen. Bisher realisierte Ansätze zum Erkennen von Gesichtsausdrücken basieren auf der Klassifikation

von trainierten Merkmalen, wie zum Beispiel Deformationsparametern [3] oder Grauwertverteilungen [4]. Neben der Notwendigkeit von großen Trainingsmengen für jede mögliche Ausprägung einer Action Unit, wird dem FACS-Coder auch der Zusammenhang zwischen der Klassifikation und der auslösenden Mimikänderung nicht deutlich erkennbar gemacht. Um Ausprägungen von Action Units ohne große Trainingsmengen klassifizieren zu können und eine nachvollziehbare Erklärungskomponente zu liefern, wird hier ein regelbasiertes Verfahren vorgestellt, das sich an den Beschreibungen des FACS orientiert.

2 Material und Methoden

Das FACS beschreibt, wie sich anatomische Positionsmerkmale im Gesicht, wie zum Beispiel Augenbrauen oder Mund, verändern müssen, damit eine bestimmte Action Unit gegeben ist. Zum Erlernen der Gesichtsausdrücke, die in Action Units kodiert werden, wird ein Regelsystem aufgebaut, welches den Beschreibungen der Positionsänderungen entspricht. Um die geforderten Positionsänderungen einer Action Unit zu erkennen, werden mimische Einzelbewegungen in Videosequenzen erkannt und über die Zeit verfolgt. Mit Hilfe einer Gitterstruktur, welche an das Gesicht angepasst wird, werden Bewegungen bestimmten Arealen im Gesicht zugeordnet (1). Die Gitterstruktur kann hierbei automatisch oder manuell an das Gesicht angepasst werden. Um Fehler aus der automatischen Anpassung zu verringern wurde in dieser Arbeit das Gitter anhand von 18 manuell gesetzten Landmarken geschätzt und angepasst. Insgesamt besteht das verwendete Gitter aus 151 Punkten. Treten mimische Einzelbewegungen auf, werden diese mittels Verfahren der Bewegungserkennung verfolgt. Ansätze hierzu sind Verfahren der Differentialbestimmung in Bildsequenzen wie der Optische Fluss [5] oder Verfolgung von Merkmalspunkten [6] über die Bildfolge. Die erkannten Bewegungen werden den anliegenden Gitterpunkten zugeordnet und das Gitter entsprechend neu angepasst. Da jedem der Gitterpunkte ein spezifisches Areal im Gesicht zugeordnet ist, kann aus der Gitterdeformation auf die aktivierte Muskelgruppe und dadurch auf die zugehörige Action Unit geschlossen werden.

Abb. 1. Skizze der an ein Gesicht angepassten Gitterstruktur (links) und Ausschnitt der Gitterstruktur an der rechten Augenbraue (rechts). Die Pfeile beschreiben die Bewegungsrichtung und -länge der Erscheinungsänderung *Augenbraue bewegt sich nach unten* (AU 4).



Beim Erstellen des Regelwerks werden Bewegungen an Gesichtsmerkmalen über die zugeordneten Gitterpunkte analysiert.

Das Regelwerk prüft für jede Erscheinungsänderung einer Action Unit folgende drei Aspekte. In Abb. 2 werden die ersten zwei Aspekte graphisch verdeutlicht.

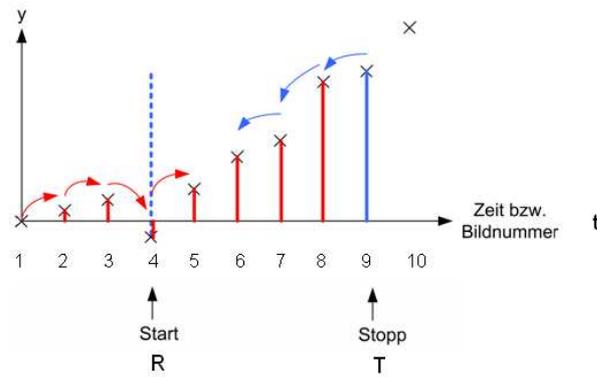


Abb. 2. Abfragepunkte des Regelwerks. Stopp- und Startmarke werden gesetzt, wenn die Bewegungslänge zwischen der Gitterposition im ersten und neunten Bild und zwischen neuntem und viertem Bild ausreichend groß ist.

1. *Bewegungsrichtung:* Die für die Entdeckung einer schmerzrelevanten mimischen Einzelbewegung betrachteten Gitterpunkte müssen sich in die durch die Action Unit beschriebene Richtung bewegen.
2. *Bewegungslänge:* Die Distanz der Bewegung des Gitterpunkts von der Ausgangsposition im ersten Bild der betrachteten Bildfolge zur Endposition in einem zeitlich späteren Bild der Bildfolge muss größer als ein entsprechend festgelegter Schwellwert θ_1 sein

$$\sum_{t=t_0}^T d_t > \theta_1$$

mit T : Bildnummer der Stopppmarke und d_t : Bewegungslänge eines Gitterpunkts zwischen Bild t und $t + 1$. Ist dies der Fall wird eine Stopppmarke T gesetzt.

3. *Bewegungsdauer:* Durch Rückwärtsprüfen der Bewegung von der Stopppmarke T bis zu einem zeitlich früheren Bild wird eine Startmarke R gesetzt. R wird zu dem Zeitpunkt oder Bild gesetzt, bei dem die Bewegungsdistanz im Intervall $I = T - R$ einen Schwellwert θ_2 überschreitet

$$\sum_{t=T}^R d_t > \theta_2$$

mit R : Bildnummer der Startmarke und $R \leq T$. Das Intervall I , das durch die Start- und Stopppmarke definiert wird, beschreibt die Dauer der betrachteten Bewegung.

3 Ergebnisse

Für diese Arbeit wurden exemplarisch vier schmerzrelevante Action Units implementiert. Dies sind die Action Units 4 (Brow Lowerer), 7 (Lid Tightener), 10 (Upper Lip Raiser) und 43 (Eyes Closed). Bisher werden nur einzeln auftretende Action Units und keine Kombinationen klassifiziert. Zur Evaluierung des Verfahrens wurde annotiertes Bildmaterial des Facial Action Coding System Manuals [1] und der Bilddatenbank von Kanade, Cohn und Tian [7] sowie eigens erstellte Bildsequenzen verwendet. Es hat sich gezeigt, dass die anatomischen Unterschiede zwischen einzelnen Personen die jeweiligen Bewegungen sehr unterschiedlich ausfallen lassen. Daher wurden die Schwellwerte der Bewegungsintensitäten für die verschiedenen Erscheinungsänderungen personenabhängig bestimmt. Da nur für die eigens erstellten Aufnahmen Informationen über die Bewegungsdauer vorhanden waren, wurde dieser Aspekt vernachlässigt. Der Schwellwert hierzu wurde so gewählt, dass das Intervall zwischen einem Bild und der maximalen Anzahl der vorhandenen Bilder der untersuchten Bildsequenz liegt. Mit dem vorhandenen Bildmaterial konnten in Experimenten 83 der 117 Testfälle erfolgreich klassifiziert werden. In den anderen Fällen wurde keine AU erkannt. Dies entspricht einer Klassifikationsrate von 70 %. Abbildung 3 zeigt ein Beispiel der Bewegungsverfolgung der Action Unit 4, welche erfolgreich klassifiziert werden konnte. Für die Berechnung des optischen Flusses wurde vorwiegend das Verfahren nach Horn und Schunck [5] verwendet. Hier konnten bei der experimentellen Untersuchung der Genauigkeit bessere Ergebnisse gegenüber dem Kanade-Lucas-Tomasi Feature Tracker [4] für Sequenzen mit Mimikbewegungen erzielt werden. Die wesentliche Fehlerquelle für eine nicht erfolgreiche Klassifikation lag in ungenauen Ergebnissen der Bewegungsanalyse. Die öffentlich verfügbaren Bilddaten [1, 7] sind nur bedingt für eine automatische Klassifikation mit diesem Ansatz geeignet. Eine sehr geringe Auflösung und verrauschte und teils unscharfe Bilder



Abb. 3. Eine Bildfolge aus [7] mit Mimikbewegungen der Action Unit 4. Erstes und letztes Bild der Action Unit sowie die erkannte Bewegung die zu einer erfolgreichen Klassifikation führte.

verhindern eine exakte Bewegungsverfolgung mit einer akzeptablen Genauigkeit. Bei Bewegungen von teilweise nur 5–10 Pixel im Bild führen bereits kleine rauschbedingte Fehler zu Distanzabweichungen von 50 % in der Bewegungserkennung. Konnte jedoch die Bewegung in den Bildsequenzen nicht genau genug erkannt und verfolgt werden, scheitert entsprechend auch das Regelsystem zur Klassifizierung. Mit dem hochwertigeren Bildmaterial konnten bessere Ergebnisse erzielt werden, jedoch standen nur wenige Aufnahmen zur Verfügung.

4 Diskussion

Es wurde ein regelbasierter Ansatz vorgestellt, der es ermöglicht Action Units des FACS zu klassifizieren ohne auf große Trainingsmengen vorklassifizierter Daten angewiesen zu sein. Zudem erlaubt dieser Ansatz den Zusammenhang zwischen kodierter Action Unit und zugrundeliegenden Änderungen von Positionsmerkmalen im Gesicht darzustellen. Beeinflusst wird das Ergebnis der Klassifikation von der Präzision der Anpassung der darunterliegenden Gitterstruktur sowie der Genauigkeit der Bewegungsverfolgung der Mimikbewegungen. Schwierigkeiten hierbei bereitet Bildmaterial, in dem das Gesicht nur sehr klein und unscharf abgebildet ist. In fortführenden Entwicklungen sollte das Regelsystem im Hinblick auf weitere Action Units und deren Erscheinungsänderungen erweitert werden. Zudem ist das gemeinsame Auftreten von mehreren Units die Regel und die Erkennung solcher Action Unit-Pattern diagnostisch besonders interessant, so dass auch hier ein weiterer Fokus gesetzt werden sollte. Eine Erweiterung des Regelsystems mittels eines Kantendetektors würde das Erkennen von Faltenbildung ermöglichen, was als Ergänzung im Regelwerk zu einer Verbesserung der Klassifizierung führen kann.

Literaturverzeichnis

1. Ekman P, Friesen WV, Hager JC. Facial Action Coding System The Manual [On CD ROM]. Nexus division of Network Information Research Corporation; 2002.
2. Kunz M, Mylius V, Schepelmann K, et al. On the relationship between verbal report and facial expression of pain. *J Pain*. 2004;5:368–76.
3. Black MJ, Yacoob Y. Recognizing facial expressions in image sequences using local parameterized models of image motion. *Int J Computer Vis*. 1997;25(1):23–48.
4. Lucey S, Ashraf AB, Cohn JF. Investigating spontaneous facial action recognition through AAM representations of the face. In: Delac K, Grgic M, editors. *Face Recognition*. Vienna, Austria: I-Tech Education & Publishing; 2007. p. 275–86.
5. Horn BKP, Schunck BG. Determining optical flow. *Artif Intell*. 1981;17:185–203.
6. Tomasi C, Kanade T. Detection and Tracking of Point Features. Carnegie Mellon University; 1991.
7. Kanade T, Cohn J, , et al. Comprehensive database for facial expression analysis. In: *Proc IEEE Int Conf Automatic Face Gesture Recogn*; 2000. p. 46–53.