

Optische Mimikanalyse

Roland Göcke

Jörg Voskamp

Fraunhofer Institut für Graphische Datenverarbeitung, Institutsteil Rostock
Abteilung Human-Centered Interaction & Technologies
J.-Jungius-Straße 11

Tel.: 0381/4024-131 / -120 Fax: 0381/4024199

Email: roland.goecke@rostock.igd.fhg.de joerg.voskamp@rostock.igd.fhg.de

Abstract. Die Ausgestaltung der Mensch-Technik-Interaktion gewinnt in heutiger Zeit immer mehr an Bedeutung. Computersysteme durchdringen mehr und mehr unseren Alltag, wobei für den Nutzer das Vorhandensein derartiger Systeme oft verborgen bleibt. Der Trend geht klar zu neuen Wegen in der Interaktion zwischen Menschen und Systemen, die natürlicher für den Menschen sind. Traditionelle Interaktionsgeräte wie Maus und Tastatur verlieren an Bedeutung und werden durch integrierte Systeme bestehend z.B. aus Spracherkennung und Gestenerkennung (Hand, Gesicht) ersetzt. Im Projekt face:)me wird ein Tool entwickelt, das die Analyse der Gesichtsmimik ermöglicht. Dazu werden bestimmte Merkmalspunkte wie die Augenecken, die Augenbrauen, die Nasenlöcher und der Mund in einer Bildfolge verfolgt und ihre Positionsveränderungen im so genannten *Face Mimic Graph* festgehalten für die spätere Analyse und Visualisierung. Zur Verfolgung der Merkmalspunkte werden Verfahren des *Template Matchings* und der Bildverarbeitung eingesetzt.

1. Motivation

Eines der am schnellsten wachsenden Gebiete in der Informatik ist das Gebiet der Mensch-Technik-Interaktion (MTI) mit seinen zahlreichen Anwendungsgebieten. Mit dem weitverbreiteten Einsatz von Informationstechnologie in vielen Gebieten und der immer stärker werdenden Durchdringung unseres täglichen Lebens mit neuen Technologien gewinnt die Frage, wie wir Menschen mit der Technik interagieren, an immer größerer Bedeutung. Zum Beispiel werden so genannte *Embedded Systems* in modernen Haushalts- und Home Entertainment-Geräten eingesetzt und finden sich ebenso in neuen Modellen der Autohersteller. Diese Systeme arbeiten oft für den Nutzer versteckt und werden kaum wahrgenommen. Sie reagieren auf das Verhalten des Nutzers in einer angebrachten Art und Weise, ohne dass der Nutzer gezwungen wird, eine spezielle Schnittstelle (*User Interface*) zu benutzen. Derartige MTI-Technologien werden auch unter den Begriffen *Ubiquitous Computing*, *Pervasive Computing* und *Ambient Intelligence* zusammengefasst.

Traditionelle Methoden der MTI, wie die Computermaus oder die Tastatur, sind für viele Anwendungen von *Embedded Systems*, z.B. in der Fahrzeugtechnologie, nicht ideal, da sie den Anwender zur Nutzung einer bestimmten Schnittstelle an einem bestimmten Ort zwingen. Dies ist häufig nicht praktikabel. Daher geht der Trend auf dem Gebiet der MTI eindeutig in Richtung natürlicher, dem Menschen nachempfunderer Interaktion, d.h. im Besonderen in Richtung von Sprachverarbeitung sowie der Erkennung bestimmter Gesten per Hand oder Kopf und von Gesichtsausdrücken.

In den vergangenen Jahren hat die automatische Beobachtung und Analyse von Gesichtsausdrücken ein starkes Forschungsinteresse erhalten, da es als eine der Schlüsseltechnologien der MTI angesehen wird. Unser Ziel ist es, neue Methoden zu entwickeln und anzuwenden, die Veränderungen im Gesichtsausdruck feststellen können, einen menschlichen oder automatischen „Beobachter“ davon in Kenntnis setzen können und eine entsprechende Reaktion des Systems auslösen können. Diese Reaktion muss in Abhängigkeit vom Verhalten des Nutzers und seines emotionalen Zustands gewählt werden. Z.B. können wir uns ein modernes Navigationssystem in einem Auto

vorstellen, das beruhigend auf den Fahrer einwirkt, wenn er oder sie gestresst ist, anstelle der derzeitigen monotonen Ansagen. Ganz allgemein gilt, dass eine derartige Technologie überall dort von besonderem Interesse ist, wo ein Beobachten des Nutzerverhaltens zu besser abgestimmten Systemreaktionen führen kann, z.B. auch in Nutzbarkeitsstudien und E-Learning-Systemen. Zum Teil kann der Gemütszustand des Anwenders vom Gesichtsausdruck abgeleitet werden oder vereinfacht gesagt, es ist möglich festzustellen, ob ein Nutzer mit einem System zufrieden ist oder nicht.

2. System-Überblick

Wir arbeiten derzeit an einem solchen System zur Mimikanalyse im Rahmen des Projektes face:)me (ehemals FAMIA). Das System findet ein Gesicht in einem Videostrom (live oder aufgezeichnet), bestimmt die Position von bestimmten Punkten im Gesicht (Augenbrauen, Augenecken, Mundwinkel usw.) und zeichnet diese Daten in einem so genannten *Face Mimic Graph (FMG)* auf für eine nachfolgende Analyse. Nun ist es ein leichtes sich vorzustellen, dass Gesichtsausdrücke stark variieren für verschiedene Menschen, ja sogar für ein und dieselbe Person zu verschiedenen Zeitpunkten. Deswegen ist es unser Ansatz, quantitative Messgrößen mit Hilfe der Positionen der Gesichtspunkte innerhalb des Gesichts und relativ zu einander zu ermitteln, anstatt ein System zu entwickeln, das nur einige wenige Gesichtsausdrücke feststellen kann. In dem die Positionsveränderungen von Punkten verfolgt werden, kann das System automatisch feststellen, wann eine Änderung des Gesichtsausdrucks und damit des Gemütszustandes möglicherweise aufgetreten ist, und eine entsprechende Handlung kann ausgelöst werden. Z. B. kann ein Testbeobachter in einer Nutzbarkeitsuntersuchung feststellen, zu welchem Zeitpunkt die Testperson nicht mit dem Testobjekt zufrieden war. Die im FMG aufgezeichneten Daten können nicht nur statistisch ausgewertet werden, sondern auch in einem Zeitdiagramm oder mit einem dreidimensionalen Kopfmodell visualisiert werden.

3. Das Finden des Gesichts

Zunächst muss festgestellt werden, ob im Kamerabild ein Gesicht vorhanden ist und wenn ja, so muss die Position ungefähr bestimmt werden. Hierfür sind verschiedene Wege möglich. In unserem z.Z. entwickelten Prototypen wird das Verfahren des *Template Matchings* angewandt [BRU93, MCK96, NEW00]. Hierfür wird ein Bereich von Interesse ausgewählt und dann mit den Grauwerten im nächsten Bild verglichen, um die neue Position des Punktes zu bestimmen. Dafür wird das Template über das Bild bewegt und die Korrelation an jeder Stelle berechnet. Wir setzen die normalisierte Kreuzkorrelation ein, da sie besser als andere Matching-Methoden (z.B. *sum of absolute differences (SAD)*) geeignet ist, unterschiedliche Beleuchtungsverhältnisse auszugleichen. Der Korrelationswert kann auch gleichzeitig als Maß für die Güte des Prozesses dienen.

Um ein Gesicht zu finden, setzen wir ein Template des gesamten Gesichts ein, jedoch mit einer 10fach niedrigeren Auflösung als im Original, um eine Bearbeitung in Echtzeit zu ermöglichen. Da sich die Gesichter verschiedener Personen in ihrem Aussehen deutlich voneinander unterscheiden, wird ein generisches Gesichtsmodell benutzt, das aus einem stark geglätteten Gesichtsbild abgeleitet wurde. Um die Echtzeitfähigkeit weiter zu erhöhen, wird der Korrelationswert für dieses Gesichtstemplate nicht an jeder Position im Bildkoordinatensystem berechnet, sondern nur an jeder zweiten auf jeder Koordinatenachse. Auf diese Weise wird die Anzahl an notwendigen Berechnungen um den Faktor 4 reduziert, ohne dass es zu erkennbaren Qualitätsverlusten bei der Bestimmung der ungefähren Position des Gesichts kommt.

Die Suche nach einem Gesicht und die Bestimmung der ungefähren Position dauert auf diese Weise weniger als 0,5s. Wenn die Position erst einmal ungefähr bekannt ist, reicht es aus, beim nächsten Mal in der Umgebung dieser Position nach der neuen Position zu suchen, da angenommen werden kann, dass keine sprunghaften Positionsänderungen in so einer kurzen Zeitspanne vorkommen. Dadurch lässt sich die Anzahl an Positionen, an denen die Korrelation berechnet werden muss, weiter deutlich reduzieren, so dass die neue Position in weniger als 0,1s gefunden wird. Unterschreitet der Korrelationswert jedoch ein bestimmtes Niveau, so wird der Suchbereich vergrößert, um sicherzustellen, dass die korrekte Position bestimmt wird.

4. Positionsbestimmung markanter Gesichtspunkte

Derzeit werden die Positionen der folgenden Punkte in einem Gesicht verfolgt (Bild 1):

- die Augenecken,

- die Augenbrauen mit Endpunkten und Mittelpunkten,
- die Nasenlöcher,
- die Lippen mit Mundwinkeln und Mittelpunkten.

Die Punkte, deren Aussehen relativ stabil ist (Augenecken, Nasenlöcher), werden ebenfalls über Template Matching gesucht. Für alle anderen Punkte werden Bildverarbeitungstechniken wie Kantendetektion und horizontale bzw. vertikale Projektion eingesetzt.

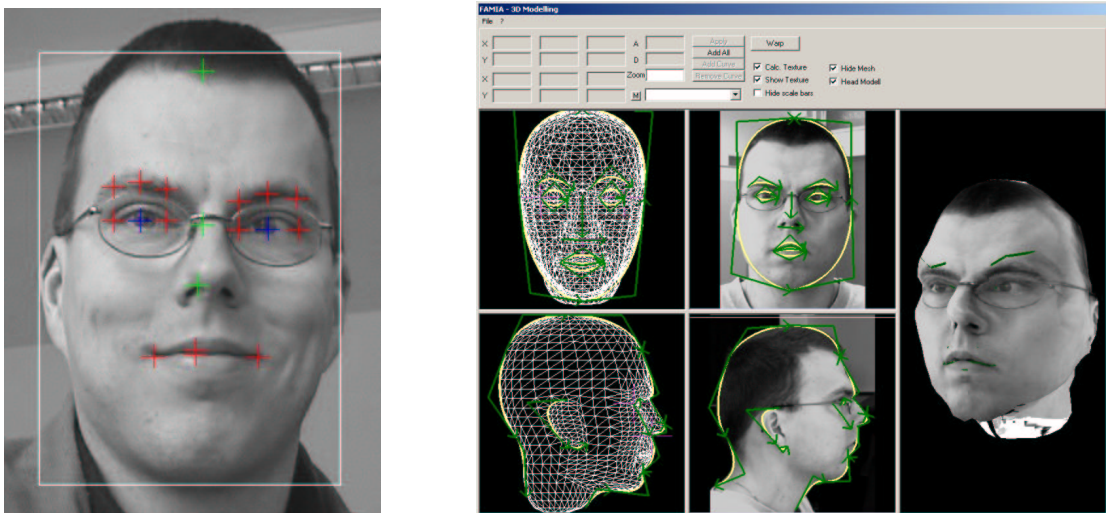


Bild 1: Links: Positionsbestimmung markanter Gesichtspunkte (rote Kreuze). Rechts: Visualisierungstool

Auf Grund der vorher bereits bestimmten ungefähren Position des Gesichts lassen sich die Suchbereiche für die einzelnen Punkte genauer abstimmen, was die Fehlerquote verringert und eine schnellere Bearbeitung ermöglicht. Zum Beispiel wissen wir, dass die Augen in der oberen Kopfhälfte zu finden sind. Es wird angenommen, dass der Kopf sich ungefähr in aufrechter Position befindet. Neigungen bis zu ca. 20 Grad werden toleriert. Für stärkere Kopfneigungen ist es denkbar, den Algorithmus so zu ändern, dass der Neigungswinkel berücksichtigt wird.

Als erstes wird die Position der Augen bestimmt, da sie ein deutlich sichtbares Feature sind. Die anderen Punkte werden danach relativ zu den Augenpositionen gesucht. Dadurch lässt sich eine Suche in nicht möglichen Gesichtsbereichen verhindern, z.B. braucht der Mund nicht oberhalb der Augen gesucht zu werden. Die Fehlerrate wird somit verringert und die Bearbeitung erfolgt schneller. Derzeit werden die Gesichtspunkte mit einer Framerate von 10Hz bestimmt.

5. Aufzeichnung der Daten und spätere Analyse

Die ermittelten Positionsdaten der markanten Gesichtspunkte werden in einem so genannten Face Mimic Graph (FMG) aufgezeichnet. Der FMG ist eine Erweiterung des bekannten *Facial Action Coding Systems (FACS)* [EKM78]. Während FACS Gesichtspunkte und Aktionen nur qualitativ festlegt, fügt der FMG die Möglichkeit hinzu, gemessene Größen aufzuzeichnen, d.h. um wieviel hat sich ein Punkt bewegt und in welcher Richtung.

Die so aufgezeichneten Daten können zu einem späteren Zeitpunkt beliebig oft und auf beliebige Weise analysiert werden. Derzeit wird an Visualisierungen gearbeitet, die die Analyse vereinfachen sollen. Dazu gehört eine Darstellung in einem Zeitdiagramm ebenso wie die Darstellung mit Hilfe eines dreidimensionalen Kopfmodells (Bild 1). Hierfür werden die gemessenen Positionsveränderungen der Gesichtspunkte aus dem FMG auf die entsprechenden Punkte des Kopfmodells abgebildet. Eine der Herausforderungen hierbei ist es, eine Bearbeitung in Echtzeit zu erreichen.

6. Zusammenfassung und Ausblick

Natürliche, auf den Menschen abgestimmte Interaktionstechniken werden in der Zukunft eine immer größere Rolle spielen. Wir entwickeln daher Methoden, um die Gesichtsmimik automatisch zu messen und anschließend zu analysieren. In der nächsten Entwicklungsstufe muss das Face Tracking robuster und adaptiver gestaltet werden. Außerdem ist geplant, die Visualisierung der aufgezeichneten Daten zu verbessern.

Literatur

- [BRU93] Brunelli,R.; Poggio,T.:
Face Recognition: Features versus Templates.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(10): 1042-1052, 1993
- [EKM78] Ekman,P.; Friesen,W.V.:
Facial action coding system: A technique for the measurement of facial movement.
Consulting Psychologists Press, Palo Alto, USA, 1978
- [MCK96] McKenna,S.; Gong,S.:
Tracking Faces.
Proceedings of the Second International Conference on Automatic Face and Gesture Recognition
FG'96, Killington, USA, S.271-276, 1996
- [NEW00] Newman,R.; Matsumoto,Y.; Rougeaux,S.; Zelinsky,A.:
Real-Time Stereo Tracking for Head Pose and Gaze Estimation.
Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture
Recognition FG'2000, Grenoble, France, S.122-128, 2000