

# Benutzeradaptive videobasierte Erfassung der Mimik als Interface für motorisch eingeschränkte Personen

Ulrich Canzler und Markus Minklai

Lehrstuhl für Technische Informatik  
Rheinisch-Westfälische Technische Hochschule (RWTH), 52074 Aachen  
Email: {canzler, minklai}@techninfo.rwth-aachen.de

**Zusammenfassung.** Im Folgenden wird ein System vorgestellt, welches in der Lage ist, vollautomatisch auf Basis einer Frontalansicht in einem mehrstufigen Prozess ein benutzeradaptives Gesichtsgraphenmodell zu erzeugen. Hierbei wird die mit Hilfe vom *Active Appearance Models* gefundene Gesichtsstruktur in ein 3D VR-Kopf umgewandelt, wobei sowohl Geometrie des AAM-Shapes als auch Textur an das virtuelle 3D-Kopfmodell angepasst werden. Auf anschließend synthetisierten Daten basierend wird ein neues, individuell dem Benutzer angepasstes Model trainiert und zur Merkmalsanalyse eingesetzt.

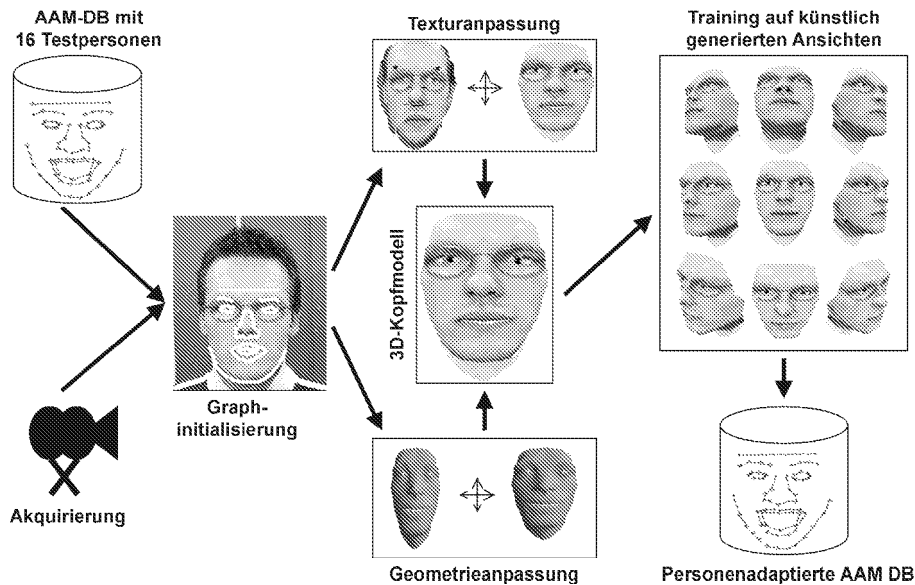
## 1 Einleitung

Durch die stark gestiegenen Rechenleistungen innerhalb der letzten zehn Jahre ergeben sich neue Einsatzbereiche für innovative Mensch-Maschine Schnittstellen.

Um behinderten Menschen ein selbstbestimmtes und aktives Leben zu ermöglichen, eröffnen sich durch die Auswertung mimischer Merkmale wie Kopfpose oder Mundform neue Möglichkeiten; dies gilt natürlich in besonderem Maße für motorisch stark eingeschränkte Menschen im Kontext interaktiver Steuerungen von Dialogsystemen.

Will man Alltagstauglichkeit, Praxisnähe und die Akzeptanz eines Systems gewährleisten, so ist von besonderer Bedeutung, dass Benutzerunabhängigkeit gewährleistet werden kann, da sich das äußere Erscheinungsbild eines Menschen innerhalb kurzer Zeiträume stark verändern kann. Eine Unabhängigkeit von Störeinflüssen wie Brillen, Bart und Frisur ist unabdingbar. Zudem muss ein System umgebungsunabhängig, d.h. weitgehend invariant gegenüber äußeren Lichteinflüssen sein. Damit ein System unmittelbar auf Eingabeinteraktionen reagieren kann und nicht für den Benutzer irritierende Totzeiten durch rechenintensive Analyse der Mimik entstehen, ist es erforderlich, dass ein System in Echtzeit mit dem Benutzer interagiert, denn nur so ist eine befriedigende Rückkopplung auf Eingaben zu gewährleisten.

Abb. 1. Modellaufbau



## 2 Systemaufbau

Im Folgenden wird die automatische benutzeradaptive Anpassung von Gesichtsglyphen beschrieben, wobei besonders auf das Erstellen eines 3D-Kopfmodells des Benutzers und auf die Gesichtsfindung eingegangen wird, die weitestgehend auf den *Active Appearance Models (AAM)* von Stegman [1] beruht.

**Akquirierung.** Eine Kamera erfasst den Gesichtsraum des Benutzers. Die akquirierte Bildsequenz wird mit Methoden der Bildverbesserung bearbeitet, um Störungseinflüsse der Umgebung wie Lichteinfall und Reflexionen zu minimalisieren. Hierzu kommen *Grayworld Constancy* und *Shadowreduction* zum Einsatz.

**Graphinitialisierung.** Nun erfolgt eine Anpassung eines aus 70 Punkten bestehenden *Active Appearance Models*, welches zuvor mit 16 Testpersonen trainiert worden ist. Das Modell vereint Geometrie- und Texturinformationen und verwendet die *Principal Component Analysis (PCA)*, so dass eine erhebliche Reduktion der Datenmenge erreicht wird. Bei der eigentlichen Graphen-anpassung handelt es sich um ein leicht zu lösendes Optimierungsproblem. Die geometrischen Informationen werden in einem sog. *Shape* festgehalten, das durch die Gleichung

$$\mathbf{k} = \bar{\mathbf{k}}_d + \Phi_k \cdot \mathbf{p}_k \quad (1)$$

beschrieben werden kann. Hierbei ist  $\mathbf{k}$  eine Shape-Instanz,  $\bar{\mathbf{k}}$  das Durchschnitts-Shape,  $\mathbf{p}_k$  die Shape-Modellparameter und  $\Phi_k$  eine Matrix ist, in der spaltenweise die Eigenvektoren der Merkmalsvektoren des Ausgangsdatenmaterials, die bei der PCA entstanden sind, stehen. Für die Texturinformationen im AAM gibt es

die analoge Gleichung:

$$\mathbf{t} = \bar{\mathbf{t}}_d + \Phi_t \cdot \mathbf{p}_t, \quad (2)$$

wobei  $\bar{\mathbf{t}}_d$  die Durchschnittstextur,  $\Phi_t$  die analoge Matrix von Eigenvektoren und  $\mathbf{p}_t$  die Modellparameter der Textur sind. Um nun beide Informationsquellen zu einem Modell zu vereinen wird eine dritte PCA durchgeführt und eine Gewichtungsmatrix  $W_k$  berechnet, um einen gemeinsamen Kontrollparameter  $\mathbf{c}_e$  zu erhalten, mit dem das gesamte Aussehen des Gesichtsgraphen kontrolliert werden kann.

$$\mathbf{p}_e = \begin{pmatrix} W_k \cdot \mathbf{p}_k \\ \mathbf{p}_t \end{pmatrix} = \Phi_e \cdot \mathbf{c}_e \quad (3)$$

Die Konstellation der aufgefundenen 70 Landmarks ermöglicht anschließend die Analyse der Mimik. Da der gefundene Graph mit Hilfe einer Datenbank ermittelt wird, in der sich der Benutzer selbst nicht befindet, kommt es meist zu einer veräuschten Anpassung des Graphen. Abhilfe hierfür schafft das in 2.3 vorgestellte Verfahren der Modellerstellung durch synthetische Ansichten.

**Das 3D-Kopfmodell.** Im folgenden wird die Erstellung eines 3D-Modells beschrieben, welches Geometrie und Textur des Benutzers übernimmt.

*Anpassung der Geometrie.* Nachdem einige einfache geometrische Verschiebungen und Skalierungen berechnet wurden, um das 3D-Kopfmodell an das AAM-Modell anzupassen, werden die inneren (d.h. die Punkte, die keinen entsprechenden Punkt im AAM-Modell besitzen) Merkmalspunkte manipuliert. Dabei sollen die Merkmalspunkte im 3D-Kopfmodell nach der Transformation an genau der Stelle liegen die sie auch im AAM-Modell, also auch in der Textur, haben. Alle anderen Punkte werden, nach der Entfernung zu den Merkmalspunkten gewichtet und in der x-y-Ebene verschoben. Über die Tiefe der Punkte liegt keine Informationen vor, so dass die z-Koordinaten unverändert bleiben. Für jeden Punkt im 3D-Kopfmodell gilt folgende Gleichung:

$$f(\vec{P}) = \vec{P} + \sum R_i(\vec{P}) \vec{\alpha}_i, \quad (4)$$

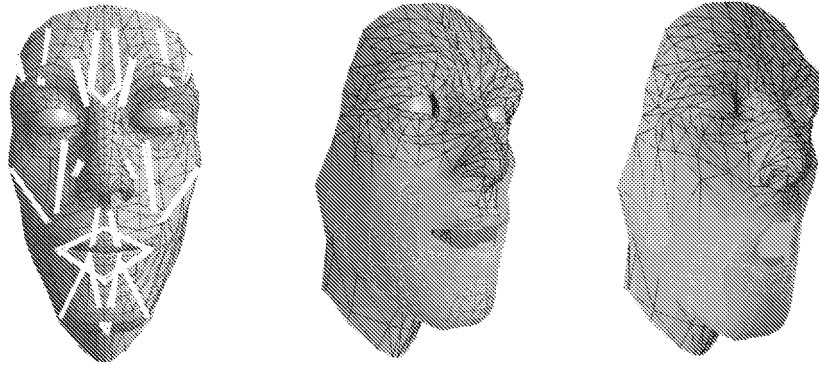
wobei

$$R_i(\vec{P}) = \frac{\left( \min_{j \neq i} |\vec{P} - \vec{P}_j| \right)^2}{|\vec{P} - \vec{P}_i|^2 + \left( \min_{j \neq i} |\vec{P} - \vec{P}_j| \right)^2} \quad (5)$$

Der Punkt  $P$  wird mit einer Gewichtung, die von der Entfernung zum jeweiligen Merkmalspunkt  $\vec{P}_i$  abhängig ist, verschoben. Setzt man in diese Gleichung die Merkmalspunkte  $\vec{P}_i$ , für welche die Endpositionen  $\vec{Q}_i$  aus dem AAM-Modell bereits bekannt sind ein, kann daraus ein Gleichungssystem abgeleitet werden, welches nach den  $\vec{\alpha}_i$  aufzulösen ist.

$$\vec{Q}_j - \vec{P}_j = \sum_{i=1}^n R_i(\vec{P}_j) \vec{\alpha}_i \quad \forall j \in \{1, 2, \dots, n\} \quad (6)$$

**Abb. 2.** 3D-Kopfmodelle mit unterschiedlichen Mundformen und Beleuchtungen



Nachdem das Gleichungssystem gelöst wurde, kann Gleichung (4) benutzt werden um alle Punkte des 3D-Kopfmodells an das AAM-Modell anzupassen.

*Anpassung der Textur.* Da die Dimension der Textur nicht genau der Größe des 3D-Kopfmodells entspricht wird ein Skalierungsfaktor benötigt, um die Größe der Textur anzupassen. Um die Texturmerkmalspunkte genau auf die 3D-Kopfmodell-Merkmalpunkte zu positionieren wird des weiteren einen Verschiebungsvektor berechnet. Die Berechnung der Skalierung für die x-Achse reduziert sich auf das Lösen der Gleichung

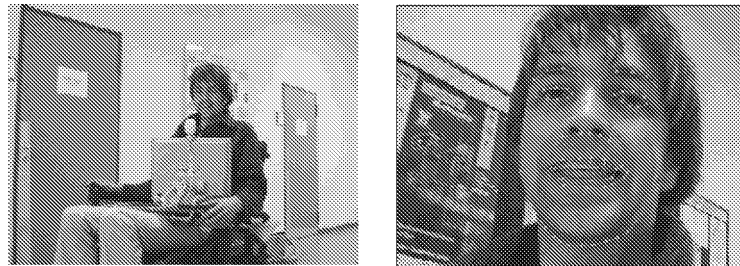
$$S_x = \frac{K_{x,\max} - K_{x,\min}}{KM_{x,\max} - KM_{x,\min}} \cdot \frac{TM_{x,\max} - TM_{x,\min}}{T_{x,\max} - T_{x,\min}}, \quad (7)$$

wobei  $K$  für das Kopfmodell,  $KM$  für die Merkmalspunkte des Kopfmodells,  $T$  für die Textur und  $TM$  für die Merkmalspunkte der Textur stehen cite549-2.

Genauso wichtig wie die Berechnung der Skalierung ist die Berechnung der Position der Textur im Kopfmodell. Dies geschieht für die x-Achse über die unten angeführte Gleichung.

$$V_x = \frac{K_{x,\min} - KM_{x,\min}}{K_{x,\max} - K_{x,\min}} \cdot \frac{TM_{x,\max} - TM_{x,\min}}{T_{x,\max} - T_{x,\min}} + \frac{TM_{x,\min} - T_{x,\min}}{T_{x,\max} - T_{x,\min}}. \quad (8)$$

**Training auf synthetisierten Ansichten.** Mit dem so generierten 3D-Kopfmodell des Benutzers werden verschiedene Posen unter verschiedenen Beleuchtungssituationen synthetisiert. Das verwendete Kopfmodell hat biomechanische Eigenschaften durch ein zugrunde liegendes anatomisch korrektes Muskelmodell, welches es ermöglicht, verschiedene Mundbilder und ganzheitliche Gesichtsmimiken zu erzeugen (s. Abb. 2). Die so künstlich generierte Variation wird dazu genutzt, ein neues benutzeradaptives Gesichtsgraphenmodell zu erzeugen, welches bzgl. Geometrie und Textur individuell an den Benutzer angepaßt ist.

**Abb. 3.** a: Gesamtansicht, b: Kameraansicht

### 3 Anwendungsszenario

Zum Testen des Systems wurde ein Szenario entwickelt, bei dem der Benutzer als Rollstuhlfahrer mit gewissen Grenzen im Blickfeld der Kamera fixiert ist. Es zeigte sich gerade bei Benutzern mit problematischen äußeren Merkmalen wie beispielsweise Tragen von Bärten und Brillen, dass dabei eine qualitativ deutliche Steigerung der Graphenanpassung erzielt werden konnte.

Die gefundenen Merkmale werden anschließend als Steuersignale an eine zweite Recheneinheit weitergegeben, welche situationsangepasst reagiert und so die eigentliche Interaktion ermöglicht, wie z.B. eine mimikgesteuerte Menüführung.

Als ein anderes Anwendungsfeld wäre die permanente Analyse des Blickes oder des Lidschlages von Personen denkbar. So können epileptische Anfälle oder Bewusstlosigkeit frühzeitig erkannt und Hilfe angefordert werden.

### 4 Zusammenfassung und Ausblick

In dieser Arbeit wurde ein System vorgestellt, welches in der Lage ist, vollautomatisch ein benutzeradaptiertes Modell aus einer einzigen Frontalansicht des Benutzers zu erzeugen. Die Lernphase ist für den Benutzer auf ein Minimum reduziert, da Seitenansichten und Variationen der Mimik durch ein biomechanisches Modell synthetisiert und anschließend trainiert werden.

Der wesentliche Vorteil der Verwendung eines benutzeradaptiven Systemaufbaus liegt in der starken Verbesserung der eigentlichen Gesichtsgraphenanpassung im späteren Interaktionsmodus des Systems, da dieser individuell auf den Benutzer abgestimmt wird. Merkmale wie Lidschlag, Mundbild oder Bewegung der Augenbrauen, die sehr stöempfindlich gegenüber persönlichen Eigenschaften wie das Tragen eines Bartes oder einer Brille sind, werden deutlich robuster verfolgt als es mit einem allgemein trainierten Systemaufbau je möglich ist.

### Literaturverzeichnis

1. Stegmann, M.B.: Active Appearance Models – Theory, Extensions & Cases, Technical University of Denmark DTU, Lyngby 2000.
2. Demann, T.: 3D-Rekonstruktion von Objekten aus Fotos an Beispielen von Gesichtern, 2000