

# Nicht-intrusive Erkennung isolierter Gesten und Gebärden

Dr.-Ing. Suat Akyol

Lehrstuhl für Technische Informatik  
Rheinisch-Westfälische Technische Hochschule Aachen (RWTH)  
Ahornstr. 55, D-52074 Aachen  
mail@akyol.de

**Abstract:** Die computergestützte Erkennung von Gesten gilt als wichtige Technologie zur Gestaltung innovativer Mensch-Maschine Schnittstellen und zur Entwicklung automatischer Gebärdenübersetzungshilfen für Gehörlose. Diese Arbeit stellt ein personenabhängig funktionierendes Gestenerkennungssystem vor, das erstmals ein großes Vokabular mit hoher Zuverlässigkeit erkennen kann. In Zahlen bedeutet dies, dass 152 verschiedene Gebärden in 97,6 % aller Experimente korrekt identifiziert werden. Dazu wird die gestikulierende Person benutzerfreundlich mittels einer Videokamera erfasst und das aufgezeichnete Bild in einem PC bezüglich der Hände des Benutzers analysiert. Anschließend werden die extrahierten Informationen über ein statistisches Klassifikationsverfahren einer Geste aus einem zuvor definierten Vokabular zugeordnet.

## 1 Einleitung

Aufgrund der steigenden Komplexität technischer Geräte werden innovative Konzepte zur Mensch-Maschine Interaktion benötigt. Derzeit gibt es zunehmende Bemühungen intuitiv bedienbare Schnittstellen durch die Erkennung und Berücksichtigung der Gestik des Benutzers zu schaffen [PSH97]. Gestenerkennung kommen z.B. in Fahrzeugen zur Bedienung von Bordsystemen zum Einsatz [ALK01]. Eine weitere Anwendungsmöglichkeit bieten die visuell-gestischen Gebärdensprachen gehörloser Menschen [ACB00]. Die Erkennung dieser eröffnet die Möglichkeit Gehörlose bei der alltäglichen Kommunikation mit Hörenden zu unterstützen.

Die Informationsübermittlung erfolgt bei der Gestik primär über die Hände, genauer gesagt über deren Form, Position, Stellung und Bewegung. Diese als manuelle Parameter bezeichneten Merkmale gilt es zunächst durch geeignete Sensoren zu erfassen. Unter dem Aspekt der Benutzerfreundlichkeit und der Alltagstauglichkeit wird dem Vorbild der menschlichen visuellen Informationsaufnahme entsprechend die Verwendung einer Videokamera zur Aufzeichnung des Benutzers favorisiert. Aus den Bildfolgen müssen die Hände mit einem angeschlossenen Rechnersystem herausgefiltert und quantitativ beschrieben werden. Anschließend sind die extrahierten Informationen einer Geste aus einem definierten Gestenvokabular zuzuordnen.

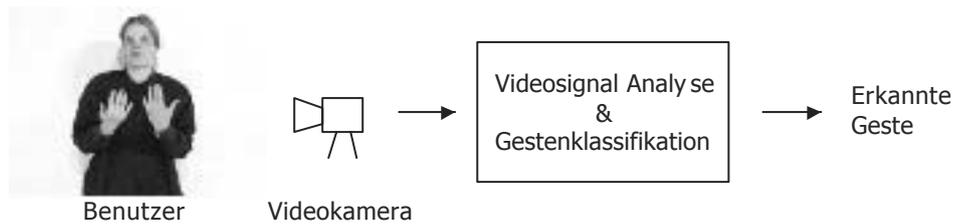


Abbildung 1: Prinzip der videobasierten Gestenerkennung

## 2 Die Problematik

Das Forschungsgebiet, dass sich mit der computergestützten Analyse von Bilddaten Zwecks Erkennung der Inhalte auseinandersetzt, wird im Allgemeinen als „Computer Vision“ (Computersehen) bezeichnet. Das Ziel ist das menschliche Sehen und Bildverstehen künstlich nachzubilden. Allerdings ist dieses Ziel noch in weiter Ferne, da die Prinzipien des menschlichen Sehens bisher nicht hinlänglich bekannt sind. Daher überwiegen pragmatische Ansätze bei der Lösung von Computer Vision Problemen.

Obgleich die Erkennung von erlernten Gesten dem Menschen trivial erscheint, ist dies in einem Computer algorithmisch nicht einfach umzusetzen. Bereits die Festlegung von Kriterien zur eindeutigen Lokalisierung der Hände in einem Bild erweist sich als schwierig, denn die Hände ändern ihre Position ständig, sind einander ähnlich und wechseln häufig die Form. Lediglich der Farbton der Haut gilt als weitestgehend konstant und daher als geeignete Eigenschaft zur Lokalisierung der Hände [JR98]. Bei Betrachtung der beobachtbaren hautfarbenen Regionen in den Beispielen von Abbildung 2 lässt sich die damit verbundene Problematik erahnen.

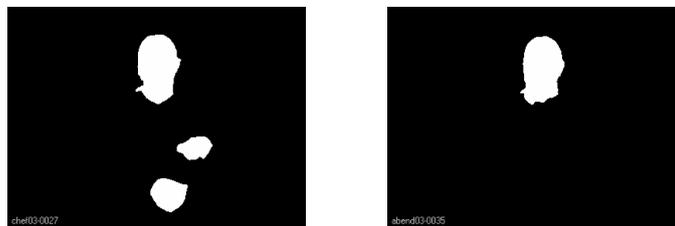


Abbildung 2: Hautfarbe als stabile Eigenschaft der Hände. Die hellen Regionen repräsentieren hautfarbene Bereiche, die dem Gesicht und den Händen entsprechen

Im linken Bild ist zwar zu erkennen welche Regionen dem Gesicht bzw. den Händen entsprechen, jedoch kann die linke oder rechte Hand nicht eindeutig zugeordnet werden. Im rechten Bild ist schwer zu sagen, ob eine, zwei oder eventuell gar keine Hand im Bild ist. Im Fall von Überlappungen der Hände miteinander oder mit dem Gesicht ist eine exakte Lokalisierung wegen der Farbgleichheit ohne weiteres kaum möglich.

Zusätzliche Schwierigkeiten bereitet die Tatsache, dass bei der wiederholten Ausführung einer Geste stets Variationen der manuellen Parameter und der Ausführungsdauer auftreten. Diese müssen jedoch als die gleiche Geste identifiziert werden können, da eine hohe Wiederholungsgenauigkeit vom Benutzer nicht zu erwarten ist. Zudem gilt, dass die Fehlerwahrscheinlichkeit bei der Erkennung in der Regel überproportional mit der Anzahl der zu unterscheidenden Gesten steigt.

### 3 Stand der Technik

Angesichts der beschriebenen Schwierigkeiten bedienen sich Wissenschaftler bezüglich der Bildanalyse bislang vereinfachender Hilfsmittel oder sie wählen einfach zu verarbeitende Gesten aus. So legen Grobel [Gro99] und Hienz [Hie00] dem Benutzer farbig markierte Handschuhe an, um dessen Hände aufgrund der eindeutigen Farben vereinfacht im Bild lokalisieren zu können. Starner et al. [SWP98] und auch Yang et al. [YAT02] hingegen verzichten auf solche Hilfsmittel, beschränken sich aber auf die Erkennung einer viel kleineren und sorgfältig ausgewählten Gestenmenge, um den nötigen Aufwand für die Bildanalyse in überschaubaren Grenzen halten zu können.

Zur Klassifikation kommen vorwiegend statistische Verfahren zum Einsatz, da diese es erlauben, die Ausführungsvielfalt einer Geste zu berücksichtigen. Dabei wird für jede zu erkennende Geste vorab ein Modell angelegt, welches im Erkennungsvorgang gewisse Variationen der beobachteten Informationen toleriert. Als Ergebnis stehen pro Modell Wahrscheinlichkeiten statt harter Pro- oder Contra Entscheidungen zur Verfügung.

Die Tabelle 1 ist eine Übersicht der Leistungsfähigkeit bisheriger Gestenerkennungssysteme. Die Vokabulargröße gibt die Anzahl der modellierten, d.h. dem System bekannten, Gesten an. Die Erkennungsrate steht für den Anteil der Experimente, bei denen die vom System erkannte Geste der tatsächlich beobachteten Geste entspricht.

Tabelle 1 Leistungsfähigkeit bisheriger Gebärdenerkennungssysteme

	<b>Grobel [Gro99] Hienz [Hie00]</b>	<b>Starner et al. [SWP98]</b>	<b>Yang et al. [YAT02]</b>
<b>Hilfsmittel</b>	Farbig markierte Handschuhe	Keine	Keine
<b>Vokabulargröße</b>	43 bzw. 262 Gesten	40 Gesten	40 Gesten
<b>Erkennungsrate</b>	97,2 bzw. 93,3 %	91,9 %	98,1 %

Die Tabelle zeigt, dass für ca. 40 Gesten eine Erkennungsrate von höchstens 98,1 % erreicht wird. Unter Verwendung von Hilfsmitteln lässt sich die Vokabulargröße steigern, die Erkennungsrate nimmt dann jedoch kontinuierlich ab. Für höchste Benutzerfreundlichkeit und Alltagstauglichkeit ist ein großes Vokabular bei hoher Erkennungsrate ohne Hilfsmittel wünschenswert.

#### 4 Zielsetzung und Konzept

Das Ziel dieser Arbeit ist eine benutzerfreundliche, d.h. Hilfsmittelfreie, Lösung für die Erkennung von Gesten zu entwickeln und die Leistungsfähigkeit des Ansatzes durch den Vergleich mit den existierenden Arbeiten auf diesem Gebiet zu verifizieren. Die Erfüllung dieses Zieles wird durch eine zuverlässige Methode für die Lokalisierung der Hände im Bild sichergestellt. Das Grundprinzip ist die Verfolgung (*engl.* Tracking) der Hände unter kontinuierlicher kumulativer Bewertung der beobachteten Informationen. Das zu diesem Zweck entworfene Verarbeitungskonzept ist in Abbildung 3 zu sehen.

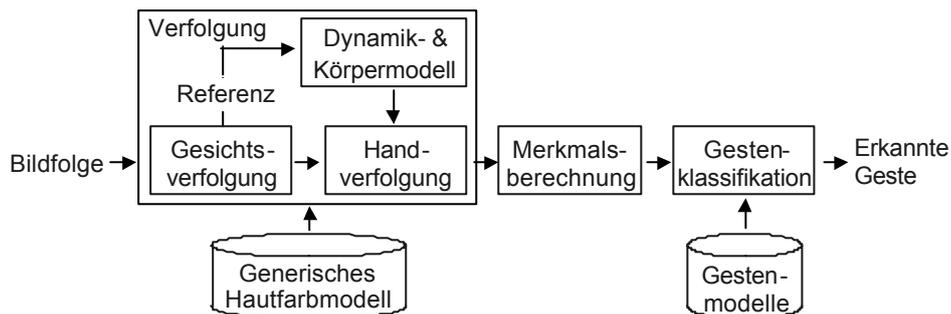


Abbildung 3: Verarbeitungskonzept für die Gestenerkennung

Die aufgezeichnete Bildsequenz durchläuft als erstes eine Verfolgungsstufe. Durch die Anwendung eines generischen Hautfarbmodells werden zunächst alle sichtbaren Hautpartien des Benutzers hervorgehoben [AA01]. Darunter wird erst die Gesichtsposition und mit dieser als Referenz die Position jeder Hand bestimmt. Dieser Reihenfolge liegt die Tatsache zugrunde, dass der Kopf weniger und langsamere Bewegungen ausführt als die Hände und daher zuverlässiger verfolgt werden kann [AZ02]. Für die Verfolgung der schnelleren Handbewegungen mit ihren abrupten Richtungsänderungen kommen Regeln in Form von Körper- und Dynamikmodellen zum Einsatz, um das Wissen über die zu erwartende Haltung und den Bewegungsablauf einzubringen.

Im Anschluss an die Verfolgungsstufe steht die Merkmalsberechnung, deren Ergebnis eine Reihe von quantitativen Maßzahlen zur Beschreibung der Handeigenschaften, genauer gesagt der manuellen Parameter (Vgl. Abschnitt 1), ist. Die Merkmale sind die Grundlage für die Klassifikation der eingehenden Informationen als eine der bekannten Gesten. Die Gestenmodelle werden vorab ebenfalls auf der Grundlage der berechneten Merkmale erstellt. Sie bilden das erkennbare Gestenvokabular.

Im Folgenden wird die Handverfolgungsstufe als signifikante Neuerung auf dem Gebiet der Gestenerkennung beschrieben. Ebenso wird die Gestenerkennungsstufe in ihrer Funktionsweise und die damit erzielte Erkennungsleistung näher erläutert.

## 5 Verfolgung von Gesicht und Händen

Es wird davon ausgegangen, dass der Benutzer mittig im Bild steht und in einer bequemen Ruhehaltung, mit den Händen am Körper herab hängend, eine Geste beginnt und beendet. Das Gesicht ist somit anfangs unverdeckt und eindeutig identifizierbar, so dass es im Verlauf als Bezugsobjekt für die Lokalisierung der Hände dienen kann. Da das Gesicht vergleichsweise groß und träge ist, lässt es sich unabhängig von den Händen durch einen Mean-Shift Tracker verfolgen. Selbst im Falle von Überlappungen durch die Hände ist die grobe Gesichtspolition damit noch zuverlässig bestimmbar. Ein zusätzliches statistisches Formmodell des Kopfes verleiht weitere Stabilität und ermöglicht die exakte Positionslokalisierung [AZ02].

Ausgehend von der Gesichtspolition werden alle Zuordnungsmöglichkeiten zwischen den Händen und allen beobachteten hautfarbenen Bildregionen gebildet. Es entstehen multiple Zuordnungshypothesen, zu denen auch einige Sonderfälle gehören, wie zum Beispiel die Möglichkeit, dass die Hände gar nicht im Bild zu sehen sind. Wenn eine Hypothese anzeigt, dass die Abbilder der Hände sich im Bild miteinander oder mit dem Abbild des Gesichts überlappen, so bewirkt ein spezielles Verfahren eine gezielte Trennung zur eindeutigen Festlegung der möglichen Handpolitionen [Ak00]. Alle zeitlich aufeinander folgenden Hypothesen werden vollständig miteinander verknüpft. Dadurch entsteht ein Lösungsraum von Zuständen und Übergängen, in dem jede Zustandsfolge einen anderen Bewegungsablauf beschreibt (Abbildung 4).

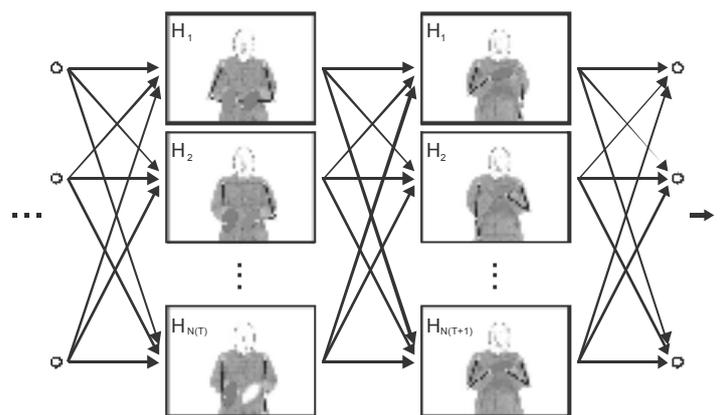


Abbildung 4: Lösungsraum der Zuordnungshypothesen und Übergänge

Zur Bestimmung der besten aller Zustandsfolgen ist es notwendig, sowohl die Hypothesen als auch die Übergänge zu bewerten. Darauf basierend lässt sich mit dem Verfahren der Dynamischen Programmierung zur Laufzeit eine eindeutige Lösung effizient finden. Diese ist jedoch nicht endgültig, so dass neue Informationen jederzeit Änderungen bei allen vorigen Teillösungen bewirken können. Bei Verfügbarkeit neuer Erkenntnisse werden somit zuvor getroffene Entscheidungen nachträglich korrigiert, sofern sie in Anbetracht der Bewertungskriterien falsch waren.

Die Bewertung eines Zustands im Lösungsraum erfolgt unter Zuhilfenahme eines einfachen Körpermodells mit heuristisch ermittelten Längenverhältnissen. Anhand der Gesichtsbreite wird das Modell skaliert und auf die hypothetischen Gesichts- und Handpositionen gelegt. Sodann kommen Bewertungsfunktionen zum Einsatz, um damit die Plausibilität sowie die Erwartung an die beobachtete Szene zu prüfen. Abbildung 5 zeigt die Kriterien und die Anwendung des Modells auf das aufgezeichnete Bild.

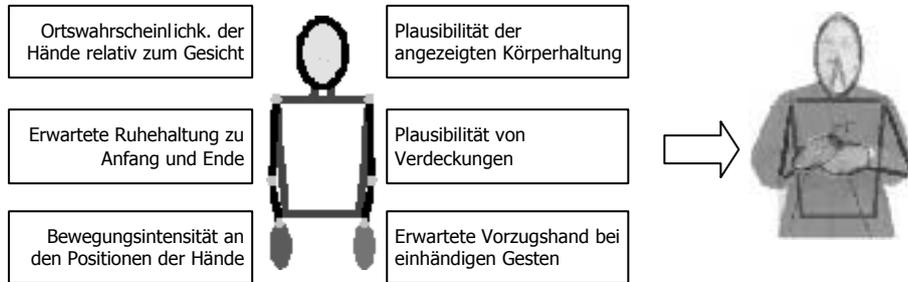


Abbildung 5: Anwendung des Körpermodells auf eine Beobachtung

Zur Bewertung der Übergänge werden die Hypothesen, wie in Abbildung 6 dargestellt, bezüglich der vermuteten Handpositionen mit Vorhersagen verglichen. Eine Vorhersage entspricht dabei einer extrapolierten Schätzung aus allen vorigen Hypothesen. Unter der vereinfachten Annahme linear beschleunigter Bewegungen ist das Schätzproblem für die zu erwartenden nächsten Handpositionen optimal durch einen Kalman Filter lösbar. Dieses Verfahren hat sein Hauptanwendungsgebiet in der Radartechnik zur Überwachung der Bewegungen im Luftverkehr, konnte sich aber auch bei der Analyse menschlicher Bewegungsabläufe mehrfach bewähren [AC99].

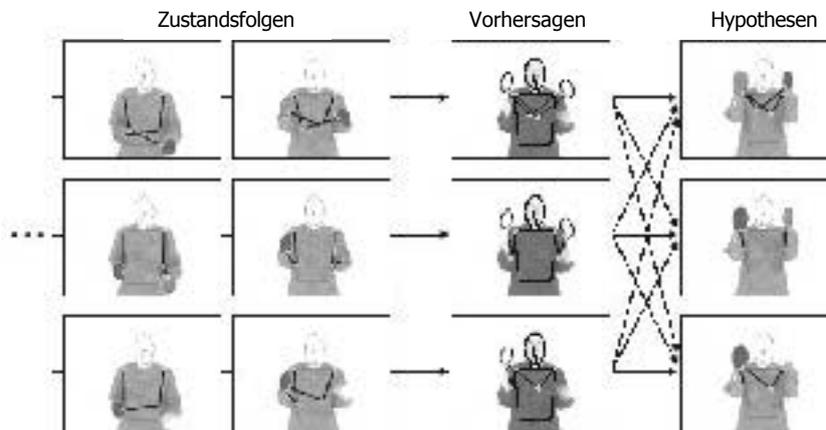


Abbildung 6: Anwendung des Dynamikmodells auf eine Zustandsfolge

Abbildung 7 zeigt ein Beispiel für die Fehlerkorrektureigenschaft des beschriebenen Verfahrens, wodurch die Berichtigung vergangener Fehlentscheidungen bewirkt wird.

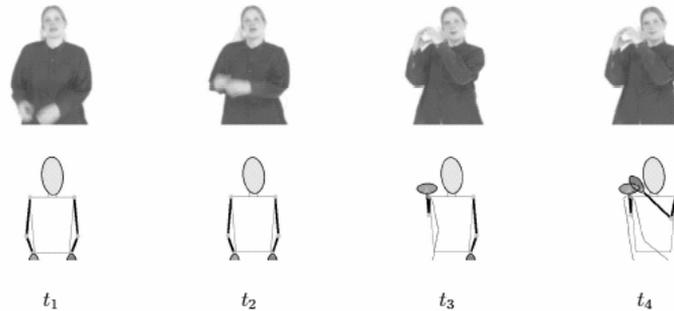


Abbildung 7: Fehlerkorrektur zur Laufzeit bei der Handpositionsbestimmung

Eine zunächst als einhändig detektierte Geste wird in diesem Fall korrigiert. Die zwischenzeitlich eingenommenen Handpositionen werden nachträglich ergänzt, wie an den eingezeichneten Bewegungsspuren zu sehen ist.

## 6 Merkmalsberechnung und Klassifikation

Im Anschluss an die Verfolgungsstufe müssen für beide Hände Merkmale berechnet werden, um die manuellen Parameter zu beschreiben. Gesten, im Gegensatz zu Handzeichen, besitzen in der Regel eine hohe Dynamik der Grundbewegung. Die dabei ausgeführten Formänderungen der Hand sind für einen Menschen wegen der Trägheit seines Sehapparats nicht im Detail bezüglich der einzelnen Finger zu erfassen. Daher gilt im Allgemeinen, dass eine Ganzheitliche Darstellung der Hand zur Unterscheidung der meisten Gesten genügt. Die folgenden einfachen Merkmale konnten in der vorliegenden Arbeit als ausreichende Beschreibung einer Hand identifiziert werden:

- Die Positionskordinaten relativ zum Gesicht
- Die Fläche relativ zum Gesicht
- Das Verhältnis von Länge zu Breite
- Der Orientierungswinkel

Die Unterscheidung komplexer Handzeichen würde die Verfügbarkeit zusätzlicher Formmerkmale voraussetzen, ist aber im Kontext dieser Arbeit nicht relevant. Hier werden lediglich die oben genannten Merkmale jeweils für die linke und rechte Hand berechnet und an die Klassifikationsstufe weitergegeben.

Zur Darstellung von Geste eignen sich Hidden Markov Modelle (HMM) sehr gut, die zum Beispiel auch in der Spracherkennung zur Modellierung von Wörtern verwendet werden [RJ93]. Bei HMM handelt es sich um stochastische endliche Automaten mit Zuständen und Übergängen. Die Zustände enthalten statistische Repräsentationen der Merkmale, die für eine Reihe von Wiederholungen der modellierten Geste beobachtet werden konnten. Die Übergänge beschreiben den zeitlichen Ablauf der Geste und gestatten diesbezüglich festgelegte Variationen. Abbildung 8 zeigt die Zuordnung der einzelnen Phasen einer Geste zu den Zuständen eines Hidden Markov Modells.

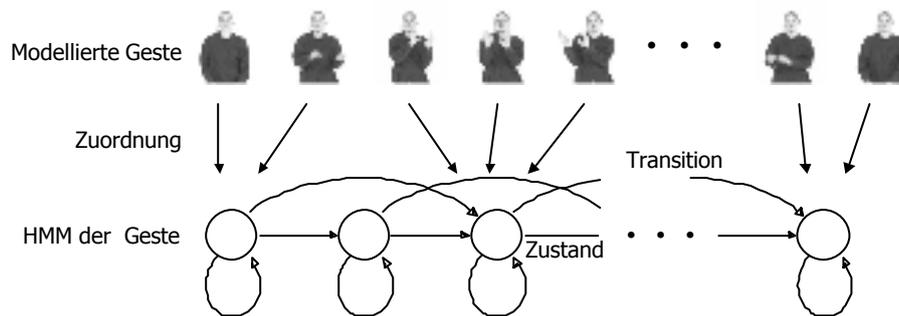


Abbildung 8: Modellierung einer Geste mit einem Hidden Markov Modell (HMM)

Eine Bank von Hidden Markov Modellen (Abbildung 9) bildet den Klassifikator. Jedes Modell wird vorab mit einer Beispielmenge derjenigen Geste parametrisiert, die sie repräsentiert. Die Bewertung einer beobachteten Folge von Merkmalen erfolgt mit allen Modellen. Das ähnlichste Modell entspricht der erkannten Geste, wobei die Ähnlichkeit im statistischen Sinne bezüglich Dauer und Ausführung einer Geste zu verstehen ist.

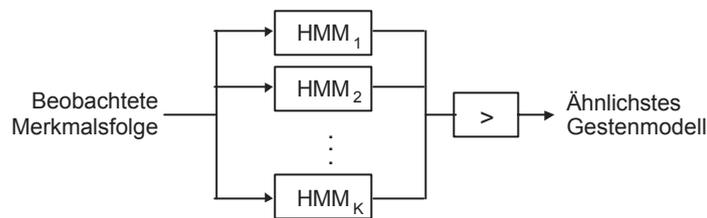


Abbildung 9: Klassifikationsprinzip mit Hidden Markov Modellen (HMM)

## 7 Experimente und Ergebnisse

Für die experimentelle Auswertung der Systemleistung wird hier ein Vokabular von 152 allgemein gebräuchlichen Gesten der Deutschen Gebärdensprache verwendet. Dadurch ist ein Anwendungsbezug gewährleistet. Jede einzelne Geste hat eine durchschnittliche Dauer von zwei Sekunden und liegt in zehn wiederholten Aufzeichnungen mit 25 Bildern pro Sekunde vor. Die Bildauflösung von  $384 \times 288$  Punkten entspricht dabei in etwa der Qualität der Kamera in einem Mobiltelefon. Die Datenbasis mit über 70.000 Einzelbilder enthält zudem zu Auswertungszwecken die manuell (d.h. von einem menschlichen Beobachter) notierten Referenzen der Gesichts- und Handpositionen. Die Ergebnisse des Systems lassen sich damit direkt vergleichen und objektiv auswerten.

Die Gesamtleistung eines Gestenerkennungssystems wird üblicherweise durch die Erkennungsrate angegeben, die es für eine definierte repräsentative Testmenge erzielt. Das Ergebnis hängt dabei von allen Verarbeitungsstufen ab, da sich in einem sequentiellen Verarbeitungsprozess Fehler stets fortpflanzen. Es ist daher sinnvoll auch die Leistung der Handverfolgungsstufe als wichtigste Komponente zu betrachten.

Die Experimente ergeben für die Handverfolgung eine Trefferrate von 97,1 % bei sehr strengen Auswertungskriterien. Es sind drei Arten von Fehlern zu beobachten. In 1,1 % aller Bilder weicht die geschätzte Position einer Hand von der tatsächlichen Position zu stark ab. In 1,5 % aller Bilder wird eine Hand fälschlicherweise dem Abbild der anderen Hand zugewiesen. In nur 0,3 % der Fälle ist eine unzutreffende Detektion einer nicht im Bild befindlichen Hand zu verzeichnen. Die Fehler rate der Handverfolgung ist insgesamt als sehr niedrig einzustufen. Eine allzu starke Beeinträchtigung der nachfolgenden Klassifikationsstufe ist nicht zu erwarten, da bei keiner Geste eine lange andauernde durchgehende Fehlerfolge auftritt.

Zur Auswertung der Klassifikationsstufe werden sogenannte leaving-one-out Tests durchgeführt. Die Erstellung der Gestenmodelle erfolgt dabei jeweils mit neun der zehn Wiederholungen einer Geste. Die nicht zur Erstellung des Modells verwendete Ausführung wird für ein einzelnes Erkennungsexperiment herangezogen. Anschließend wird zyklisch gewechselt, so dass pro Geste 10 Experimente möglich sind.

Von den insgesamt 1520 Erkennungsexperimenten liefern 97,6 % das richtige Ergebnis. Im direkten Vergleich mit den in Abschnitt 3 erwähnten Arbeiten ist dieses Resultat als signifikante Verbesserung zu betrachten, zumal trotz des Verzichts auf Hilfsmittel mit 152 Gesten ein vergleichsweise großes Vokabular verarbeitet wird.

## 8 Zusammenfassung und Ausblick

Das in dieser Arbeit erzielte Resultat für die Erkennung von über 150 Gesten aus Videoaufzeichnungen stellt das weltweit bislang beste Ergebnis dar. Dies wird vor allem durch eine zuverlässige und robuste Methodik zur Verfolgung der Hände als Hauptinformationsträger der Gestik sichergestellt. Für die Entwicklung und Optimierung der Verfolgungsstufe ist vor allem das im Rahmen dieser Arbeit angefertigte und in seiner Art einzigartige Datenmaterial mit den in Hunderten von Arbeitsstunden erzeugten Referenzangaben von unschätzbarem Wert.

Obgleich die Experimente benutzerabhängig getestet wurden, ist eine Erweiterung auf den benutzerunabhängigen Fall analog zu der Entwicklung in der Spracherkennung möglich. Gleiches gilt für die kontinuierliche Verarbeitung, bei der es gilt, die Gesten in ganzen Satzgebilden zu erkennen und dabei die sogenannten Koartikulationseffekte an den Gestenübergängen zu berücksichtigen.

Für die Weiterentwicklung der Gestenerkennung ist besondere Aufmerksamkeit auf die Vergrößerung des Vokabulars auf Tausende von Gesten zu legen. Nur dadurch lässt sich gewährleisten, dass die Vokabelvielfalt von Gebärdensprachen vollständig abgedeckt wird. Dies setzt die Extraktion zusätzlicher Informationen aus dem Bild voraus. So könnten zum einen die Handdetails unter Verwendung von Handmodellen und rekonstruktiver Analyse bestimmt werden. Andererseits ist es denkbar die Mimik des Benutzers auszuwerten und im Erkennungsprozess zu nutzen. Die Erkenntnisse dieser Arbeit bieten eine gute Basis für die Anwendung weiterführender Analyseverfahren.

## Literaturverzeichnis

- [AA01] Akyol, S.; Alvarado, P.: Finding Relevant Image Content for Mobile Sign Language Recognition. In (Hamza, M.H. Hrsg.): IASTED Intl. Conf. Signal Processing, Pattern Recognition and Applications (SPPRA), Rhodes, Greece, 2001; S. 48-52.
- [AC99] Aggarwal, J.K.; Cai Q.: Human Motion Analysis: A Review. *Computer Vision and Image Understanding*, 73 (3), 1999; S. 428-440.
- [ACB00] Akyol, S.; Canzler, U.; Bauer, B.: Gesture and Mimic Interpretation for Sign Language Recognition. In: Proceedings of the 4th International Student Conference on Electrical Engineering (Poster 2000), Prague, Czech Republic, 2000; S. IC1.
- [Ak00] Akyol, S.: Nicht-intrusive Erkennung isolierter Gesten und Gebärden. Doktorarbeit, Technische Informatik RWTH Aachen, 2003.
- [ALK01] Akyol, S.; Libuda, L.; Kraiss, K.-F.: Multimodale Benutzung adaptiver Kfz-Bordsysteme. In (Jürgensohn, T.; Timpe, K.-P., Hrsg.): *Kraftfahrzeugführung*, Springer, 2001; S. 137-154.
- [AZ02] Akyol, S.; Zieren, J.: Evaluation of ASM Head Tracker for Robustness against Occlusion. In (Arabnia, H.R.; Mun Y., Hrsg.): *Proc. Int. Conference on Imaging Science, Systems, and Technology (CISST)*, Band I, Las Vegas, 2002; S. 28-34.
- [Gro99] Grobel, K.: Videobasierte Gebärdenspracherkennung mit Hidden-Markov-Modellen. Doktorarbeit, Technische Informatik RWTH Aachen. *Fortschritts-Berichte VDI 10/592*. VDI Verlag, Düsseldorf, 1999.
- [Hie00] Hienz, H.: Erkennung kontinuierlicher Gebärdensprache mit Ganzwortmodellen. Doktorarbeit, Technische Informatik RWTH Aachen. Shaker Verlag, Aachen, 2000.
- [JR98] Jones, M.J.; Rehg, J.M.: Statistical Color Models with Application to Skin Detection. *Technischer Bericht CRL 98/11*, Compaq Cambridge Research Lab, Dezember 1998.
- [PSH97] Pavlovic, V.I.; Sharma, R.; Huang, T.: Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19 (7), 1997; S. 677-695.
- [RJ93] Rabiner, L.; Juang, B.-H.: *Fundamentals of speech recognition*. Prentice Hall, 1993.
- [SWP98] Starner, T.; Weaver, J.; Pentland, A.: Real-time American sign language recognition using desk and wearable computer based video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20 (12), 1998; S. 1371-1375.
- [YAT02] Yang, M.-H.; Ahuja, N.; Tabb, M.: Extraction of 2D Motion Trajectories and its Application to Hand Gesture Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24 (8), 2002; S. 1061-1074.

## Dr.-Ing. Suat Akyol

Der Autor dieses Beitrags, Dr.-Ing. Suat Akyol, wurde am 9. März 1973 in Mönchengladbach als Sohn türkischer Gastarbeiter geboren. Nach der Erlangung der allgemeinen Hochschulreife im Jahr 1992 studierte er Elektrotechnik an der RWTH Aachen und erhielt zum Abschluss den akademischen Grad eines Dipl.-Ing. Im Jahr 1998 nahm er eine Stelle als Wissenschaftlicher Angestellter am Lehrstuhl für Technische Informatik der RWTH an. Dort war er auf den Gebieten der Mensch-Maschine Kommunikation, der Bildverarbeitung und der Mustererkennung tätig und fertigte in diesem Kontext eine Dissertation an. Am 11. Juli 2003 legte er seine Doktorprüfung ab und erhielt dafür eine Auszeichnung.