
Zusammenfassung der Diplomarbeit
“Appearance-Based Gesture Recognition”
von
Dipl.-Inform. Philippe Dreuw
26.04.2005

Betreuer: Dipl.-Inform. Daniel Keysers
Gutachter: Prof. Dr.-Ing. Hermann Ney
Prof. Dr. Thomas Seidl
Datum der Abgabe: 11. Januar 2005
erteilte Note: sehr gut

Einleitung

In der Diplomarbeit wurden erscheinungsbasierte Bildmerkmale zur Erkennung von Gesten in Video untersucht. Bisherige Arbeiten im Bereich der Gestenerkennung benötigten oft aufwendige Vorverarbeitungsschritte und Merkmale. So musste z.B. erst die Hand aus dem Video segmentiert werden, bzw. mussten Personen farbige Handschuhe oder Daten-Handschuhe tragen.

Gute Ergebnisse im Bereich der Objekterkennung und Handschrifterkennung suggerieren jedoch, dass dieser fehlerbehaftete Vorverarbeitungsprozess nicht nötig ist und stattdessen das ganze Bild ohne jegliche Segmentierung verwendet werden kann. In der Diplomarbeit konnte gezeigt werden, dass

- Segmentierung von Händen oder tragen von farbigen Handschuhen nicht nötig ist, und man mit Hilfe erscheinungsbasierter Bildmerkmale exzellente Resultate in der Videoanalyse von Gesten erzielen kann
- angepasste Distanzmaße, die die Variabilität in Bildern berücksichtigen, herkömmlichen Distanzmaßen überlegen sind und zur Gestenerkennung benutzt werden können
- Hidden Markov Modelle in Kombination mit solchen Distanzmaßen in der Gestenerkennung auf völlig unterschiedlichen Datenbanken sehr gute Fehlerraten erzielen und anderen Ansätzen überlegen sind

Zu diesem Zweck wurde eine neue Video-Datenbank mit Gesten des deutschen Fingeralphabet aufgenommen. Weiterhin wurde im Rahmen dieser Diplomarbeit ein neues Tracking Verfahren entwickelt. Mittels dynamischer Programmierung wurde eine Verfolgung von Objekten, z.B. der Hand, ermöglicht. Theoretische Grundlagen zur Kombination von Tracking und Erkennung in einem Hidden Markov Modell wurden erstmals in der Diplomarbeit vorgestellt. In [Dreuw & Keysers⁺ 05] wurde bereits ein Teil der Erkenntnisse aus der Diplomarbeit publiziert.

Die Gebärdensprache ist die natürliche Sprache Gehörloser. Ein Hauptgrund zur Forschung im Bereich der Gesten- und Gebärdenspracherkennung ist die Entwicklung von Anwendungen für Gehörlose, die sehr wichtig und hilfreich für deren Alltag sind. Ideal wäre ein Übersetzungssystem, das die Kommunikation mit Gehörlosen ermöglichen würde. Man stelle sich ein System für Blinde und Gehörlose vor ...

Ein weiterer Grund ist die Entwicklung von neuen Eingabegeräten im Bereich der Mensch-Maschine-Kommunikation. Ein Hauptvorteil der Benutzung von Gesten ist die Möglichkeit der Kommunikation, ohne physikalisch mit der Maschine in Kontakt treten zu müssen. Der Ersatz von Tastaturen, Joysticks und Mäusen durch gestengesteuerte Eingabegeräte ermöglicht eine körperlich uneingeschränkte Steuerung durch den Benutzer, z.B. durch die von einer Kamera aufgezeichneten Zeige-Gesten. Der Einsatz solcher Kommunikationsgeräte erfreut sich immer größerer Beliebtheit, vor allen Dingen im Bereich Spiele- oder der Automobilindustrie: die Steuerung eines Navigationssystems durch Gesten erscheint leichter, als während der Fahrt nach Bedienknöpfen zu suchen.

Aufbau eines Systems zur Gestenerkennung

Bei den meisten Ansätzen zur Gestenerkennung wird versucht, das Bild in seine Bestandteile zu zerlegen (z.B. Hände, Arme, Kopf), um diese dann als Merkmale zu benutzen. In einem erscheinungsbasierten Ansatz zur Gestenerkennung wird hingegen das Originalbild oder beliebige Transformationen desselben (z.B. Verzerrungen, Kanten-Filterungen, Unterausschnitte, ...) als Merkmal benutzt. Dadurch kann eine eventuell fehlerbehaftete Segmentierung der Bilder vermieden werden.

Bildmerkmale. Die Untersuchung von erscheinungsbasierten Bildmerkmalen zur Erkennung von Gesten war ein wesentlicher Bestandteil der Diplomarbeit. Bei der Arbeit mit Grauwertbildern (z.B. Infrarotbilder wie in Abb. 1(a)) kann z.B. ein Originalbild als Merkmal benutzt werden. Anhand von solchen Originalbild-Merkmalen oder deren räumlichen Ableitungen (z.B. Sobel gefilterte Bilder) wurden bereits hervorragende Ergebnisse erzielt, die denen anderer Ansätze deutlich überlegen sind.

Die Berechnung von Differenzbildern (siehe Abb. 1(c)) stellt eine der einfachsten Methoden zur Bewegungserkennung in Bildsequenzen dar. Bewegung ist ein wichtiges erscheinungsbasiertes Merkmal in Bildsequenzen, die den Zusammenhang zwischen lokalen Eigenschaften und zeitlicher Variation repräsentiert.

Bewegungsbilder werden dazu benutzt, um darzustellen, *wie* sich etwas in einem Bild bewegt im Gegensatz zu *wo*. Das Resultat entspricht einem Bild, in dem sich kürzlich bewegende Pixel heller erscheinen als jene, die sich seit längerem oder garnicht bewegt haben. Diese Bilder können dann mit gespeicherten Darstellungen bekannter Bewegungen verglichen werden. Abb. 1(d)-(e) zeigt ein Bild einer Bildsequenz mit seinen entsprechenden Bewegungsbild.

Hautfarbbilder werden anhand ihrer Hautfarbwahrscheinlichkeiten erstellt und bilden ein wichtiges Merkmal bei der Erkennung von Gesten und Gebärden. Abb. 1(f)-(h) zeigt

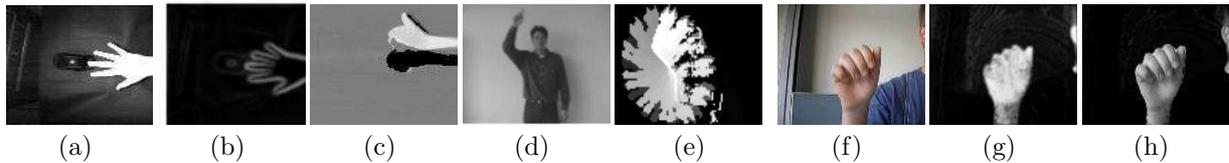


Abb. 1. Erscheinungsbasierte Bildmerkmale: (a) **Originalbild** sowie (b) nach Anwendung eines Sobel-Filters, (c) **Differenzbilder** erster Ordnung. **Bewegungsbilder** als Bildmerkmale: (d) das Originalbild einer Bildsequenz mit (e) seinem entsprechenden Bewegungsbild. **Hautfarbbilder** als Bildmerkmale: (f) Originalbild, (g) Hautfarbwahrscheinlichkeiten als Bild und (h) Originalbild nach Anwendung der Hautfarbwahrscheinlichkeit als Schwellwert.

einige mögliche Bildmerkmale, die anhand von Hautfarbwahrscheinlichkeiten erstellt wurden.

Hidden Markov Modelle. Ein weiterer wesentlicher Bestandteil war die Untersuchung von Hidden Markov Modellen (HMM), die ein leistungsfähiges Werkzeug zur statistischen Modellierung zeitlicher Signale darstellen. Sie ermöglichen die zeitliche Anpassung einer Beobachtungssequenz an eine Referenzsequenz und haben ihre Qualitäten im Bereich der Spracherkennung, der Gesten- und Gebärdenspracherkennung und der Analyse menschlicher Bewegungen bereits ausgiebig unter Beweis gestellt. Die Ergebnisse in der Diplomarbeit zeigen erstmals, dass die Erkenntnisse aus der Spracherkennung nicht immer mit denen aus der Gestenerkennung übereinstimmen. Sie stellen wichtige Erkenntnisse für die weitere Erkennung von Gesten und Gebärden mit Hidden Markov Modellen dar.

Distanzmaße. In der Bildverarbeitung werden Bild-Distanzmaße, die die Variabilität in Bildern berücksichtigen, bereits erfolgreich zur Klassifikation von Bildern und in der Handschrifterkennung eingesetzt.

Neu auf dem Gebiet der Gestenerkennung war die erstmalige Untersuchung der Kombination von Hidden Markov Modellen mit Bild-Distanzmaßen zur Erkennung von Gesten. In der Diplomarbeit wurde gezeigt, dass durch die Verwendung von Tangentendistanz oder Image Distortion Modellen, die invariant gegenüber kleinen affinen Transformationen sind, die Fehlerrate auf einem Test-Korpus mit 14 dynamischen Gesten und 140 Testsequenzen von 5.7% auf 1.4% gesenkt werden kann. Ebenfalls waren diese Distanzmaße herkömmlichen Distanzmaßen auf allen anderen getesteten Datenbanken der Gestenerkennung überlegen.

Tracking. Wenn detaillierte Information über sich bewegende Objekte in Videos verlangt wird, benötigt man Methoden, um diese Objekte zu verfolgen oder voneinander unterscheiden zu können. Bei der Betrachtung bekannter Tracking Methoden entstand die Idee, mittels dynamischer Programmierung ein neuartiges Tracking Verfahren. Es basiert auf der gleichen Methodik, wie sie in der Spracherkennung zur zeitlichen Anpassung von Beobachtungsfolgen benutzt wird. Hierbei erfolgen die Tracking Entscheidungen erst am einer Beobachtungssequenz. Der Vergleich mit einem bekannten Standardverfahren zeigt die Überlegenheit des neuen Verfahrens beim Tracking in stark verrauschten Bildsequenzen. Weiterhin wurden theoretische Grundlagen entwickelt, um Tracking und Erkennung in einem Hidden Markov Modell zu kombinieren.

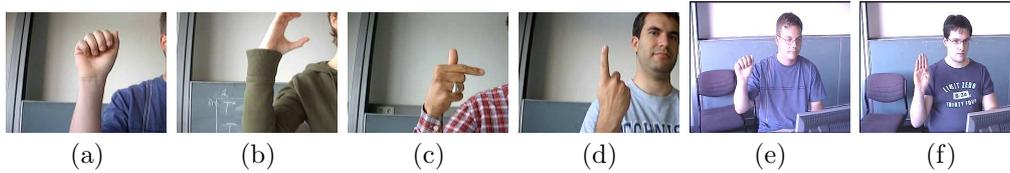


Abb. 2. Einige Beispiel des DGS Fingeralphabetes aus der neuen i6-Gesture Datenbank : (a)-(d) aufgezeichnet mit einer Webcam mit (a) "A", (b) "C", (c) "T", and (d) "1". (e)-(f) wurden mit einem Camcorder aufgenommen mit (e) "A" and (f) "B".

Datenbanken und Experimente

Die Auswertung der untersuchten Bildmerkmale sowie der Kombination von Hidden Markov Modellen mit neuen Bild-Distanzmaßen konnte auf drei völlig unterschiedlichen Datenbanken (gestengesteuertes Auto-Navigationssystem (siehe Abb. 1(a)), Roboter Steuerung (siehe Abb. 1(d)) und Fingeralphabet der DGS¹ (siehe Abb. 2) konkurrenzfähige Ergebnisse oder Verbesserungen gegenüber herkömmlichen Verfahren erzielen. Eine neue Datenbank² mit DGS Gebärden wurde speziell zu diesem Zweck im Rahmen der Diplomarbeit aufgezeichnet und ist nun für weitere Forschungszwecke auf dem Gebiet der Gebärdensprache frei verfügbar.

Zusammenfassung und Schlussfolgerungen

In der Diplomarbeit wurde gezeigt, dass einfache erscheinungsbasierte Merkmale in Kombination mit neuen Bild-Distanzmaßen zur Erkennung von Gesten benutzt werden können. Der Gebrauch von Tangentendistanz und Image Distortion Modellen als angepasste Distanzmaße zur Modellierung von Bildvariabilitäten in Kombination mit Hidden Markov Modellen wurde untersucht. Auf völlig unterschiedlichen Datenbanken konnten Verbesserungen im Vergleich zu herkömmlichen Distanzmaßen erzielt werden. Der Systemaufbau wurde ebenfalls bereits erfolgreich zur Erkennung von Gebärdensprache getestet.

Diese innovative Kombination wurde in dieser Form bisher noch nie untersucht und zeigt die Möglichkeiten dieses Ansatzes. Gesten oder Gebärden können auch ohne fehlerbehaftete Segmentierung oder das Tragen von farbigen Handschuhen erkannt werden, um gestengesteuerte Anwendungen zu entwickeln und Gehörlosen Menschen den Alltag zu erleichtern.

Literatur

[Dreuw & Keysers⁺ 05] P. Dreuw, D. Keysers, T. Deselaers, H. Ney: Gesture Recognition Using Image Comparison Methods. In *International Gesture Workshop 2005*, Lecture Notes in Computer Science, in press, Île-de-Berder, France, May 2005.

¹Deutsche Gebärdensprache

²<http://www-i6.informatik.rwth-aachen.de/~dreuw/database.html>