Wissensbasierte Gesichtsmodellierung und Kamerasteuerung zur Analyse von Patientengesichtern

A. Gebhard[1], U. Ahlrichs, D. Paulus

Lehrstuhl für Mustererkennung (LME, Informatik 5)
Martensstr. 3, Universität Erlangen–Nürnberg, 91058 Erlangen
Tel.: +49 (9131) 85–27824 — Fax: +49 (9131) 303811
gebhard@informatik.uni-erlangen.de
http://www5.informatik.uni-erlangen.de/Persons/ge

Übersicht In diesem Beitrag wird ein wissensbasiertes Bildanalysesystem zur Diagnoseunterstützung und Rehabilitation von Gesichtslähmungen vorgestellt. Die grundlegende Struktur für die Wissensbasis ist ein semantisches Netz, von dem die zentralen Komponenten beschrieben werden. Für das vorliegende System sind speziell die aktiven Komponenten der Bildanalyse von Bedeutung, beispielsweise Kameraaktionen. Die Kontrolle der Wissensbasis dient zur Auswahl der Kameraparameter und Aktionen ebenso wie zur Objekterkennung.

Keywords: Wissensbasierte Bildanalyse, aktives Sehen, Gesichtslähmungen

1 Motivation und medizinische Problematik

Die Fazialisparese (Gesichtslähmung) ist die häufigste isoliert auftretende Lähmung überhaupt. Allein in der Klinik und Poliklinik für Hals–, Nasen– und Ohrenkranke (HNO) der Universität Erlangen–Nürnberg werden pro Jahr über 100 Patienten mit neu auftretenden Gesichtsnervenlähmungen beobachtet. Die Diagnose dieser Lähmungen erfolgt aufgrund von Beobachtungen, die ein Arzt vom Patienten gewinnt, während dieser spezifische mimische Bewegungen ausführt. Aus diesen Beobachtungen wird mittels zweier in der Medizin etablierten Indizierungssysteme [6, 14] eine Diagnose erstellt. Ein Teil der Diagnose wird daher aufgrund subjektiver Bewertungen gestellt. Der Wunsch nach einer objektiven Diagnoseunterstützung von Seiten der Ärzte erfordert die Detektion und genaue Lokalisation von markanten Punkten im Gesicht. Durch automatische Analyse der Gesichtsmerkmale Stirn, Augen und Mund soll aufgrund von Maßzahlen eine Diagnoseunterstützung realisiert werden, die dem behandelnden Arzt ein objektives Werkzeug zuf Verfügung stellt.

Die Grundlagen dieses System werden in Abs. 2 vorgestellt und in einen allgemeinen Rahmen eingeordnet. Wissensbasis und Kontrolle werden in Abs. 3 beschrieben. Die Funktionsfähigkeit des Systems wird in Abs. 4 durch Experimente belegt, bevor in Abs. 5 eine Zusammenfassung erfolgt.

[1] Diese Arbeit wurde unterstützt durch die *Deutsche Forschungsgemeinschaft* im Rahmen des Sonderforschungsbereichs 603.

2 Grundlagen

Eine automatische Auswertung von Gesichtsbildern kann die in Abs. 1 vorgestellte Diagnoseaufgabe unterstützen, indem sie eine objektivierte Beurteilung der Gesichtsmerkmale und eine Klassifikation in Lähmungsklassen gestattet [1]. Bei eventuell notwendigen Rehabilitationsübungen kann der Betreuungsaufwand durch medizinisches Fachpersonal reduziert werden, indem der Patient nach einer Einführung selbständig vor dem Rechner und der Kamera die empfohlenen Übungen durchführt.

Das hier vorgestellte Projekt ist Teil eines Gesamstsystems zur wissensbasierten aktiven Szenenanalyse, wie es in [9] vorgestellt wurde. Hierin werden zwei wichtige Aspekte der Bildanalyse vereinigt, die in der Verarbeitungsstrategie des aktiven Sehens [2] und der wissensbasierten Bildanalyse [13] bestehen.

Als Wissensrepräsentationsformalismus dient hier der ERNEST–Formalismus [8, 12] für semantische Netze, der in zahlreichen Problemen der Mustererkennung eingesetzt wurde, beispielsweise zur Analyse von Verkehrsszenen [4]. In [9] wurde am Beispiel der Exploration einer Büroszene beschrieben, wie *ein* semantisches Netz zur Repräsentation von Objektmodellen *und* Kameraaktionen eingesetzt werden kann.

Klassische wissensbasierte Bildanalyse kann in vielen Systemen gefunden werden, beispielsweise in SPAM [7]. In keinem dieser Systeme sind Kameraaktionen repräsentiert. Arbeiten zur Aktionsrepräsentation können z.B. in [10] gefunden werden; dabei werden Bayes–Netze eingesetzt. Beispiele für weitere einheitliche Repräsentationen von Aktionen und Modellen sind im Zusammenhang mit Situationslogik und Entscheidungsnetzen in z.B. in [11].

Der Stand der Kunst zur Gesichtsbildanalyse ist beispielsweise in [3] zusammengefaßt. Im Zusammenhang mit medizinischen Anwendungen befaßt sich [5] mit der Analyse von Gesichtsbildern.

3 Wissensbasis und Kontrolle des Systems

In Bild 1 ist die Wissensbasis, mit deren Hilfe die Analyse des Lähmungsgrades durchgeführt wird, dargestellt. Das Zielkonzept des Netzes ist das Konzept "Diagnose", das nach Abschluß der Analyse die Information über den Lähmungsgrad enthält. Die Diagnose setzt sich aus Teildiagnosen ("LähmungAuge", "LähmungMund" und "LähmungStirn") für die einzelnen Gesichtsregionen zusammen. Das Wissen über Gesichter findet sich im Konzept "Gesicht" und dessen Bestandteile "linkesAuge", "rechtesAuge" "Stirn" und "Mund" wieder. Abgesehen vom Konzept "Stirn" werden die Konzepte auf der Abstraktionsebene der Bilder durch eine Farbregion konkretisiert. Als Kameraaktion wird das Konzept "veränderZoom" integriert, dessen Aufgabe es ist, eine neue Brennweiteneinstellung zu berechnen, wenn eine Farbregion im Bild zu klein ist. Die Kameraaktion wird mit der Berechnung des Konzepts "explRegSeg" bei Bedarf ausgeführt, nämlich genau dann, wenn die Farbregion für eine Analyse der Gesichtsteile zu klein ist. Sie konkurriert sozusagen mit der Instantiierung der Konzepte, die die Gesichtsteile repräsentiert.

Zur Instantiierung der Konzepte der Wissensbasis werden problemabhängige Methoden eingesetzt. So wird bei der Instantiierung des Konzepts "Auge" überprüft, ob die

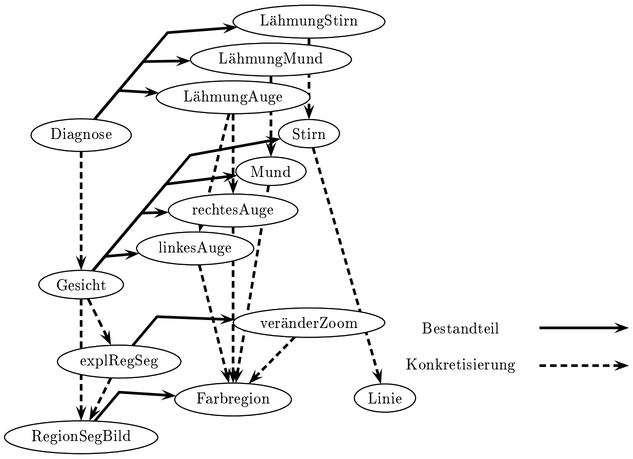

Bild1. Repräsentation von Wissen über Gesichter und Kameraaktionen

Größenverhältnisse denen eines Auges entsprechen, ob die Farbverteilung der erwarteten entspricht, etc.

Ist das Konzept "Auge" instantiiert, so kann die Instantiierung von "LähmungAuge" erfolgen. Dazu werden an der Stelle der Augenposition im Farbbild Merkmale extrahiert. Anhand von medizinischen Wissen und der Symptome kann dann der Lähmungsgrad bestimmt werden.

Sind alle Gesichtsmerkmale analysiert, so kann der Gesamtlähmungsgrad ermittelt und ein Diagnosevorschlag gemacht werden.

Während der Analyse von Gesichtsbildern bildet das in der Wissensbasis repräsentierte Wissen die Grundlage für die Interpretation der Bilder. Gesucht ist hierbei eine Interpretation die optimal zu den Daten paßt und kompatibel ist mit dem in der Wissensbasis gespeicherten Wissen. Aufgrund von Fehlern in der Segmentierung und Mehrdeutigkeiten in der Wissensbasis (Modalitäten) wird das Finden der Interpretation auf ein Suchproblem reduziert, das mit Suchalgorithmen wie zum Beispiel dem A^*–Algorithmus oder iterativen Optimierungsverfahren gelöst werden kann.

Die Instantiierung der Wissensbasis erfolgt momentan datengetrieben. Dazu wird zunächst modellgetrieben ein sogenannter Instantiierungspfad gebildet, durch den die Instantiierungsreihenfolge vorgegeben wird.

Im Bild 3 sind Regionen dargestellt, die aus Bild 2 segmentiert wurden und dem semantischen Netz zur Analyse übergeben werden.

4 Experimente

Experimente wurden bislang mit einem Teilnetz des vorgestellten semantischen Netzes durchgeführt. Gefordert wurde die Instantiierung eines Gesichts, das aus einem Mund und zwei Augen besteht.

 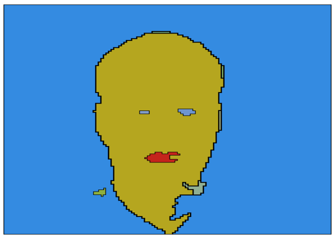

Bild2. Original **Bild3.** Regionensegmentierung

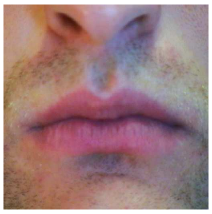 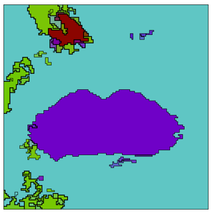  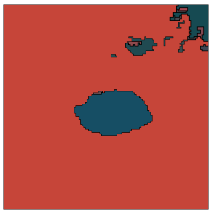

Bild4. Mund **Bild5.** Segmente **Bild6.** Auge **Bild7.** Segmente

In den Bildern 4,5,6,7 werden die Originale und die automatisch segmentierten Regionen gezeigt, die dem semantischen Netz übergeben und von ihm zur Instantiierung verwendet werden.

Die Experimente wurden mit 9 Personen durchgeführt. Teilweise trugen die Personen Brillen oder Bart. Die Beleuchtung der beiden Gesichtshälften war unterschiedlich. Die rechte Gesichtshälfte wurde von einer Glühbirne beleuchtet, die linke von Tageslicht. Daraus ergaben sich starke Qualitätsunterschiede in der Augensegmentierung.

Die Kameraaktionen sind bislang in [9] konzeptionell realisiert, hier jedoch nur konzipiert und noch nicht im System implementiert. Daher wurde die Kamera von Hand geführt und dem System die so generierten Bilder übergeben.

Erkennungswahrscheinlichkeit Auge (Tageslicht)	0,54
Erkennungswahrscheinlichkeit Auge (Glühbirne)	0,77
Erkennungswahrscheinlichkeit Mund	0,88
Dauer Vorverarbeitung	4s
Dauer Instantiierung	10s

Tabelle1. Erste Ergebnisse

Tabelle 1 verdeutlicht, daß die Qualität einer regionbasierten Segmentierung von Gesichtsmerkmalen, speziell den Augen, stark von der Beleuchtung (Helligkeit *und* Temperatur) abhängt. Die gemessenen Zeiten wurden auf einer SGI O2 (R10000, 250 MHz) ermittelt.

5 Zusammenfassung

In diesem Beitrag stellen wir einen zentralen Teil eines Systems zur Diagnoseunterstützung von Patienten mit Gesichtslähmungen vor. Es handelt sich um das automatische Generieren von Nahaufnahmen der Gesichtsmerkmale Stirn, Augen, Nase und Mund. Die Implementierung erfolgte wissensbasiert und setzte auf dem semantischen Netzwerkformalismus ERNEST auf.

In einem regionensegmentierten Gesichtsbild werden die möglichen Regionen der Gesichtsmerkmale angezoomt. Vom semantischen Netz wird wiederum verifiziert, ob sich in den Bilddaten die gewünschten Gesichtsmerkmale befinden. Ist das der Fall, so können mit anschließenden Methoden Merkmale aus den Bildern extrahiert werden, aus denen dann anhand medizinischen Wissens eine Diagnose automatisch berechnen werden kann.

References

1. U. Ahlrichs, D. Paulus, S. Wolf: *Objektivierung der Beurteilung von Gesichtsasymmetrien durch Bildanalyse*, in T. Lehmann, I. Scholl, K. Spitzer (Hrsg.): *Workshop Bildverarbeitung für die Medizin*, Aachen, 1996, S. 125–130.
2. J. Aloimonos, I. Weiss, A. Bandyopadhyay: *Active Vision*, *International Journal of Computer Vision*, Bd. 2, Nr. 3, 1988, S. 333–356.
3. G. Chow, X. Li: *Towards A System For Automatic Facial Feature Detection*, *Pattern Recognition*, Bd. 26, 1993, S. 1739–1755.
4. V. Fischer: *Parallelverarbeitung in einem semantischen Netzwerk für die wissensbasierte Musteranalyse*, Dissertation, Technische Fakultät, Universität Erlangen–Nürnberg, Erlangen, 1995.
5. R. Herpers, H. Rodax, G. Sommer: *A neural network identifies faces with morphological syndrome*, in S. Andreassen (Hrsg.): *Artificial Intelligence in Medicine*, IOS Press, Amsterdam, 1993, S. 481–485.
6. J. W. House: *Facial Nerve Grading Systems*, *Laryngoscope*, Bd. 93, 1983, S. 1056–1069.
7. D. McKeown, W. Harvey, J. McDermott: *Rule-Based Interpretation of Aerial Imagery*, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Bd. 7, Nr. 5, 1985, S. 570–585.
8. H. Niemann, G. Sagerer, S. Schröder, F. Kummert: *ERNEST: A Semantic Network System for Pattern Understanding*, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Bd. 9, 1990, S. 883–905.
9. D. Paulus, U. Ahlrichs, B. Heigl, H. Niemann: *Wissensbasierte aktive Szenenanalyse*, in P. Levi (Hrsg.): *Mustererkennung 1998*, Springer, Heidelberg, September 1998, S. 185–192.
10. R. Rimey: *Control of Selective Perception using Bayes Nets and Decision Theory*, Department of Computer Science, College of Arts and Science, University of Rochester, Rochester, New York, 1993.
11. S. J. Russell, P. Norvig: *Artificial Intelligence. A Modern Approach*, Prentice-Hall, Englewood Cliffs, NJ, 1995.
12. G. Sagerer: *Darstellung und Nutzung von Expertenwissen für ein Bildanalysesystem*, Springer, Berlin, 1985.
13. G. Sagerer, H. Niemann: *Semantic Networks for Understanding Scenes*, Advances in Computer Vision and Machine Intelligence, Plenum Press, New York and London, 1997.
14. E. Stennert, C. H. Limberg, K. P. Frentrup: *Parese- und Defektheilungs-Index*, *HNO*, Bd. 25, 1977, S. 238–245.