From: FIfF-Kommunikation, 19(1), 27–30, 2002.

Technik und Leistungsfähigkeit automatischer Gesichtserkennung

Rolf P. Würtz

Institut für Neuroinformatik
Ruhr-Universität, 44780 Bochum
http://www.neuroinformatik.ruhr-uni-bochum.de/PEOPLE/rolf/
Rolf.Wuertz@neuroinformatik.ruhr-uni-bochum.de

Über den Autor

Rolf Würtz erhielt sein Diplom in Mathematik mit Nebenfächern Physik und angewandte Informatik 1986 an der Universität Heidelberg. Nach einem 2-jährigen Forschungsaufenthalt am Max-Planck-Institut für Hirnforschung in Frankfurt war er wissenschaftlicher Angestellter am neugegründeten Institut für Neuroinformatik der Ruhr-Universität Bochum. Im Jahre 1994 promovierte er an der Fakultät für Physik und Astronomie über ein neuronales System zur Gesichtserkennung. Es folgte eine zweijährige Postdoc-Zeit an der Fakultät für Informatik der Universität Groningen/Niederlande. Seit 1997 ist Rolf Würtz leitender wissenschaftlicher Mitarbeiter und Dozent am Institut für Neuroinformatik in Bochum. Seine Forschungsinteressen beinhalten visuelle Objekterkennung und Robotik.

Zusammenfassung

Nach einem allgemeinen Überblick über Aufgaben und Probleme des Computersehens wird diskutiert, warum es schwierig ist, die Leistungsfähigkeit solcher Systeme obejktiv zu messen. Es folgt eine Einführung in die Technik der automatischen Gesichtserkennung. Danach werden die Ergebnisse des FERET Tests beschrieben, bei dem Gesichtserkennungssysteme unabhängig auf einer großen Datenbasis verglichen wurden. Die Schlussfolgerung ist, dass die derzeitigen Systeme für den praktischen Einsatz noch nicht tauglich sind, es aber durch die intensive Forschungs- und Entwicklungstätigkeit recht bald werden könnten.

1 Einleitung

Durch die geplante Einführung biometrischer Merkmale und eine erhohte Nachfrage nach automatisierter Überwachung ist die automatische Personenerkennung in den vergangenen Monaten ins Zentrum des Interesses gerückt. Unabhängig von den dadurch aufgeworfenen gesellschaftlichen, rechtlichen und politischen Fragen soll dieser Artikel den Stand der Technik skizzieren und die Leistungsfähigkeit der existierenden Systeme bewerten. Dazu ist es erforderlich, einige allgemeine Bemerkungen zu Problematik und Technologie des automatischen Bildverstehens oder Computersehens zu machen.

Die Aufgabe dieses Wissenschaftszweiges ist es, aus Kamerabildern sinnvolle Rückschlüsse auf Objekte in der Welt zu ziehen. Die Anwendungen sind vielfältig und umfassen z.B. industrielle Qualitätkontrolle, automatische Auswertung von Satellitenbildern, intuitive Mensch-Maschine-Schnittstellen, Serviceroboter und Überwachung. Innerhalb der Informatik ist Computersehen für die Künstliche Intelligenz von eminenter Bedeutung, da es letztendlich Maschinen in die Lage versetzen soll, Information aus der Umwelt autonom, d.h. ohne menschliche Mithilfe, aufzunehmen und weiter zu verarbeiten.

Dieses Gebiet hat sich in den vergangenen Jahren stark auf die Verarbeitung von Bildern und Bildfolgen von Personen konzentriert. Der Anteil solcher Publikationen an allen Publikationen im Computersehen ist zwischen 1991 und 1999 von unter 2% auf über 9% angestiegen - zusätzlich zur allgemeinen Steigerung der jährlichen Gesamtzahl der Publikationen. Zum Einen haben sowohl die verfügbare Hardware als auch die Methoden in diesem Zeitraum eine starke Leistungssteigerung erlebt. Zweitens ist ein großer Teil dieser Forschung drittmittelfinanziert, was bedeutet, dass es ein deutliches politisches Interesse an dieser Forschung gibt. Drittens haben sich die politische Landschaft und das öffentliche Meinungsbild so verändert, dass computergestützten Überwachungsmethoden nur noch mit geringem Misstrauen begegnet wird. Besonders in Großbritannien, aber in zunehmendem Maße auch in Deutschland, überwiegt die Hoffnung auf Kriminalitätsbekämpfung durch Videoüberwachung die Furcht vor unangemessener Kontrolle durch öffentliche oder private Instanzen.

Innerhalb der Informatik nimmt das Computersehen (wie alle Versuche, aus Sensordaten symbolische Information zu extrahieren), eine Sonderstellung ein, da die Aufgaben sehr schlecht zu formalisieren sind. Oft sind Aussagen aus den Bildern nur mit Hilfe von zusätzlichen Annahmen möglich. Das menschliche Gehirn ist sehr gut in der Lage, plausible Annahmen zu ma-

chen, um zu einer richtigen Interpretation der Bilddaten zu kommen. Folglich gibt es grundsätzlich nur eine Möglichkeit, die Leistungsfähigkeit eines Systems zum Computersehen zu evaluieren. Bilder oder Bildfolgen müssen unter möglichst kontrollierten Bedingungen aufgenommen werden, und ein Mensch muss entscheiden, welche Aussagen über die Bilder zutreffend sind. Keinesfalls kann die Bewertung eines Algorithmus ohne Rückgriff auf menschliche Wahrnehmung geleistet werden.

Diese Situation erschwert objektive Leistungsmessungen ungemein. Normalerweise erfolgt die Auswahl der Daten, auf denen ein neu entwickelter Algorithmus getestet wird, durch dieselbe Arbeitsgruppe, die auch die Entwicklung durchführt. Da die Leistungsgrenzen eines Algorithmus nur schwer publizierbar sind und alle wissenschaftlichen Einrichtungen unter starkem Publikationsdruck stehen, gibt es eine Tendenz, Fehlschläge unter den Tisch fallen zu lassen. Daher ist zu erwarten, dass die Qualität von Algorithmen bei Anwendung auf neues Datenmaterial hinter der veröffentlichten Leistungsfähigkeit zurückbleibt. Aus den Entwicklungsabteilungen von Firmen sind keine objektiven Zahlen über die Leistung von Systemen verfügbar. Die zugehörigen Werbeabteilungen unterliegen einem noch weit höheren Zwang zur Schönfärberei als wissenschaftliche Institutionen. Generell kann man daher davon ausgehen, dass veröffentlichte Leistungsdaten im Computersehen das unter günstigsten Bedingungen erreichbare Optimum widerspiegeln.

2 Funktionsweise von Gesichtserkennungssystemen

Die technische Problematik der Identifikation von Personen soll nun anhand der Gesichtserkennung aus Kamerabildern aufgezeigt werden. Für andere Biometrieverfahren, wie z.B. Fingerabdruckerkennung, Iriserkennung, Sprecheridentifikation, Erkennung der Gangart ist die Grundproblematik die gleiche, die Lösungsansätze aber z.T. sehr verschieden.

Das Problem der Erkennung von Personen aus Kamerabildern besteht darin, dass ein und dasselbe Gesicht sehr unterschiedliche Bilder erzeugen kann, je nach Aufnahmebedingungen. Die Unterschiede können u.a. auf folgende Veränderungen zwischen zwei Aufnahmen zurückgeführt werden:

- Position des Gesichtes im Bild
- Abstand von der Kamera bzw. Vergrößerungsfaktor

- Kopfausrichtung
- Gesichtsausdruck
- Hintergrund
- teilweise Verdeckung
- Beleuchtung
- Kameratyp und -einstellung

All diese Veränderungen müssen durch ein Gesichtserkennungssystem aufgefangen werden, um die Identität der Gesichter in zwei verschiedenen Bildern verlässlich festzustellen. Man spricht daher von *invarianter Erkennung*.

Die meisten dieser Veränderungen können durch Lösung des sog. *Korrespondenzproblems* [8] aus den Bildern „herausgerechnet“ werden. Es bezeichnet die Aufgabe, zu jedem Punkt in einem Bild den zugehörigen Punkt im anderen zu finden, also z.B. das linke Auge dem linken Auge zuzuordnen. Die größte Schwierigkeit besteht darin, dass das Abbild einzelner Punkte bei verschiedenen Aufnahmen recht verschieden ausfallen kann. Dies gilt insbesondere bei Beleuchtungsänderungen, anderer Kamera und bei Kopfdrehungen. Andererseits können verschiedene Punkte relativ ähnlich ausfallen, sodass nur mit Hilfe zusätzlicher Annahmen, wie z.B. ungefährer Anordnung verschiedener Punkte sinnvolle Schätzungen der wahren Korrespondenzen gemacht werden können [1, 8]. Diese Korrespondenzschätzungen werden durch alle Veränderungen im Bild mehr oder weniger gestört, und das Zusammenkommen mehrerer Veränderungen, wie z.B. Kopfdrehung und Beleuchtung führen häufig dazu, dass sie grob falsch werden.

Wenn das Korrespondenzproblem näherungsweise gelöst ist, kann die Bildinformation in der Nähe der jeweils zusammengehörigen Punkte verglichen werden. Die Gesamtähnlichkeit aller Punktepaare ergibt dann eine Ähnlichkeit der beiden Gesichter. Bei hoher Ähnlichkeit kann geschlossen werden, dass beide Bilder dieselbe Person zeigen. Diese Entscheidung wird *Verifikation* genannt, da entschieden werden kann, ob die behauptete Identität zu einer gegebenen Person mit den Bilddaten übereinstimmt.

Da die Struktur aller Gesichter sehr ähnlich ist, kann das Korrespondenzproblem auch für Gesichter verschiedener Personen gelöst werden. Die Gesamtähnlichkeit kann dann direkt als Ähnlichkeit der Gesichter interpretiert werden. Auf diese Weise kann jedes Gesicht in einer Datenbank mit einem

aktuellen Kamerabild verglichen und so die *Identität* einer Person festgestellt werden. Diese Aufgabe ist schwieriger als die Verifikation.

In beiden Fällen muss eine Entscheidungsschwelle willkürlich gesetzt werden, so dass das System Identität zwischen den Bildern feststellt, falls die Ähnlichkeit über dieser Schwelle liegt. Liegt die Schwelle sehr niedrig, werden viele Bilder als gleich beurteilt, die in Wirklichkeit verschiedene Personen zeigen. Diese Entscheidungen werden als „falsch positiv“ bezeichnet. Liegt die Schwelle sehr hoch, werden korrekte Identitäten durch das System zurückgewiesen („falsch negative“ Entscheidung). Der mit der Entscheidungsschwelle zu vergleichende Wert wird meist in komplizierterer Weise berechnet als hier skizziert, der Gebrauch einer Schwelle ist jedoch allen Verfahren gemeinsam. Das zu erwartende Verhältnis der falsch positiven zu falsch negativen Entscheidungen kann mit ihrer Hilfe eingestellt werden.

Da die Korrespondenzschätzung aufwendig und fehleranfällig ist, ist es für die Effizienz eines Verfahrens entscheidend, dass die Datenbasis zur Identifikation so organisiert ist, dass die Anwendung der Korrespondenzsuche minimiert und der eigentliche Vergleich sehr schnell und effizient ausgeführt werden kann. Zwei wesentliche Techniken sind hier *Eigengesichter* [6] und *Bündelgraphen* [7].

Diese Verfahren sind nur auf Bilder anwendbar, die ausser dem Gesicht wenig Hintergrund enthalten. In praktischen Szenarien müssen sie durch weitere Komponenten ergänzt werden, die aus Bildern oder Videosequenzen für Erkennung oder Identifikation geeigneten Teilbereiche herauszufiltern. eignen. Ein solches System, dessen Kern die Bündelgraphentechnik ist, ist in [2] beschrieben.

In dieser Einführung wurden grob die Probleme umrissen, die jedes Gesichtserkennungssystem lösen muss. Gegenwärtig gibt es mehrere konkurrierende Verfahren, die das Schwergewicht auf die eine oder andere Anforderung legen. Es dürfte deutlich geworden sein, dass ein System zum praktischen Einsatz eine komplexe Kombination mehrerer Verfahren sein muss, und seine Gesamtqualität nur mit sehr langwierigen Testreihen in echten Umgebungen festgestellt werden kann.

3 Unabhängige Leistungsuntersuchungen

Da Software zur Personenerkennung normalerweise nicht öffentlich verfügbar ist, die Leistungsfähigkeit aber von einer Unzahl von Details abhängt, sind

nach wissenschaftlichen Standards objektive Leistungsmessungen derzeit kaum zu erreichen. Die Verfügbarkeit von Gesichtsdatenbanken macht es möglich, dass verschiedene Entwicklergruppen ihre Algorithmen auf ein und demselben Datensatz testen. Dies behebt die Schwierigkeit, die Leistung eines Algorithmus zu bewerten nur unvollkommen, da im Laufe der Entwicklung viele kleine Entscheidungen gefällt und Parameter optimiert werden. Dadurch wird der Algorithmus im Laufe der Entwicklung wissentlich oder unbewusst darauf optimiert, mit den bekannten Daten gut zu funktionieren. Die Erkennungsraten fallen bei Anwendung auf neue Daten teilweise drastisch.

Eine Anstrengung, diese Probleme für den Fall der Personenidentifikation aus Gesichtsbildern zu lösen, wurde von dem US-amerikanischen Verteidigungsministerium in den Jahren 1994 bis 1997 unternommen. Im Rahmen des „Counterdrug Technology Development Program“ wurde eine Datenbank von 14126 Gesichtsbildern von 1199 Personen erstellt. Ein kleiner Teil wurde an verschiedene Forschergruppen gegeben, um deren Systeme zu entwickeln. Nach abgeschlossener Entwicklung wurden die Systeme in kontrollierter Weise auf die Gesamtdaten angewandt und die Ergebnisse von einer unabhängigen Stelle ausgewertet [4].

Um die verschiedenen Modalitäten so gut wie möglich abzubilden, wurden die Testbilder eingeteilt in solche, die am gleichen Tag wie das Datenbankbild derselben Person aufgenommen wurden, und zwar mit identischer (*FB*) bzw. veränderter Beleuchtung (*FC*), Bilder die in mehreren Tagen Abstand (*duplicate I*) und mit mindestens einem Jahr Abstand (*duplicate II*) zu den Datenbankbildern aufgenommen wurden.

Das erste Ergebnis der Studie war, dass die meisten Systeme überhaupt nicht in der Lage waren, das Korrespondenzproblem zufriedenstellend zu lösen. Daher wurde ein vereinfachter *teilautomatischer* Test angeboten, bei dem zusätzlich zu den Bildern die handmarkierten Augenpositionen gegeben waren. Die Erkennungsraten in diesem Test variierten in folgenden Bereichen *FB*:87% - 97%; *FC*: 33% - 82%; *duplicate I*: 33% - 60%; *duplicate II*: 9% - 54%.

Am praxisnahen *vollautomatischen* Test ohne handmarkierte Zusatzinformation haben überhaupt nur zwei Forschungsgruppen teilgenommen, und zwar ein Team vom MIT und ein gemeinsames von der Universität Bochum und der University of Southern California in Los Angeles. Deren Ergebnisse sind: *FB*:87%, 95%; *duplicate I*: 50%, 67%. Die eigene Eitelkeit verbietet es zu verschweigen, dass die jeweils besseren Werte von dem an der Ruhr-Universität Bochum und USC gemeinsam entwickelten Verfahren erreicht

wurden. Die Werte der vollautomatischen Tests für die schwierigen Fälle *FC* und *duplicate II* werden in der Studie nicht genannt.

Über die einfachere *Verifikationsaufgabe* wird in [3] berichtet, dass bei einer Einstellung der Erkennungsschwelle auf 2% falsch positiver Entscheidungen folgende Fehlerraten gemessen wurden: *FB*: 0,4%; *FC*: 9%; *duplicate I*: 11%; *duplicate II*: 43%. Das zeigt die Leistungsfähigkeit in nicht sicherheitskritischen Bereichen auf, wo 2% Durchlässe unter falschen Voraussetzungen nicht bedrohlich sind. Anders ausgedrückt: geht man davon aus, dass pro Tag 10000 Zugänge durch Verifikation kontrolliert werden, von denen 100 Betrugsversuche sind, so werden - nach dem Stand der Technik von 1996 - zwei der Betrugsversuche gelingen und ca. 1000 Personen abgewiesen werden. Dies ist kein akzeptables Szenario für ein vollautomatisches System.

Alle Ergebnisse des Tests zeigen, dass Gesichtserkennung bei kurz hintereinander im gleichen Raum mit der gleichen Kamera und der gleichen Beleuchtung aufgenommen Bildern sehr gut funktioniert. Das Problem der Beleuchtungsvariation ist technisch völlig ungelöst. Das gleiche gilt für die wenig kontrollierbaren Veränderungen, die Gesichter von Personen im Laufe von Jahren durchmachen. Die menschliche Erkennung ist durch eine Unmenge von Erfahrung optimiert, über diese Veränderungen hinwegzusehen und Personen invariant zu erkennen. Für das Computersehen bleibt hier ein sehr weites Forschungsfeld zu bearbeiten.

Als Beispiel eines direkten Praxistests sei noch das Beispiel eines installierten Überwachungssystems in der Stadt Tampa in Florada angeführt. Die Amerikanische Bürgerrechtsvereinigung (ACLU) berichtet in einer kürzlich erschienenen Veröffentlichung [5], dass dieses System dazu gedacht war, bekannte Verbrecher auf Videofilmen von Straßen und öffentlichen Plätzen zu identifizieren. Die Erkennungsprotokolle sind nach amerikanischem Recht öffentlich, und die Ergebnisse werden von der ACLU wie folgt zusammengefasst. „Das System aus mehreren Dutzend Kameras wurde am 29.06.2001 installiert. Es hat kein einziges der gespeicherten Gesichter von Verdächtigen korrekt wiedererkannt, geschweige denn zu einer Verhaftung geführt. Es wurde am 11.08.2001 abgeschaltet und wurde seither nicht mehr benutzt. Während der Laufzeit hat es sehr viele falsch positive Meldungen produziert, einschließlich der - für Menschen eindeutig erkennbaren - Verwechslung von Frauen und Männern.“ Natürlich ist unbekannt, wie viele der Gesuchten tatsächlich jemals an der Kamera vorbeigekommen sind, aber es darf doch geschlossen werden, dass das System noch nicht ganz praxisreif war. Die Verwechslung von Männern und Frauen weist auch auf die Tatsache hin,

dass die Funktionsweise technischer Gesichtserkennung sich sehr stark von der menschlichen unterscheidet.

4 Diskussion

Wie im ganzen Gebiet des Computersehens fallen derzeitige Systeme zur automatischen Personenerkennung noch sehr weit hinter dem technisch wünschenswerten zurück. Insbesondere der Einsatz im Freien oder in Räumen mit nennenswerten Fensterflächen muss heute noch am Problem der Beleuchtungsvariation scheitern. Für dessen Lösung ist mehr Rechenleistung oder einfache Systemoptimierung nicht ausreichend, vielmehr bedarf es deutlicher Durchbrüche in der Methodik. Systeme zur Verifikation, d.h. zur Überprüfung einer behaupteten Identität, haben wohl den höchsten Reifegrad erreicht. Dennoch sind sie für einen vollautomatischen Einsatz höchstens bedingt geeignet.

Die oben zitierten Zahlen sprechen eine deutliche Sprache. Sie belegen, dass automatische Gesichtserkennung eine aufstrebende Technologie ist, die in den vergangenen Jahren beachtliche Erfolge verzeichnen konnte. Andererseits ist sie noch weit davon entfernt, unter realen Bedingungen robust zu funktionieren.

Zweifellos sind die Erkennungssysteme in den seit dem FERET-Test vergangenen Jahren deutlich verbessert worden. Verlässliche Zahlen hierzu liegen nicht vor, da ein guter Teil der Optimierungen von Firmen an eigenen Systemen gemacht werden, die unabhängigen Tests erst nach Kauf zugänglich sind. Ergebnisse von neueren unabhängigen Tests sind meines Wissens nicht publiziert.

Für die Diskussion über die Einführung von Biometriesystemen sind diese Resultate zweischneidig. Einerseits dürfte klar sein, dass derzeit nur wenig Vertrauen in Gesichtserkennungssysteme gesetzt werden sollte. Andererseits wäre es falsch, die notwendige gesellschaftliche Diskussion um die Abwägung zwischen Privatsphäre und Sicherheit mit Hinweis auf mangelnde Funktionsfähigkeit zu vertagen. Ich habe nur sehr geringe Zweifel daran, dass diese Funktionsfähigkeit weiter rapide ansteigen wird. Wenn leistungsfähige Personenerkennungssysteme überall preisgünstig zu haben sind, könnte es leicht zu spät sein darüber nachzudenken, ob, wo und unter welchen Umständen die Bürger sie haben wollen.

Literatur

[1] M. Lades, J. C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Würtz und W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42(3):300–311, 1993.

[2] H. S. Loos. Suchbilder – Computer erkennt Personen in Echtzeit. *c't*, (15):128–131, 2000.

[3] P. J. Philips, A. Martin, C. Wilson und M. Przybocki. An introduction to evaluating biometric systems. *IEEE Computer*, 33(2):56–63, 2000.

[4] P. J. Philips, H. Moon, S. A. Rizvi und P. J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.

[5] J. Stanley und B. Steinhardt. Drawing a blank: The failure of facial recognition technology in Tampa, Florida. Technischer Bericht, American Civil Liberties Union, 2001. http://www.aclu.org/issues/privacy/drawing_blank.pdf.

[6] M. Turk und A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[7] L. Wiskott, J.-M. Fellous, N. Krüger und C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.

[8] R. P. Würtz. Object recognition robust under translations, deformations and changes in background. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):769–775, 1997.