

# Verfahren zur Objekt- und Gesichtserkennung: Möglichkeiten und Grenzen

## Object Recognition and Face Detection: Capabilities and Limits

Prof. Dr. Klaus Jung  
Hochschule für Technik und Wirtschaft Berlin  
Wilhelminenhofstraße 75A, 12459 Berlin  
Tel.: +49 30 5019-2877, Fax: +49 30 5019-48-2877  
E-Mail: klaus.jung@htw-berlin.de, Internet: <http://home.htw-berlin.de/~jungk>

### Zusammenfassung:

Automatisierte Verfahren zur Objekterkennung werden hinsichtlich ihrer Zielsetzung und der zugrundeliegenden Technologie klassifiziert. Die Verbesserung der Effizienz solcher Verfahren in den letzten 15 Jahren wird anhand verschiedener Untersuchungen dargestellt. Ein Verfahren der Objektklassifizierung auf Grundlage von Support Vector Machines wird exemplarisch näher beschrieben.

### Abstract:

This article classifies different approaches in object detection and recognition towards their functionality and underlying technology. The improvement in efficiency of such systems over the last 15 years is documented by results of various evaluations. An algorithm for object classification based on Support Vector Machines is presented in more detail.

### Einleitung

Verfahren zur automatisierten Objekterkennung kommen seit Mitte der 70er Jahre auf Computern zur Anwendung. Akzeptable Ergebnisse hinsichtlich der Trefferquote werden im Bereich der Gesichtserkennung jedoch erst seit Ende der 80er Jahre erzielt. Seitdem wurden die Verfahren immer weiter verbessert [1] und erlauben dank der gestiegenen Prozessorleistungen praktische Anwendungen. Besondere Bedeutung hat dabei die Erkennung von Gesichtern. Zum einen liegt das an der Einführung der Biometrie in Sicherheitsprozessen, zum anderen ist die Erkennung von Gesichtern die derzeit am weitesten entwickelte Technik in Rahmen der Erkennung von Objekten im allgemeinen.

Dieser Artikel soll einen Überblick über die Einordnung verschiedener Verfahren zur Objekterkennung geben, den Stand der Technik aufzeigen und über die Erkennung von Gesichtern hinaus Ansätze zur Erkennung beliebiger Objekte vorstellen.

### Zielsetzungen

Das deutsche Wort „Erkennung“ fasst unterschiedliche Zielsetzungen im Themengebiet der Objekterkennung zusammen. Im englischen wird dabei in der Regel unterschieden zwischen

- *Object Detection*: Das Auffinden eines Objekts auf einem Bild. Wo befinden sich Gesichter auf dem Bild? Sind es mehrere?
- *Object Recognition*: Das Identifizieren, Wiedererkennen eines Objekts. Um welche Person handelt es sich?
- *Object Classification*: Eine Variante des Identifizierens: Die Zuordnung des Objekts zu einer von mehreren zuvor definierten Objektklassen. Ist es ein Baum oder ein Auto?

Für all diese Fragestellungen gibt es entsprechende Verfahren. Die Schwierigkeit der Aufgaben ist durchaus vergleichbar. So ist es nicht notwendig „einfacher“ ein Gesicht zu finden als es einer Person zuzuordnen.

## Verfahren

Die den Verfahren zugrunde liegenden Algorithmen können sehr unterschiedlich sein. Zunächst lassen sich diese in nicht-holistische (musterbasierte) und holistische Ansätze einordnen.

Musterbasierte Ansätze nutzen Domänenwissen zur Erkennung der Objekte. Dies bedeutet, dass gewisse spezifische Eigenschaften der zu erkennen Objekte im Verfahren fest implementiert sind. Anwendung findet dies vor allem bei der Gesichtserkennung. Die Form und die Bestandteile (Augen, Nase, Mund) eines Gesichts sind bekannt und werden zur Erkennung herangezogen. Dies kann z.B. durch Analyse der geometrischen Anordnung der separat erkannten Bestandteile geschehen [2]. Die Bestandteile werden dazu als Knoten (*fiducial points*) in einem Graphen angeordnet. Solche Ansätze lassen sich nur schwer auf andere Objekte übertragen, haben aber den Vorteil, dass sie von der Perspektive des aufgenommenen Objektes in bestimmten Grenzen unabhängig sind. So lassen sich damit Gesichter nicht nur in Frontalansicht, sondern auch im Halbprofil erkennen.

Holistische Ansätze besitzen kein spezielles Wissen über die zu erkennenden Objekte. Sie sind damit für die Erkennung beliebiger Objekte geeignet, müssen jedoch vorher trainiert werden [3], [4]. Das zu erkennende Bild wird dabei zunächst als ein Vektor von Grau- oder Farbwerten aufgefasst. Die Ähnlichkeit zweier Bilder über den Abstand dieser Vektoren bezüglich einer geeigneten Norm zu definieren, ist nicht praktikabel. Statt dessen werden gewisse Eigenschaften (*Features*) aus dem Bild extrahiert, die ihrerseits einen Vektorraum (*Feature Space*) bilden, der zur Analyse benutzt wird. Ziel ist es dabei, Gemeinsamkeiten zwischen ähnlichen Bildern aufzufinden. Zum Einsatz kommen statistische Verfahren wie die Diskriminantenanalyse, Hauptkomponentenanalyse oder die Berechnung von „*simple features*“ wie in [3] beschrieben.

Für die Trainingsphase bei der *Object Detection* sind Positiv-Bilder (Objekt vorhanden) und Negativ-Bilder (kein Objekt) notwendig. Beim Training der *Object Recognition* bzw. *Object Classification* müssen jeweils Bilder aus den verschiedenen Klassen vorgegeben werden. Dabei unterscheiden sich sowohl die Anzahl der Trainingsbilder als auch der Rechenaufwand zum Erstellen eines Detektors. So werden für die in [3] beschriebene und in OpenCV [5] implementierte *Object Detection* mehrere tausend Bilder benötigt, um einen guten Detektor zu erzeugen. Entsprechend ergeben sich auch auf moderner Rechnerhardware Trainingszeiten von mehreren Stunden. Verfahren der *Object Classification*, die auf Support Vector Machines [6] basieren, kommen dagegen mit wenigen Trainingsbildern pro Klasse aus und erreichen Trainingszeiten von wenigen Minuten (abhängig von der Anzahl der Klassen).

Die Rechenzeiten für die Erkennung bewegen sich bei allen genannten Verfahren im Sekundenbereich.

## Objektklassifizierung mit Support Vector Machines

In diesem Abschnitt wird ein Verfahren zur *Object Classification* vorgestellt, wie es vom Autor verwendet wird und derzeit für Desktop-Rechner als auch für mobile Geräte (iPhone) implementiert ist. Das Verfahren lässt sich zur Klassifikation beliebiger Objekte einsetzen, zeichnet sich durch geringe Anforderungen an die Rechenleistung und Robustheit gegenüber Veränderungen der Perspektive der Objektansicht aus, ist derzeit aber noch anfällig gegenüber verschiedenen Beleuchtungssituationen und dem Hintergrund, vor dem sich die Objekte befinden.

Aus den Trainings- und Testbildern werden zunächst *low level features* berechnet. Einige stammen aus der visuellen Bildsortierung, wie sie bei der Firma pixolution GmbH zum Einsatz kommt [7], andere wurden für diese Anwendung speziell hinzugefügt. Allen gemeinsam ist ihr statistischer Charakter. So werden die im Bild vorkommenden Farben, die Kantenaktivitäten als auch verschiedene im Bild vorhandene Frequenzanteile analysiert. Der verwendete *Feature Space* hat ca. 300 Dimensionen.

Mit einem Trainingsset von 20 Bildern pro Klasse wird eine Support Vector Machine [6] trainiert. Abbildung 1 zeigt einen Ausschnitt aus einem typischen Trainingsset.

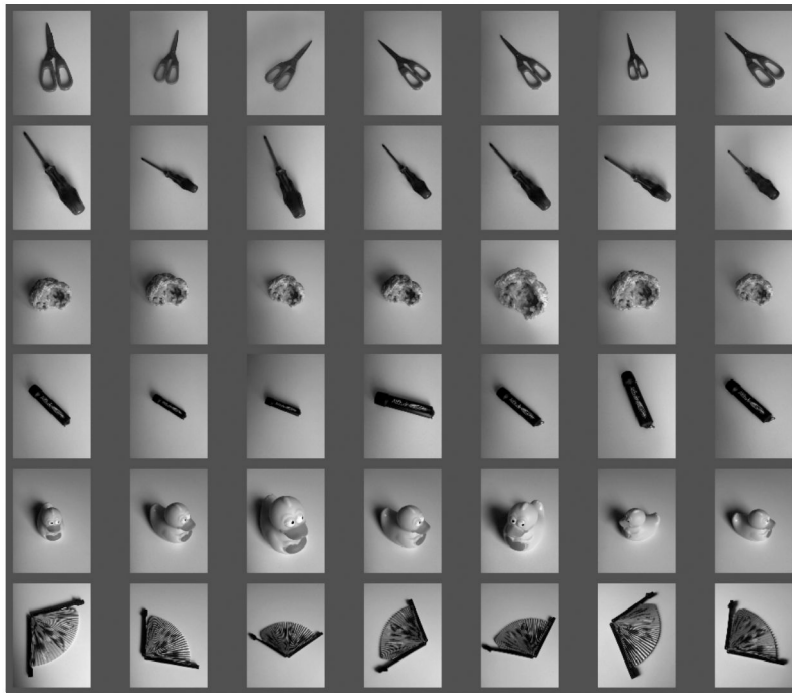


Abbildung 1: Ausschnitt aus einem Trainingsset für eine Objektklassifizierung

Die grundlegende Idee der Klassifizierung ist, dass die Feature-Vektoren aller Bilder derselben Klasse „nahe“ beieinander liegen, während unterschiedliche Klassen einen gewissen Abstand untereinander aufweisen. Nur lassen sich diese Bereiche im *Feature Space* leider nicht so einfach durch ebene Flächen (Geraden im 2-dimensionalen) trennen. Abbildung 2 zeigt eine typische Situation in zwei Dimensionen [6]. Die Punkte stellen die Feature-Vektoren der Trainingsbilder aus drei verschiedenen Klassen dar. Durch die Hintergrundfarbe ist die Einteilung des gesamten Raumes in die drei Klassen sichtbar gemacht.

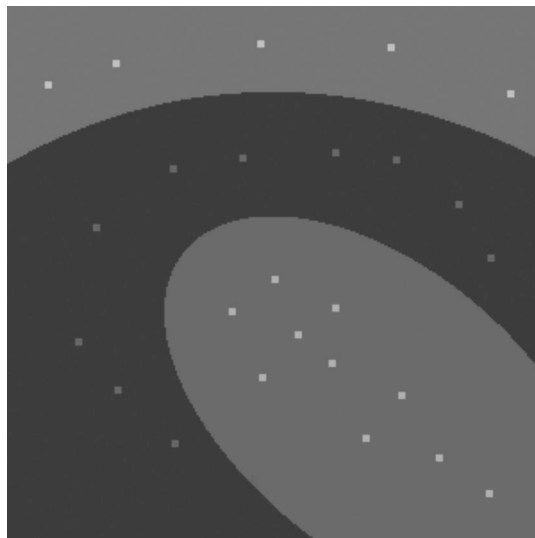


Abbildung 2: Unterteilung eines 2-dimensionalen *Feature Space* in drei Klassen

Optimale, nicht-lineare Trennflächen für die gegebenen Trainingsvektoren zu finden, das ist die Aufgabe einer Support Vector Machine. Die zugrundeliegende mathematische Theorie wird in [8] beschrieben. Support Vector Machines eignen sich zur Klassifikation beliebiger auch abstrakter Daten. Darüber hinaus gibt es Varianten zur Berechnung von Regressionen und zur Abschätzung des Trägers hochdimensionaler Verteilungen (*distribution estimation*).

Sobald die Bereiche im *Feature Space* den verschiedenen Klassen zugeordnet sind, ist die Klassifikation eines Testbildes einfach zu berechnen. Aus dem Testbild werden die gleichen Features wie beim Training extrahiert und der entsprechenden Klasse zugeordnet.

Der Vorteil des hier vorgestellten Verfahrens liegt in seiner Robustheit gegenüber Veränderungen der Objektperspektive und dem schnellen Trainieren auf Grundlage weniger Trainingsbilder. Ein großer Nachteil ist die Tatsache, dass jedes zu erkennende Bild immer einer Klasse zugeordnet wird, wie es einem Klassifikationsverfahren zueigen ist. Das Resultat der Klassifikation ist also nie: „Dieses Objekt kenne ich nicht.“

### Effizienz von Verfahren zur Objekterkennung

Von der dem Menschen eigenen schnellen Erkennung von Objekten und Personen unter verschiedensten Umgebungsbedingungen sind die computergestützten automatisch en Verfahren noch weit entfernt. Die meisten Verfahren brauchen klar definierte und „saubere“ Bedingungen. Das bedeutet: Gute Lichtverhältnisse bei der Aufnahme der Bilder, wenig Rauschen, ggf. hohe Auslösung, eine fest vorgegebene Aufnahmegeometrie (für die Passfotos der neuen biometrischen Ausweise darf man nicht mehr lächeln, Halbprofilaufnahmen sind nicht mehr erlaubt), unter Umständen einen homogenen Hintergrund. Aus Bildern von Straßenszenen die Autos verschiedener Marken zu erkennen, ist heutzutage sicherlich noch nicht möglich. Auf der anderen Seite haben die Verfahren in den letzten 10 Jahren deutliche Fortschritte in ihren Erkennungsraten gemacht, vor allem im Bereich der Gesichtserkennung. Durch den gleichzeitigen Anstieg der Rechenleistung der Hardware finden solche Verfahren inzwischen ihren Weg in die Consumer Elektronik. Eine *Face Detection* gehört inzwischen zur Standardausstattung jeder Digitalkamera. Ein Programm zum Sortieren der Urlaubsfotos nach den darauf abgebildeten Personen findet man bereits in vielen Bildverwaltungsprogrammen, wie z.B. iPhoto.

Abbildung 3 zeigt die Zusammenfassung einer systematischen Untersuchung von kommerziellen Verfahren zur Gesichtserkennung über die Jahre hinweg. Im Jahr 2006 wurden Verfahren von 22 Unternehmen, Universitäten und Forschungseinrichtungen getestet. Auf einem vorgegebenen Testsatz von Bildern werden Gesichter erkannt. Dabei werden die Verfahren so eingestellt, dass sie im Mittel weniger als 0,1% aller Gesichter falsch erkennen (*false accept rate*). Gemessen wird dann der Anteil an Gesichtern, der nicht erkannt wurden (*false reject rate*).

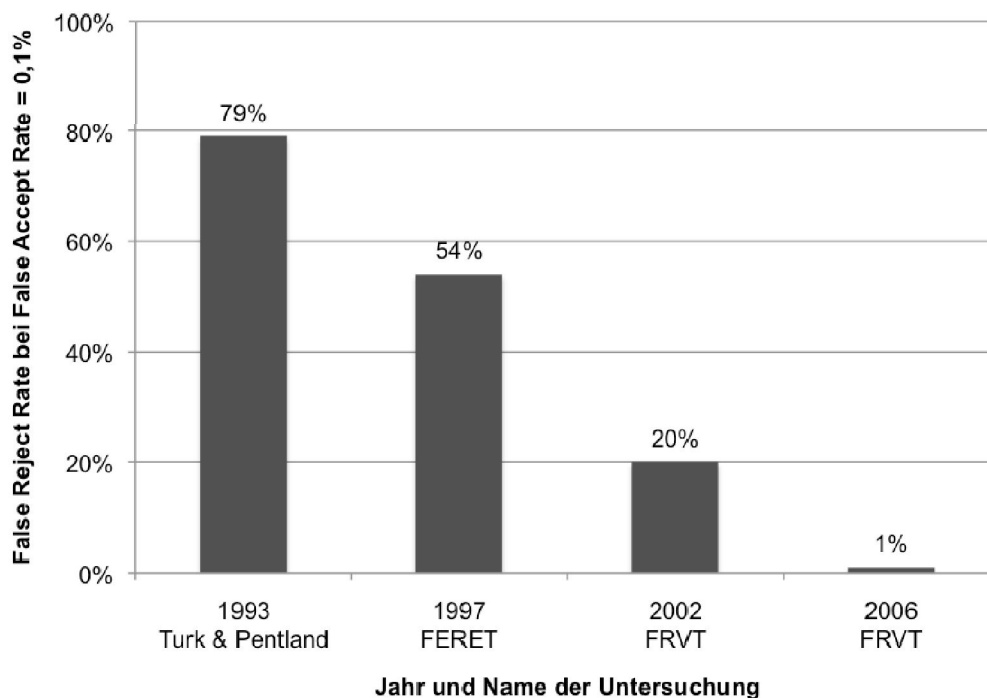


Abbildung 3: Fehlerrate von state-of-the-art Gesichtserkennungsverfahren, dokumentiert durch verschiedene Untersuchungen. Quelle: [1]

## Fazit

Seit dem Aufkommen erster Verfahren zur automatischen Objekterkennung in den 70er Jahren wurden die Verfahren stetig, teils drastisch verbessert. Aber selbst im Jahr 2002 lag die Fehlerrate bei den Gesichtserkennungsverfahren noch bei rund 20% [1], was einen Einsatz in sicherheitsrelevanten Bereichen praktisch ausschließt. Anders stellt sich die Situation in den letzten Jahren da, wo Verfahren mit unter 1% Fehlerrate unabhängig verifiziert werden konnten.

Von derart kleinen Fehlerraten sind Verfahren zur Erkennung beliebiger Objekte heutzutage allerdings noch weit entfernt. Vielversprechende Ansätze ergeben sich durch den Einsatz von Support Vector Machines zur Klassifizierung.

## Literatur

- [1] P.J. Phillips, W.T. Scruggs, A.J. O'Toole, P.J. Flynn, K.W. Bowyer, C.L. Schott, M. Sharpe: „FRVT 2006 and ICE 2006 Large-Scale Experimental Results“ in *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(5), 2010
- [2] L. Goldmann, U. Mönich, T. Sikora: „Robust Face Detection Based on Components and Their Topology“ in *Proc. SPIE, Vol. 6077, 60771V*, 2006
- [3] P. Viola, M. Jones: „Rapid Object Detection Using a Boosted Cascade of Simple Features“ in *Proc. Conf. on Computer Vision and Pattern Recognition*, 2001
- [4] H. Schneiderman, T. Kanade: „A Histogram-based Method for Detection of Faces and Cars“ in *ICIP00: International Conference on Image Processing*, 2000
- [5] G. Bradski, Adrian Kaehler: „Learning OpenCV Computer Vision with the OpenCV Library“, O'Reilly, 2008
- [6] C. Chang, C. Lin: „LIBSVM - A Library for Support Vector Machines“, [www.csie.ntu.edu.tw/~cjlin/libsvm](http://www.csie.ntu.edu.tw/~cjlin/libsvm) aufgerufen am 25.09.2010
- [7] pixolution GmbH: „Visuelle Bildsortierung“, [www.pixolution.de/sites/VisualSorting\\_de.html](http://www.pixolution.de/sites/VisualSorting_de.html) aufgerufen am 25.09.2010
- [8] J. Shawe-Taylor, N. Cristianini: „Kernel Methods for Pattern Analysis“, Cambridge University Press, 2004