Invariante Objekterkennung

Daniel Keysers
keysers@informatik.rwth-aachen.de

Lehrstuhl für Informatik VI
RWTH Aachen
D-52056 Aachen

Betreuer der Arbeit:	Prof. Dr.-Ing H. Ney
Art der Arbeit:	Diplom-Arbeit
Fachbereiche der GI:	Bildverstehen (1.0.4), Maschinelles Lernen (1.1.3)

Zusammenfassung
Invarianz der Erkennungsmethode gegenüber Transformationen spielt eine große Rolle in der Objekterkennung, da schon kleine Veränderungen eines Bildes wesentlichen Einfluss auf die Klassifikation haben können. Die vorliegende Arbeit fasst die Diplomarbeit des Autors auf diesem Gebiet zusammen und geht insbesondere auf die so genannte Tangentendistanz im Zusammenhang mit statistischen Klassifikatoren ein. Die durchgeführten Experimente zeigen, dass der gewählte Ansatz die Erkennungsleistung signifikant verbessert.

1 Einleitung

In der Mustererkennung beschäftigt man sich mit der Entwicklung von Klassifikatoren, die Entscheidungsfunktionen implementieren: $x \in \mathbb{R}^D \longmapsto r(x) = k \in \{1, \ldots, K\}$. Im Spezialfall der Objekterkennung in Bildern besteht dabei der Merkmalsvektor $x \in \mathbb{R}^{I \times J}$ aus einem zweidimensionalen, diskreten Intensitätsmuster, das einer der vorgegebenen Klassen $1, \ldots, K$ zugeordnet werden soll. Hierbei besitzt man oft a-priori Wissen über Transformationen, die die Klassenzugehörigkeit der Objekte nicht verändern. So stellt beispielsweise ein gedrehtes Bild eines Objektes in den meisten Fällen immer noch dasselbe Objekt dar. Dieses Wissen möchte man ausnutzen, um die Entscheidung des Klassifikators zu verbessern, indem man Invarianz von r bzgl. der betrachteten Transformationen anstrebt. Hierzu existieren verschiedene Ansätze, die auf zahlreichen Datensammlungen evaluiert werden können. Als Gütemaß dient dabei die (empirische) Fehlerrate, d.h. das Verhältnis der Klassifikationsfehler zur Gesamtzahl der Klassifikationen.

2 Invariante Objekterkennung

Statistische Mustererkennung. Die gewählte Entscheidungsfunktion basiert häufig auf einer Diskriminantenfunktion $g(x, k)$ in der Art, dass der Klassifikator diejenige Klasse k zu gegebenem x auswählt, die die Diskriminante maximiert. Dabei werden für g z.B. künstliche neuronale Netze oder polynomielle Funktionen benutzt. In der statistischen Mustererkennung wird $g(x, k) \in \{p(k|x),\ p(k)p(x|k),\ \log[p(k)p(x|k)]\}$ verwendet (Bayes'sche Entscheidungsregel, beweisbar optimal für bekannte Verteilungen und das Ziel minimale Fehlerrate). Dabei müssen passende Modelle für die Wahrscheinlichkeitsdichten gewählt werden, deren Parameter an Hand der gegebenen Trainingsdaten bestimmt werden. Wählt man z.B. Gauß'sche Kernel Densities zur Modellierung der klassenbedingten Wahrscheinlichkeiten $p(x|k)$, so ergibt sich die Entscheidungsfunktion unter geeigneten Annahmen als Funktion der Euklidischen Distanzen $\|x - x_{nk}\|$, wobei x_{nk} das Trainingsdatum n aus Klasse k bezeichnet. Eine Möglichkeit, Invarianz zu erreichen, besteht daher in der Verwendung von invarianten Distanzmaßen, die die Euklidische Distanz ersetzen können.

Ansätze zur Invarianz. Unter der Voraussetzung, dass die Transformationen bekannt sind (etwa affine, projektive, Strichdicke, Helligkeit ...) können verschiedenen Ansätze zur Invarianz verfolgt werden. Eine triviale, aber selten handhabbare Methode besteht darin, alle Transformationen der zu vergleichenden Daten zu erzeugen und die minimale Distanz zu betrachten. Andere Möglichkeiten beinhalten Normalisierung (alle Daten werden in ein kanonisches Format gebracht, z.B. auf eine bestimmte Größe skaliert), Extraktion invarianter Merkmale, Datenvervielfältigung (Hinzufügen von transformierten Daten zu den Trainingsdaten bzw. kombinierte Entscheidung an Hand von transformierten Testdaten [1]) und die schon erwähnten invarianten Distanzmaße, auf die im Folgenden kurz eingegangen werden soll.

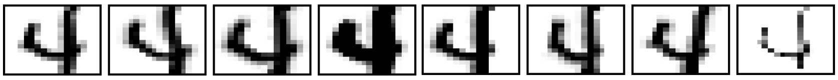

Abbildung 1: Beispiele zur Tangentenapproximation (affine Transformationen und Strichdicke)

Invariante Distanzmaße. Zu den invarianten Distanzmaßen gehört die von SIMARD 1993 eingeführte Tangentendistanz [8]. Betrachtet man alle von einer Menge von Transformationen und einem Datum erzeugten Daten, so bilden diese eine Mannigfaltigkeit (vorzustellen als Oberfläche) im hochdimensionalen Merkmalsraum. Eine ideal invariante Distanz zwischen zwei Merkmalsvektoren kann dann definiert werden als minimale Distanz zwischen den von ihnen erzeugten Mannigfaltigkeiten, welche im Allgemeinen allerdings schwer zu bestimmen ist. Daher kann man sich auf die Tangenten an die Mannigfaltigkeiten beschränken, die von den partiellen Ableitungen der Transformationen nach den Parametern in den gegebenen Punkten aufgespannt werden. Diese lineare Näherung erlaubt die effiziente Bestimmung der minimalen Distanz und erreicht für kleine Transformationen eine gute Beschreibung, wie in Abbildung 1 gezeigt (Originalbild links). Genauere Näherungen sind z.B. durch Methoden ähnlich dem Newton-Verfahren möglich, bringen jedoch einen höheren Aufwand mit sich.

Während die Tangentendistanz globale Transformationen des Bildes berücksichtigt (das ganze Bild wird z.B. gedreht), können auch lokale Transformationen eine Rolle spielen. Eine Möglichkeit, Invarianz gegenüber lokalen Transformation zu erreichen besteht in der Verwendung des Image Distortion Model, auch bekannt als Hausdorff-Distanz oder lokale Perturbation. Dabei wird in der Berechnung des lokalen, pixelweisen Distanzanteils der am besten passende Pixel des Referenzbildes in einer lokalen Umgebung gesucht und die Gesamtdistanz ergibt sich durch Summation.

Diese Methode erlaubt, genau wie die Tangentendistanz, zahlreiche Erweiterungen und eine Kombination der beiden Methoden ist effizient zu erreichen, indem bei einer Distanzberechnung zunächst der Tangentenansatz zur Registrierung verwendet wird und dann das Distortion Model angewendet wird [5]. Die theoretische Betrachtung der Distanzmaße im probabilistischen Kontext erlaubt Rückschlüsse auf die Eigenschaften der zu Grunde liegenden Verteilungen und andererseits ist eine Herleitung der Tangentendistanz aus der Annahme von Variation in den Daten möglich [4, 2]. Weiterhin lassen sich die beiden Ansätze als Spezialfälle eines generischen beschreiben.

3 Experimente

Datensammlungen. Der Großteil der Experimente wurde auf der US Postal Service Datensammlung durchgeführt, die Bilder handgeschriebener, segmentierter und normalisierter Ziffern von US Briefumschlägen enthält (16×16 Pixel, 256 Graustufen, 7.291 Trainingsbilder, 2.007 Testbilder). Ein Vorteil dieser Sammlung ist, dass viele publizierte Ergebnisse verschiedener Forschergruppen existieren, was einen fairen Vergleich der Verfahren ermöglicht. Tabelle 1 zeigt einen Überblick über die erzielten Erkennungsleistungen auf USPS. Weitere Experimente wurden mit der ähnlichen, aber größeren NIST Datensammlung und auf einer Datenbank mit 1.617 Röntgenbildern des IRMA Projektes der RWTH Aachen (Image Retrieval in Medical Applications [3]) durchgeführt.

Ergebnisse. In den Experimenten zeigte sich, dass die Verwendung der Tangentendistanz, die Vervielfachung der Daten und die Benutzung von Kernel Densities die Fehlerrate erheblich senken konnten. Da verschiedenen Klassifikatoren entwickelt wurden, konnten deren Ergebnisse zu einer gemeinsamen Entscheidung kombiniert werden („Bagging“), was die Fehlerrate weiter auf das sehr gute Ergebnis

Tabelle 1: Stand der Technik USPS

*: Trainingsdaten um 2.400 maschinengedruckte Ziffern erweitert

Methode		Fehlerrate USPS [%]
Menschliche Fehlerrate (SIMARD'93 [8, 7])		2.5
Neuronales Netz (LeNet1, LECUN'90 [7])		4.2
Invariante Support Vektoren (SCHÖLKOPF'98 [6])		3.0
Tangentendistanz (SIMARD'93 [8])		*2.5
Boosting (DRUCKER'93 [7])		*2.6
Diese Arbeit:	1-NN, Euklidische Distanz	5.6
	TD, 1-NN	3.3
	TD, Erweiterungen	2.4
	TD, Erweiterungen, Bagging	2.2

von 2.2% senkte. Das Image Distortion Model brachte keine Verbesserungen bei der Ziffernerkennung, jedoch deutliche bei der Klassifiaktion von Röntgenbildern, wo jedoch die Datenvervielfachung keine Verbesserungen erzielte. Dieses Beispiel illustriert, dass die Wahl der Methode zur Klassifikation im Allgemeinen stark von den Datencharakteristiken abhängt.

4 Zusammenfassung und Ausblick

Invariante Distanzmaße wie Tangentendistanz und Image Distortion Model stellen einen effektiven Ansatz zur invarianten Objekerkennung dar. Dies zeigt sich in den guten erzielten Erkennungsergebnissen auf verschiedenen Datensammlungen. Die probabilistische Interpretation der Invarianz erlaubt eine theoretische Beschreibung und Gewinnung von Einsichten, z.B. zur Schätzung von Tangentenvektoren. Der Erfolg der verschiedenen Methoden hängt stark von den zu erkennenden Daten ab und ein wichtiger Tradeoff besteht bei der Stärke der erlaubten Transformationen zwischen zugelassenen Annäherungen an die „richtigen“ und an die „falschen“ Referenzdaten.

Als Themen zukünftiger Arbeit lassen sich z.B. die Verwendung diskriminativer Verfahren und die Übertragung der Methoden auf Bildindizierung, bewegte Bilder und Spracherkennung anführen. Auch eine eventuelle exakte Beschreibung der Mannigfaltigkeiten und die Betrachtung anderer Invarianzmodelle, wie z.B. Warping-Modelle, bieten weitere Ansatzpunkte. Ein weiteres Thema ist die Verfolgung eines holistischen Ansatzes in der Erkennung von Objekten unbekannter Lage in Bildern.

Literatur

[1] J. Dahmen, D. Keysers, M. O. Güld, and H. Ney. Invariant Image Object Recognition using Mixture Densities. In *Proceedings 15th International Conference on Pattern Recognition*, Barcelona, Spain, September 2000. In press.

[2] J. Dahmen, D. Keysers, M. Pitz, and H. Ney. Structured Covariance Matrices for Statistical Image Object Recognition. In *22. DAGM Symposium Mustererkennung 2000*, Springer, Kiel, Germany, September 2000. In press.

[3] J. Dahmen, T. Theiner, D. Keysers, H. Ney, T. Lehmann, and B. Wein. Classification of Radiographs in the 'Image Retrieval in Medical Applications' System (IRMA). In *Proceedings of the 6th International RIAO Conference on Content-Based Multimedia Information Access, Paris, France*, pages 551–566, April 2000.

[4] D. Keysers, J. Dahmen, and H. Ney. A Probabilistic View on Tangent Distance. In *22. DAGM Symposium Mustererkennung 2000*, Springer, Kiel, Germany, September 2000. In press.

[5] D. Keysers, J. Dahmen, T. Theiner, and H. Ney. Experiments with an Extended Tangent Distance. In *Proceedings 15th International Conference on Pattern Recognition*, Barcelona, Spain, September 2000. In press.

[6] B. Schölkopf, P. Simard, A. Smola, and V. Vapnik. Prior Knowledge in Support Vector Kernels. In M. I. Jordan, M. J. Kearns, and S. A. Solla, editors, *Advances in Neural Inf. Proc. Systems*, volume 10. MIT Press, pages 640–646, 1998.

[7] P. Simard, Y. Le Cun, J. Denker, and B. Victorri. Transformation Invariance in Pattern Recognition — Tangent Distance and Tangent Propagation. In G. Orr and K.-R. Müller, editors, *Neural networks: tricks of the trade*, volume 1524 of *Lecture Notes in Computer Science*, Springer, Heidelberg, pages 239–274, 1998.

[8] P. Simard, Y. Le Cun, and J. Denker. Efficient Pattern Recognition Using a New Transformation Distance. In S. Hanson, J. Cowan, and C. Giles, editors, *Advances in Neural Inf. Proc. Systems*, volume 5, Morgan Kaufmann, San Mateo CA, pages 50–58, 1993.