

# Variabilitätsmodellierung für die Bilderkennung

Daniel Keyzers

Forschungsbereich Bildverstehen und Mustererkennung  
Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) GmbH  
daniel.keyzers@dfki.de

**Abstract:** Dem Computer das Verstehen von Bildern zu ermöglichen ist eine große Herausforderung. Durch die geeignete Behandlung typischer Veränderungen in Bildern kann man die Erkennung von Objekten in vielen Fällen verbessern. In der hier vorgestellten Dissertation werden Modelle zur Beschreibung von Variabilität in Bildern für die erscheinungsbasierte Klassifikation von Objekten untersucht. Die Modelle bestimmen dabei die Ähnlichkeit zwischen zwei gegebenen Bildern, die für die Klassifikation verwendet wird. Als theoretisches Ergebnis wird erstmals gezeigt, dass die Bestimmung der besten flexiblen Abbildung zwischen zwei Bildern für ein zweidimensionales Modell zur Klasse der NP-harten Probleme gehört. In der praktischen Anwendung stellt sich andererseits als entscheidend für niedrige Fehlerraten heraus, dass ein geeigneter Kontext der Bildpixel bei der Abbildung berücksichtigt wird. Durch die Hinzunahme von Kontext werden auch mit weniger komplexen Modellen sehr gute Fehlerraten erreicht. Die Anwendung der vorgestellten Methoden wird vor allem für die Klassifikation handgeschriebener Zeichen und die Kategorisierung von medizinischen Bildern untersucht, wobei in beiden Fällen Ergebnisse erzielt werden, die im Vergleich mit denen anderer Forschergruppen sehr gut abschneiden.

## 1 Einführung

Für die automatische Erkennung von Bildinhalten ist die geeignete Modellierung der typischen Veränderungen wichtig, denen Objekte in Bildern unterliegen. Ein menschlicher Betrachter kann ein Objekt auch dann leicht erkennen, wenn es leicht verändert in einem Bild erscheint. Einem Computer muss man diese Fähigkeit durch die geeignete Modellierung dieser Variabilität vermitteln. Abbildung 1 zeigt verschiedene Beispiele handgeschriebener Ziffern ‘2’, die sich in ihrem Erscheinungsbild deutlich unterscheiden. Diese soll ein Klassifikationssystem –etwa zur Erkennung von Postleitzahlen– jedoch alle in dieselbe Klasse einordnen. Zur geeigneten Behandlung von Variabilität gibt es je nach konkreter Aufgabenstellung verschiedene Ansätze. Die hier vorgestellte Dissertation [Key06]



Abbildung 1: Beispiele für Bildvariabilität: handgeschriebene Ziffern ‘2’ aus dem USPS-Datensatz.

behandelt einige dieser Ansätze und zeigt unterschiedliche Verbesserungsmöglichkeiten.

Zunächst werden lineare Variabilitätsmodelle diskutiert und ihre Beziehung zu Gauß-Verteilungen dargestellt. Diese Beziehung erlaubt es, bekannte Schätzverfahren zur Bestimmung der Vektoren, die die Variabilität repräsentieren, einzusetzen. Darüber hinaus wird der Zusammenhang zwischen dem diskriminativen ‚Maximum Entropy‘-Ansatz und Gauß-Verteilungen behandelt, der dann verwendet wird um die neuartige ‚Maximum Entropy Linear Discriminant Analysis‘ herzuleiten. Weiterhin werden diskrete Verformungsmodelle der Ordnung null, eins und zwei untersucht, die Bildpixel auf Bildpixel abbilden. Dabei wird die Ordnung der Modelle durch die Art der angesetzten zweidimensionalen Verformungseinschränkungen bestimmt. Es wird erstmals gezeigt, dass die Bestimmung der besten Abbildung zwischen zwei Bildern für das Modell der Ordnung zwei zur Klasse der NP-harten Probleme gehört. Für die diskreten Modelle ist es entscheidend, dass ein geeigneter Kontext der Pixel verwendet wird, um niedrige Fehlerraten zu erreichen. Durch die Hinzunahme von Kontext werden auch mit den weniger komplexen Modellen geringerer Ordnung sehr gute Fehlerraten erreicht. Des Weiteren werden Modelle zur Behandlung von starker Variabilität in der Objekterkennung erörtert, die auf der Verwendung von lokalen Bildteilen basieren, und die in Kombination mit diskriminativen Trainingsverfahren sehr konkurrenzfähige Ergebnisse erlauben.

Die Anwendung der vorgestellten Methoden wird vor allem für die Klassifikation handgeschriebener Zeichen und die Kategorisierung von medizinischen Bildern untersucht, wobei in beiden Fällen hervorragende Ergebnisse erzielt werden. Insbesondere wird eine Fehlerquote von 0,52% auf der bekannten Datenbank MNIST erzielt und eine von 12,6% auf der Datenbank IRMA-10.000, welche die niedrigste in der 2005 durchgeführten ImageCLEF-Evaluation war. Es wird außerdem gezeigt, dass die Variabilitätsmodelle auch die Resultate für die erscheinungsbasierte Erkennung von Gesten und Gebärdensprache verbessert, was die breite Anwendbarkeit der beschriebenen Modelle unterstreicht.

Im Folgenden werden einige dieser Ergebnisse zusammenfassend dargestellt. Auf eine wissenschaftliche Einordnung in den Kontext anderer Veröffentlichungen wird in dieser Kurzfassung aus Platzgründen weitgehend verzichtet; eine ausführliche Darstellung der Zusammenhänge findet sich in der Dissertationsschrift selbst [Key06].

## 2 Bildklassifikation und invariante Distanzfunktionen

Aufgabe der Mustererkennung oder Klassifikation [DHS01] ist es, aus gegebenen Daten eine optimale Entscheidung abzuleiten. In der hier betrachteten Aufgabenstellung wird aus einem Bild (den Daten) ein Merkmalsvektor  $x$  extrahiert, der im einfachsten Fall die Bildwerte selbst enthalten kann. Die Entscheidungen, die hier betrachtet werden, sind die Zuordnung zu einer Klasse  $k \in \{1, \dots, K\}$ . Optimalität wird dabei durch das Erreichen der niedrigsten Fehlerrate, d.h. der Anzahl der falschen Klassenzuordnungen definiert. Die geringste erwartete Fehlerrate wird erreicht, wenn die Zuordnung entsprechend der Bayes'schen Entscheidungsregel getroffen wird:  $x \mapsto \hat{k} = \arg \max_k P(k|x)$ . Die Wahrscheinlichkeitsverteilung  $P(k|x)$  ist nun in der Praxis unbekannt, so dass sie (oder ge-



Abbildung 2: Zu klassifizierendes Testbild (links) und zwei Referenzbilder. Nach Euklidischer Distanz ist das Testbild näher am ersten Referenzbild. Eine Distanzfunktion, die invariant gegenüber Veränderungen der Strichdicke ist, ergibt hingegen, dass die Distanz zum zweiten, richtigen, Referenzbild kleiner ist.

eignete abgeleitete Größen) aus gegebenen Trainingsdaten gelernt werden müssen. Die verschiedenen Möglichkeiten zur Modellierung und zum Training werden dann auf unabhängigen Testdaten evaluiert. Hierbei besitzt man oft a-priori Wissen über Variationen, die die Klassenzugehörigkeit der Objekte nicht verändern. Zur Behandlung der auftretenden Variabilität gibt es verschiedene Ansätze:

- Normalisierung: die Transformation der Daten in eine kanonische Darstellung unter Eliminierung der Variabilität
- invariante Merkmale: die Extraktion von Merkmalsvektoren  $x$  aus den Daten, die sich unter der Variabilität möglichst wenig ändern
- virtuelle Daten: Erzeugung von zusätzlichen variierten Trainingsdaten aus den gegebenen, um den Klassifikator unempfindlicher gegen die Variabilität zu machen
- invariante Distanzfunktionen: Verwendung einer Distanzfunktion zwischen Test- und Referenzdaten, die sich unter der Variabilität möglichst wenig ändert

In diesem Beitrag beschränken wir uns auf den letzten der genannten Ansätze und die unter bestimmten Modellannahmen ableitbare sogenannte Nächste-Nachbar-Regel (NNR) als Entscheidungsregel

$$x \mapsto \arg \min_k \left\{ \min_n d(x, \mu_{kn}) \right\}$$

für gegebene Trainingsdaten  $\mu_{kn}, n = 1, \dots, N_k$  der Klassen  $k = 1, \dots, K$ . Dabei wird ein neues Testbild derjenigen Klasse zugeordnet, der auch das am nächsten gelegene Trainingsbeispiel angehört. Diese Entscheidungsregel erlaubt es uns, verschiedene Ansätze zur Modellierung von Variabilität in der Distanzfunktion  $d(\cdot, \cdot)$  unterzubringen und diese zu vergleichen. Die Idee einer invarianten Distanzfunktion ist in Abbildung 2 veranschaulicht.

Wir betrachten die folgende Darstellung einer invarianten Distanzfunktion, die die meisten praktisch verwendeten Ansätze einschließt:

$$d(x, \mu) = \min_{\alpha \in \mathcal{A}} \left\{ d'(x, t(\mu, \alpha)) \right\}$$

Die invariante Distanz  $d(x, \mu)$  ergibt sich dabei als die minimale Distanz  $d'$  zwischen dem Testbild  $x$  und allen Transformationen (Variationen)  $t(\mu, \alpha)$  der Referenz  $\mu$  mit zulässigen Transformationsparametern  $\alpha$ . Die einfachere Distanz  $d'$  wird hier als Euklidische Distanz gewählt. Die Wahl der zulässigen Transformationen  $(t, \mathcal{A})$  bestimmt nun die Menge



Abbildung 3: Beispiele zur Tangentenapproximation (affine Transformationen und Strichdicke).

der modellierten Transformationen sowie die Komplexität der Minimierung. Dabei ist das Ziel, eine Distanzfunktion zu finden, die möglichst gut zwischen verschiedenen Bildklassen unterscheidet, also nicht nur niedrige Distanzen für Bilder der gleichen Klasse liefert, sondern gleichzeitig große Distanzen für Bilder verschiedener Klassen ergibt.

Zu den invarianten Distanzmaßen gehört die von Simard eingeführte Tangentendistanz [SLD93]. Betrachtet man alle durch eine Menge von Transformationen und einem Bild erzeugbaren Bilder, so bilden diese eine Mannigfaltigkeit 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 3 gezeigt (Originalbild links). Bei Bestimmung der Tangentenvektoren durch  $v_l = \partial t(\mu, \alpha) / \partial \alpha_l$  ergibt sich die Tangentendistanz als

$$d(x, \mu) = \min_{\alpha \in R^L} \left\{ \|x - (\mu + \sum_l \alpha_l v_l)\|^2 \right\}$$

Eine statistische Untersuchung der Tangentendistanz ergibt eine Darstellung durch Gaußmodelle, deren Kovarianzmatrix eine bestimmte Struktur aufweist [KMND04]. Die sich daraus ergebenden Schätzverfahren können dazu genutzt werden, diese Variabilitätsmodelle auch dann zu schätzen, wenn die tatsächlichen Transformationen der Daten nicht einfach zu ermitteln sind, etwa in der Spracherkennung [KMDN01].

Während die Tangentendistanz globale Transformationen des Bildes als lineare Funktion im Merkmalsraum betrachtet, können auch lokale Transformationen eine wichtige Rolle spielen. Diese können zwar auch linearisiert betrachtet werden, jedoch führt dies zu einer unhandlichen Beschreibung, so dass hier nichtlineare Modelle erfolgreicher sind [KDG07]. Wir betrachten dazu ein Testbild  $x = \{x_{ij}\}$  und ein Referenzbild  $\mu = \{\mu_{rs}\}$ . Zwischen diesen wird eine Verformungsabbildung  $(r_{11}^{IJ}, s_{11}^{IJ}) : (i, j) \mapsto (r_{ij}, s_{ij})$  eingeführt, die gewisse Einschränkungen erfüllen muss:  $(r_{11}^{IJ}, s_{11}^{IJ}) \in \mathcal{M}$ . Die sich ergebende Distanzfunktion ist dann

$$d(x, \mu) = \min_{(r_{11}^{IJ}, s_{11}^{IJ}) \in \mathcal{M}} \left\{ d'(x, \mu_{(r_{11}^{IJ}, s_{11}^{IJ})}) \right\}$$

wobei  $\mu_{(r_{11}^{IJ}, s_{11}^{IJ})}$  das entsprechend der Abbildung verformte Referenzbild bezeichnet. Hierbei handelt es sich also um diskrete Verformungsmodelle deren Ordnung von der Gestaltung der Menge  $\mathcal{M}$  der zulässigen Abbildungen abhängt [KDN04, KGN04]. In den durchgeführten Experimenten hat sich dabei herausgestellt, dass es für diese diskreten Modelle



Abbildung 4: Illustration der Reduktion des 3SAT-Problems auf ein Problem des Findens einer optimalen Abbildung zwischen zwei Bildern zum Beweis der NP-Vollständigkeit des letzteren.

entscheidend ist, dass ein geeigneter Kontext der Pixel verwendet wird, um niedrige Fehlerraten zu erreichen. Dies kann etwa durch die Verwendung der örtlichen Ableitungen in einer kleinen Umgebung jedes Pixels geschehen. Durch die Hinzunahme von Kontext werden auch mit den weniger komplexen Modellen geringerer Ordnung sehr gute Fehlerraten erreicht. Im Folgenden werden vier mögliche Modelle kurz informell dargestellt.

**2DW** Beim 2-Dimensional Warping wird gefordert, dass ein benachbarter Pixel in der Abbildung nicht zu stark von der Position abweichen darf, die ihm als direkter Nachbar gehören würde. Dies wird für horizontale und für vertikale Nachbarn gefordert und es ergibt sich so ein Modell der Ordnung 2, dessen Minimierung zur Klasse der NP-vollständigen Probleme gehört.

**P2DHMM** Beim Pseudo-2-Dimensionalen Hidden Markov Modell wird zur Verringerung der Komplexität die Abhängigkeit zwischen den Bildspalten aufgehoben, so dass sich ein Modell der Ordnung 1 ergibt.

**P2DHMDM** Beim Pseudo-2-Dimensionalen Hidden Markov Distortion Modell wird die starre Abbildung ganzer Spalten aufeinander im P2DHMM durch das Erlauben von Abweichungen auf benachbarte Spalten (jedoch weiterhin ohne Abhängigkeiten) flexibilisiert. Dieses Modell gehört ebenfalls der Ordnung 1 an.

**IDM** Im Image Distortion Modell wird die Modellierung der relativen Abhängigkeiten vollständig aufgehoben und nur eine maximale absolute Abweichung vorgegeben. Die Minimierung ist dadurch in diesem Modell der Ordnung 0 sehr einfach.

In der Arbeit wird erstmals gezeigt, dass die Bestimmung der besten Abbildung zwischen zwei Bildern für Modelle der Ordnung zwei (die Abhängigkeiten in beiden Bilddimensionen berücksichtigen) zur Klasse der NP-harten Probleme gehört [KU03]. Die Beweisidee basiert auf der Reduktion des 3SAT-Problems auf das Abbildungsproblem und ist in Abbildung 4 illustriert. Ausgehend von einer Formel in 3-KNF (einer Instanz des 3SAT-Problems) wird der Abhängigkeitsgraph gezeichnet und es werden daraus zwei Bilder konstruiert, die eine Abbildung unter 2DW-Beschränkungen genau dann zur Distanz 0 erlauben, wenn die Formel erfüllbar ist.

In Abbildung 5 sind einige Beispielabbildungen unter den angesprochenen Verformungsmodellen dargestellt. Dabei wurden zur Illustration Gesichter verwendet, wobei Fotos eines Zwillingspaars als die gleiche Klasse gelten sollen. Zu jedem Modell wird das beste verformte Bild sowie das entstehende Verformungsgitter angegeben. Man erkennt, dass die Verformung unter 2DW tatsächlich den gestellten Anforderungen entspricht, sowie die charakteristische Spaltenstruktur des P2DHMM, die im P2DHMDM gelockert wird, und

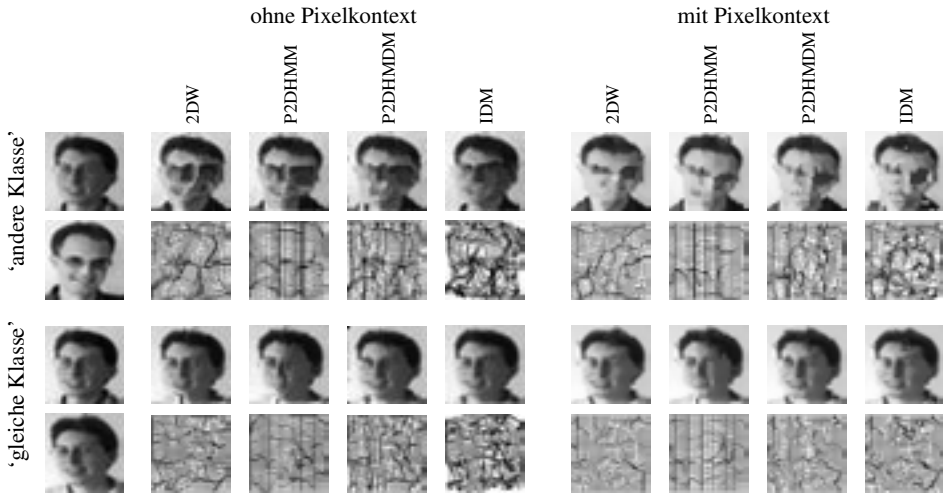


Abbildung 5: Beispiele für nichtlineare Verformungsabbildungen (siehe Text).

die deutlich zufälliger wirkende Struktur des IDM. Es ist zu sehen, dass die Abbildung zwischen Bildern der gleichen Klasse in allen Modellen gute Ähnlichkeit erreicht, dass bei verschiedenen Klassen eine hohe Distanz aber nur erreicht wird, wenn Modelle hoher Ordnung oder eine Berücksichtigung des Pixelkontextes verwendet werden. Dies bedeutet, dass Modelle geringer Ordnung mit Pixelkontext eine gute Alternative zu den komplexen Modellen hoher Ordnung sind, was sich in den quantitativen Experimenten bestätigt.

### 3 Experimentelle Ergebnisse

In diesem Abschnitt werden experimentelle Ergebnisse auf zwei von vielen anderen Forschern als Benchmark verwendeten Datensätzen vorgestellt. Für die Verformungsmodelle wurde die NNR unter Berücksichtigung der drei nächsten Nachbarn und ein Pixelkontext der Größe  $3 \times 3$  der Gradienten (18 Werte pro Pixel) verwendet. Tabelle 1 vergleicht die erzielten Ergebnisse mit den besten aus der Literatur bekannten Ergebnissen anderer Forschergruppen. Es zeigt sich, dass insbesondere das P2DHMDM und das in der Implementierung sehr einfache IDM auf beiden Datensätzen sehr gute Ergebnisse erzielen. Das komplexere 2DW-Modell erzielt schlechtere Ergebnisse, es wurden jedoch auf Grund der Komplexität hier Näherungsverfahren verwendet. Das beste Ergebnis auf MNIST erzielt eine Kombination von vier verschiedenen Verfahren, was zeigt, dass selbst sehr ausgereifte Methoden, die alle ähnlich gute Ergebnisse erzielen, unterschiedliche Fehler machen und daher in der Kombination gewinnen können. Ein sehr anschauliches Beispiel für diesen Effekt kann man auch bei der als ‚local patches + tangent distance‘ gekennzeichneten Methode sehen, die die oben beschriebene Tangentendistanz mit einer Klassifikation durch lokale Bildteile verbindet [KPNV02]. Hier ist eine Kombination deshalb besonders er-

Tabelle 1: Kurzstatistik der Datensätze und Ergebnisse (Fehlerraten). Aus Platzgründen wird für die Referenzen der mit [\*\*\*] gekennzeichneten Ergebnisse anderer Forschergruppen auf [Key06] verwiesen. (SVM = Support Vector Machine)

Name	Beispielbilder	Größe	# Train	# Test
USPS		16x16	7 291	2 007
MNIST		28x28	60 000	10 000

Verfahren/USPS	FR[%]
no matching, 1-NN	5.6
SVM + invariant features [***]	3.5
invariant SVM [***]	3.0
tangent distance [Key06]	3.0
2DW [Key06]	2.7
extended SVM training [***]	2.5
P2DHMM [Key06]	2.5
tangent distance, virtual data [Key06]	2.4
IDM [Key06]	<b>2.4</b>
extended SVM training [***]	2.2
Hungarian matching [Key06]	2.2
local patches + tangent distance [Key06]	2.0
P2DHMDM [Key06]	<b>1.9</b>

Verfahren/MNIST	FR[%]
no matching, 1-NN	3.1
tangent distance, virtual data [Key06]	1.0
biol. inspired features, SVM [***]	0.72
invariant SVM [***]	0.68
shape context matching (*) [***]	0.63
extended SVM training [***]	0.60
biol. inspired features [***]	0.59
invariant SVM (*) [***]	0.56
IDM (*) [Key06]	<b>0.54</b>
P2DHMDM [Key06]	<b>0.52</b>
preprocessing, SVM [***]	0.42
distortions+, neural net (*) [***]	0.42
combination of (*) [Key06]	0.35

folgreich, weil die beiden Verfahren sehr komplementäre Eigenschaften haben, wie dies in Abbildung 6 illustriert ist. Im Allgemeinen hat die Verwendung von lokalen Bildteilen bei der Klassifikation von Objekten in Fotografien in der jüngeren Vergangenheit eine besondere Bedeutung erlangt und kann insbesondere in Kombination mit diskriminativen Trainingsverfahren sehr konkurrenzfähige Ergebnisse erlauben [DKN05].

Die dargestellten Verformungsmodelle lassen sich auch bei der Erstellung geeigneter Referenzdaten verwenden. Abbildung 7 zeigt die sich ergebenden Referenzen. Die Referenzen, die sich unter Berücksichtigung der Verformung ergeben, erscheinen deutlich schärfer. Das Ergebnis von 4.9% Fehlerrate, dass sich mit diesen Referenzen ergibt, ist eines der besten bekannten Ergebnisse für lediglich ein Referenzbild pro Klasse.

### 4 Weitere Anwendungen

Die beschriebenen Verfahren lassen sich auch in weiteren Anwendungen, die hier kurz vorgestellt werden, vorteilhaft zur Modellierung von Variabilität in Bildern einsetzen.

- Bei der Kategorisierung von medizinischen Bildern [KDN<sup>+</sup>03] konnten mit dem



Abbildung 6: Beispiele für Ziffern, die bei Verwendung der Tangentendistanz richtig klassifiziert werden, aber nicht bei Verwendung von lokalen Teilbildern (links, besondere Variabilität in Strichdicke und affinen Veränderungen) und umgekehrt (rechts, fehlende und zusätzliche Teile).



USPS Fehlerraten [%]

Distanz mit Verformung	Verformung in Mittelwertberechnung	
	nein	ja
nein	18.6	26.1
ja	25.3	4.9

Abbildung 7: Referenzen der 10 Klassen des USPS-Datensatzes mit und ohne Berücksichtigung von Verformungen sowie die mit diesen 10 Referenzen erzielten Fehlerraten.

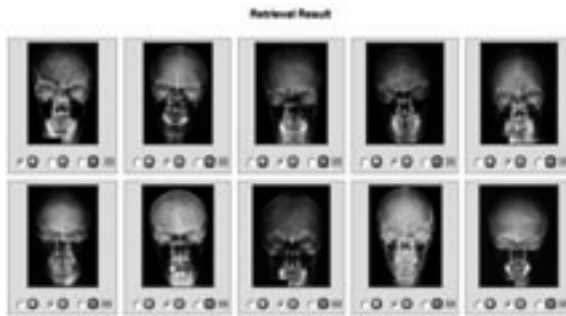


Abbildung 8: Beispiel zu Image Retrieval in Medical Applications (IRMA).

IDM bei der 2005 durchgeführten internationalen ImageCLEF-Evaluation die besten Ergebnisse erzielt werden. Die Arbeiten bauen in diesem Zusammenhang auf einem Projekt zum Image Retrieval in Medical Applications (IRMA) auf, das eine inhaltsbasierte Suche in medizinischen Bildarchiven zum Ziel hat (vgl. Abbildung 8).

- Bei der Gestenerkennung konnten die Ergebnisse sowohl unter Verwendung der Tangentendistanz als auch des IDM deutlich verbessert werden. In dem durchgeführten Experiment war das Ziel 14 dynamische Gesten der Hand in Videodaten zu unterscheiden.
- Bei der Erkennung von Wörtern der amerikanischen Gebärdensprache in Videos konnten ebenfalls mit Tangentendistanz und IDM deutliche Verbesserungen erzielt werden. Hier ergab sich eine weitere Verbesserung durch die Verwendung der Tangentendistanz für die Distanzberechnung der lokalen Pixelkontexte innerhalb des IDM.
- Bei der Unterscheidung verschiedener Ausprägungen von roten Blutzellen wurden



sehr gute Ergebnisse durch die Kombination der Methode der lokalen Bildteile und der Tangentendistanz erzielt.

- Bei der Erkennung von handgeschriebenem Fließtext durch Hidden Markov Modelle führt die Tangentendistanz zur Modellierung der vertikalen Variabilität innerhalb einer Zeile zu deutlichen Verbesserungen. Hier hilft insbesondere die statistische Interpretation bei der Integration in bestehende Modelle.

## 5 Zusammenfassung

In dieser Kurzdarstellung wurden lineare und nichtlineare Modelle der Variabilität von Bildern und ihr Einsatz in der Bildklassifikation vorgestellt. Dabei stellte sich heraus, dass voll zwei-dimensionale Modell der Verformung einen hohen Rechenaufwand erfordern. Die Verwendung von lokalen Kontextinformation in der diskreten Abbildung zwischen zwei Bildern kann jedoch zu sehr guten Ergebnissen auch mit weniger komplexen Modellen führen. Dabei ergeben sich hervorragende Ergebnisse für die Erkennung von handgeschriebenen Ziffern und von medizinischen Bildern und deutliche Verbesserungen auch für Videodaten. Für die Behandlung der Bildvariabilität bei der Objekterkennung in allgemeinen Fotografien erscheint insbesondere die Verwendung von Methoden vielversprechend, die die Klassifikation an Hand von lokalen Bildteilen durchführen.

## Danksagung

Die hier vorgestellte Dissertation wäre ohne die fachliche Unterstützung vieler Personen nicht zu Stande gekommen. An dieser Stelle möchte ich nur kurz einigen davon danken: Prof. Hermann Ney für die Betreuung und Ermöglichung der Arbeit, Thomas Deselaers für verschiedenste Beiträge, sowie Jörg Dahmen, Christian Gollan, Philippe Dreuw, Tobias Kölsch, Morteza Zahedi und Walter Unger.

## Literatur

- [DHS01] Richard Duda, Peter Hart und David Stork. *Pattern Classification*. John Wiley & Sons, Inc., New York, 2. Auflage, 2001.
- [DKN05] Thomas Deselaers, Daniel Keyzers und Hermann Ney. Discriminative Training for Object Recognition using Image Patches. In *CVPR 2005, Int. Conf. on Computer Vision and Pattern Recognition*, Bd. II, Seiten 157–162, San Diego, CA, Juni 2005.
- [KDG07] Daniel Keyzers, Thomas Deselaers, Christian Gollan und Hermann Ney. Deformation Models for Image Recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2007. In press.

- [KDN<sup>+</sup>03] Daniel Keysers, Jörg Dahmen, Hermann Ney, Berthold Wein und Thomas Lehmann. Statistical Framework for Model-based Image Retrieval in Medical Applications. *J. Electronic Imaging*, 12(1):59–68, Januar 2003.
- [KDN04] Daniel Keysers, Thomas Deselaers und Hermann Ney. Pixel-to-Pixel Matching for Image Recognition using Hungarian Graph Matching. In *DAGM 2004, Pattern Recognition, 26th DAGM Symposium*, Bd. 3175 von *Lecture Notes in Computer Science*, Seiten 154–162, Tübingen, August 2004.
- [Key06] Daniel Keysers. *Modeling of Image Variability for Recognition*. Dissertation, RWTH Aachen, März 2006.
- [KGN04] Daniel Keysers, Christian Gollan und Hermann Ney. Local Context in Non-linear Deformation Models for Handwritten Character Recognition. In *ICPR 2004, 17th Int. Conf. on Pattern Recognition*, Bd. IV, Seiten 511–514, Cambridge, UK, August 2004.
- [KMDN01] Daniel Keysers, Wolfgang Macherey, Jörg Dahmen und Hermann Ney. Learning of Variability for Invariant Statistical Pattern Recognition. In *12th European Conf. on Machine Learning*, Bd. 2167 von *Lecture Notes in Computer Science*, Seiten 263–275, Freiburg, September 2001.
- [KMND04] Daniel Keysers, Wolfgang Macherey, Hermann Ney und Jörg Dahmen. Adaptation in Statistical Pattern Recognition using Tangent Vectors. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(2):269–274, Februar 2004.
- [KPNV02] Daniel Keysers, Roberto Paredes, Hermann Ney und Enrique Vidal. Combination of Tangent Vectors and Local Representations for Handwritten Digit Recognition. In *SPR 2002, Int. Workshop on Statistical Pattern Recognition*, Jgg. 2396 of *Lecture Notes in Computer Science*, Seiten 538–547, Windsor, Ontario, Canada, August 2002.
- [KU03] Daniel Keysers und Walter Unger. Elastic Image Matching is NP-complete. *Pattern Recognition Letters*, 24(1–3):445–453, Januar 2003.
- [SLD93] Patrice Simard, Yann Le Cun und John Denker. Efficient Pattern Recognition Using a New Transformation Distance. In S.J. Hanson et al., Hrsg., *Advances in Neural Information Processing Systems 5*, Seiten 50–58. Morgan Kaufmann, 1993.



**Daniel Keysers** studierte Informatik an der RWTH Aachen und der Universidad Complutense de Madrid. Nach dem Abschluss als Diplom-Informatiker mit Auszeichnung im Juni 2000 war er als wissenschaftlicher Mitarbeiter am Lehrstuhl für Sprachverarbeitung und Mustererkennung der RWTH Aachen tätig, sowie als Visiting Researcher an der Universidad Politécnica de Valencia. Im März 2006 schloss er mit der Note Summa Cum Laude seine Dissertation an der RWTH Aachen ab. In 2006 verbrachte er einen Forschungsaufenthalt bei Microsoft Research in Redmond und wurde im gleichen Jahr mit dem Olympus-Preis der Deutschen Arbeitsgemeinschaft für Mustererkennung ausgezeichnet. Seit Mai 2005 arbeitet er im Forschungsbereich Bildverstehen und Mustererkennung des Deutschen Forschungszentrums für Künstliche Intelligenz in Kaiserslautern als Senior Researcher, wo sein Forschungsschwerpunkt in der Dokumentbildanalyse und der statistischen Mustererkennung liegt.