

Invariante Merkmale für die Bilderkennung

von Ilja Bezrukov

Betreuer: Daniel Keyzers

Inhalt

1	Einleitung	2
2	Grundlagen	2
3	Übersicht über Methoden zur Erzeugung von invarianten Merkmalen	3
3.1	Normalisierung	4
3.2	Differentiale Invarianten	4
3.3	Integrale Invarianten	4
4	Erzeugung von Invarianten durch Integration	4
4.1	Beispiel: Invariante Integration für eindimensionale Translation	5
4.2	Berechnung der invarianten Merkmale für die Gruppe der Kongruenzen	7
5	Histogramme von lokalen Merkmalen	9
5.1	Motivation	9
5.2	Beispiel	9
6	Reduzierung der Laufzeit durch nichtdeterministische Schätzung	9
6.1	Motivation	9
6.2	Schätzung von globalen invarianten Merkmalen	10
6.3	Schätzung von Histogrammen von lokalen Merkmalen	11
7	Anwendungen	11
7.1	Analyse von 3D-Objekten	11
7.2	Inhaltsbasierte Bildsuche	12
7.3	Beispiel: SIMBA	12
8	Zusammenfassung	14

Zusammenfassung

Invariante Merkmale eignen sich gut für die Erkennung und Klassifizierung von Objekten, die aus unterschiedlichen Positionen aufgenommen wurden. Dieser Beitrag gibt einen Überblick über die existierenden Verfahren zur Berechnung von invarianten Merkmalen und erläutert ihre Erzeugung durch Integration[1]. Bei diesem Verfahren müssen die Bilder nicht segmentiert werden, was eine automatische Verarbeitung ermöglicht. Invariante Merkmale können in Histogrammen dargestellt werden, was sie besonders für die inhaltsbasierte Suche interessant macht. Die Berechnung lässt sich durch eine nichtdeterministische Abschätzung auf eine konstante Laufzeit reduzieren. Als ein Anwendungsbeispiel wird SIMBA[3], ein System für inhaltsbasierte Bildsuche vorgestellt.

Keywords: Invariante Merkmale, Invarianten, Inhaltsbasierte Bildsuche, Integrale Invarianten, Musterrererkennung

1 Einleitung

Bei vielen Anwendungsgebieten der Bilderkennung ist es notwendig, bestimmte Objekte in einem Bild zu erkennen, wobei die Aufnahmeposition und die Lage der Objekte keine Rolle spielt. Ein Ansatz, dieses Problem zu lösen, ist die Verwendung von invarianten Merkmalen. Bei der invarianten Bilderkennung versucht man, Objekte unabhängig von der Position, aus der die Aufnahme erfolgt ist, zu identifizieren. Dazu werden aus dem Bild Merkmale extrahiert, die invariant gegenüber Transformationen des abgebildeten Objektes sind. Die komplexen Transformationen, die entstehen, wenn ein dreidimensionales Objekt auf eine zweidimensionale Bildebene projiziert wird, werden dabei durch einfachere approximiert.

In diesem Beitrag werden die bekannten Verfahren zur Berechnung von invarianten Merkmalen kurz erläutert, und die Methode der integralen Invarianten im Detail vorgestellt. Die letztgenannte Methode hat gegenüber den anderen Methoden den Vorteil, dass sie sich leicht auf dreidimensionale Daten übertragen lässt, und zum Beispiel bei der Klassifizierung von dreidimensionalen CT-Scans eingesetzt werden kann.

2 Grundlagen

Zuerst soll formal festgelegt werden, wann zwei Objekte gleich sind. Der Objektraum wird dazu in Äquivalenzklassen eingeteilt, so dass zwei Objekte gleich bzw. äquivalent sind, gdw. eine Transformation $g \in G$ existiert, die ein Objekt in das andere überführt:

$$X_1 \cong X_2 \Leftrightarrow \exists g \in G : X_2 = g(X_1)$$

Da es sich bei Bildern um zweidimensionale Objekte handelt, ist es sinnvoll, G als eine Gruppe von zunächst beliebigen Koordinatentransformationen zu betrachten.

Die Gruppenelemente können durch Parametervektoren λ parametrisiert werden. Jedem Vektor λ entspricht dann ein Gruppenelement g . Die Dimension von λ bezeichnet die Zahl der Freiheitsgrade. Zum Beispiel kann die Gruppe der Kongruenzen durch drei Parameter beschrieben werden: Verschiebung in x- und y-Richtung und Drehung. Also ist die Zahl der Freiheitsgrade= $\dim(\lambda) = 3$

Möchte man herausfinden, ob zwei Objekte äquivalent unter G sind, dann wäre der naive Ansatz, alle möglichen Transformationen zu testen. In der Praxis würde das zu einer sehr schlechten Laufzeit führen, weil die Anzahl der möglichen Transformationen exponentiell mit der Anzahl der Freiheitsgrade wächst. Ein besserer Weg ist es, Merkmale zu finden, die bei der Transformation des Objektes invariant bleiben. Anhand dieser Merkmale kann dann festgestellt werden, ob 2 Objekte äquivalent sind.

Die grundlegende Idee ist es, eine Abbildung T zu finden, welche in der Lage ist, die invarianten Merkmale eines Objektes zu extrahieren. Alle Objekte aus einer Äquivalenzklasse sollen von T auf einen Punkt im Merkmalsraum abbildet werden:

$$X_1 \cong X_2 \Rightarrow T(X_1) = T(X_2) \tag{1}$$

An diese Abbildung werden bestimmte Anforderungen gestellt, damit sie für reale Anwendungen benutzt werden kann, denn z.B. erfüllt bereits die Abbildung $F(x) = 0 \forall x$ die Gleichung (1):

- Vollständigkeit, um Mehrdeutigkeiten bei der Abbildung in den Merkmalsraum zu vermeiden. Die Menge der Objekte, die von T auf denselben Punkt, wie ein Objekt x abgebildet werden, sei definiert als

$$I_T(x) := \{x' | T(x') = T(x)\}$$

Die Menge der Objekte in einer Äquivalenzklasse bezüglich einer Gruppe G , also aller Objekte, die sich aus x durch Anwendung von Transformationen aus der Gruppe G generieren lassen, wird als Orbit bezeichnet:

$$G(x) := \{x' | x' \cong x\}$$

Aus der Gleichung (1) kann man folgern, dass

$$G(x) \subseteq I_T(x)$$

Eine Abbildung ist vollständig, wenn die beiden Mengen gleich sind, und die Gleichung (1) zu einer Äquivalenzbeziehung wird:

$$X_1 \cong X_2 \Leftrightarrow T(X_1) = T(X_2)$$

Alle Äquivalenzklassen werden von T auf unterschiedliche Punkte im Merkmalsraum abgebildet. Abbildungen mit einem hohen Vollständigkeitsgrad, d.h. mit $|G(x)| \approx |I_T(x)|$ zu finden, kann in der Praxis nicht einfach sein, weshalb die Forderung nach Vollständigkeit je nach Anwendungsgebiet reduziert werden kann zu der Forderung, eine bestimmte Menge von Orbits trennen zu können. Vollständige Abbildungen sind grundsätzlich nicht-linear.

- Stetigkeit: Ähnliche Objekte sollen auf benachbarte Punkte im Merkmalsraum abgebildet werden. Der Wert der Abbildung T soll sich bei geringen Unterschieden zwischen den Objekten ebenfalls nur gering verändern.
- Unterscheidungsleistung: Die Abbildung sollte robust gegenüber Bildstörungen sein.
- Verhalten gegenüber Bildstörungen: Die Abbildung sollte gleichzeitig leicht zu interpretierende Ergebnisse in Bezug auf Bildstörungen. liefern
- Komplexität: die Abbildung sollte effizient berechenbar sein. Die hier im Detail vorgestellte Methode der integralen Invarianten besitzt eine lineare Komplexität, die sogar mit Hilfe der nichtdeterministischen Schätzung zu konstanter Laufzeit reduziert werden kann.

3 Übersicht über Methoden zur Erzeugung von invarianten Merkmalen

Nachdem im letzten Abschnitt die mathematischen Grundlagen und gewünschte Eigenschaften von invarianten Merkmalen vorgestellt wurden, wird in diesem Abschnitt ein Überblick über die Methoden zur Erzeugung von Invarianten gegeben.

Laut [1] können alle Methoden, mit deren Hilfe man systematisch invariante Merkmale berechnen kann, in drei Kategorien eingeteilt werden:

1. Normalisierung
2. Differentiale Invarianten
3. Integrale Invarianten

Nachfolgend werden diese drei Ansätze kurz erläutert.

3.1 Normalisierung

Die Idee bei der Methode der Normalisierung ist es, einen Standardvertreter aus jeder Äquivalenzklasse zu finden. Dazu existieren verschiedene Verfahren, z.B. die affine Normalisierung. Bei diesem Verfahren wird der minimale Umfang des Objektes bei konstanter Fläche gesucht, es wird sozusagen versucht, das Objekt einem Kreis ähnlich werden zu lassen. Bei dieser Methode ist eine Objektkontur erforderlich, d.h. das Bild muss vorher segmentiert werden. Nach der Anwendung der affinen Normalisierung ist noch die Anwendung einer weiteren Methode erforderlich, weil die Objekte noch skaliert, verschoben oder rotiert sein können.

Die Nachteile bei der Normalisierung sind die Notwendigkeit einer Objektkontur sowie die mögliche Uneindeutigkeit bei der Suche nach Standardvertretern. Der Vorteil ist die Behandlung von komplexeren Transformationen, als es bei der Methode der integralen Invarianten möglich ist.

3.2 Differentiale Invarianten

Das Prinzip bei der Methode der differentialen Invarianten ist die Unveränderlichkeit der invarianten Merkmale bei kleinen Transformationen auszunutzen. Bei dieser Methode entstehen komplizierte Differentialgleichungen, die zu lösen sind, so dass sie in der Praxis selten angewendet wird.

3.3 Integrale Invarianten

Die Idee bei der Methode der integralen Invarianten ist es, über die gesamte Gruppe der Transformationen zu integrieren. Es wird eine Funktion f auf allen Objekten des Orbits ausgewertet, und anschließend normiert. Man berechnet dazu das Integral über den gesamten Orbit. Dieses Integral ist unabhängig von dem Parametervektor λ , und deshalb invariant:

$$T(x) = \frac{1}{|G|} \int_G f(gx) dg$$

Die Vorteile dieser Methode in der Praxis sind die effiziente Berechenbarkeit und keine Notwendigkeit der Segmentierung, der Nachteil ist Beschränkung auf einfache Transformationen.

4 Erzeugung von Invarianten durch Integration

In diesem Abschnitt wird im Detail auf die Erzeugung von integralen Invarianten bei Bildern eingegangen. Der Einfachheit halber werden nur Grauwertbilder behandelt. Es ist auch möglich, Farbbilder zu behandeln, wobei die Merkmale dann für jeden Farbkanal einzeln berechnet werden.

Da die verwendeten Indizes sowohl reell als auch ganzzahlig sein können, wird die folgende Notation für den Grauwert an einer Koordinate $n = (n_0, n_1)$ eingeführt:

$$X(n) \quad n \in \mathbf{R}^2 \text{ oder } n \in \mathbf{N}^2$$

Die Indizes werden als periodisch fortgesetzt betrachtet, es wird eine Modulo-Operation bezüglich der Bilddimensionen angewendet. Die zu integrierende Funktion f wird im Folgenden als Kernfunktion bezeichnet.

Als Transformationsgruppe G wird die Gruppe der Kongruenzen behandelt, so dass für eine Koordinate n gilt:

$$(gX)(n) = X(n'), \quad \text{mit } n' = \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix} n + \begin{pmatrix} t_0 \\ t_1 \end{pmatrix}$$

Die hier vorgestellten integralen Invarianten sind nicht invariant gegenüber komplizierteren Transformationen, insbesondere der Skalierung, aber robust gegenüber kleineren Veränderungen der Größe und Form.

Die Konstruktion einer invarianten Abbildung $T[f](X)$ für eine Kernfunktion f und ein Grauwertbild X geschieht durch die folgende Integration:

$$T[f](X) := \frac{1}{|G|} \int_G f(gX) dg \quad (2)$$

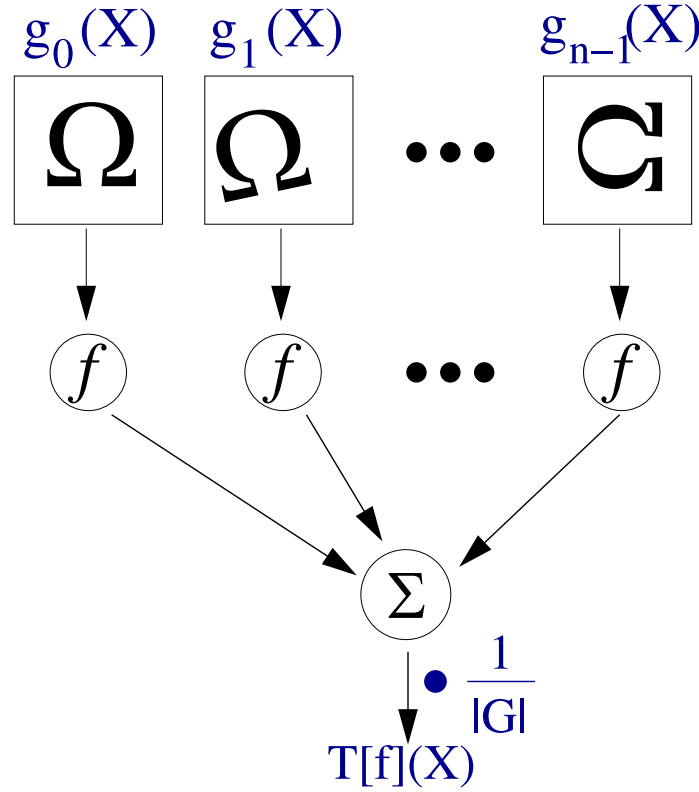


Abbildung 1: Die anschauliche Bedeutung der Gleichung (2) für eine Gruppe mit n Elementen

Die anschauliche Bedeutung der Formel ist in Abbildung 1 dargestellt. Die Kernfunktion f wird auf alle Objekte im Orbit angewendet, die Ergebnisse werden addiert und normiert. Diese Technik wird als „group averaging“ bezeichnet, mit $T[f](X)$ als „group average“ oder Gruppenmittelwert. In der Praxis arbeitet man mit diskreten Bildkoordinaten und das Integral wird durch eine Summe ersetzt. Für die Gruppe der Kongruenzen wird Gleichung (2) wie folgt umgeschrieben:

$$T[f](X) = \frac{1}{2\pi N_0 N_1} \int_{n_0=0}^{N_0} \int_{n_1=0}^{N_1} \int_{\phi=0}^{2\pi} f(gX) d\phi dn_1 dn_0 \quad (3)$$

wobei N_0 und N_1 die Dimensionen des Bildes sind.

Die Methode der integralen Invarianten lässt sich auch auf dreidimensionale Daten anwenden. Die dreidimensionalen euklidischen Transformationen können durch einen dreidimensionalen Translationsvektor t und drei Winkel ϕ, ψ, τ parametrisiert werden. Das Integral in Gleichung 3 wird erweitert zu

$$T[f](X) = \frac{1}{8\pi^3 N_0 N_1 N_2} \int_{n_0=0}^{N_0} \int_{n_1=0}^{N_1} \int_{n_2=0}^{N_2} \int_{\phi=0}^{2\pi} \int_{\psi=0}^{2\pi} \int_{\tau=0}^{2\pi} f(gX) d\tau d\psi d\phi dn_2 dn_1 dn_0$$

Aufmerksamkeit verdient schließlich die Wahl der Kernfunktion f . Für diese werden Monome vom Grad $\leq G$ verwendet, weil es mit diesen möglich ist, vollständige Merkmalsmengen zu konstruieren.

Im nächsten Abschnitt soll die invariante Integration an einem einfachen Beispiel vorgeführt werden.

4.1 Beispiel: Invariante Integration für eindimensionale Translation

Es wird die Gruppe G der eindimensionalen zyklischen Translationen betrachtet. Jedes $g \in G$ kann durch einen Parameter n , der die Länge der Verschiebung ausdrückt, parametrisiert werden. Auf Signalen der Länge N sieht das Integral über die Gruppe dann wie folgt aus:

$$T[f](x) = \frac{1}{N} \int_{n=0}^N f(gx) dt$$

Orbit	Signale
O_0	(0,0,0,0)
O_1	(0,0,0,1) (0,1,0,0) (0,0,1,0) (0,0,0,1)
O_2	(0,0,1,1) (0,1,1,0) (1,1,0,0) (1,0,0,1)
O_3	(0,1,0,1) (1,0,1,0)
O_4	(0,1,1,1) (1,1,1,0) (1,1,0,1) (1,0,1,1)
O_5	(1,1,1,1)

Tabelle 1: Orbits der äquivalenten Signale.

Die Signale sind in der Praxis diskret, so dass das Integral durch eine Summe ersetzt werden kann:

$$T[f](x) = \frac{1}{N} \sum_{n=0}^N f(gx)dt \quad (4)$$

Als ein konkretes Beispiel sollen nun binäre Signale der Länge 4 betrachtet werden:

$$x = (x_0, x_1, x_2, x_3)^T \quad x_i \in \{0, 1\}$$

Als nächstes sollen die Orbits aller äquivalenten Signale angegeben werden. Da es nur 16 verschiedene Signale gibt, lässt sich das leicht bewerkstelligen. Die Orbits sind in Tabelle 1 angegeben. Um eine vollständige Merkmalsmenge zu erhalten, soll der Gruppenmittelwert aller Monome $x_0^{e_0} x_1^{e_1} x_2^{e_2} x_3^{e_3}$ mit der Eigenschaft $\sum_{i=0}^3 e_i \leq 4$ berechnet werden. Da über die Gruppe der zyklischen Verschiebungen gemittelt werden soll, brauchen Monome, die sich nur durch eine zyklische Verschiebung der e_i unterscheiden, nicht berücksichtigt zu werden. Außerdem gilt $x_i^{e_i} = x_i \quad \forall e_i > 0$, da es sich um binäre Signale handelt. Unter Beachtung dieser Einschränkungen erhält man die folgenden Monome:

$$\begin{aligned} f_0(x) &= x_0 \\ f_1(x) &= x_0 x_1 \\ f_2(x) &= x_0 x_2 \\ f_3(x) &= x_0 x_1 x_2 \\ f_4(x) &= x_0 x_1 x_2 x_3 \end{aligned}$$

Die Gruppenmittelwerte für ein Signal x sind wie folgt:

$$\begin{aligned} T[f_0](x) &= \frac{1}{4} \sum_{n=0}^3 f_0(g_n x) \\ &= \frac{1}{4} (x_0 + x_1 + x_2 + x_3) \\ T[f_1](x) &= \frac{1}{4} (x_0 x_1 + x_1 x_2 + x_2 x_3 + x_3 x_0) \\ T[f_2](x) &= \frac{1}{4} (x_0 x_2 + x_1 x_3 + x_2 x_0 + x_3 x_1) \\ &= \frac{1}{2} (x_0 x_2 + x_1 x_3) \\ T[f_3](x) &= \frac{1}{4} (x_0 x_1 x_2 + x_1 x_2 x_3 + x_2 x_3 x_0 + x_3 x_0 x_1) \\ T[f_4](x) &= (x_0 x_1 x_2 x_3) \end{aligned}$$

Die Berechnung erfolgte, indem die Gleichung (4) auf ein Signal x angewendet wurde. Die Ergebnisse für die verschiedenen Orbits sind in Tabelle 2 zusammengefasst. Man kann erkennen, dass bereits zwei invariante Abbildungen, z.B. $T[f_0](x)$ und $T[f_1](x)$ genügen, um eine vollständige Merkmalsmenge zu erhalten. Die Orbits lassen sich damit eindeutig unterscheiden.

Orbits	$T[f_0](x)$	$T[f_1](x)$	$T[f_2](x)$	$T[f_3](x)$	$T[f_4](x)$
O_0	0	0	0	0	0
O_1	1/4	0	0	0	0
O_2	1/2	1/4	0	0	0
O_3	1/2	0	1/2	0	0
O_4	3/4	1/2	1/2	1/4	0
O_5	1	1	1	1	1

Tabelle 2: Merkmale für die Orbits aus der Tabelle 1.

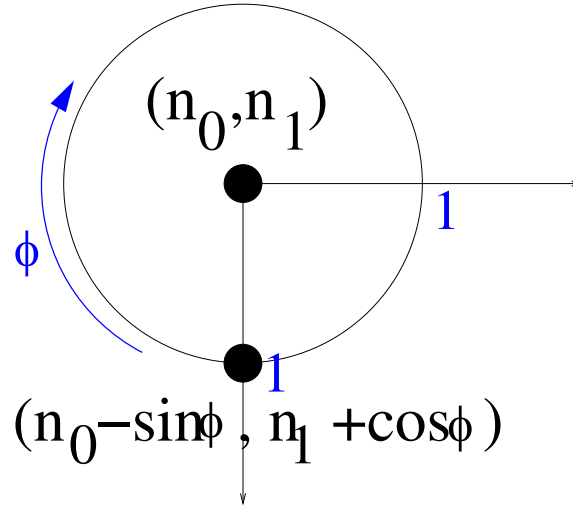


Abbildung 2: Berechnung der Kernfunktion $X(0,0) \cdot X(0,1)$

4.2 Berechnung der invarianten Merkmale für die Gruppe der Kongruenzen

In diesem Abschnitt wird die Berechnung der invarianten Merkmale für die Gruppe der Kongruenzen an einem Beispiel erklärt und einige Eigenschaften der integralen Invarianten besprochen.

Das Integral über die Gruppe der Kongruenzen für eine Kernfunktion f und ein Bild mit den Dimensionen N_0, N_1 hat die in Gleichung (3) gegebene Form. Wenn man $f = X(0,0) \cdot X(0,1)$ setzt, ergibt sich daraus:

$$[f](X) = \frac{1}{2\pi N_0 N_1} \int_{n_0=0}^{N_0} \int_{n_1=0}^{N_1} \int_{\phi=0}^{2\pi} X(n_0, n_1) \cdot X(n_0 - \sin \phi, n_1 + \cos \phi) d\phi dn_1 dn_0 \quad (5)$$

Die Wahl der Koordinaten ist in Abbildung 2 veranschaulicht. Bei der Diskretisierung der Koordinaten werden die Integrale wieder durch eine Summe ersetzt.

Die Berechnung von $T[f](X)$ lässt sich in zwei Teile gliedern: zuerst wird das innere Integral über ϕ ausgewertet. n_0 und n_1 bleiben dabei konstant, und $X(0,1)$ nimmt die Grauwerte der Punkte an, die auf einem Kreis mit dem Radius 1 um (n_0, n_1) liegen. Die Auswertung der beiden äußeren Integrale bedingt, dass die Berechnung für jeden Punkt im Bild wiederholt wird.

Diese Vorgehensweise lässt sich in einer Zweischritt-Strategie darstellen:

1. Berechne das Integral über den Winkel ϕ , d.h. suche alle Pixel, die den Abstand 1 von (n_0, n_1) haben, und multipliziere ihren Grauwert mit dem Grauwert von $X(n_0, n_1)$.
2. Wiederhole diesen Prozess für alle (n_0, n_1) mit $0 \leq n_0 < N_0, 0 \leq n_1 < N_1$, addiere die Ergebnisse und normiere anschließend

Aus dieser Strategie folgt, dass nur Grauwerte in einer bestimmten Entfernung von (n_0, n_1) Einfluss auf das Ergebnis im ersten Schritt haben. Diese Entfernung wird als Supportgröße bezeichnet. Der Support

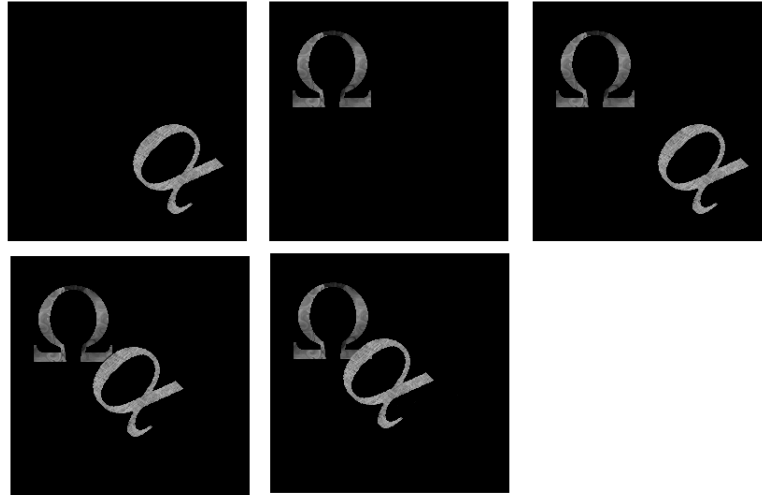


Abbildung 3: Zwei verschieden texturierte Objekte, kombiniert mit und ohne Überlappung.

Bild	$T[X(0, 0) \cdot X(0, 7)]$
alpha	418,697
gamma	210,689
alpha & omega	629,386
alpha & omega verdeckt I	637,347
alpha & omega verdeckt II	620,521

Tabelle 3: Ergebnisse der Berechnung von invarianten Merkmalen auf Objekte aus Abbildung 3.

$S(f)$ einer Kernfunktion f ist definiert als die Menge aller Pixelkoordinaten, die das Ergebnis von $f(X)$ beeinflussen. In diesem Beispiel ist $S(f) = \{(0, 0), (0, 1)\}$. Die Supportgröße $s(S(f))$ ist der Radius des Kreises, aus dem Werte für die Berechnung von $f(X)$ im ersten Schritt entnommen werden.

In diesem Beispiel ist die Supportgröße 1. Kernfunktionen, die auf einem lokalen Bereich operieren, werden als FLS-Funktionen, „functions of local support“, bezeichnet.

Man schreibt sie als f_r , wobei r die Supportgröße ist. Die Supportgröße ist wichtig, wenn man Szenen verarbeiten will, in denen mehrere, unabhängig voneinander transformierte Objekte vorkommen. Bei Szenen mit einem homogenen Hintergrund können invariante Merkmale gebildet werden, solange der Abstand der Objekte zueinander größer ist, als die Supportgröße der verwendeten Kernfunktion.

Leichte Überlappungen verändern das Ergebnis nur geringfügig, und sind in der Praxis daher erlaubt.

Eine weitere, nützliche Eigenschaft ist die Additivität. Berechnet man eine invariante Abbildung T auf zwei Bildern X_1 und X_2 , die je ein Objekt vom Typ 1 und 2 enthalten, so gilt für eine Abbildung X , die beide Objekte zusammen enthält:

$$T(X) = T(X_1) + T(X_2) \quad (6)$$

Die Voraussetzung für die Gültigkeit der Additivitätseigenschaft ist ein Hintergrund mit der Intensität 0. Die Gültigkeit ist leicht einsichtig: Die lokale Auswertung der Kernfunktion ergibt 0 für alle Pixel (n_0, n_1) , die nicht Teil von Objekt 1 in X_1 bzw. Objekt 2 in X_2 sind. Fügt man beide Objekte in ein Bild ein, ist das Ergebnis der Auswertung von T auf X die Summe der beiden Einzelauswertungen.

Auch hier gilt, dass leichte Überlappungen zwischen den Objekten das Ergebnis nur geringfügig beeinflussen.

Für die in Abbildung 3 dargestellten Objekte wurden die invarianten Merkmale basierend auf der Kernfunktion $X(0, 0) \cdot X(0, 7)$ berechnet. Die Ergebnisse sind in Tabelle 3 aufgelistet. Durch die Überdeckung ändert sich das Ergebnis bei ersten Überlappung nur um 1,3 %, bei der zweiten Überlappung um 0,8 %. Die hier vorgestellten invarianten Merkmale sind zudem robust gegenüber topologischen Veränderungen.

Bei einer geringfügigen Deformation des Objektes ändert sich das Gesamtergebnis ebenfalls nur wenig.

5 Histogramme von lokalen Merkmalen

In diesem Abschnitt wird auf die Bildung von Histogrammen als Alternative zur globalen Mittelwertbildung eingegangen.

5.1 Motivation

Die in den vorigen Abschnitten vorgestellten Merkmale lassen sich in der Praxis nur bedingt verwenden. Größere Bildüberdeckungen und Bildänderungen können aufgrund der Bildung des Mittelwertes über das gesamte Bild nur ungenügend behandelt werden. Daher ist es erstrebenswert, Merkmale zu konstruieren, die mehr von den lokalen Bildeigenschaften erhalten. Eine Möglichkeit besteht darin, Histogramme zu verwenden, anstatt einen Mittelwert zu bilden.

In realen Anwendungen wird, wie bereits erwähnt, bei der Berechnung der invarianten Abbildungen die Integration durch eine Summe ersetzt. Gleichung (3) läßt sich daher approximieren durch

$$T[f](X) \approx \frac{1}{PN_0N_1} \sum_{n_0=0}^{N_0-1} \sum_{n_1=0}^{N_1-1} \sum_{p=0}^{P-1} f(g(X))|_{\phi=p \cdot \frac{2\pi}{P}} \quad (7)$$

Anstatt der Bildung des globalen Mittelwerts kann man die Summen durch eine Histogrammoperation ersetzen. Die Invarianz bleibt dabei erhalten. Alternativ kann man die innere Summe stehen lassen, und nur die beiden äußeren Summen ersetzen. Man erhält dann ein Histogramm von lokalen rotationsinvarianten Merkmalen.

Histogramme haben den Vorteil, dass mehrere verschiedene Merkmale in einem mehrdimensionalen Histogramm kombiniert werden können. Man könnte so zwei oder mehrere invariante Abbildungen, die auf verschiedenen Kernfunktionen basieren, in einem Histogramm kombinieren.

Histogramme erlauben eine flexiblere Suche, es ist z.B. möglich, einen Teilbereich eines Bildes als Suchmuster anzugeben und als Ergebnis das zugehörige Gesamtbild zu finden. Man kann auch die gewöhnlichen Farbhistogramme mit den Merkmals-Histogrammen kombinieren, indem man jeden Farbkorb zusätzlich in Merkmalskörbe unterteilt.

Um die Diskretisierung zu glätten, kann man Fuzzy-Histogramme[1] verwenden, die eine stetige Korbzuweisungsfunktion benutzen.

5.2 Beispiel

Basierend auf der bereits aus Abschnitt 4.2 bekannten Funktion $f(X) = X(0,0) \cdot X(0,1)$ wurde ein Histogramm von lokalen Merkmalen zur Abbildung 4 berechnet, wobei die innere Summe über den Winkel gebildet wurde. Zum Vergleich wurde ein Histogramm von dem auf 80% der ursprünglichen Größe verkleinerten Bild berechnet. Beide Histogramme sind in Abbildung 5 dargestellt. Sie haben jeweils 32 Körbe und wurden so normiert, dass die Summe der Körbe 1 ist. Der Übersichtlichkeit halber wurde der erste Korb in der Darstellung weggelassen, da er wegen des schwarzen Bildhintergrundes sehr viele Treffer aufweist.

6 Reduzierung der Laufzeit durch nichtdeterministische Schätzung

In diesem Abschnitt wird auf die Reduzierung der Komplexität von linear auf konstant mit Hilfe einer nichtdeterministischen Schätzung eingegangen.

6.1 Motivation

Obwohl die Laufzeit bei der Berechnung der invarianten Merkmale linear bezüglich der Anzahl der Pixel und der Supportgröße ist, kann dies für viele Anwendungen dennoch zu hoch sein. Für einen 3-D-Datensatz der Größe 128x128x128 und einer Supportgröße von 5 wird die Kernfunktion bereits $30 \cdot 10^9$ Mal ausgewertet. Eine Reduzierung der Laufzeit ist daher wünschenswert, zumal insbesondere im Bereich der



Abbildung 4: Ein Bild, mit schwarzem Hintergrund und hellem Objekt.

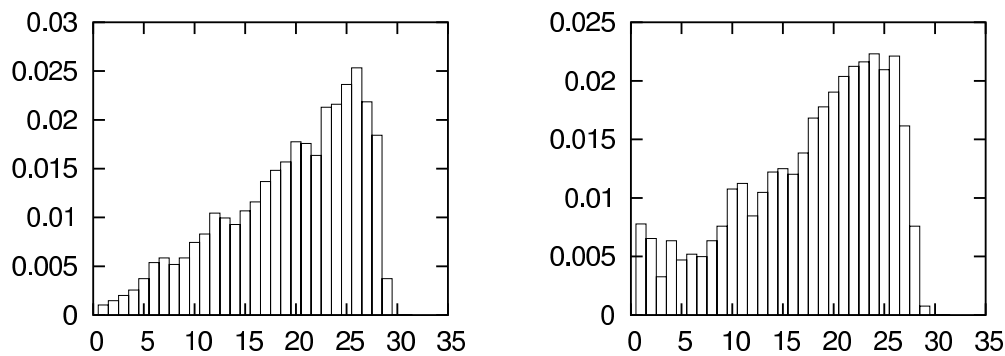


Abbildung 5: Histogramme der Funktion $X(0,0) \cdot X(0,1)$, angewendet auf das Objekt in Abbildung 4 (links) und auf die 80%-Skalierung (rechts).

inhaltsbasierten Bildsuche schnelle Antwortzeiten gewünscht sind.

Die Reduzierung kann man mit Hilfe einer nichtdeterministischen Schätzung durchführen und erreicht so eine Komplexität von $O(1)$. Die Schätzungen von globalen invarianten Merkmalen und von Histogrammen werden in getrennten Abschnitten behandelt.

6.2 Schätzung von globalen invarianten Merkmalen

Das Integral aus Gleichung (3) lässt sich durch die Monte-Carlo-Methode approximieren. Mit Hilfe der Monte-Carlo-Methode lassen sich Integrale vom Typ

$$I = \int_X \frac{1}{|X|} f(x_1, x_2, \dots, x_k) dx_1 dx_2 \dots dx_k \quad \text{mit} \quad |X| = \int_X dx_1 dx_2 \dots dx_k \quad (8)$$

näherungsweise lösen. Der Algorithmus dafür lautet wie folgt:

- für ein Integral der Dimension k erzeuge eine Menge in X gleichverteilter Zufallszahlen $x_{11}, x_{12}, \dots, x_{1k}, x_{21}, x_{22}, \dots, x_{Nk}$
- werte für $n = 1, \dots, N$ die Funktion f aus: $y_n = f(x_{n1}, \dots, x_{nk})$
- Schätze I ab durch $y = \frac{1}{N} \sum_{n=1}^N y_n$

Um das Integral (3) zu approximieren, muss die Kernfunktion f auf einer Menge von zufälligen Tripeln (n_0, n_1, ϕ) ausgewertet und der Mittelwert aus diesen Ergebnissen berechnet werden.

Die erforderliche Anzahl der Samples N , um einen Fehler von höchstens ϵ mit einer Überschreitungswahrscheinlichkeit δ zu erhalten, kann man bestimmen, indem man den zentralen Grenzwertsatz anwendet.

Dieser besagt, dass die Summe vieler unabhängiger Zufallsvariablen normalverteilt ist. Nach Integration der gauß'schen Dichtefunktion von $-\epsilon$ bis ϵ erhält man das folgende Ergebnis:

$$\Phi\left(\frac{\epsilon\sqrt{N}}{\sqrt{V(Y_n)}}\right) \geq 1 - \delta/2 \quad (9)$$

$V(Y_n)$ ist wie üblich die Varianz von Y_n und

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp^{-\frac{1}{2}x^2} dx$$

Die Gleichung (9) soll auf ein konkretes Beispiel angewendet werden: Sei $\epsilon = 0.01$, $\delta = 5\%$, $f(gX) \in [0, 1]$. Für die Varianz $V(Y_n)$ läßt sich die folgende grobe Abschätzung machen:

$$V(Y_n) = E(Y_n^2) - E(Y_n)^2 \leq 1$$

Damit erhält man $\Phi(0.01\sqrt{N}) \geq 0.975$, also $N \geq 38.416$, unabhängig von den Bilddimensionen und der Supportgröße.

Für ein Bild der Größe 128x128 Pixel und die Supportgröße 5 erhält man bereits eine Beschleunigung der Berechnung um den Faktor 100 (38.416 statt 4.200.000 Auswertungen von f). Bei größeren Bildern und 3D-Daten ist der Beschleunigungsfaktor entsprechend höher.

6.3 Schätzung von Histogrammen von lokalen Merkmalen

Die Histogrammverteilung kann aus der relativen Trefferfrequenz der einzelnen Körbe geschätzt werden. Die Histogrammverteilung wird als eine Bernoulli-Serie der Länge N aufgefaßt. Wenn h_n die relative Trefferfrequenz für einen Korb n ist, und p_n die Wahrscheinlichkeit, dann gilt nach dem zentralen Grenzwertsatz für große N :

$$P(|h_n - p_n| \leq \epsilon) \approx 2\Phi\left(\frac{\epsilon\sqrt{N}}{\sqrt{p_n(1-p_n)}}\right) - 1$$

Wegen der Monotonie von Φ und $p_n(1-p_n) \leq 0.25$ lässt sich die folgende Abschätzung machen:

$$\Phi(2\epsilon\sqrt{N}) \geq 1 - \frac{\delta}{2}$$

Für $\epsilon = 0.01$ und $\delta = 5\%$ erhält man für N die Forderung $N \geq 9604$. Dieses Ergebnis liefert eine Fehlerabschätzung für einen Korb. Eine einfache Fehlerabschätzung für das gesamte Histogramm lässt sich durchführen, indem der Fehler für den einzelnen Korb mit der Anzahl der Körbe multipliziert wird. Für Histogramme mit vielen Körben ist diese Abschätzung ungenau, da die geforderte Zahl der Samples N quadratisch mit dem Kehrwert der Fehlerschranke wächst. Für eine genauere Abschätzung sei auf [2] verwiesen.

7 Anwendungen

In diesem Abschnitt soll auf die Möglichkeiten der Anwendung von invarianten Merkmalen eingegangen werden. Globale Merkmale werden bei der Detektion von Defekten in Texturen und der Analyse von medizinischen 3D-Daten angewendet. Eine weitere Anwendung aus dem 3D-Bereich ist die Klassifizierung von 3D-Scans von Pollen. Histogramme von lokalen Merkmalen werden vorwiegend in inhaltsbasierter Bildsuche eingesetzt, da sie robuster gegenüber Bildveränderungen sind und flexiblere Vergleichsmöglichkeiten erlauben.

7.1 Analyse von 3D-Objekten

Im medizinischen Bereich können die globalen invarianten Merkmale eingesetzt werden, um Volumenbilder zu vergleichen, die von Computer-Tomographen generiert wurden. Auf diese Weise lassen sich zwei zu verschiedenen Zeitpunkten aufgenommene Datensätze eines Patienten, die möglicherweise ungleich ausgerichtet und verschoben sind, miteinander vergleichen. Die Ergebnisse können anschließend interpretiert werden, so dass zum Beispiel versucht werden kann, morphologische Prozesse festzustellen. Dazu ist im

Gegensatz zu herkömmlichen Techniken keine Segmentierung notwendig, allerdings ist Sorgfalt bei der Definition der Metriken, nach denen die Ergebnisse interpretiert werden, vonnöten. Neben der Anwendung in der Medizin sind auch andere Bereiche denkbar, in denen 3D-invariante Merkmale verwendet werden können.

7.2 Inhaltsbasierte Bildsuche

Im Bereich der inhaltsbasierten Bildsuche lassen sich die in diesem Beitrag vorgestellten Histogramme von lokalen Merkmalen gut verwenden, vor allem wenn eine Automatisierung erforderlich ist. Das ist insbesondere dann der Fall, wenn die Datenmenge sehr groß ist, so dass eine manuelle Verarbeitung nicht möglich ist, und das Bildmaterial inhomogen ist, so dass Segmentierungstechniken nicht anwendbar sind. Die häufig verwendeten Farbhistogramme haben den Nachteil, dass alle Informationen über die textuellen Eigenschaften verloren gehen, während Methoden, die auf Segmentierung aufbauen, nur in Spezialfällen einsetzbar sind.

Histogramme von lokalen Merkmalen erweitern die gewöhnlichen Farb/Graustufenhistogramme um Informationen über die lokale Nachbarschaft. Der histogrammbasierte Ansatz ermöglicht es, Bilder zu finden, wenn nur ein Teil als Suchmuster vorliegt und ist robust gegenüber Objektüberschneidungen oder Verdeckungen. Zum Vergleich von Histogrammen können verschiedene Methoden verwendet werden, z.B. der χ^2 -Test und der Durchschnitt zweier Histogramme.

Der χ^2 -Test gibt Auskunft darüber, wie weit sich zwei Verteilungen unterscheiden. Für ein Suchmuster q und einen Kandidaten h lautet die Formel für den symmetrischen χ^2 -Test wie folgt:

$$\chi^2(q, h) = \sum_n \frac{(q_n - h_n)^2}{q_n + h_n}$$

Ein Beispiel für die Anwendung ist in Abbildung 6 angegeben. Als Suchmuster wurde das Bild unten in der Abbildung verwendet. Die Menge der Kandidaten bestand aus 4 gleichen verschobenen und gedrehten, und zwei anders texturierten Objekten. Für die 4 gedrehten und verschobenen Instanzen liefert der χ^2 -Test Werte zwischen 109 und 333, während für die anders texturierten Objekte das Ergebnis über 2000 liegt.

Der Durchschnitt zweier Histogramme liefert den gemeinsamen Teil von beiden zurück. Diese Vergleichsmethode sollte verwendet werden, wenn das Suchmuster kleiner ist, als die Kandidaten, bzw. nur als ein Ausschnitt vorliegt. Der Durchschnitt ist wie folgt definiert:

$$\cap(q, h) = \frac{\sum_n \min(q_n, h_n)}{\sum_n q_n}$$

Die beiden Histogramme sollten bei dieser Vergleichsmethode in einer nichtnormierten Form vorliegen. Je nach Bedarf kann das Gewicht mehr auf lokale oder auf globale Ähnlichkeit gelegt werden, indem die Supportgröße der Kernfunktion verändert wird. Kleine Supportgrößen bedeuten lokale Ähnlichkeit, große bedeuten globale.

Die Erkennungsrate kann durch Verwendung von kombinierten Farb-Merkmal-Histogrammen erhöht werden. Dazu wird jeder Farbkorb h_n unterteilt in K Merkmalkörbe $h_{n,0}, h_{n,1} \dots$

Mit den hier vorgestellten Methoden können skalierte Bilder nur eingeschränkt behandelt werden, und die Ergebnisse verschlechtern sich, wenn der Größenunterschied wächst. Ein möglicher Ausweg wäre das Vergleichen mit mehreren verschiedenen skalierten Versionen eines Bildes.

7.3 Beispiel: SIMBA

In diesem Abschnitt wird SIMBA vorgestellt, ein System zur inhaltsbasierten Bildsuche, das vom Institut für Mustererkennung und Bildverarbeitung an der Universität Freiburg entwickelt wurde. Die Abkürzung SIMBA steht für Search Images By Appearance. SIMBA verwendet die in diesem Beitrag vorgestellten Histogramme von invarianten Merkmalen, ergänzt um Texturmerkmale, die invariant gegenüber Helligkeitsveränderungen im Bild sind.

Bei diesen Merkmalen ist die Kernfunktion f wie folgt definiert:

$$f(X) = \text{rel}(X(x_1, y_1) - X(x_2, y_2))$$

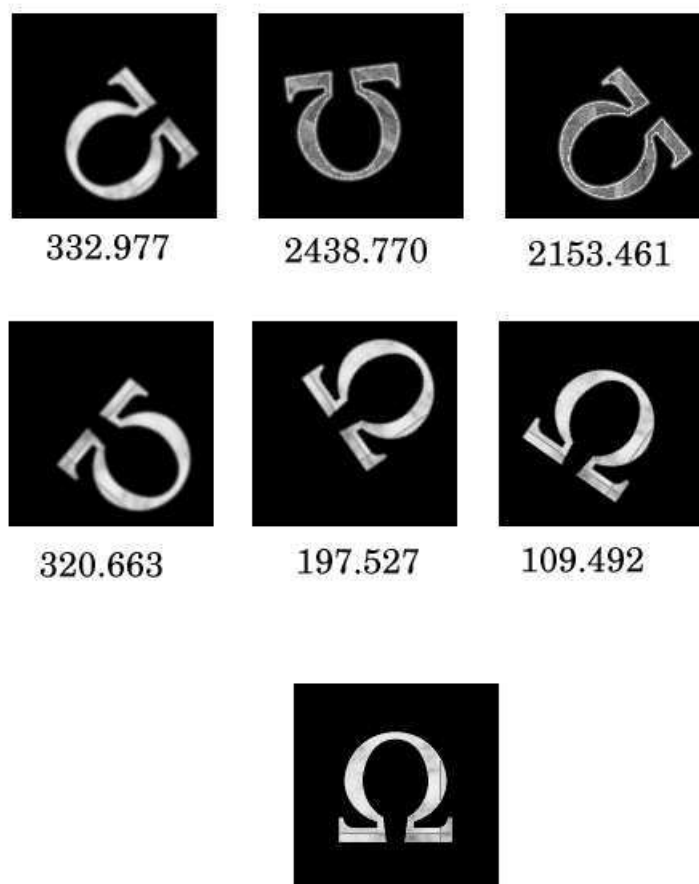


Abbildung 6: Beispiel für die Ergebnisse eines χ^2 -Tests mit $f = X[0, 0] * X[0, 1]$.

mit

$$rel(\delta) = \begin{cases} 1 & \delta \leq -\epsilon \\ \frac{1}{2\epsilon}(\epsilon - \delta) & -\epsilon \leq \delta \leq \epsilon \\ 0 & \epsilon < \delta \end{cases}$$

für ein $\epsilon > 0$. Die *rel*-Funktion vergleicht zwei Grauwerte w_1 und w_2 und liefert 0 zurück, falls $w_1 < w_2$, 1 falls $w_2 < w_1$ und einen Wert zwischen 0 und 1, wenn die beiden Grauwerte ungefähr gleich sind. Die Toleranz für die Gleichheit wird durch die Wahl von ϵ bestimmt. Für jede lokale Auswertung wird ein dreikörbiges Histogramm erstellt, das die Pixel in der Nachbarschaft in die Klassen “heller”, “gleichhell” und “dunkler” einteilt. Anschließend werden die lokalen Auswertungen in ein gemeinsames Histogramm überführt.

SIMBA ist als ein Client-Server-System implementiert. Der Client kann ein beliebiger Rechner mit Internet-Zugang sein. Das Suchmuster braucht nicht über das Internet gesendet zu werden, sondern es reicht aus, die lokal berechneten Merkmale an den Server weiterzuleiten. Der Server hält die Merkmale von allen in der Datenbank vorhandenen Bildern parat. Mit dem Suchmuster wird eine gewichtete “nächster-Nachbar”-Abfrage durchgeführt und die Treffer zusammen mit der Trefferquote werden an den Client zurückgeliefert.

Der Prototyp des Systems mit einem Web-Interface ist unter <http://simba.informatik.uni-freiburg.de> zu finden.

Das System erlaubt es, die Gewichtung der Texturmerkmale und der farbbasierten Merkmale zu verändern. Je nach Suchmuster liefert die eine oder die andere Strategie bessere Ergebnisse. In Abbildung 7 ist ein Beispiel für ein Suchergebnis angegeben. In der im Internet verfügbaren Version sind die Texturmerkmale noch nicht implementiert, so dass es nicht möglich war, diese zu testen.



Abbildung 7: Beispiel für eine Suchabfrage mit dem Web-Interface von SIMBA.

8 Zusammenfassung

In diesem Beitrag wurde die Erzeugung von invarianten Merkmalen und ihre Anwendung bei der Klassifikation von Bildern vorgestellt. Invariante Merkmale lassen sich mit Hilfe von drei verschiedenen Ansätzen berechnen: der affinen Normalisierung, der Differentiation und der Integration. Die ersten zwei Ansätze wurden kurz beschrieben, die integralen Invarianten wurden ausführlich vorgestellt.

Die Bildung von integralen Invarianten erfolgt durch Integration über alle Elemente einer Gruppe von Transformationen, in der Praxis der Gruppe der Kongruenzen. Integrale Invarianten lassen sich effizient berechnen, mittels Schätzung sogar in konstanter Zeit, und erfordern keine vorsegmentierten Daten, so dass eine automatisierte Verarbeitung möglich ist. Nachteile sind die eingeschränkte Einsetzbarkeit bei Skalierung und die starke Abhängigkeit von den Grauwerten des Bildes.

Als Alternative zu invarianten Merkmalen, die global für das gesamte Bild berechnet werden, wurden Histogramme von lokalen Merkmalen vorgestellt. Diese kommen besser mit Objektüberdeckung zurecht und ermöglichen eine Suche basierend auf Bildteilen.

Invariante Merkmale können in verschiedenen Bereichen der Bilderkennung eingesetzt werden, z.B. bei der inhaltsbasierten Bildsuche oder bei der Analyse von 3D-Daten. Die in diesem Beitrag vorgestellten Verfahren sind in SIMBA implementiert, einem System für inhaltsbasierte Bildsuche, welches als ein praktisches Beispiel vorgestellt wurde.

Literatur

- [1] H. Burkhardt und S. Siggelkow. Invariant features for discriminating between equivalence classes. In: I. Pitas und C. Kotropoulos, *Nonlinear Model-Based Image/Video Processing and Analysis*. John Wiley & Sons, 2001
- [2] S. Siggelkow und M. Schael. Fast estimation of invariant features. In: W. Förstner, J. M. Buhmann, A. Faber und P. Faber, *Mustererkennung DAGM 1999, Informatik aktuell*, Seiten 181-188, Bonn, September 1999
- [3] S. Siggelkow, M. Schael, und H. Burkhardt. SIMBA - Search Images By Appearance. In B. Radig and S. Florczyk, editors, *Pattern Recognition, DAGM, LNCS 2191*, Seiten 9-16, München, September 2001.
- [4] H. Schulz-Mirbach: Invariant features for gray scale images. In G. Sagerer, S. Posch und F. Kummer, *Mustererkennung DAGM 1995, Informatik aktuell*, Seiten 1-14, Bielefeld, 1995